



CENTRAL RESERVE BANK OF PERU

## University Extension Course - Summer 2025 Time Series Econometrics<sup>1</sup>

Prof.: Dr. Ricardo Quineche Uribe<sup>2</sup>  
T.A. Rita Huarancca Delgado<sup>3</sup>

### Contents

<b>1 Complex Numbers</b>	<b>5</b>
1.1 Introduction . . . . .	5
1.2 Polar form . . . . .	5
1.3 Exponential Form . . . . .	8
1.4 Modulus (Absolute value) . . . . .	9
1.5 Unit circle . . . . .	10
1.6 Convergence of Series of Complex Numbers . . . . .	10
1.7 Complex Polynomial . . . . .	11
1.8 The Derivative . . . . .	13
1.8.1 The Derivative function . . . . .	13
1.8.2 The derivative at a point, $z_0$ . . . . .	15
1.9 Complex-differentiable function . . . . .	16
1.9.1 What if the function is not continuous? . . . . .	16
1.10 Regular Function . . . . .	17
1.11 Convergence of a Power Series . . . . .	17
1.11.1 Ratio test . . . . .	18
1.11.2 Root test . . . . .	19
1.12 Taylor's Theorem . . . . .	20
1.13 Complex Random Variable . . . . .	20
1.13.1 Expectation . . . . .	20
1.13.2 Variance . . . . .	21
1.13.3 Pseudo-variance . . . . .	22
1.13.4 Covariance and Complementary Covariance . . . . .	22
1.13.5 Uncorrelatedness . . . . .	22
1.13.6 Orthogonality . . . . .	22
<b>2 Complex functions</b>	<b>23</b>
2.1 The natural logarithm of a complex number . . . . .	23
2.1.1 Principal Logarithm . . . . .	23
2.1.2 When is the function continuous? . . . . .	25
2.1.3 Complex differentiability of natural logarithm . . . . .	26
2.2 The natural exponential of a complex number . . . . .	29
2.2.1 Complex differentiability . . . . .	30
2.3 The complex valued function: $(1 - \alpha z)^{-1}$ . . . . .	32

<sup>1</sup>This content cannot be shared, uploaded, or distributed. This content has not been written for publication.

<sup>2</sup>The contact email address is rquinezheu@pucp.pe

<sup>3</sup>The contact email address is rita.h.delgado@gmail.com

2.3.1	Complex differentiability . . . . .	32
2.3.2	Taylor Expansion . . . . .	34
2.3.3	Convergence of the Taylor Expansion . . . . .	34
<b>3</b>	<b>Lag polynomial</b>	<b>36</b>
<b>4</b>	<b>Convergence of Random Variables</b>	<b>37</b>
4.1	Convergence of a Sequence of Numbers . . . . .	38
4.2	Types of Convergence for Sequences of Random Variables . . . . .	38
4.3	Convergence in Probability . . . . .	39
4.3.1	Some intuition . . . . .	40
4.3.2	Markov's Inequality . . . . .	40
4.4	Convergence in probability of a Infinite Sum of Random Variables: Absolutely Summable . .	41
4.5	Convergence in $q^{\text{th}}$ moment . . . . .	46
4.5.1	Convergence in 1st moment . . . . .	47
4.5.2	Convergence in 2nd moment . . . . .	47
4.6	Convergence in second moment of a Infinite Sum of Random Variables: Absolutely Summable	48
4.6.1	Usefulness of this exercise . . . . .	48
4.6.2	Proof . . . . .	49
4.7	Convergence in second moment of an Infinite Sum of Random Variables: Square Summable .	55
4.7.1	Usefulness of this exercise . . . . .	55
4.7.2	Proof . . . . .	55
<b>5</b>	<b>Stationary Time Series</b>	<b>55</b>
5.1	Strict Stationarity . . . . .	55
5.2	Covariance Stationarity . . . . .	56
5.3	The Autocovariance and Autocorrelation Function . . . . .	56
5.4	Complex-Valued Stationary Time Series . . . . .	57
5.5	The Complex-Valued Autocovariance Function . . . . .	57
5.6	Stationarity of an Infinite sum of random variables (Real-valued/complex-valued) . . . . .	58
5.6.1	Proof . . . . .	58
<b>6</b>	<b>Spectral Representation</b>	<b>59</b>
6.1	Frequency Domain . . . . .	59
6.2	Frequency Domain versus Time Domain . . . . .	59
6.3	Advantages . . . . .	60
6.4	Fourier Transform - Some Intuition . . . . .	61
6.4.1	Winding the original series around a circle . . . . .	61
6.4.2	The winding frequency . . . . .	61
6.4.3	The center of mass of the winding graph . . . . .	61
6.4.4	The center of mass as a function of the winding frequency . . . . .	62
6.4.5	Winding the original series around a circle on the Complex Plane . . . . .	62
6.5	Fourier Transform - Formal Definition . . . . .	69
6.6	Fourier Inversion . . . . .	71
6.6.1	Proof . . . . .	71
6.7	Alternative Conventions for formal definitions . . . . .	71
6.7.1	Putting the the factor of $\frac{1}{2\pi}$ in the Fourier transform instead of in its inverse . . . . .	71
6.7.2	Splitting the factor of $\frac{1}{2\pi}$ evenly between the Fourier transform and its inverse . . . . .	72
6.8	Lag Operator Calculus and Fourier Transforms . . . . .	72
6.8.1	Case 1 . . . . .	72
6.8.2	Case 2: AR(m) . . . . .	74
6.8.3	Case 3: MA(n) . . . . .	74

6.9	Spectral Density . . . . .	74
6.9.1	Some Intuition . . . . .	74
6.9.2	Formal Definition . . . . .	75
6.9.3	Approximation of the spectral density by the second moment of the Fourier transform . . . . .	76
6.9.4	Finding the autocovariance: Fourier Inverse . . . . .	77
6.9.5	Lag Operator Calculus, Stationarity and the Spectral density . . . . .	79
<b>7</b>	<b>AR(P)</b>	<b>82</b>
7.1	The lag polynomial . . . . .	82
7.2	Solving the difference equation . . . . .	82
7.3	Is the inverse of the lag polynomial well defined? . . . . .	83
7.3.1	The roots of the characteristic polynomial . . . . .	83
7.3.2	An alternative characteristic polynomial: The reflected polynomial . . . . .	84
7.4	Is the function a regular one? . . . . .	85
7.4.1	Inverting $\pi(L)$ . . . . .	85
7.4.2	Is the characteristic polynomial $\pi(z)^{-1}$ analytic on the unit circle, $ z  = 1$ ? . . . . .	87
7.5	Is the process stationary? . . . . .	89
7.6	The Spectral Density . . . . .	90
7.7	Impulse response Function . . . . .	90
<b>8</b>	<b>Testing for the presence of Unit Root</b>	<b>91</b>
8.1	Motivation . . . . .	91
8.2	The most common unit root tests . . . . .	92
8.3	The Dickey-Fuller (DF) Statistic . . . . .	93
8.4	The Augmented Dickey Fuller - ADF . . . . .	93
8.5	The Phillips Perron Test . . . . .	94
8.6	The $ADF - GLS$ test . . . . .	94
8.7	The $M$ statistics . . . . .	95
<b>9</b>	<b>VAR (p)</b>	<b>95</b>
9.1	Motivation . . . . .	95
9.2	Companion Form Representation of an AR(p) Model . . . . .	96
9.2.1	Stationarity of the Stacked VAR(1) . . . . .	97
9.3	The VAR(1) model with 2 variables . . . . .	97
9.3.1	Matrix Notation . . . . .	98
9.3.2	Assumptions on the errors . . . . .	98
9.3.3	Solving the system of equations: Inverting the matrix . . . . .	99
9.3.4	The roots of the characteristic polynomial and the eigenvalues . . . . .	101
9.3.5	Stationary Solution . . . . .	102
9.4	VAR(1) with $n$ variables . . . . .	102
9.5	The VAR(2) model . . . . .	102
9.5.1	Matrix Notation . . . . .	102
9.5.2	Solving the system of equations: Inverting the matrix . . . . .	103
9.5.3	Stationary Solution . . . . .	105
9.6	The VAR(p) model . . . . .	105
9.6.1	Stationary Solution . . . . .	105
<b>10</b>	<b>Structural Vector Autoregressions</b>	<b>106</b>
10.1	Motivation . . . . .	106
10.2	The Reduced VAR . . . . .	106
10.3	The Structural VAR . . . . .	107
10.4	From a SVAR to an RVAR . . . . .	107

10.5	The Identification Problem . . . . .	108
10.6	Reduced form to structure . . . . .	108
10.6.1	A note about $B_0$ . . . . .	109
10.6.2	Identification of $R$ . . . . .	109
10.7	Identification by Short Run Restrictions . . . . .	110
10.8	Identification by Long Run Restrictions . . . . .	111
10.9	Identification from Heteroskedasticity . . . . .	115
<b>11</b>	<b>Cointegrated VAR</b>	<b>116</b>
11.1	Motivation . . . . .	116
11.2	$I(0)$ process . . . . .	116
11.3	$I(d)$ process . . . . .	116
11.4	Cointegrated process with cointegrating vector $\beta$ . . . . .	117
11.4.1	Example . . . . .	117
11.5	Spurious Regression . . . . .	117
11.6	Cointegrated VAR(1) . . . . .	118
11.6.1	Finding the integrated order of the process . . . . .	118
11.6.2	Is it cointegrated? . . . . .	119
11.7	Rank decomposition . . . . .	119
11.8	The Vector Error Correction model . . . . .	120
11.9	Cointegrated VAR( $p$ ) . . . . .	121
11.9.1	VECM representation . . . . .	121
11.10	The rank of $\Pi$ . . . . .	123
<b>12</b>	<b>Estimation of the Cointegrated VAR(<math>P</math>)</b>	<b>125</b>
12.1	Residual-Based Tests for Cointegration . . . . .	125
12.2	Dynamic OLS . . . . .	126
12.3	Johansen's Methodology for Modeling Cointegration . . . . .	127
12.3.1	Likelihood Ratio Tests for the Number of Cointegrating Vectors . . . . .	127
12.3.2	Johansen's Trace Statistic . . . . .	127
12.3.3	Johansen's Maximum Eigenvalue Statistic . . . . .	128
12.3.4	Estimation of the Cointegrated VECM . . . . .	128
12.3.5	Reduced Rank Regression . . . . .	129
12.3.6	Hypothesis Testing about the coefficients . . . . .	131
<b>13</b>	<b>Further Research on Cointegration</b>	<b>131</b>
<b>Appendices</b>		<b>132</b>
<b>Appendix A Some matrix algebra facts</b>		<b>132</b>
A.1	Eigenvalues and Eigenvectors . . . . .	132
A.2	Eigenvalues and the characteristic polynomial . . . . .	134
A.3	Eigen Decomposition "Matrix Diagonalization" . . . . .	134
A.4	Features of Eigen Decomposition . . . . .	135
A.5	Cholesky Decomposition . . . . .	136
A.5.1	Example of a variance-covariance matrix . . . . .	136

# 1 Complex Numbers

## 1.1 Introduction

$$z = x + iy \quad (1)$$

- $i = \sqrt{-1}$
- $\operatorname{Re}(z) = x$ : denotes the real parts of  $z$
- $\operatorname{Im}(z) = y$ : denotes the imaginary parts of  $z$
- **The absolute value, or modulus** of a complex number is denoted by:

$$|z| = \sqrt{x^2 + y^2}$$

Notice that the modulus of a complex number is always a **real number** and in fact it will never be negative.

- **The complex conjugate** of a complex number is the number with an equal real part and an imaginary part equal in magnitude, but opposite in sign.

$$\bar{z} = x - iy$$

An important identity:

$$z\bar{z} = |z|^2$$

Thus, the product of a complex number and its conjugate is a **real number**.

- **The modulus of a complex conjugate** of a complex number is the same as the modulus of the original complex number:

$$|z| = |\bar{z}|$$

$$\sqrt{x^2 + y^2} = \sqrt{x^2 + (-y)^2}$$

## 1.2 Polar form

- We can think of complex numbers as vectors as in Figure 1
- $r$  denotes the length of the vector
- $r$  is the hypotenuse
- $\varphi$  is the angle from the positive real axis to the vector representing  $z$
- $\varphi$  is usually measured in the standard unit called “radians”.
- The numeric value is given by the angle in radians, and is positive if measured counterclockwise, and negative if measured clockwise
- $-\pi < \varphi \leq \pi$

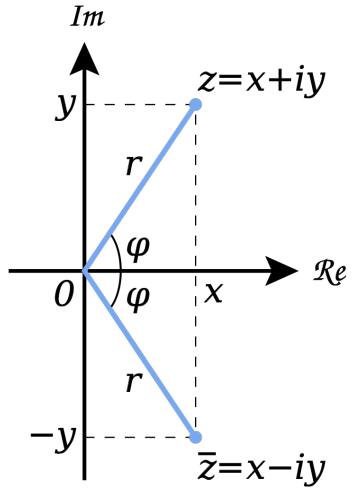


Figure 1: Complex plane

- $\varphi$  is also an element of  $\theta$ , which is **the argument of a complex number** and it is defined as the angle inclined from the real axis in the direction of the complex number represented on the complex plane

$$\theta = \arg z$$

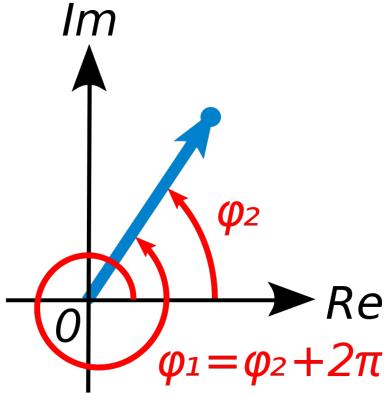


Figure 2:  $\arg(z)$

- It can be seen that the argument of any non-zero complex number has many possible values: firstly, as a geometrical angle, it is clear that whole circle rotations do not change the point, so angles differing by an integer multiple of  $2\pi$  radians (a complete circle) are the same, as reflected by Figure 2
- Because a complete rotation around the origin leaves a complex number unchanged, there are many choices which could be made for  $\theta$  by circling the origin any number of times
- Thus,  $\arg z$  is a multi-valued (set-valued) function in the sense that there are a set of values which includes  $\varphi$

- The argument of  $z$  can be any of the infinite possible values of  $\theta$  each of which can be found by solving

$$\tan \theta = \frac{y}{x}$$

Thus,

$$\theta = \tan^{-1} \frac{y}{x}$$

$$\theta = \arctan \frac{y}{x}$$

The inverse tangent is a **multivalued function**

- Thus, whenever we see  $\arg z$ , we know that the function is not **unique**
- We would like to see  $\varphi$  in a function instead of  $\arg z$
- When a well-defined function is required, then the usual choice, known as the **principal value**,  $\text{Arg}(z)$ :

$$\text{Arg}(z) = \varphi = \arg(z) - 2\pi n \mid n \in \mathbb{Z} \wedge -\pi < \arg(z) - 2\pi n \leq \pi$$

This represents an angle of up to half a complete circle from the positive real axis in either direction.

- Likewise, the argument of  $z$  in terms of the principal value is

$$\arg(z) \in \{\text{Arg}(z) + 2\pi n \mid n \in \mathbb{Z}\}$$

- $x$  is the adjacent
- $iy$  is the opposite
- Our aim in this section is to write complex numbers in terms of a distance from the origin and a direction (or angle) from the positive horizontal axis.
- From **Pythagoras**, we have:

$$r^2 = x^2 + (y)^2$$

$$r = \sqrt{x^2 + (y)^2}$$

- $r$  is the absolute value (or modulus) of the complex number
- Let's forget about the  $i$  for a moment. Basic trigonometry gives us:

Sine

$$\sin \varphi = \frac{y}{r} = \frac{\text{opposite}}{\text{hypotenuse}}$$

Cosine

$$\cos \varphi = \frac{x}{r} = \frac{\text{adjacent}}{\text{hypotenuse}}$$

Tangent

$$\tan \varphi = \frac{y}{x} = \frac{\text{opposite}}{\text{adjacent}}$$

Cosecant

$$\csc \varphi = \frac{r}{y} = \frac{\text{hypotenuse}}{\text{opposite}}$$

Secant

$$\sec \varphi = \frac{r}{x} = \frac{\text{hypotenuse}}{\text{adjacent}}$$

Cotangent

$$\cot \varphi = \frac{x}{y} = \frac{\text{adjacent}}{\text{opposite}}$$

- Thus, we have:

$$\tan \varphi = \frac{y}{x}$$

$$x = r \cos \varphi$$

$$y = r \sin \varphi$$

- Multiplying the last expression throughout by  $i$  gives us:

$$yi = ir \sin \varphi$$

- So we can write the polar form of a complex number as:

$$z = x + yi = r(\cos \varphi + i \sin \varphi) \quad (2)$$

$$z = x + yi = |z|(\cos \varphi + i \sin \varphi) \quad (3)$$

But, this also holds for  $\theta$ . So a polar representation in terms of  $\theta$  is also valid:

$$z = x + yi = r(\cos \theta + i \sin \theta)$$

- So, it is a parametric representation of a circle of radius  $r$

### 1.3 Exponential Form

- Actually, it is an alternative way of writing the **polar form**
- From **Euler's formula, for any real number  $\theta$ :**

$$e^{i\theta} = \cos \theta + i \sin \theta \quad (4)$$

where  $e$  is the base of the natural logarithm,  $i$  is the imaginary unit

- Euler's formula, named after Leonhard Euler, is a mathematical formula in complex analysis that establishes the fundamental relationship between the trigonometric functions and the **complex exponential function**.
- As  $\theta$  is a real number, from (2), we have that the **exponential form** of a complex number,  $z$  with modulus  $r$  and  $\arg z = \theta$  is given by:

$$z = re^{i\theta} \quad (5)$$

- where  $\theta = \arg z$  and so we can see that, much like the polar form, there are an infinite number of possible exponential forms for a given complex number
- Thus it is not unique unless we limit  $\theta$  to the principal value

- Also, because any two arguments for a given complex number differ by an integer multiple of  $2\pi$ , which is the measure in radians of a complete circle, we will sometimes write the exponential form as:

$$z = re^{i(\varphi+2\pi n)} \quad n = 0, \pm 1, \pm 2, \dots$$

where  $\varphi$  is the **principal value** of the argument

## 1.4 Modulus (Absolute value)

- Let's start by defining the absolute value function for a **real number**. For any real number  $x$ , the absolute value or modulus of  $x$  is denoted by  $|x|$  is defined as:

$$|x| = \begin{cases} x, & \text{if } x \geq 0 \\ -x, & \text{if } x < 0 \end{cases}$$

$$|x| = \sqrt{x^2}$$

- when  $x$  itself is negative ( $x < 0$ ), then its absolute value is necessarily positive  $|x| = -x > 0$ . Since  $x$  is negative we get  $-1$  times a negative, which gives me a positive number
- The absolute value of  $x$  is thus always either positive or zero, but never negative
- In the second representation, the **square root symbol** represents the unique **positive square root** (when applied to a positive number)
- The real absolute value cannot be directly applied to complex numbers.
- However, the geometric interpretation of the absolute value of a real number as its distance from 0 can be generalised.
- The absolute value of a complex number is defined by the Euclidean distance of its corresponding point in the complex plane from the origin.
- This can be computed using the **Pythagorean theorem**. From Figure 1, it is easy to see that the absolute value or modulus of  $z$  is:

$$\begin{aligned} |z| &= \sqrt{[\operatorname{Re}(z)]^2 + [\operatorname{Im}(z)]^2} = \sqrt{x^2 + y^2} \\ |z| &= r \end{aligned} \tag{6}$$

, notice that the **square root symbol** represents the unique **positive square root** (when applied to a positive number), which means that the **modulus of a complex number is always a real positive number**

- Notice that the absolute value (or modulus) of the complex number is  $r$  in the polar form
- To get the value of  $r$  we can use (5) and the modulus of both sides and then do a little simplification as follows:

$$|z| = |r e^{i\theta}| = |r| |e^{i\theta}| = |r| |\cos \theta + i \sin \theta| = \sqrt{r^2 + 0} \sqrt{\cos^2 \theta + \sin^2 \theta} = r$$

since  $\cos^2 \theta + \sin^2 \theta = 1$  by the fundamental Pythagorean identity.

- When the imaginary part  $y$  is zero, this coincides with the definition of the absolute value of the real number  $x$ .

## 1.5 Unit circle

- The unit circle is the set of complex numbers whose modulus,  $|z|$ , is 1
- On the complex plane they form a circle centered at the origin with a radius of 1
- It includes the value :
  - 1 on the right extreme
  - $i$  at the top extreme
  - $-1$  at the left extreme
  - $-i$  at the bottom extreme
- On its exponential form it can be written as:

$$z = e^{i\theta}$$

since the modulus of  $z$  is 1 ( $|z| = r = 1$ ). Its conjugate:

$$\bar{z} = \overline{e^{i\theta}} = \cos \theta - i \sin \theta = e^{-i\theta}$$

We can also find its modulus. Recall from the identity with its conjugate:

$$|z| = (z\bar{z})^{1/2}$$

$$|e^{i\theta}| = \sqrt{e^{i\theta}e^{-i\theta}} = 1$$

- The interior of the unit circle is called the **open unit disk**
- While the interior of the unit circle combined with the unit circle itself is called the **closed unit disk**.

## 1.6 Convergence of Series of Complex Numbers

- Sequence of complex numbers:  $\{z_i\}_{i=1}^{\infty}$
- Infinite series of complex numbers:  $\sum_{i=1}^{\infty} z_i$
- A series  $\sum_{m=1}^{\infty} z_m$  converges to  $A \in \mathbf{C}$  if

$$\lim_{n \rightarrow \infty} \sum_{m=1}^n z_m = A$$

that is, if

$$\lim_{n \rightarrow \infty} \left| \sum_{m=1}^n z_m - A \right| = 0$$

- **Theorem 1** If  $\sum_{m=1}^{\infty} z_m$  converges then

$$\lim_{m \rightarrow \infty} z_m = 0,$$

Thus, the element  $z_m$  inside  $\sum_{m=1}^{\infty} z_m$  goes to zero as long as  $m$  goes to infinity.

- **Cauchy's Criterion.**  $\sum_{m=1}^{\infty} z_m$  converges if and only if

$$\left| \sum_{m=k}^n z_m \right| \rightarrow 0$$

as  $k, n \rightarrow \infty$

- **Absolute Convergence.**  $\sum_{m=1}^{\infty} z_m$  converges absolutely if  $\sum_{m=1}^{\infty} |z_m|$  converges.

## 1.7 Complex Polynomial

A complex polynomial of degree  $n$  is a function of the form

$$P(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$$

$$P(z) = \sum_{k=0}^n a_k z^k$$

- $n$  is a nonnegative integer,  $n \geq 1$ ,
- The  $a_k$  are complex numbers not all zero
- $z$  is a complex variable
- Any polynomial of degree  $n$  has precisely  $n$  roots
- A root is a value of  $z$  such that  $P(z) = 0$
- **The Fundamental Theorem of Algebra:** For any polynomial of degree  $n$ , we can rewrite the polynomial in terms of its roots,  $z_i$ :

$$P(z) = a_n (z - z_1)(z - z_2) \cdots (z - z_n)$$

- **Reflected (Reciprocal) polynomial:**

$$\tilde{P}(z) = a_n + a_{n-1}z + a_{n-2}z^2 + \cdots + a_0 z^n$$

That is, the coefficients of  $\tilde{P}(z)$  are the coefficients of  $P(z)$  in reverse order. Notice that:

$$\tilde{P}(z) = (a_n z^n + a_{n-1} z^{n-1} + a_{n-2} z^{n-2} + \cdots + a_0) z^n$$

Thus, we can rewrite the reflected polynomial as:

$$\tilde{P}(z) = z^n P(z^{-1})$$

- Let's denote  $\lambda_i$  as the roots in the characteristic polynomial  $\tilde{P}(z)$
- We know that the both polynomials the original and the reflected one are related. So, are the roots also related?

- Indeed the roots are related. When finding the roots for  $\tilde{P}(z)$ , we are finding the values of  $z$  such that:

$$\tilde{P}(z) = 0$$

, which is the same as finding the values of  $z$  such that:

$$z^n P(z^{-1}) = 0$$

Since the values for  $z$  cannot be zero, then we are finding the values of  $z$  such that:

$$P(z^{-1}) = 0$$

Let's denote  $z_i$  as the roots of the original polynomial,  $P(z)$ . Given this notation, the roots of  $P(z^{-1})$  should be  $z_i^{-1}$  since  $P(z^{-1})$  is  $P(z)$  using the inverse variable. Therefore, we have that:

$$\lambda_i = z_i^{-1} \tag{7}$$

- Notice that:

$$P(z^{-1}) = a_n (z - z_1^{-1})(z - z_2^{-1}) \cdots (z - z_n^{-1})$$

or

$$P(z^{-1}) = a_n (z^{-1} - z_1)(z^{-1} - z_2) \cdots (z^{-1} - z_n)$$

and that

$$\tilde{P}(z) = a_0 (z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_n)$$

$$\tilde{P}(z) = a_0 (z - z_1^{-1})(z - z_2^{-1}) \cdots (z - z_n^{-1})$$

Thus,

$$\tilde{P}(z^{-1}) = a_0 (z - \lambda_1^{-1})(z - \lambda_2^{-1}) \cdots (z - \lambda_n^{-1})$$

$$\tilde{P}(z^{-1}) = a_0 (z^{-1} - \lambda_1)(z^{-1} - \lambda_2) \cdots (z^{-1} - \lambda_n)$$

- Further, notice that  $P(z)$  is also the reflected polynomial of  $\tilde{P}(z)$ :

$$P(z) = z^n \tilde{P}(z^{-1})$$

Thus,

$$P(z) = z^n a_0 (z^{-1} - \lambda_1)(z^{-1} - \lambda_2) \cdots (z^{-1} - \lambda_n)$$

$$P(z) = a_0 z^n (z^{-1} - \lambda_1) (z^{-1} - \lambda_2) \cdots (z^{-1} - \lambda_n)$$

$$P(z) = a_0 (z^{-1} - \lambda_1) z (z^{-1} - \lambda_2) z \cdots (z^{-1} - \lambda_n) z$$

$$P(z) = a_0 (1 - \lambda_1 z) (1 - \lambda_2 z) \cdots (1 - \lambda_n z)$$

- We know that  $\lambda_i = z_i^{-1}$ . Thus, we can also write:

$$P(z) = a_0 \left(1 - \frac{1}{z_1} z\right) \left(1 - \frac{1}{z_2} z\right) \left(1 - \frac{1}{z_3} z\right) \cdots \left(1 - \frac{1}{z_p} z\right) \quad (8)$$

$$P(z) = a_0 (1 - z_1^{-1} z) (1 - z_2^{-1} z) \cdots (1 - z_n^{-1} z)$$

## 1.8 The Derivative

### 1.8.1 The Derivative function

We can define the derivative function of  $f(z)$  as:

$$f'(z) = \lim_{\Delta z \rightarrow 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}$$

- Notice that the derivative is a limit, so for the derivative function to exist we need the limit to exist
- Knowing the set of values for which the function is **not continuous** is relevant since if the function is not continuous at some values of  $z$ , then its derivative does not exist at those values.
- Recall that  $z = x + iy$ . Since  $i$  is a constant, a change on  $z$  can only be triggered by a change in  $x$  or  $y$ . Thus:

$$\Delta z = \Delta x + i\Delta y$$

- Since  $z = x + iy$ ,  $f(z)$  can be rewritten such that:

$$f(z) \equiv f(x, y) = u(x, y) + iv(x, y)$$

So,  $u(\cdot)$  and  $v(\cdot)$  are functions of real numbers

- Thus, if we are only considering a change on the real axis,  $x$ , we have that:

$$f'(z) = \lim_{\Delta x \rightarrow 0} \frac{f(z + \Delta x) - f(z)}{\Delta x}$$

In terms of  $u$  and  $v$ :

$$f'(z) = \lim_{\Delta x \rightarrow 0} \frac{u(x + \Delta x, y) - u(x, y)}{\Delta x} + \lim_{\Delta x \rightarrow 0} \frac{iv(x + \Delta x, y) - iv(x, y)}{\Delta x}$$

$$f'(z) = \lim_{\Delta x \rightarrow 0} \frac{u(x + \Delta x, y) - u(x, y)}{\Delta x} + i \lim_{\Delta x \rightarrow 0} \frac{v(x + \Delta x, y) - v(x, y)}{\Delta x}$$

By using the definition of partial derivative for  $u$  and  $v$ , we have that

$$f'(z) = u_x(x, y) + iv_x(x, y)$$

- On the other hand, if we are only considering a change on the imaginary axis,  $y$ , we have that:

$$f'(z) = \lim_{\Delta y \rightarrow 0} \frac{f(z + i\Delta y) - f(z)}{i\Delta y}$$

,  $i\Delta y$  takes into account the fact that a change in  $y$  affects  $z$  in  $+i\Delta y$ . In terms of  $u$  and  $v$  we have that the change in  $y$  can be isolated in the sense that we do not need to express it as  $+i\Delta y$ :

$$f'(z) = \lim_{\Delta y \rightarrow 0} \frac{u(x, y + \Delta y) - u(x, y)}{i\Delta y} + \lim_{i\Delta y \rightarrow 0} \frac{iv(x, y + \Delta y) - iv(x, y)}{i\Delta y}$$

$$f'(z) = \lim_{\Delta y \rightarrow 0} \frac{u(x, y + \Delta y) - u(x, y)}{i\Delta y} + \lim_{\Delta y \rightarrow 0} \frac{i(v(x, y + \Delta y) - v(x, y))}{i\Delta y}$$

**TRICK:** Notice that  $iy$  is not an argument in  $u$  or  $v$ . We need the denominator to be  $\Delta y$  so the limit expression has the meaning of partial derivative. Thus, we proceed as follows:

$$f'(z) = \lim_{\Delta y \rightarrow 0} \frac{i(u(x, y + \Delta y) - u(x, y))}{i(i\Delta y)} + \lim_{\Delta y \rightarrow 0} \frac{i(v(x, y + \Delta y) - v(x, y))}{i\Delta y}$$

$$f'(z) = \lim_{\Delta y \rightarrow 0} \frac{i(u(x, y + \Delta y) - u(x, y))}{-1\Delta y} + \lim_{\Delta y \rightarrow 0} \frac{v(x, y + \Delta y) - v(x, y)}{\Delta y}$$

$$f'(z) = \lim_{\Delta y \rightarrow 0} \frac{-i(u(x, y + \Delta y) - u(x, y))}{\Delta y} + \lim_{\Delta y \rightarrow 0} \frac{v(x, y + \Delta y) - v(x, y)}{\Delta y}$$

$$f'(z) = -i \lim_{\Delta y \rightarrow 0} \frac{u(x, y + \Delta y) - u(x, y)}{\Delta y} + \lim_{\Delta y \rightarrow 0} \frac{v(x, y + \Delta y) - v(x, y)}{\Delta y}$$

By using the definition of partial derivative for  $u$  and  $v$ , we have that

$$f'(z) = -iu_y(x, y) + v_y(x, y)$$

- Thus we have two expressions for  $f'(z)$ :

$$f'(z) = u_x(x, y) + iv_x(x, y)$$

$$f'(z) = -iu_y(x, y) + v_y(x, y)$$

Both can be used, however they can only be used if the derivative exists

- Those two expressions helps us to get the **Cauchy-Riemann equations** by equating the real part and the imaginary part, respectively:

$$u_x(x, y) = v_y(x, y)$$

$$v_x(x, y) = -u_y(x, y)$$

If both equations hold, then the derivative exist.

- Example:

$$f(x + yi) = (x + yi)^2 = (x^2 - y^2) + i2xy$$

Notice that,

$$u(x, y) = x^2 - y^2 \text{ and } v(x, y) = 2xy$$

Thus,

$$u_x = 2x = v_y$$

$$u_y = -2y = -v_x.$$

Further, the derivative of  $f(z)$  is clearly  $f'(z) = 2z$ , so:

$$f'(x + iy) = 2(x + iy) = 2x + i2y = u_x + iv_x = v_y - iu_y$$

### 1.8.2 The derivative at a point, $z_0$

The derivative of  $f(z)$  at  $z_0$  is defined by<sup>4</sup>

$$f'(z_0) = \lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

provided this limit exists.

- where  $z_0$  is a complex number and  $z \rightarrow z_0$  means  $|z - z_0| \rightarrow 0$ .
- This limit must be the same no matter how  $|z - z_0| \rightarrow 0$ .
- Since this is the complex plane,  $|z - z_0| \rightarrow 0$  is not only considering the values coming from the left and from the right as in the definition of the existence of a limit on  $\mathbb{R}$ .  $|z - z_0|$  is the modulus, so it is considering all the values inside a circle with radius  $|z - z_0|$  on the complex plane
- Since it is possible to approach  $z_0$  from many directions on the complex plane, the differentiability of functions of complex numbers is in a sense stricter than the differentiability of functions of real numbers
- The function  $f(z)$  is said to be differentiable at  $z_0$  if its derivative at  $z_0$  exists.

---

<sup>4</sup>The difference between this formula and the previous one is that this one is evaluating the derivative function at  $z_0$ , while the former is just the definition of a derivative function

## 1.9 Complex-differentiable function

Let  $z = x + iy$  and  $f(z) = u(x, y) + iv(x, y)$  on some region  $G$  containing the point  $z_0$ . If  $f(z)$  satisfies the Cauchy-Riemann equations and has continuous first partial derivatives in the neighborhood of  $z_0$ , then  $f'(z_0)$  exists and is given by

$$f'(z_0) = \lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

and the function is said to be complex differentiable (or, equivalently, analytic or holomorphic).

**Cauchy-Riemann equations:** We usually have  $f(z)$ , but we know that  $z = x + iy$ . Thus, the idea is to rewrite  $f(z)$  such that:

$$f(z) \equiv f(x, y) \equiv u(x, y) + iv(x, y)$$

If  $f$  is complex differentiable, then the Cauchy-Riemann equations must hold:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$$

and

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$$

*Note:* For using these equations, we usually take advantage of the polar or exponential representation of  $z$ . However it is not so easy to do.

### 1.9.1 What if the function is not continuous?

It is convenient to analyze if the function is **continuous** at some points or not since:

$$\text{Differentiability} \implies \text{Continuity, so } \neg \text{Continuity} \implies \neg \text{Differentiability.}$$

This will play an important role when analyzing the complex natural logarithm function.

#### Example

- Let's analyze the simple case  $y = 0$ , so  $z = x \in \mathbb{R}$
- Take for example the very simple function:

$$f(x) = \begin{cases} x + 1 & x \geq 0 \\ x & x < 0 \end{cases}$$

- It is discontinuous at  $x = 0$
- The limit  $\lim_{x \rightarrow 0} f(x)$  is 0, which is not equal to the limit when  $x$  approaches to zero from the RHS (positive numbers), 1.
- However if we apply the derivative formula like robots, we have that

$$f'(x) = 1$$

for any  $x$ , so we can wrongly conclude the function is differentiable at  $x = 0$

- Recall that we require the derivative to exist at  $x = 0$ , so we should use the definition of derivative:

$$\lim_{x \rightarrow 0} \frac{f(x) - f(0)}{x}$$

To analyze if it exists we need to analyze if the limit when approaching to zero from the positive and negative numbers. Analyzing the limit when approaching to zero from the negative numbers:

$$\lim_{x \rightarrow 0^-} \frac{f(x) - f(0)}{x} = \lim_{x \rightarrow 0^-} \frac{x - 1}{x} = \infty$$

So, the limit does not exist. Therefore, the derivative at  $x = 0$  does not exist.

## 1.10 Regular Function

- A function  $f$  of the complex number  $z$  is analytic at a point  $z_0$  if its derivative exists not only at  $z_0$  but also at each point  $z$  in some neighborhood of  $z_0$ .
- A **regular (or holomorphic or analytic) function** is defined as a complex-valued differentiable function on an open set  $\mathbb{D}$  of  $\mathbb{C}$ . That is, a function is regular on a region  $\mathbb{D}$  of  $\mathbb{C}$ , if the complex-valued function is complex differentiable at every point in the set  $\mathbb{D}$  of  $\mathbb{C}$
- If a function  $f$  is analytic at a point, then its derivatives of all orders exist (**infinitely differentiable**) and are themselves analytic there.
- The **main result** about a regular function is that at any point of the domain of definition,  $\mathbb{D}$ , the regular function can be expanded in a **Taylor's series** that converges in the largest open disk that does not contain any singularity.
- Another fundamental result is that for two regular functions the composition  $f(g(z))$  is again regular provided the range of  $g$  is in the domain of  $f$ .
- If two functions  $f(z)$  and  $g(z)$  are analytic in a domain  $D$ , then their sum and their product are both analytic in  $D$ .
- The quotient  $\frac{f(z)}{g(z)}$  is also analytic in  $D$  provided that  $g(z) \neq 0$  for any  $z$  in  $D$ .
- An **entire function** is a function that is analytic at each point in the entire complex plane.
- Every polynomial is an entire function.
- Hence the quotient  $\frac{P(z)}{Q(z)}$  of two polynomials is analytic in any domain throughout which  $Q(z) \neq 0$ .

## 1.11 Convergence of a Power Series

A power series in one variable is an infinite series of the form

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n = a_0 + a_1 (z - z_0)^1 + a_2 (z - z_0)^2 + \dots$$

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

- $a_n$  represents the complex coefficient of the  $n$ th term
- $a_n$  is independent of  $z$  and may be expressed as a function of  $n$  (e.g.,  $a_n = \frac{1}{n!}$ )

- $z_0$  is the center of the series and is a complex constant
- If there is a number  $r > 0$ , then the series converges absolutely to an analytic function for  $|z - z_0| < r$ .
- The series diverges for  $|z - z_0| > r$
- $r$  is called the radius of convergence
- The disk  $|z - z_0| < r$  is called the disk of convergence
- The theorem doesn't say what happens when  $|z - z_0| = r$ .
- If  $r = \infty$  the function  $f(z)$  is entire.
- If  $r = 0$  the series only converges at the point  $z = z_0$ . In this case, the series does not represent an analytic function on any disk around  $z_0$ .
- The derivative is given by term-by-term differentiation

$$f'(z) = \sum_{n=0}^{\infty} n a_n (z - z_0)^{n-1}$$

The series for  $f'$  also has radius of convergence  $r$

- How do we find  $r$ ? Often (not always) we can find  $r$  using the Ratio test and the Root test. Both are two standard tests from calculus on the convergence of infinite series.

### 1.11.1 Ratio test

Consider the series

$$\sum_{n=0}^{\infty} c_n$$

If  $L = \lim_{n \rightarrow \infty} |c_{n+1}/c_n|$  exists, then:

- If  $L < 1$  then the series converges absolutely
- If  $L > 1$  then the series diverges.
- If  $L = 1$  then the test gives no information.

#### Example

- Consider the geometric series

$$f(z) = 1 + z + z^2 + z^3 + \dots$$

- The limit of the absolute ratios of consecutive terms is

$$L = \lim_{n \rightarrow \infty} \frac{|z^{n+1}|}{|z^n|} = |z|$$

- Thus,  $L$  exists and for the series to converge, we need  $L < 1$ .
- Thus, the geometric series converges when  $|z| < 1$
- Thus, the radius of convergence is 1

- $|z - 0| < 1$  is the disk of convergence
- We know this converges to  $1/(1 - z)$
- Note, the disk of convergence ends exactly at the singularity  $z = 1$

**Example**

- Consider the geometric series

$$f(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

- The limit of the absolute ratios of consecutive terms is

$$L = \lim_{n \rightarrow \infty} \frac{|z^{n+1}| / (n+1)!}{|z^n| / n!} = \lim_{n \rightarrow \infty} \frac{|z|}{n+1} = 0$$

- Thus,  $L$  exists and for the series to converge, we need  $L < 1$ .
- However,  $L = 0$ , so it always converges no mattering the value of  $z$
- Since  $L < 1$  this series converges for every  $z$ .
- the radius of convergence for this series is  $\infty$ . That is,  $f(z)$  is entire.

### 1.11.2 Root test

Consider the series

$$\sum_0^{\infty} c_n$$

If  $L = \lim_{n \rightarrow \infty} |c_n|^{1/n}$  exists, then:

- If  $L < 1$  then the series converges absolutely.
- If  $L > 1$  then the series diverges.
- If  $L = 1$  then the test gives no information.

**Example**

- Consider the geometric series

$$1 + z + z^2 + z^3 + \dots$$

- The limit of the  $n$  th roots of the terms is

$$L = \lim_{n \rightarrow \infty} |z^n|^{1/n} = \lim_{n \rightarrow \infty} |z| = |z|$$

- Thus, the root test agrees that the geometric series converges when  $|z| < 1$ .

## 1.12 Taylor's Theorem

**Theorem 2** Let  $f(z)$  be an **analytic** everywhere inside a region  $G$ . Let  $z_0 \in G$ . Then,

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

where the series converges on any disk  $|z - z_0| < r$  contained in  $G$ . Furthermore, we have formulas for the coefficients:

$$f(z) = f(z_0) + \frac{f'(z_0)}{1!} (z - z_0) + \frac{f''(z_0)}{2!} (z - z_0)^2 + \dots$$

That is, the Taylor series of a real or complex-valued function  $f(z)$  that is infinitely differentiable (regular) at a real or complex number  $z_0$ .

In the more compact sigma notation, this can be written as

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x - a)^n$$

where  $f^{(n)}(a)$  denotes the  $n$  th derivative of  $f$  evaluated at the point  $a$ . (The derivative of order zero of  $f$  is defined to be  $f$  itself and  $(x - a)^0$  and  $0!$  are both defined to be 1.)

- The special case of series when  $z_0 = 0$  is called the **Maclaurin series**.
- Notice that it holds with equality because it is not an approximation but an equality, but the function must be infinity differentiable (which means that the function is regular). That is why we have an infinite sum.
- If the function is not infinite differentiable we can have a approximation of order  $n$
- **$n^{th}$  Taylor polynomial:** The partial sum formed by the  $n$  first terms of a Taylor series is a polynomial of degree  $n$ . Taylor polynomials are **approximations of a function**, which become generally better when  $n$  increases.

## 1.13 Complex Random Variable

- Complex random variables are a generalization of real-valued random variables to complex numbers
- Complex random variables can always be considered as pairs of real random variables: their real and imaginary parts
- Therefore, the distribution of one complex random variable may be interpreted as the joint distribution of two real random variables
- Some concepts of real random variables have a straightforward generalization to complex random variables—e.g., the definition of the mean of a complex random variable. Other concepts are unique to complex random variables

### 1.13.1 Expectation

- The expectation of a complex random variable is defined based on the definition of the expectation of a real random variable:

$$E[z] = E[x] + iE[y]$$

- Note that the expectation of a complex random variable does not exist if  $E[x]$  or  $E[y]$  does not exist.
- if the complex random variable  $Z$  has a probability density function  $f_Z(z)$ , then the expectation is given by

$$E[z] = \int_{\mathbb{C}} z \cdot f_z(z) dz$$

- If the complex random variable  $Z$  has a probability mass function  $p_Z(z)$ , then the expectation is given by

$$E[z] = \sum_z z \cdot p_z(z)$$

### Properties

- Whenever the expectation of a complex random variable exists, taking the expectation and complex conjugation commute:

$$\overline{E[z]} = E[\bar{z}]$$

- The expected value operator  $E[\cdot]$  is linear in the sense that:

$$E[az + bw] = aE[z] + bE[w]$$

for any complex coefficients  $a, b$  even if  $z$  and  $w$  are not independent.

### 1.13.2 Variance

- The variance is defined as

$$\text{Var}[z] = E[|z - E[z]|^2] = E[|z|^2] - |E[z]|^2$$

### Properties

- The variance is always a nonnegative real number
- It is equal to the sum of the variances of the real and imaginary part of the complex random variable:

$$\text{Var}[z] = \text{Var}(x) + \text{Var}(y)$$

- The variance of a linear combination of complex random variables may be calculated using the following formula:

$$\text{Var} \left[ \sum_{k=1}^N a_k Z_k \right] = \sum_{i=1}^N \sum_{j=1}^N a_i \overline{a_j} \text{Cov}[Z_i, Z_j]$$

for any complex coefficients  $a_k$

### 1.13.3 Pseudo-variance

- The pseudo-variance is a special case of the pseudo-covariance and is given by

$$E[(z - E[z])^2] = E[z^2] - (E[z])^2$$

- Unlike the variance of  $z$ , which is always real and positive, the pseudo-variance of  $z$  is in general complex.

### 1.13.4 Covariance and Complementary Covariance

- The covariance between two complex random variables  $z$  and  $w$  is defined as:

$$K_{zw} = \text{Cov}[z, w] = E[(z - E[z])(\bar{w} - E[\bar{w}])] = E[z\bar{w}] - E[z]E[\bar{w}]$$

- Notice the **complex conjugation of the second factor** in the definition
- In contrast to real random variables, we also define a **pseudo-covariance** (also called complementary variance):

$$J_{zw} = \text{Cov}[z, \bar{w}] = E[(z - E[z])(w - E[w])] = E[zw] - E[z]E[w]$$

### Properties

$$\begin{aligned} \text{Cov}[z, w] &= \overline{\text{Cov}[w, z]} && \text{(Conjugate symmetry)} \\ \text{Cov}[\alpha z, w] &= \alpha \text{Cov}[z, w] && \text{(Sesquilinearity)} \\ \text{Cov}[z, \alpha w] &= \bar{\alpha} \text{Cov}[z, w] \\ \text{Cov}[z_1 + z_2, W] &= \text{Cov}[z_1, W] + \text{Cov}[z_2, W] \\ \text{Cov}[z, w_1 + w_2] &= \text{Cov}[z, w_1] + \text{Cov}[z, w_2] \\ \text{Cov}[z, z] &= \text{Var}[z] \end{aligned}$$

### 1.13.5 Uncorrelatedness

Two complex random variables  $z$  and  $w$  are called uncorrelated if their covariance is zero:

$$K_{zw} = J_{zw} = 0$$

Thus,

$$E[z\bar{w}] = E[z]E[\bar{w}]$$

$$E[zw] = E[z]E[w]$$

### 1.13.6 Orthogonality

As with real random variables, complex quantities are said to be orthogonal if:

$$E[z\bar{w}] = 0$$

As always, it does not imply covariance zero unless the mean of the variables are zero.

## 2 Complex functions

### 2.1 The natural logarithm of a complex number

In complex analysis, a complex logarithm of the non-zero complex number  $z$  is

$$w = \ln z$$

- When is it undefined? If  $z$  is a real number, then we know that  $w$  is undefined for  $z \leq 0$
- Thus, we would be tempted to say that this function is undefined when  $z \leq 0$ , but since  $z$  is complex, it is not that simple.
- From the Polar form, we know that  $z = |z|e^{i\theta}$ . Thus,

$$\ln z = \ln |z|e^{i\theta}$$

$$\ln z = \ln |z| + \ln e^{i\theta}$$

$$\ln z = \ln |z| + i\theta \ln e$$

$$\ln z = \ln |z| + i \arg(z)$$

Thus, **ln  $z$**  is **defined** if

$$|z| > 0$$

- This means that we want the radius of  $z$  on the complex plane to be greater than zero
- Since  $|z|$  is always positive by definition, then we allow  $z < 0$
- That is why we say that **ln  $z$**  is **defined** if  $z \in \mathbb{C} \setminus \{0\}$ , which means that the function is defined for all complex numbers (positive and negatives) except for 0
- However, we have a problem:  $\arg(z)$  is **not unique**, which means that **ln** is not the inverse of the exponential function.
- For the function to be **single-valued**, we need to define the **Principal Logarithm**

#### 2.1.1 Principal Logarithm

In complex analysis, the principal logarithm of the non-zero complex number  $z$  is

$$\text{Ln } z = \ln |z| + i \arg(z), \quad -\pi < \arg(z) \leq \pi$$

That is, the the principal logarithm is the complex logarithm but using the principal value,  $\text{Arg}(z)$ . Thus, another way to write is

$$\text{Ln } z = \ln |z| + i \text{Arg}(z)$$

- Thus, the function is unique

- Now, I have an expression for the inverse function of  $e^z$

**Example 1**

- Let's find the principal logarithm for  $z = i$

$$\ln i = \ln |i| + i \operatorname{Arg}(i)$$

- Recall

$$z = x + iy$$

$$|z| = \sqrt{x^2 + y^2}$$

Thus,

$$|z| = |0 + i(1)| = \sqrt{0^2 + 1^2} = 1$$

- Notice that  $\operatorname{Arg}(i)$  can be found directly on the complex plane to be  $\frac{\pi}{2}$ . However, we can also find it by:

$$\operatorname{Arg}(i) = \tan^{-1} \frac{y}{x}$$

$$\operatorname{Arg}(i) = \arctan \frac{1}{0}$$

$$\operatorname{Arg}(i) = \frac{\pi}{2}$$

- Thus,

$$\ln i = \ln 1 + i \frac{\pi}{2}$$

$$\ln i = i \frac{\pi}{2}$$

**Example 2**

- Let's find the principal logarithm for  $z = 1 + i$

$$\ln(1 + i) = \ln |1 + i| + i \operatorname{Arg}(1 + i)$$

- Recall

$$z = x + iy$$

$$|z| = \sqrt{x^2 + y^2}$$

Thus,

$$|z| = |1 + i(1)| = \sqrt{1^2 + 1^2} = \sqrt{2}$$

- Notice that  $\text{Arg}(i)$  can be found directly on the complex plane to be  $\frac{\pi}{4}$ . However, we can also find it by:

$$\text{Arg}(i + 1) = \tan^{-1} \frac{y}{x}$$

$$\text{Arg}(i + 1) = \arctan \frac{1}{1}$$

$$\text{Arg}(i + 1) = \frac{\pi}{4}$$

- Thus,

$$\text{Ln}(1 + i) = \ln \sqrt{2} + i \frac{\pi}{4}$$

### 2.1.2 When is the function continuous?

We care about continuity because if the function is not continuous for some values of  $z$ , then it is not differentiable for those values of  $z$ .

*On a rough intuitive level, continuous means “looks connected as you zoom in”, and differentiable means “looks like a line segment as you zoom in”. It can’t look like a line segment without looking connected.*

To have a single-value function, we will focus on the principal logarithm. Thus, for  $z \in \mathbb{C} \setminus \{0\}$ , we have that the principal logarithm is given by:

$$\text{Ln } z = \ln |z| + i \text{Arg}(z)$$

The function  $f : \mathbb{C} \setminus \{0\} \rightarrow \mathbb{C}$  given by  $f(z) = \text{Ln } z$  is continuous at all  $z$  except those along the negative real axis.

- Let’s analyze it by its components
- $\ln |z|$  is clearly continuous for all  $z \in \mathbb{C} \setminus \{0\}$  since the modulus operator,  $|.|$ , is always positive and we are already excluding the case  $z = 0$ , so the argument of  $\ln(.)$  is always greater than 0.
- The question is if the second component is continuous or not.
- **The continuity of  $\text{Arg}(z)$** 
  - $\text{Arg}(z)$  is noncontinuous on any point on the negative real axis by its definition.
  - Recall that  $\text{Arg}(z)$  is an angle such that:

$$-\pi < \text{Arg}(z) \leq \pi$$

So, the angle can never be  $-\pi$ . This is exactly what ensures that the  $\text{Arg}$  is single-valued because if we have a  $y = 0$  and  $x < 0$ , the angle is always  $\pi$  and never  $-\pi$ . Allowing  $-\pi$  will yield to have two values for this case,  $\pi$  and  $-\pi$ .

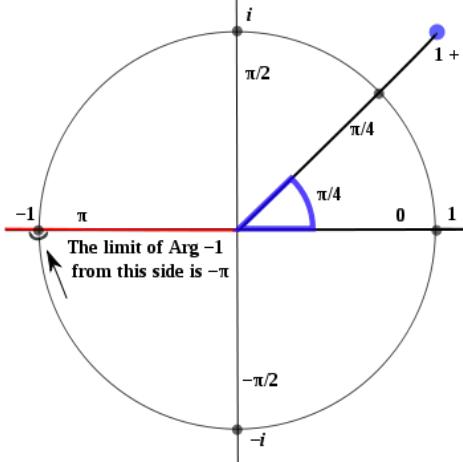


Figure 3:  $\text{Arg}(z)$

- To better understand the locations of  $\pi$  and  $-\pi$  let's take a look at Figure 3
- Now, when analyzing the limits we should take into account the sign of the angle since it indicates if measured counterclockwise ( $\text{Arg}(z) < 0$ ) or not
- Let's go to Figure 3 and see the point on the unit circle for the blue line. Let's suppose it is  $(x_0, y_0)$ . We can see that by fixing  $x = x_0$ , if approaching  $y$  from above to  $y_0$ , the limit for  $\text{Arg}(\cdot)$  is  $\pi/4$ . When approaching  $y$  from below to  $y_0$ , the limit is also  $\pi/4$
- From Figure 3, we can see that when  $x < 0$ , and  $y$  goes to zero from above, the limit of the angle is  $\pi$
- However, when  $x < 0$ , and  $y$  goes to zero from below, the limit of the angle is  $-\pi$
- Therefore,  $\text{Arg}(z)$  is discontinuous at each point on the nonpositive real axis:
  - Let  $z = x_0 + iy$  for some  $x_0 < 0$  fixed
  - If  $y$  approaches to 0 from above, then  $\text{Arg}(z) \downarrow \pi$ ,
  - Whereas if  $y$  approaches to zero from the bottom, then  $\text{Arg}(z) \uparrow -\pi$
- Thus, we have that for  $\text{Ln}(z)$  is continuous and well-defined for all  $z \in \mathbb{C} \setminus \mathbb{L}$ , in which:

$$\mathbb{L} = \{z \in \mathbb{C} : \text{Re}(z) \leq 0 \text{ and } \text{Im}(z) = 0\}$$

- Thus, we see that we need to exclude more than 0 in order to get  $\text{Ln}(z)$  to be continuous

### 2.1.3 Complex differentiability of natural logarithm

Recall that if the function is not continuous for some points, then its derivative does not exist and therefore it is not differentiable at those values. Thus, we need to restrict the set of values for  $z$  in  $f(z)$

To have a single-value continuous well-defined function, we will focus on the principal logarithm for  $z \in \mathbb{C} \setminus \mathbb{L}$ , we have that the principal logarithm is given by:

$$\text{Ln } z = \ln |z| + i \text{Arg}(z)$$

- Recall  $z = x + iy$ , so to find the derivative, we prefer to analyze

$$f(x, y) = u(x, y) + iv(x, y)$$

- For rewriting  $f(z)$  as  $f(x, y)$ , we proceed to replace  $|z|$  and  $\text{Arg}(z)$ :

$$\ln z = \ln \sqrt{x^2 + y^2} + i \tan^{-1} \frac{y}{x}$$

$$\ln z = \frac{1}{2} \ln (x^2 + y^2) + i \tan^{-1} \frac{y}{x}$$

Then, we have that:

$$f(x, y) = u(x, y) + iv(x, y)$$

$$u(x, y) = \frac{1}{2} \ln (x^2 + y^2)$$

$$v(x, y) = \tan^{-1} \frac{y}{x}$$

- Remember we have that

$$f'(z) = u_x(x, y) + iv_x(x, y)$$

Also, notice that for  $y = \arctan(x)$ :

$$\frac{d(\arctan(x))}{dx} = \frac{1}{1+x^2}$$

and that for  $y = \arctan\left(\frac{1}{x}\right)$ , we differentiate using the chain rule, which states that  $\frac{d}{dx}[f(g(x))]$  is  $f'(g(x))g'(x)$  where  $f(x) = \arctan(x)$  and  $g(x) = \frac{1}{x}$ . Thus,

$$\frac{dy}{dx} = \frac{1}{1 + \left(\frac{1}{x}\right)^2} \frac{d}{dx} \left[ \frac{1}{x} \right]$$

Thus, we have

$$u_x(x, y) = \frac{1}{2} \left( \frac{2x}{x^2 + y^2} \right)$$

$$u_x(x, y) = \frac{x}{x^2 + y^2}$$

$$v_x(x, y) = \frac{1}{1 + \left(\frac{y}{x}\right)^2} \left( -\frac{y}{x^2} \right)$$

$$v_x(x, y) = \frac{x^2}{x^2 + y^2} \left( -\frac{y}{x^2} \right)$$

$$v_x(x, y) = -\frac{y}{x^2 + y^2}$$

- Therefore we have:

$$f'(z) = u_x(x, y) + i v_x(x, y)$$

$$\frac{d(\ln z)}{dz} = \frac{x}{x^2 + y^2} + i \left( -\frac{y}{x^2 + y^2} \right)$$

$$\frac{d(\ln z)}{dz} = \frac{x - iy}{x^2 + y^2}$$

**TRICK:** Use the notable product: difference of two squares to decompose the denominator:

$$(x + iy)(x - iy) = (x^2 - (iy)^2) = (x^2 - (\sqrt{-1})^2 y^2) = (x^2 - (-1y^2)) = (x^2 + y^2)$$

Thus, we have:

$$\frac{d(\ln z)}{dz} = \frac{x - iy}{(x + iy)(x - iy)}$$

$$\frac{d(\ln z)}{dz} = \frac{x - iy}{(x + iy)(x - iy)}$$

$$\frac{d(\ln z)}{dz} = \frac{1}{(x + iy)}$$

Recall:  $z = x + iy$ , so:

$$\frac{d(\ln z)}{dz} = \frac{1}{z}$$

- So it does not exist for  $z = 0$ . Now, we only need to check if the derivative does not exist for any other values
- We proceed to check the **Cauchy-Riemann equations**:

$$u_x(x, y) = v_y(x, y)$$

$$v_x(x, y) = -u_y(x, y)$$

- We have that:

$$u_y(x, y) = \frac{1}{2} \left( \frac{2y}{x^2 + y^2} \right)$$

$$u_y(x, y) = \frac{y}{x^2 + y^2}$$

$$v_y(x, y) = \frac{1}{1 + \left(\frac{y}{x}\right)^2} \left( \frac{1}{x} \right)$$

$$v_y(x, y) = \frac{x^2}{x^2 + y^2} \left( \frac{1}{x} \right)$$

$$v_y(x, y) = \frac{x}{x^2 + y^2}$$

- Thus we have that the **Cauchy-Riemann equations** hold for any number of  $x$  and  $y$ :

$$u_x(x, y) = \frac{x}{x^2 + y^2} = v_y(x, y) = \frac{x}{x^2 + y^2}$$

$$v_x(x, y) = -\frac{y}{x^2 + y^2} = -u_y(x, y) = \frac{-y}{x^2 + y^2}$$

Thus, the derivative exists as long as  $x^2 + y^2 \neq 0$ , which is implied by  $z \neq 0$ , which is also implied by  $|z| > 0$  that comes from the condition for the natural logarithm to be defined

- Therefore,

1. the natural logarithm of  $z$  is **complex differentiable** for all  $z \in \mathbb{C} \setminus \mathbb{L}$
2. the natural logarithm of  $z$  is **analytic-regular-holomorphic** for all  $z \in \mathbb{C} \setminus \mathbb{L}$

## 2.2 The natural exponential of a complex number

In complex analysis, the exponential of complex number  $z$  is

$$f(z) = e^z$$

- When is it undefined? If  $z \in \mathbb{R}$ , we know it is always defined
- Would that change now that  $z \in \mathbb{C}$ ?
- Recall  $z = x + iy$ , so

$$f(z) = e^{x+iy}$$

$$f(z) = e^x e^{iy}$$

From **Euler's Theorem**, since  $y \in \mathbb{R}$ :

$$e^{iy} = \cos y + i \sin y$$

Sine and Cosine are defined over every real number and the  $e^x$  is defined for any  $x \in \mathbb{R}$ . Thus  $f(z)$  is defined for any  $x, y \in \mathbb{R}$

- Thus, the natural exponential function of a complex number is always defined for any  $z \in \mathbb{C}$

### 2.2.1 Complex differentiability

For any  $z \in \mathbb{C}$ , we have that the natural exponential function is a well-defined function:

$$f(z) = e^z$$

- Recall  $z = x + iy$ , so to find the derivative, we prefer to analyze

$$f(x, y) = u(x, y) + iv(x, y)$$

- For rewriting  $f(z)$  as  $f(x, y)$ , we proceed to replace  $z$ :

$$f(x, y) = e^x e^{iy}$$

From **Euler's Theorem**, since  $y \in \mathbb{R}$ :

$$e^{iy} = \cos y + i \sin y$$

Then, we have that:

$$f(x, y) = e^x (\cos y + i \sin y)$$

$$f(x, y) = e^x (\cos y) + i(e^x \sin y)$$

$$f(x, y) = u(x, y) + iv(x, y)$$

$$u(x, y) = e^x (\cos y)$$

$$v(x, y) = e^x \sin y$$

- Remember we have that

$$f'(z) = u_x(x, y) + iv_x(x, y)$$

Also, notice that:

$$\frac{d}{dx} \sin(x) = \cos(x)$$

$$\frac{d}{dx} \cos(x) = -\sin(x)$$

Thus, we have

$$u_x(x, y) = e^x (\cos y)$$

$$v_x(x, y) = e^x \sin y$$

- Therefore we have:

$$f'(z) = u_x(x, y) + i v_x(x, y)$$

$$\frac{d(e^z)}{dz} = e^x (\cos y) + i e^x \sin y$$

$$\frac{d(e^z)}{dz} = e^x (\cos y + i \sin y)$$

Thus, we have:

$$\frac{d(e^z)}{dz} = e^x e^{iy}$$

$$\frac{d(e^z)}{dz} = e^{x+iy}$$

Recall:  $z = x + iy$ , so:

$$\frac{d(e^z)}{dz} = e^z$$

- As the derivative of the natural exponential function, it is defined for all  $z \in \mathbb{C}$ . To double-check, we proceed to check the Cauchy-Riemann equations
- We proceed to check the **Cauchy-Riemann equations**:

$$u_x(x, y) = v_y(x, y)$$

$$v_x(x, y) = -u_y(x, y)$$

- We have that:

$$u_y(x, y) = e^x (-\sin y)$$

$$v_y(x, y) = e^x (\cos y)$$

- Thus we have that that the **Cauchy-Riemann equations** hold for any number of  $x$  and  $y$ :

$$u_x(x, y) = e^x(\cos y) = v_y(x, y) = e^x(\cos y)$$

$$v_x(x, y) = e^x(\sin y) = -u_y(x, y) = -e^x(-\sin y)$$

Thus, the derivative exists for any  $\in \mathbb{C}$

- Therefore,

1. the natural exponential of  $z$  is **complex differentiable** for all  $z \in \mathbb{C}$
2. the natural exponential of  $z$  is **analytic-regular-holomorphic** for all  $z \in \mathbb{C}$

## 2.3 The complex valued function: $(1 - \alpha z)^{-1}$

Consider  $\alpha \in \mathbb{C}$ , so as general as possible. This function is useful for analyzing and *AR* process with  $\rho = \alpha$ :

$$f(z) = \frac{1}{1 - \alpha z}$$

- When is it undefined? If  $z \in \mathbb{R}$ , we know it is not defined when  $z = \frac{1}{\alpha}$
- Would that change now that  $z \in \mathbb{C}$ ?
- No, since the only way  $f(z)$  is not defined is when the denominator is 0, which only occurs when  $z = \frac{1}{\alpha}$

### 2.3.1 Complex differentiability

For any  $z \in \mathbb{C} \setminus \{\alpha^{-1}\}$ , we have the following well-defined function:

$$f(z) = \frac{1}{1 - \alpha z}$$

- As usual, we would like to find a way to rewrite  $f(z)$  as

$$f(x, y) = u(x, y) + iv(x, y)$$

However, this time it is not easy!

- **TRICK:** to analyze the differentiability of  $f(z)$  we will use the natural exponential and natural logarithm:

$$f(z) = e^{\ln f(z)}$$

Thus,

$$f(z) = e^{\ln \frac{1}{1 - \alpha z}}$$

$$f(z) = e^{\ln 1 - \ln(1 - \alpha z)}$$

$$f(z) = e^{-\ln(1 - \alpha z)}$$

- The natural exponential function,  $e^z$ , is complex differentiable for all  $z \in \mathbb{C}$
- However, the natural logarithm,  $\ln z$ , is not complex differentiable for all  $z \in \mathbb{C}$
- so, we only need to analyze the differentiability of  $\ln(1 - \alpha z)$
- To get a single-value function, we analyze the principal logarithm:

$$\text{Ln}(1 - az) = \ln|1 - az| + i \operatorname{Arg}(1 - az)$$

Thus,  $\ln(1 - \alpha z)$  is **defined** if

$$|1 - \alpha z| > 0$$

Recall  $|.|$  is always positive. Thus we only need:

$$1 \neq \alpha z$$

That is,

$$z \neq \frac{1}{\alpha}$$

- To find the derivative we can follow the same analysis as in 2.1.2 by analyzing:

$$\ln(z^*) \equiv \ln(1 - \alpha z)$$

, in which:

$$z^* = (1 - \alpha x) + i(-\alpha y)$$

- We know that  $\text{Ln}(z^*)$  is not continuous nor well defined for

$$\mathbb{L}_{z^*} = \{z^* \in \mathbb{C} : \operatorname{Re}(z^*) \leq 0 \text{ and } \operatorname{Im}(z^*) = 0\}$$

Recall that real part of  $z^*$  is  $1 - \alpha x$ . So,

$$\operatorname{Re}(z^*) \leq 0$$

$$1 - \alpha x \leq 0$$

$$x \geq \frac{1}{\alpha}$$

Recall that  $z = x + iy$ , so the real part of  $z$  is  $x$ . From above we have that:

$$x \geq \frac{1}{\alpha}$$

$$\operatorname{Re}(z) \geq \frac{1}{\alpha}$$

The imaginary part being zero in  $z^*$  is the same as the imaginary part being zero in  $z$ . So, by rewriting  $\mathbb{L}$  in terms of  $z$ , we have that:

$$\mathbb{L} = \{z \in \mathbb{C} : \operatorname{Re}(z) \geq \frac{1}{\alpha} \text{ and } \operatorname{Im}(z) = 0\}$$

Thus,  $\ln(1 - \alpha z)$  is well defined and continuous for all  $z \in \mathbb{C} \setminus \mathbb{L}$

- We know that  $\ln z^*$  is complex differentiable for all  $z^* \in \mathbb{C} \setminus \mathbb{L}_{z^*}$ .
- Thus, we have that  $\ln(1 - \alpha z)$  is complex differentiable for all  $z \in \mathbb{C} \setminus \mathbb{L}$
- That is,  $\ln(1 - \alpha z)$  is analytic for all  $z \in \mathbb{C} \setminus \mathbb{L}$
- Therefore,  $f(z) = \frac{1}{1 - \alpha z}$  is analytic for all  $z \in \mathbb{C} \setminus \mathbb{L}$

### 2.3.2 Taylor Expansion

For any  $z \in \mathbb{C} \setminus \mathbb{L}$ , we have the following function:

$$f(z) = \frac{1}{1 - \alpha z}$$

- is well defined
- is analytic
- As  $f(z)$  is analytic everywhere inside  $\mathbb{C} \setminus \mathbb{L}$ , we can use Taylor's Theorem to express it as a power series with  $z_0 \in \mathbb{C} \setminus \mathbb{L}$ :

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!} (z - z_0)^n$$

where the series converges on any disk  $|z - z_0| < r$  contained in  $\mathbb{C} \setminus \mathbb{L}$

- Recall that  $r$  is the radius of convergence and we would like to find it so that we know that power series converges

### 2.3.3 Convergence of the Taylor Expansion

- We are interested in finding  $r$  so we will use the Ratio Test
- First we need to choose  $z_0 \in \mathbb{C} \setminus \mathbb{L}$
- Since  $0 \in \mathbb{C} \setminus \mathbb{L}$ , we select  $z_0 = 0$
- Now, we can proceed to use the ratio test, but we can use the same trick as when analyzing the differentiability of  $f(z)$ .

- So to find  $r$  when  $z_0 = 0$  for  $f(z)$  is the same as finding it for

$$g(z) = \ln(1 - \alpha z)$$

- We know that  $g(z)$  is analytic everywhere inside  $\mathbb{C} \setminus \mathbb{L}$ , so we can use the Taylor expansion around  $z_0 = 0$ :

$$\ln(1 - \alpha z) = -\frac{\alpha}{1}z^1 - \frac{\alpha^2}{2}z^2 - \frac{\alpha^3}{3}z^3 - \frac{\alpha^4}{4}z^4 \dots - \frac{\alpha^n}{n}z^n + \dots$$

So,

$$\ln(1 - \alpha z) = \sum_{n=1}^{\infty} -\frac{\alpha^n}{n}z^n$$

- To apply the Ratio test, we need to find the limit of the absolute ratios of consecutive terms,  $L$ :

$$L = \lim_{n \rightarrow \infty} \left| \frac{\frac{\alpha^{n+1}}{n+1}z^{n+1}}{\frac{\alpha^n}{n}z^n} \right|$$

$$L = \lim_{n \rightarrow \infty} \left| \frac{\alpha n}{n+1} z \right|$$

I can take out  $|\alpha z|$  since it does not depend on  $n$ , so it is not affected by the limit:

$$L = |\alpha z| \lim_{n \rightarrow \infty} \left| \frac{n}{n+1} \right|$$

Now, I can use L'Hôpital's rule<sup>5</sup>

$$L = |\alpha z| \lim_{n \rightarrow \infty} \left| \frac{1}{1} \right|$$

$$L = |\alpha z| \lim_{n \rightarrow \infty} |1|$$

$$L = |\alpha z| |1|$$

$$L = |\alpha z|$$

So, the limit,  $L$ , exists. Recall for convergence, we need

$$L < 1$$

---

<sup>5</sup> :  $\lim_{x \rightarrow c} f(x) = \lim_{x \rightarrow c} g(x) = 0$  or  $\pm \infty$ , and  $g'(x) \neq 0$  for all  $x$  in  $I$  with  $x \neq c$ , and  $\lim_{x \rightarrow c} \frac{f'(x)}{g'(x)}$  exists, then  $\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c} \frac{f'(x)}{g'(x)}$

, which can only be achieved if:

$$|\alpha z| < 1$$

That is, if:

$$|\alpha||z| < 1$$

$$|z| < |\alpha|^{-1}$$

- Thus, we have that the series  $g(z)$  and therefore the series  $f(z)$  converge for any  $z$  such that  $|z - 0| < |\alpha|^{-1}$ .
- One big difference between this function and  $\frac{1}{1-z}$  is that the convergence can be achieved when  $|z| = 1$  (i.e.,  $z$  is on the **unit circle**) as long as

$$|a| < 1$$

This is an extremely important result as we will see when analyzing the lag operator.

### 3 Lag polynomial

- A lag polynomial is a polynomial in which the variable is the **lag operator** denoted by  $L$
- The **lag operator** or **backshift operator**,  $B$ , operates on an element of a time series to produce the previous element:

$$Ly_t = y_{t-1}$$

- The lag operator (as well as backshift operator) can be raised to arbitrary integer powers so that:

$$L^2y_t = y_{t-2}$$

$$L^{-1}y_t = y_{t+1}$$

- The lag polynomial will be useful when analyzing the properties of a time series.
- For

$$\sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

we can express it as:

$$= \sum_{j=-\infty}^{\infty} \psi_j L^j X_t$$

$$= \psi(L)X_t$$

So,  $\psi(L)$  is the lag polynomial.

- As it is a polynomial, we will use all the tools learned in Section 1 and 2. So we would like to see for instance if we can apply Taylor's Theorem:

$$\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j = \frac{1}{1 - \rho L}$$

- As the time series is a random variable, we will need to use the concept of convergence of random variables that we will learn in Section 4.
- The idea is to find the conditions for the lag polynomial such that a time series is **stationary**
- In short, we will try to express a times series in terms of a lag polynomial. As the time series is a random variable, we will need to analyze the convergence in probability or in moments. As we will see in Section 5, certain features of the moments of the variable can tell us if the variables is **stationary** (e.g., second moment convergence is a necessary but not a sufficient condition for the series to be **stationary**). Then, we will come up with what conditions we need for the lag polynomial such that the time series is **stationary**

## 4 Convergence of Random Variables

- In some situations, we would like to see if a sequence of random variables  $X_1, X_2, X_3, \dots$  "converges" to a random variable  $X$
- That is, we would like to see if  $X_n$  gets closer and closer to  $X$  in some sense as  $n$  increases
- For example, suppose that we are interested in knowing the value of a random variable  $X$ , but we are not able to observe  $X$  directly
- Instead, you can do some measurements and come up with an estimate of  $X$  : call it  $X_1$ .
- You then perform more measurements and update your estimate of  $X$  and call it  $X_2$ .
- You continue this process to obtain  $X_1, X_2, X_3, \dots$  Your hope is that as  $n$  increases, your estimate gets better and better.
- That is, you hope that as  $n$  increases,  $X_n$  gets closer and closer to  $X$ . In other words, you hope that  $X_n$  converges to  $X$ .
- In fact, limit theorems (the weak law of large numbers (WLLN) and the central limit theorem (CLT)) use the concept of convergence.
- The WLLN states that the average of a large number of i.i.d. random variables converges in probability to the expected value.
- The CLT states that the normalized average of a sequence of i.i.d. random variables converges in distribution to a standard normal distribution
- In this section, we will develop the theoretical background to study the convergence of a sequence of random variables in more detail
- In particular, we will define different types of convergence.

## 4.1 Convergence of a Sequence of Numbers

- If we have a sequence of real numbers  $a_1, a_2, a_3, \dots$ , we can ask whether the sequence converges
- For example, the sequence

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots, \frac{n}{n+1}, \dots$$

is defined as

$$a_n = \frac{n}{n+1}, \quad \text{for } n = 1, 2, 3, \dots$$

We say that a sequence  $a_1, a_2, a_3, \dots$  converges to a limit  $L$  if  $a_n$  approaches  $L$  as  $n$  goes to infinity.

- That is, a sequence  $a_1, a_2, a_3, \dots$  converges to a limit  $L$  if

$$\lim_{n \rightarrow \infty} a_n = L$$

That is, for any  $\epsilon > 0$ , there exists an  $N \in \mathbb{N}$  such that

$$|a_n - L| < \epsilon, \quad \text{for all } n > N$$

## 4.2 Types of Convergence for Sequences of Random Variables

- Unlike the non-random variables, there are different types of convergence when dealing with random variables
- Consider a sequence of random variables  $X_1, X_2, X_3, \dots$ , i.e.,  $\{X_n, n \in \mathbb{N}\}$ .
- There are four types of convergence:
  - Convergence in distribution
  - Convergence in probability
  - Convergence in r-th moment
  - Almost sure convergence
- A sequence might converge in one sense but not in another one.
- Some of these convergence types are "stronger" than others and some are "weaker."
- Figure 4 summarizes how these types of convergence are related

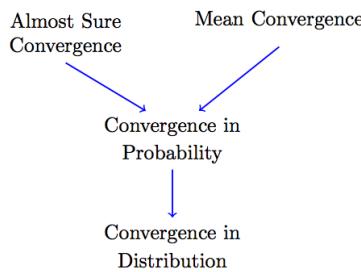


Figure 4: Relation between different types of convergence

### 4.3 Convergence in Probability

- Let  $\{X_n : n \geq 1\}$  be a random sequence of vectors on  $\mathbb{R}^k$
- That is,  $k$  elements in each vector. That is,  $k$  random variables.
- When  $k = 1$ , we are in the simple case of a sequence of random variables (Each vector is one dimensional, which means that each vector is only one variable)
- Let  $X$  be a random vector on  $\mathbb{R}^k$
- $X_n$  is said to converge in probability to  $X$ , denoted by:

$$X_n \xrightarrow{\text{P}} X,$$

if, as  $n \rightarrow \infty$

$$\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$$

for any

$$\varepsilon > 0$$

where  $|\cdot|$  is the usual Euclidean norm (also called the  $L^2$  norm), which gives the ordinary distance from the origin to the point  $X$ —a consequence of the Pythagorean theorem:

$$|X| = \sqrt{x_1^2 + \cdots + x_k^2}$$

, in which  $\sqrt{\cdot}$  denote the positive root

- Equivalently,  $X_n$  is said to converge in probability to  $X$ , if

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

for any

$$\varepsilon > 0$$

- Equivalently,  $X_n$  is said to converge in probability to  $X$ , if for any  $\delta, \varepsilon > 0$ , there exists  $N \in \mathbb{N}$  such that,

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \delta$$

for all  $n \geq N$

- It is worth pointing out that **Markov's inequality** is very useful in many proofs involving convergence in probability.

### 4.3.1 Some intuition

- Let's say  $X$  is a constant, something finite
- Let's say  $X_n$  is simple sum:

$$X_n = \sum_{i=1}^n x_i$$

- As  $n$  goes to infinity, we have an infinite sum
- So, we would like to know if the infinity sum converges
- That is, we are interesting in knowing if the infinite sum is finite or not
- Since, the variable inside the sum is a random variable, now we have to think on what kind of convergence are we checking
- Recall, we still want the infinite sum to converge to something finite.
- We cannot be sure of it because we have a random variable!
- However we can talk in probabilities.
- Thus, we would like to check if the infinite sum converges to something finite as  $n$  goes to infinity with probability close to 1.
- In other words, we would like to check if the probability of the infinite sum being different than something finite goes to 0 as  $n$  goes to infinity.
- That is exactly the definition of convergence in probability!
- You might ask why the Euclidean norm?
- Well, it is just part of the definition. If we are checking the difference between two elements. This could be either positive or negative.
- Thus, the Euclidean norm appears there basically to ensure that the distance between the two elements is positive
- That is why  $\epsilon$  is any positive real number
- We are just saying, the probability of the distance between two elements being greater than any positive number should go to zero as  $n$  increases
- In other words, the probability of the distance between two elements being equal to zero should go to 1 as  $n$  increases.

### 4.3.2 Markov's Inequality

- For any random variable  $X$ ,

$$\mathbb{P}(|X| > \varepsilon) \leq \frac{\mathbb{E}[|X|^q]}{\varepsilon^q}, \quad \forall q, \varepsilon > 0$$

where  $|\cdot|$  is the usual Euclidean norm

- If the RHS goes to 0 as  $n$  goes to infinity, then the LHS does it too.
- Thus, we can see directly how the convergence in the first moment,  $q = 1$ , implies convergence in probability
- That is why Markov's inequality is useful in many proofs involving convergence in probability.

## 4.4 Convergence in probability of a Infinite Sum of Random Variables: Absolutely Summable

**Theorem 3** If  $\{X_t\}$  is any sequence of random variables such that

$$\sup_t E |X_t| < \infty,$$

and if

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty,$$

then the series

$$Y_t = \psi(L)X_t = \sum_{j=-\infty}^{\infty} \psi_j L^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

converges.

### Proof

- So, we would like to know if the infinity sum converges.
- That is, we are interesting in knowing if the infinite sum is finite or not.
- Since, the variable inside the sum is a random variable, now we have to think on what kind of convergence are we checking.
- Recall, we still want the infinite sum to converge to something finite.
- We cannot be sure of it because we have a random variable
- However we can talk in probabilities
- Thus, we would like to check if the infinite sum converges to something finite as  $n$  goes to infinity with probability close to 1.
- That is exactly the definition of convergence in probability
- Thus, we are checking for convergence in probability
- More precisely for this exercise, we are checking if, as  $n \rightarrow \infty$

$$\mathbb{P} \left( \left| \sum_{j=-n}^n \psi_j X_{t-j} - X \right| > \varepsilon \right) \rightarrow 0$$

for a finite constant  $X$  and any

$$\varepsilon > 0$$

- By Markov's inequality we know that:

$$\mathbb{P} \left( \left| \sum_{j=-n}^n \psi_j X_{t-j} - X \right| > \varepsilon \right) \leq \frac{\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} - X \right|^q \right]}{\varepsilon^q}, \quad \forall q, \varepsilon > 0$$

- So, we will take advantage of the fact that we are given that:

$$\sup_t E |X_t| < \infty$$

and

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

which are clearly first moments conditions.

- Thus, we will analyze the behavior of the RHS of Markov's inequality as  $n$  goes to infinity by setting  $q = 1$
- Recall that if we find that the RHS goes to zero, then the LHS goes to zero and we have that the infinite sum converge to something finite with a probability close to one
- From the RHS of Markov's inequality, we have:

$$\frac{\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} - X \right| \right]}{\varepsilon}$$

Since the denominator is not related to  $n$ , let's focus on the numerator:

$$\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} - X \right| \right]$$

By Triangle inequality, we have that:

$$\left| \sum_{j=-n}^n \psi_j X_{t-j} + (-X) \right| \leq \left| \sum_{j=-n}^n \psi_j X_{t-j} \right| + |-X|$$

Since the absolute value always gives a positive value, we have that  $|a| = |-a|$ ,

$$\left| \sum_{j=-n}^n \psi_j X_{t-j} + (-X) \right| \leq \left| \sum_{j=-n}^n \psi_j X_{t-j} \right| + |X|$$

By taking expectation on both sides, we have that:

$$\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} + (-X) \right| \right] \leq \mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} \right| + |X| \right]$$

Since the expectation is linear:

$$\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} + (-X) \right| \right] \leq \mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} \right| \right] + \mathbb{E} [|X|]$$

So, if we make the RHS goes to zero as  $n$  goes to infinity, we are done. So let's focus on the RHS:

$$\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} \right| \right] + \mathbb{E} [|X|]$$

Recall that  $X$  denotes something finite and it can be anything as long as  $X < \infty$ . As  $n$  goes to infinity we have:

$$\mathbb{E} \left[ \left| \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right| \right] + \mathbb{E} [|X|]$$

So, for this to go to zero we need:

$$\mathbb{E} \left[ \left| \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right| \right] = -\mathbb{E} [|X|]$$

, which is impossible. Thus for this case, the triangle inequality is not helpful. However, we could use the **Reverse Triangle Inequality**:

$$\left| \sum_{j=-n}^n \psi_j X_{t-j} \right| - |X| \leq \left| \sum_{j=-n}^n \psi_j X_{t-j} - X \right|$$

By taking expectation on both sides, we have that:

$$\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} \right| \right] - \mathbb{E} [|X|] \leq \mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} + X \right| \right]$$

More precisely, since  $X$  is anything finite, let  $X$  be such that:

$$\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} + -X \right| \right] = \mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} \right| \right] - \mathbb{E} [|X|]$$

So, for this to go to zero we need:

$$\mathbb{E} \left[ \left| \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right| \right] = \mathbb{E} [|X|]$$

Given that  $\mathbb{E} [|X|]$  is anything finite such that the equality in the reserves triangle inequality holds, we are actually only requiring

$$\mathbb{E} \left[ \left| \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right| \right] = \text{Finite}$$

That is,

$$\mathbb{E} \left[ \left| \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right| \right] < \infty$$

By **Triangle inequality**:

$$\mathbb{E} \left[ \left| \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right| \right] \leq \mathbb{E} \left[ \sum_{j=-\infty}^{\infty} |\psi_j X_{t-j}| \right]$$

Thus, we only need to check if the RHS of the above is finite:

$$\mathbb{E} \left[ \sum_{j=-\infty}^{\infty} |\psi_j X_{t-j}| \right] < \infty$$

Recall that  $|ab| = |a||b|$ , so we have that

$$\mathbb{E} \left[ \sum_{j=-\infty}^{\infty} |\psi_j X_{t-j}| \right] = \mathbb{E} \left[ \sum_{j=-\infty}^{\infty} |\psi_j| |X_{t-j}| \right]$$

Now, we have to analyze:

$$\mathbb{E} \left[ \sum_{j=-\infty}^{\infty} |\psi_j| |X_{t-j}| \right]$$

Since,  $\psi_j$  is not random:

$$= \sum_{j=-\infty}^{\infty} |\psi_j| \mathbb{E}[|X_{t-j}|]$$

We would like to have that  $X_t$  is i.i.d with mean  $\mu < \infty$ , so  $\mathbb{E}[|X_{t-j}|] = \mathbb{E}[|X_t|] = \mu$  and we could factor  $\mathbb{E}[|X_{t-j}|]$  out of the sum. If so, we would easily have that the above is finite given that  $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ . This example is a more general case in which we do not have i.i.d, but we do have that:

$$\sup_t \mathbb{E}[|X_t|] < \infty$$

, which means that the max possible value for the expectation of  $X_t$  for any  $t$  is finite. Thus, we can assume as an extreme case that all the  $X_t$  take that value. So we have that:

$$\sum_{j=-\infty}^{\infty} |\psi_j| \mathbb{E}[|X_{t-j}|] \leq \sum_{j=-\infty}^{\infty} |\psi_j| \sup_t \mathbb{E}[|X_t|]$$

so we can factor out the expectation of the infinite sum (since we are assuming all the expectations are taking the supreme value, so it is no longer indexed to the infinite sum):

$$\sum_{j=-\infty}^{\infty} |\psi_j| \mathbb{E}[|X_{t-j}|] \leq \sup_t \mathbb{E}[|X_t|] \sum_{j=-\infty}^{\infty} |\psi_j|$$

Recall by assumption we have that:

$$\sup_t E|X_t| < \infty$$

and

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

Therefore, we have:

$$\sup_t \mathbb{E}[|X_t|] \sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

which implies

1.

$$\sum_{j=-\infty}^{\infty} |\psi_j| \mathbb{E}[|X_{t-j}|] < \infty$$

2.

$$\mathbb{E} \left[ \sum_{j=-\infty}^{\infty} |\psi_j X_{t-j}| \right] < \infty$$

3.

$$\mathbb{E} \left[ \left| \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right| \right] < \infty$$

$$\mathbb{E} \left[ \left| \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right| \right] = \text{Finite}$$

4. As long as  $X$  is something finite such that:

$$\mathbb{E} \left[ \left| \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right| \right] = \mathbb{E}[|X|]$$

5. Then, we have, as  $n$  goes to infinity:

$$E \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} \right| \right] - \mathbb{E}[|X|] = 0$$

6. So, as  $n$  goes to infinity

$$\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} \right| \right] - \mathbb{E}[|X|] \rightarrow 0$$

7. So, as  $n$  goes to infinity

$$\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} \right| - |X| \right] \rightarrow 0$$

8. So, as  $n$  goes to infinity

$$\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} - X \right| \right] \rightarrow 0$$

9. So, as  $n$  goes to infinity

$$\frac{\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} - X \right| \right]}{\varepsilon} \rightarrow 0$$

10. So, as  $n$  goes to infinity

$$\mathbb{P} \left( \left| \sum_{j=-n}^n \psi_j X_{t-j} - X \right| > \varepsilon \right) \rightarrow 0$$

- Thus, we have that the infinite sum converges in probability to something finite

#### 4.5 Convergence in $q^{\text{th}}$ moment

- Let  $\{X_n : n \geq 1\}$  be a sequence of random vectors with  $k$  elements
- Let  $X$  be a random vectors on  $\mathbb{R}^k$
- $X_n$  is said to converge in  $q^{\text{th}}$  moment to  $X$  such that  $q \geq 1$ , denoted:

$$X_n \xrightarrow{L^q} X$$

if, as  $n \rightarrow \infty$

$$\mathbb{E} [|X_n - X|^q] \rightarrow 0$$

- By **Jensen's inequality** we can show that convergence in higher moments imply convergence in lower moments

#### 4.5.1 Convergence in 1st moment

- $X_n$  is said to converge in 1st moment to  $X$  such denoted:

$$X_n \xrightarrow{L^1} X$$

if, as  $n \rightarrow \infty$

$$\mathbb{E}[|X_n - X|] \rightarrow 0$$

- Let's say  $X$  is something finite, then we have that  $X_n$  converges in first moment to something finite
- Thus, we have that:

$$\mathbb{E}[|X_n|] < \infty$$

as  $n$  goes to infinity.

- Recall that we define expectations by sums or integrals only if they are **absolutely summable** or integrable.
- Therefore, we have that a random variable  $Y$  has expectation only if  $|Y|$  has expectation:

$$\mathbb{E}[Y] = \sum_y y p(y) < \infty$$

if and only if

$$\mathbb{E}[|Y|] = \sum_y |y| p(y) < \infty$$

- Thus, we have that

$$\mathbb{E}[X_n] < \infty$$

as  $n$  goes to infinity.

#### 4.5.2 Convergence in 2nd moment

- This convergence is also known as **mean square convergence** or **convergence on  $L^2$**
- $X_n$  is said to converge in 2nd moment to  $X$  such denoted:

$$X_n \xrightarrow{L^2} X$$

if, as  $n \rightarrow \infty$

$$\mathbb{E}[|X_n - X|^2] \rightarrow 0$$

- Let's say  $X$  is something finite, then we have that  $X_n$  converges in second moment to something finite.
- Thus, we have that the second moment of  $X_n$  exists.
- This convergence carries out a great implication: **the variance exist.**
- 2nd moment convergence implies 1st moment convergence. Thus, the first moment is also finite.
- The variance of  $X_n$  is given by:

$$\text{Var}(X_n) = E[(X_n)^2] - (E[X_n])^2$$

By definition of the expectation, we know that the expectation of  $X_n^2$  exists if and only if the expectation of  $|X_n|^2$  exists. Thus, we have that the first component of the variance exists. The second component exists because convergence of second moment implies the expectation of  $|X_n|$  exists which implies that expectation of  $X_n$  exists and so the square of this expectation.

## 4.6 Convergence in second moment of a Infinite Sum of Random Variables: Absolutely Summable

**Theorem 4** *If  $\{X_t\}$  is any sequence of random variables such that*

$$\sup_t E |X_t|^2 < \infty$$

*and if*

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty,$$

*then, the series*

$$Y_t = \psi(L)X_t = \sum_{j=-\infty}^{\infty} \psi_j L^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

*converges in mean square.*

### 4.6.1 Usefulness of this exercise

- This example is similar to the one in the previous section.
- It differs on the fact that now we are given that the second moment of  $X_t$  is finite instead of the first moment.
- We know that the second moment being finite implies that the first moment is also finite.
- Thus, given the above conditions, this Theorem states that  $Y_t$  has a finite second and first moment even though it is an infinite sum of random variables.

### 4.6.2 Proof

- In this case we want to show that the infinite sum converges in mean square.
- That is, we want to show that the infinite sum converges in second moment to something finite.
- We will denote that something finite as  $K$ .
- Thus, we want to check if the the following holds:

$$\mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} - K \right|^2 \right] \rightarrow 0$$

as  $n$  goes to infinity.

- This trick is the same use in Proposition 3.1.1. in Brockwell and Davis (1991)
- Let's say  $K$  is the finite variance if it exists. So we are actually only requiring

$$\mathbb{E} \left[ \left| \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right|^2 \right] = \text{Finite}$$

Now recall the definition of having a finite second moment:  $E|Z|^2 < \infty$ .

Notice that  $E|Z|^2 = E(Z)^2$  only for  $Z \in \mathbb{R}$ .

If  $Z \in \mathbb{C}$ , then we cannot say that because of the definition of a modulus.

I will explain it in more detail in the following lines.

Thus, we are looking for:

$$\mathbb{E} \left[ \left| \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \right|^2 \right] < \infty$$

Again, we would love to use here the beautiful **Triangle inequality**, but this time we have the Euclidean norm to the power of 2, so we cannot.

For this case, we will use the beautiful property of the  $|.|^2$

- Recall that the **Euclidean norm** or  $L - 2$  norm (or distance) of a k-dimensional vector  $x$  is given by

$$|x| = \sqrt{\sum_{i=1}^k x_i^2}$$

, in which  $x_i$  denotes an element of the k-dimensional vector. Thus, the square of the  $L-2$  norm is:

$$|x|^2 = \left( \sqrt{\sum_{i=1}^k x_i^2} \right)^2 = \sum_{i=1}^k x_i^2$$

, so it is like cancelling the absolute value.

However, that is only useful if  $x$  is a Real number.

If we have that  $x$  is a complex number, then the modulus is such that it only contains the real parts of the complex number:

$$|z| = \sqrt{\sum_{i=1}^k |z_i|^2} = \sqrt{z_1\bar{z}_1 + \cdots + z_k\bar{z}_k}$$

since:

$$|z|^2 = z\bar{z}$$

- Notice that in this case, we have a one dimensional vector.

Let's allow for the random variable and for the coefficients to be complex numbers.

So using the definition of Euclidean norm is the  $L - 2$  norm, we have that a direct application of the above:

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} \right|^2 \right] &= \mathbb{E} \left[ \left( \sum_{j=-n}^n \psi_j X_{t-j} \right) \left( \sum_{j=-n}^n \bar{\psi}_j \bar{X}_{t-j} \right) \right] \\ &= \mathbb{E} \left[ \sum_{j=-n}^n \sum_{k=-n}^n \psi_j \bar{\psi}_k (X_{t-j} \bar{X}_{t-k}) \right] \\ &= \sum_{j=-n}^n \sum_{k=-n}^n \psi_j \bar{\psi}_k \mathbb{E}[X_{t-j} \bar{X}_{t-k}] \end{aligned}$$

Notice that the last term looks like a covariance so we would like to use the **Covariance Inequality**. It is not a covariance because it does not include the means.

- The **Covariance Inequality** for real random variables states that:

$$(\text{Cov}(Y, X))^2 \leq \text{Var}(X) \text{Var}(Y)$$

, which comes directly by applying **Cauchy-Schwarz inequality**, which can be written using only the inner product:

$$|\langle X, Y \rangle|^2 \leq \langle X, X \rangle \langle Y, Y \rangle$$

or by taking the square root of both sides of the above inequality, the Cauchy-Schwarz inequality it can be written using the norm and inner product :

$$|\langle X, Y \rangle| \leq \|X\| \|Y\|.$$

when defining the inner product as:

$$\langle X, Y \rangle = E(XY)$$

and therefore the norm as:

$$\|X\| = \sqrt{\langle X, X \rangle} = \sqrt{E(XX)}$$

Specifically, it gives us:

$$|E(XY)|^2 \leq E(X^2) E(Y^2)$$

but we can easily redefine  $x$  and  $y$  such that  $\mu = E(X)$  and  $\nu = E(Y)$ , then:

$$\begin{aligned} |\text{Cov}(X, Y)|^2 &= |\mathbf{E}((X - \mu)(Y - \nu))|^2 \\ &= |\langle X - \mu, Y - \nu \rangle|^2 \\ &\leq \langle X - \mu, X - \mu \rangle \langle Y - \nu, Y - \nu \rangle \\ &= E((X - \mu)^2) E((Y - \nu)^2) \\ &= \text{Var}(X) \text{Var}(Y) \end{aligned}$$

For the case of **complex random variables**, we also proceed to use the **Cauchy-Schwarz inequality** when defining the inner product as<sup>6</sup>:

$$\langle Z, W \rangle = E(ZW)$$

so,

$$\langle Z, \bar{Z} \rangle = E(Z\bar{Z}) = E(|Z|^2)$$

and the norm as:

$$\|Z\| = \sqrt{\langle Z, \bar{Z} \rangle} = \sqrt{E(Z\bar{Z})} = \sqrt{E(|Z|^2)}$$

In, this case we have:

$$|E[Z\bar{W}]|^2 \leq E[|Z|^2] E[|W|^2]$$

So,

$$|E[Z\bar{W}]| \leq \sqrt{E[|Z|^2] E[|W|^2]}$$

---

<sup>6</sup>It can also be shown using Triangle inequality and Hölder's inequality

- Recall we want to know if the following is finite (converges):

$$\sum_{j=-n}^n \sum_{k=-n}^n \psi_j \bar{\psi}_k \mathbb{E}[X_{t-j} \bar{X}_{t-k}]$$

We know that the above converges if the following converges:

$$\sum_{j=-n}^n \sum_{k=-n}^n |\psi_j \bar{\psi}_k| \left| \mathbb{E}[X_{t-j} \bar{X}_{t-k}] \right|$$

Now by applying the **Cauchy-Schwarz inequality** for complex random variables, we have that:

$$\sum_{j=-n}^n \sum_{k=-n}^n |\psi_j \bar{\psi}_k| \left| \mathbb{E}[X_{t-j} \bar{X}_{t-k}] \right| \leq \sum_{j=-n}^n \sum_{k=-n}^n |\psi_j \bar{\psi}_k| \sqrt{\mathbb{E}[|X_{t-j}|^2] \mathbb{E}[|X_{t-k}|^2]}$$

We would like to have that  $X_t$  is i.i.d with  $E|X_t|^2 = \theta < \infty$ , so  $\mathbb{E}[|X_{t-j}|^2] = \mathbb{E}[|X_{t-k}|^2] = \theta$  and we could factor it out of the sum.

However, we do not have an i.i.d process, since this is a more general case, but we do have that:

$$\sup_t E[|X_t|^2] < \infty$$

, which means that the max possible value for  $E|X_t|^2$  for any  $t$  is finite.

Thus, we can assume as an extreme case that all the  $E|X_t|^2$  take that extreme value value:

$$\begin{aligned} \sum_{j=-n}^n \sum_{k=-n}^n |\psi_j \bar{\psi}_k| \sqrt{\mathbb{E}[|X_{t-j}|^2] \mathbb{E}[|X_{t-k}|^2]} &\leq \sum_{j=-n}^n \sum_{k=-n}^n |\psi_j \bar{\psi}_k| \sqrt{\sup_t E[|X_t|^2] \sup_t E[|X_t|^2]} \\ \sum_{j=-n}^n \sum_{k=-n}^n |\psi_j \bar{\psi}_k| \sqrt{\mathbb{E}[|X_{t-j}|^2] \mathbb{E}[|X_{t-k}|^2]} &\leq \sum_{j=-n}^n \sum_{k=-n}^n |\psi_j \bar{\psi}_k| \sup_t E[|X_t|^2] \end{aligned}$$

so we can factor out the expectation of the infinite sum (since we are assuming all the expectations are taking the supreme value, so it is no longer indexed to the infinite sum):

$$\sum_{j=-n}^n \sum_{k=-n}^n |\psi_j \bar{\psi}_k| \sqrt{\mathbb{E}[|X_{t-j}|^2] \mathbb{E}[|X_{t-k}|^2]} \leq \sup_t E[|X_t|^2] \sum_{j=-n}^n \sum_{k=-n}^n |\psi_j \bar{\psi}_k|$$

Recall that  $|z|^2 = z\bar{z}$ , so:

$$\left| \sum_{j=-n}^n \psi_j \right|^2 = \left( \sum_{j=-n}^n \psi_j \right) \left( \sum_{j=-n}^n \bar{\psi}_j \right) = \sum_{j=-n}^n \sum_{k=-n}^n \psi_j \bar{\psi}_k$$

Thus, we can rewrite the RHS such that:

$$\sum_{j=-n}^n \sum_{k=-n}^n \psi_j \bar{\psi}_k \sqrt{\mathbb{E}[|X_{t-j}|^2] \mathbb{E}[|X_{t-k}|^2]} \leq \sup_t E[|X_t|^2] \left| \sum_{j=-n}^n \psi_j \right|^2$$

$$\sum_{j=-n}^n \sum_{k=-n}^n \psi_j \bar{\psi}_k \sqrt{\mathbb{E}[|X_{t-j}|^2] \mathbb{E}[|X_{t-k}|^2]} \leq \sup_t E[|X_t|^2] \left| \sum_{j=-n}^n \psi_j \right| \left| \sum_{j=-n}^n \psi_j \right|$$

by **Triangle Inequality**:

$$\left| \sum_{j=-n}^n \psi_j \right| \leq \sum_{j=-n}^n |\psi_j|$$

Thus,

$$\begin{aligned} \sup_t E[|X_t|^2] \left| \sum_{j=-n}^n \psi_j \right| \left| \sum_{j=-n}^n \psi_j \right| &\leq \sup_t E[|X_t|^2] \sum_{j=-n}^n |\psi_j| \sum_{j=-n}^n |\psi_j| \\ \sup_t E[|X_t|^2] \left| \sum_{j=-n}^n \psi_j \right| \left| \sum_{j=-n}^n \psi_j \right| &\leq \sup_t E[|X_t|^2] \left( \sum_{j=-n}^n |\psi_j| \right)^2 \end{aligned}$$

So, we just need to check if the RHS is finite as  $n$  goes to infinity:

$$\sup_t E[|X_t|^2] \left( \sum_{j=-\infty}^{\infty} |\psi_j| \right)^2$$

Recall we are given that:

$$\sup_t E[|X_t|^2] < \infty$$

and

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

, which implies

$$\left( \sum_{j=-\infty}^{\infty} |\psi_j| \right)^2 < \infty$$

Therefore, we have:

$$\sup_t E[|X_t|^2] \left( \sum_{j=-\infty}^{\infty} |\psi_j| \right)^2 < \infty$$

which implies that as  $n$  goes to infinity:

1.

$$\sup_t E [|X_t|^2] \left| \sum_{j=-n}^n \psi_j \right|^2 < \infty$$

2.

$$\sum_{j=-n}^n \sum_{k=-n}^n \psi_j \bar{\psi}_k \sqrt{E [|X_{t-j}|^2] E [|X_{t-k}|^2]} < \infty$$

3.

$$\sum_{j=-n}^n \sum_{k=-n}^n \psi_j \bar{\psi}_k \sqrt{E [|X_{t-j}|^2] E [|X_{t-k}|^2]} < \infty$$

4.

$$\sum_{j=-n}^n \sum_{k=-n}^n \psi_j \bar{\psi}_k |E [X_{t-j} \bar{X}_{t-k}]| < \infty$$

5.

$$\sum_{j=-n}^n \sum_{k=-n}^n \psi_j \bar{\psi}_k E [X_{t-j} \bar{X}_{t-k}] < \infty$$

6.

$$E \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} \right|^2 \right] < \infty$$

$$\lim_{n \rightarrow \infty} E \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} \right|^2 \right] = \text{Finite}$$

7. Since  $K$  denotes the finite variance, then, we have, as  $n$  goes to infinity:

$$E \left[ \left| \sum_{j=-n}^n \psi_j X_{t-j} - K \right|^2 \right] \rightarrow 0$$

or

$$\sum_{j=-n}^n \psi_j X_{t-j} \xrightarrow{L^2} X$$

- Thus, we have that the infinite sum converges in mean square/second moment/L-2
- Thus, it converges in the first moment, which implies convergence in probability.
- Notice that for the variance of the infinite sum to be finite we haven't needed  $\sum_{j=-\infty}^{\infty} |\psi_j|^2 < \infty$ , it was enough to have  $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$  as long as the second moment of  $X_t$  is finite.<sup>7</sup>

---

<sup>7</sup>This is because any absolute-summable series is square-summable.

## 4.7 Convergence in second moment of an Infinite Sum of Random Variables: Square Summable

It seems that checking for an infinite sum to be absolutely summable is the only way we can get convergence in second moment.

However, it is not the case!

Actually for convergence in second moment of the infinite sum we need a weaker condition: the infinite sum to be square summable.

**Theorem 5** *If  $\{X_t\}$  is any sequence of random variables such that*

$$\sup_t E |X_t|^2 < \infty$$

*and if*

$$\sum_{j=-\infty}^{\infty} |\psi_j|^2 < \infty,$$

*then the series*

$$Y_t = \psi(L)X_t = \sum_{j=-\infty}^{\infty} \psi_j L^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

*converges in mean square.*

### 4.7.1 Usefulness of this exercise

- This condition will be relevant when dealing with fractionally integrated series or long memory processes
- A long memory process can be expressed as an infinite sum that is not absolutely summable when  $d > 0$ . Then, we could wrongly conclude that it does not have a finite variance when  $d > 0$ , which implies that the series is not stationary (having a finite second moment is a condition for stationarity).
- However, for  $0.5 > d > 0$ , the infinite sum is square summable, which let us conclude that the process has a finite second moment.

### 4.7.2 Proof

For a more advanced time series econometric course!

## 5 Stationary Time Series

### 5.1 Strict Stationarity

- The time series  $\{X_t, t \in \mathbb{Z}\}$  is said to be strictly stationary if the joint distributions of  $(X_{t_1}, \dots, X_{t_k})'$  and  $(X_{t_1+h}, \dots, X_{t_k+h})'$  are the same for all positive integers  $k$  and for all  $t_1, \dots, t_k, h \in \mathbb{Z}$
- Strict stationarity means intuitively that the graphs over two equal-length time intervals of a realization of the time series should exhibit similar statistical characteristics. For example, the proportion of ordinates not exceeding a given level  $x$  should be roughly the same for both intervals.

## 5.2 Covariance Stationarity

The time series  $\{X_t, t \in \mathbb{Z}\}$ , with index set  $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ , is said to be stationary if

1. For all  $t \in \mathbb{Z}$ , the second moment exists:

$$E |X_t|^2 < \infty$$

2. For all  $t \in \mathbb{Z}$ , the mean is independent of  $t$ :

$$EX_t = m$$

The first condition, implies that first moment exists. However this condition requires it to be independent of  $t$  and constant over time, which implies that the mean is the same for all  $X_t, \forall t \in \mathbb{Z}$

3. Let's define the autocovariance function as a function of two variables:

$$\gamma_x(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - EX_r)(X_s - EX_s)], \quad r, s \in T$$

Then, for all  $r, s$  and  $t \in \mathbb{Z}$ , the autocovariance between  $X_t$  and  $X_{t+s}$  must only be a function of the distance between the two observations,  $s$ :

$$\gamma_x(t, t+s) = g(s)$$

, such that:

$$\gamma_x(r, s) = \gamma_x(r+t, s+t) = g(s-r) = g(r-s)$$

since the covariance between  $X_{r+t}$  and  $X_{s+t}$  is only a function of the distance between both observations,  $s-r$ , which is the same as the distance between the observations  $X_r$  and  $X_s$ . Since both covariance have the same distance, then they should be equal in order for the process to be stationary.

## 5.3 The Autocovariance and Autocorrelation Function

- If  $\{X_t, t \in \mathbb{Z}\}$  is stationary then

$$\gamma_x(r, s) = \gamma_x(r-s, 0)$$

for all  $r, s \in \mathbb{Z}$ .

- Thus, it is convenient to redefine the autocovariance function of a stationary process as the function of just one variable:

$$\gamma_x(h) \equiv \gamma_x(h, 0) = \text{Cov}(X_{t+h}, X_t)$$

for all  $t, h \in \mathbb{Z}$

- The function  $\gamma_x(\cdot)$  will be referred to as the autocovariance function of  $\{X_t\}$
- $\gamma_x(h)$  as the value of the autocovariance function of  $\{X_t\}$  at lag  $h$ .

- The autocorrelation function (acf) of  $\{X_t\}$  is defined analogously as the function whose value at lag  $h$  is:

$$\rho_x(h) \equiv \frac{\gamma_x(h)}{\gamma_x(0)} = \text{Corr}(X_{t+h}, X_t)$$

for all  $t, h \in \mathbb{Z}$

## 5.4 Complex-Valued Stationary Time Series

- Processes encountered in practice are nearly always **real-valued**
- However, it is mathematically simpler in **spectral analysis** to treat them as special cases of complex-valued processes.

The process  $\{X_t\}$  is a complex-valued stationary process if

1. For all  $t \in \mathbb{Z}$ , the second moment exists:

$$E|X_t|^2 < \infty$$

2. For all  $t \in \mathbb{Z}$ , the mean is independent of  $t$ :

$$E[X_t] = m$$

3. For all  $t \in \mathbb{Z}$ , the autocovariance is independent of  $t$ :

$$\begin{aligned} \text{Cov}[X_{t+h}, X_t] &= E[(X_{t+h} - E[X_{t+h}])(\overline{X_t - E[X_t]})] \\ &= E[X_{t+h}\bar{X}_t] - E[X_{t+h}]E[\bar{X}_t] \end{aligned}$$

Given that for the process to be stationary we need condition 2:  $E X_{t+h} = E \bar{X}_t = m$ , then for condition 3, we only need:

$$E(X_{t+h}\bar{X}_t) \quad \text{to be independent of } t$$

This is what makes it different from our usual definition with real-valued numbers: the fact that when dealing with complex-valued random variables, the covariance formula requires the conjugate of the variables.

## 5.5 The Complex-Valued Autocovariance Function

- The autocovariance function  $\gamma(\cdot)$  of a **complex-valued stationary process**  $\{X_t\}$  is:

$$\gamma(h) = E(X_{t+h}\bar{X}_t) - E X_{t+h} E \bar{X}_t$$

for all  $t, h \in \mathbb{Z}$

- The function  $\gamma_x(\cdot)$  will be referred to as the autocovariance function of  $\{X_t\}$  and has the following properties<sup>8</sup>:

$$\begin{aligned}\gamma(0) &\geq 0 \\ |\gamma(h)| &\leq \gamma(0) \quad \text{for all integers } h \\ \gamma(\cdot) &\text{ is a Hermitian function (i.e. } \gamma(h) = \overline{\gamma(-h)}\end{aligned}$$

## 5.6 Stationarity of an Infinite sum of random variables (Real-valued/complex-valued)

**Theorem 6** If

1.  $\{X_t\}$  is a stationary process with autocovariance function  $\gamma_X(\cdot)$

2.

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty,$$

3.

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

Then, then for  $X \in \mathbb{R}$ , the process  $\{y_t\}$  is stationary with autocovariance function:

$$\gamma_Y(h) = \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h - j + k)$$

For  $X \in \mathbb{C}$  with zero mean, the process  $\{Y_t\}$  is stationary with autocovariance function:

$$\gamma_Y(h) = E(Y_{t+h} \bar{Y}_t) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \bar{\psi}_k \gamma_X(h - j + k), \quad h = 0, \pm 1, \dots$$

### 5.6.1 Proof

- If  $\{X_t\}$  is stationary, we have that

$$E |X_t| = c$$

where  $c$  is finite and independent of  $t$ , and that its second moment is finite. Thus we can apply Theorem 3 and 4

- From the Theorem 3 and Theorem 4, we know that

$$\sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$$

---

<sup>8</sup>A **Hermitian function** is a complex function with the property that its complex conjugate is equal to the original function with the variable changed in sign:  $f(x) = f(-x)$

converges in second and first moment. Thus,  $Y_t$  has a finite second moment, which is the first requirement to be stationary.

- Now we need to check its first moment. We know that its first moment is finite, but is it also time independent?:

$$E[Y_t] = \lim_{n \rightarrow \infty} \sum_{j=-n}^n \psi_j E[X_{t-j}] = \left( \sum_{j=-\infty}^{\infty} \psi_j \right) E[X_t] = \left( \sum_{j=-\infty}^{\infty} \psi_j \right) c$$

Yes, it is!. Notice that finding its mean is much more simpler than Theorem 3 because we are not finding  $E|Y_t|$ . We know it exists so  $E[Y_t]$  must exist, and we just have to find it in the usual easy way.

- Notice that we could also find the variance of  $Y_t$  by finding the variance of the infinite sum, which is straightforward and easy given that in this case  $\text{Var}(X_t) = \sigma^2$
- Now, we just have to find the autocovariance function for  $Y_t$ :

$$\begin{aligned} E(Y_{t+h} Y_t) &= \lim_{n \rightarrow \infty} E \left[ \left( \sum_{j=-n}^n \psi_j X_{t+h-j} \right) \left( \sum_{k=-n}^n \psi_k X_{t-k} \right) \right] \\ &= \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \left( \gamma_X(h-j+k) + (E[X_t])^2 \right) \end{aligned}$$

Thus,  $E[Y_t]$  and  $E(Y_{t+h} Y_t)$  are both finite and independent of  $t$ . The autocovariance function  $\gamma_Y(\cdot)$  of  $\{Y_t\}$  is given by

$$\gamma_Y(h) = E(Y_{t+h} Y_t) - E[Y_{t+h}]E[Y_t] = \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h-j+k)$$

## 6 Spectral Representation

### 6.1 Frequency Domain

- Frequency is the number of occurrences of a repeating event per unit of time (e.g. beats per second).
- The frequency domain refers to the analysis of mathematical functions or signals with respect to frequency, rather than time.
- In signal processing, a signal is a function that conveys information about a phenomenon. In electronics and telecommunications, it refers to any time varying voltage, current or electromagnetic wave that carries information

### 6.2 Frequency Domain versus Time Domain

- A time-domain graph shows how a signal changes over time.
- A frequency-domain graph shows how much of the signal lies within each given frequency band over a range of frequencies.
- The idea here is to realize that when observing an intensity (e.g. pressure) of a series over time, we are only looking at the final product but we are not looking at its decomposition.

- Let me explain this idea with a sound. Figure 5 shows the evolution of the air pressure of 3 sounds. A sound signal represents variations in air pressure over time. The dotted line represents the equilibrium of each sound
- The sounds D294 (294 beats per second) and A440 (440 beats per second) are pure sines waves. Thus we can characterize their evolution over time by the sine function.
- The top panel on Figure 5 represents the air pressure generated when both sounds are played at once. The resulting graph shows that at any point of time the pressure difference will be the sum of the pressure difference of each note.
- This is something complicated to think about. At some points the peaks match up with each other, resulting in a really high pressure. At other points, they tend to cancel out.
- Thus, what you get is a wave-ish pressure versus time graph because it is the combination of two pure frequencies.
- However, the resulting series it is not a pure sine wave.
- Now, imagine that you add more notes, then the resulting series is even more complicated.
- Recall we only see the final series, so is there a way to decompose a final signal (series) into the pure frequencies that make it up?
- Well, for that we need a mathematical machine that treats signals with a given frequency differently from how it treats other signals.
- An example is the **Fourier transform**, which converts a time function into a sum or integral of sine waves of different frequencies, each of which represents a frequency component.

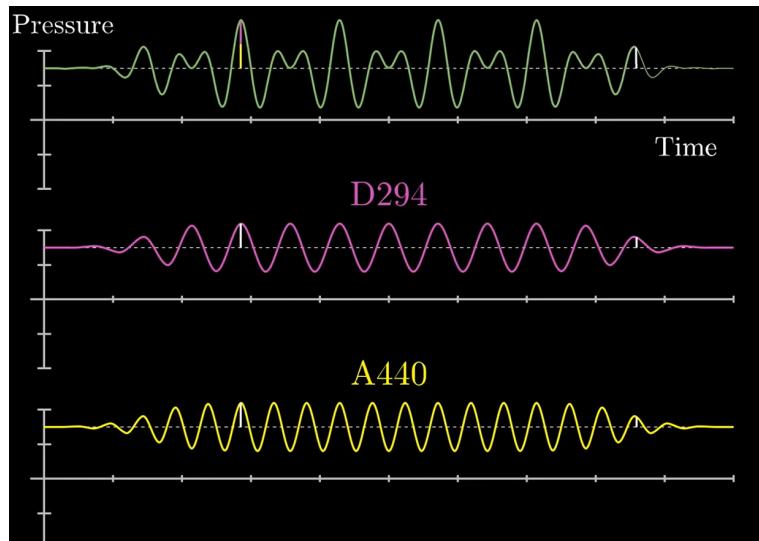


Figure 5: Air Pressure of sound over time

### 6.3 Advantages

- The simplification of the mathematical analysis.

## 6.4 Fourier Transform - Some Intuition

### 6.4.1 Winding the original series around a circle

- To start let's analyze the pure signal with a lowly 3 beats per second (Figure 6)
- The idea is to take the usual series on the time domain and wrap it up around a circle like in Figure 7
- This wrapping up is such that we have a vector (represented by the white arrow) where at each point in time,  $t$ , its length is equal to the height of the original series at  $t$
- Thus, high points of the original graph correspond to a greater distance from the origin on the circular graph. Consequently, low points end up closer to the origin
- The circle in Figure 7 is drawn in such a way that moving forward 2 seconds in time corresponds to a single rotation around the circle. That is easy to see since every one second there are 3 beats and on the circle, and there are 6 times that the vector reaches its maximum length. So, a single rotation around the circle lasts 2 seconds
- Let's define a **cycle** as the rotation of the vector on the circle. Thus, in this case we have that the vector on the circle is rotating at 0.5 cycle per second (or 1 cycle per 2 seconds, i.e., 1 rotation every 2 seconds.)
- Thus, at the moment we have two frequencies.
- There's the frequency of our signal, which goes up and down, three times per second. And then, separately, there's the frequency with which we're wrapping the graph around the circle, which at the moment is half cycle per second.

### 6.4.2 The winding frequency

- Notice that we can adjust that second frequency however we want.
- Figure 8 represents wrapping it around faster. So in this graph the petals are fatter because we have that one cycle is reached in 0.81 seconds, which means that there should only be 2.5 beats in one circle. We see a lot of petals because we are representing the total 4.5 seconds of the original graph.
- Figure 9 represents wrapping it around slower (1 cycle in 5 seconds, so we should expect 15 petals in one circle. We have space for 15 petals but we only have 12.5 petals because we are representing the 4.5 seconds of the original graph).
- Thus, the choice of winding frequency determines what the wrapped up graph looks like.
- Since the signal has 3 beats per second, we can do a funny thing!. We can wrap it up around the circle such that it has a frequency of 3 cycles per second, which implies that it only has one petal in a circle. Figure 10 depicts this case: the winding frequency matches the frequency of our signal (three beats per second).
- In this case, the petal indicates that all the high points on the graph happen on the right side of the circle and all of the low points happen on the left
- This is an interesting feature and we can take advantage of that in our attempt to build a frequency-unmixing machine.

### 6.4.3 The center of mass of the winding graph

- Imagine that the graph on the circle has some kind of mass to it, like a metal wire. We can represent the center of mass of that wire by a little dot like in Figure 11

- As we change the frequency, and the graph winds up differently, that center of mass kind of wobbles around a bit.
- For most of the winding frequencies, the peaks and valleys are all spaced out around the circle in such a way that the center of mass stays pretty close to the origin.
- However, when the winding frequency is the same as the frequency of our signal, we have in this case that all of the peaks are on the right and all of the valleys are on the left, so the center of mass is unusually far to the right (See Figure 12)

#### 6.4.4 The center of mass as a function of the winding frequency

- To capture the above, let's draw some kind of plot that keeps track of where that center of mass is for each winding frequency. Figure 13, 14, 15, 16, 17 and 18 depicts this.
- Of course, the center of mass is a two-dimensional thing, and requires two coordinates to fully keep track of, but for the moment, let's only keep track of the  $x$  coordinate.
- From Figure 13, 14, 15, 16, 17 and 18, we can observe that:
  - For a frequency of 0, when everything is bunched up on the right, this  $x$  coordinate is relatively high
  - Then, as you increase that winding frequency, and the graph balances out around the circle, the  $x$  coordinate of that center of mass goes closer to 0 and it just kind of wobbles around a bit.
  - But then, at three beats per second, there's a spike as everything lines up to the right.
- Notice that in this example we have that the original series is not around zero. This is why we have a big number for frequency zero (The cosine wave is shifted up). If we allow the original series to be around zero, then the spike for frequency zero does not show up and we only have a spike on frequency 3. See Figure 19
- Thus, we can clearly see how the winding frequency affects the center of mass
- Those graphs with the center of mass as a function of the winding frequency is an **almost Fourier Transform** of the original signal
- This is super important! Imagine having as original series the sum of a 3 and a 2 beats per second signals. Then, if we only look at the final series, we are not able to identify the pure signals. However, if we are able to wrap it up in around circle and then make a graph of its center of mass as a function of the winding frequency, we will be able to detect the spikes of the pure signals. See Figure 20
- Now what's going on here with the two different spikes, is that if you were to take two signals and then apply this Almost-Fourier transform to each of them individually, and then add up the results, what you get is the same as if you first added up the signals, and then applied this Almost-Fourier transform. See Figure 21

#### 6.4.5 Winding the original series around a circle on the Complex Plane

- Now, what is missing for the **Fourier Transform**?
- Well, until now we were focusing of the  $x$  value of the center of mass. However, the circumference is graphed in a two dimensions plane. So what about the  $y$  value of the center of mass?
- Well, we can think of it as the **Complex Plane**. Then, the center of mass is a complex number that has both real and imaginary part
- Complex numbers lend themselves to really nice descriptions of things that have to do with winding (a twisting movement or course) and rotation.

- For example, **Euler's Formula** tells us that if you take  $e$  to the power of some real number times  $i$ ,

$$z = e^{i\theta}$$

, you're gonna land on the point that you get if you were to walk that number of units around a circle with radius 1, counterclockwise starting on the right. See Figure 22

- So, imagine you wanted to describe rotating at a rate of one cycle per second:

- Recall that  $\pi = 3.1416\dots$  and the full length of circumference of a circle with radius 1 is  $2\pi$ . Thus we have that 1 cycle can be denoted by  $2\pi$ .
- Then, you could start by taking the expression  $e^{2\pi t i}$ , where  $t$  is the amount of time that has passed
- Then, decompose  $t$  in  $f$  times  $t$  so  $f$  captures the winding frequency. If  $f = 1$ , then we are in the case of 1 cycle per unit time  $t$  (second, in this example)
- For example, if  $f = \frac{1}{10}$ , this vector makes one full turn every 10 seconds since the time has to increase all the way to 10 before the full exponent looks like  $2\pi i$ . So, we have 1 cycle every 10 seconds or 1/10 cycle every 1 second

- Remember that the idea is to wrap up the series around a circle on the complex plane
- We can make it by using  $e^{2\pi f t i}$
- Let's say that the original series can be described by  $g(t)$ , then we have that the wrapping up around the circle on the complex plane is given by:

$$g(t)e^{2\pi f t i} \quad \text{for counterclockwise rotation}$$

or

$$g(t)e^{-2\pi f t i} \quad \text{for clockwise rotation, which is the convention}$$

See Figure 23

- This is amazing, since this really small expression is a super elegant way to encapsulate the whole idea of winding a graph around a circle with a variable frequency  $f$ .
- Now, remember we want to express the measure of mass a function of the frequency. How can we obtain the center of mass?
- Well, we need to consider all the observed data (i.e., for all time) and let the only exogenous variable in the function  $\phi(\cdot)$  be the frequency per unit of time.
- Thus, we could take the sum (for discrete time) or integral (for continuous time) of that expression when the number of observations goes to infinity.

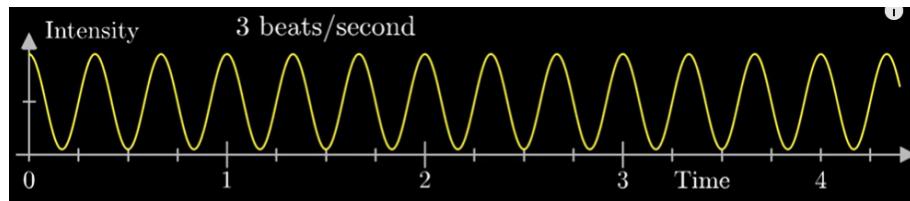


Figure 6: Pure signal with 3 beats per second on the time domain

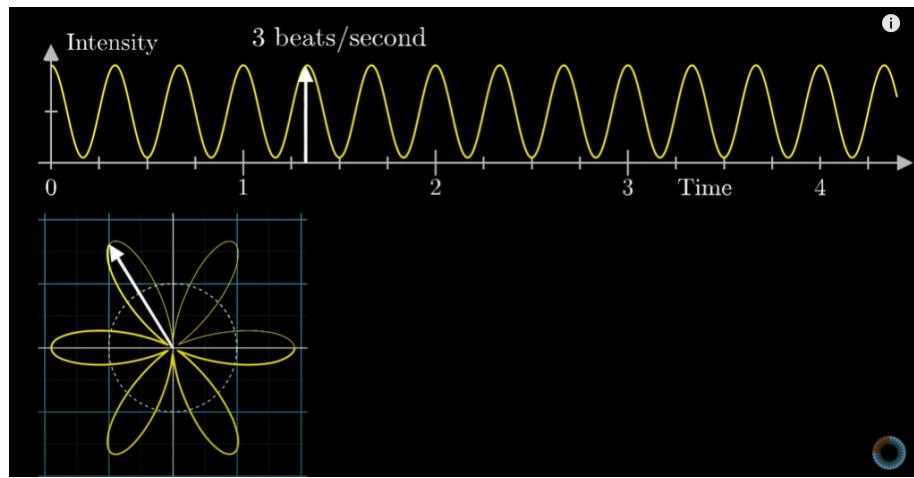


Figure 7: Wrapping the pure signal with 3 beats per second up around a circle

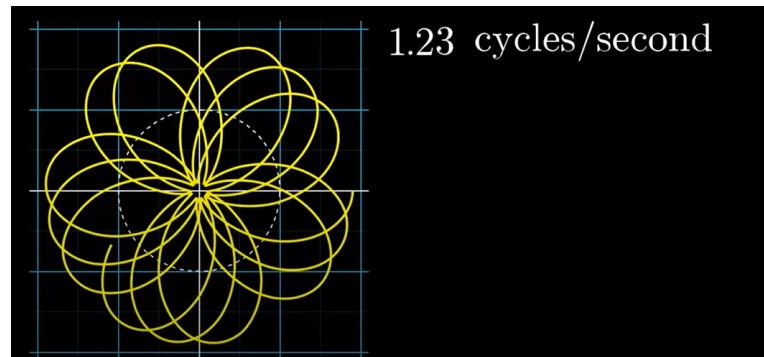


Figure 8: Wrapping the pure signal with 3 beats per second up around a circle faster

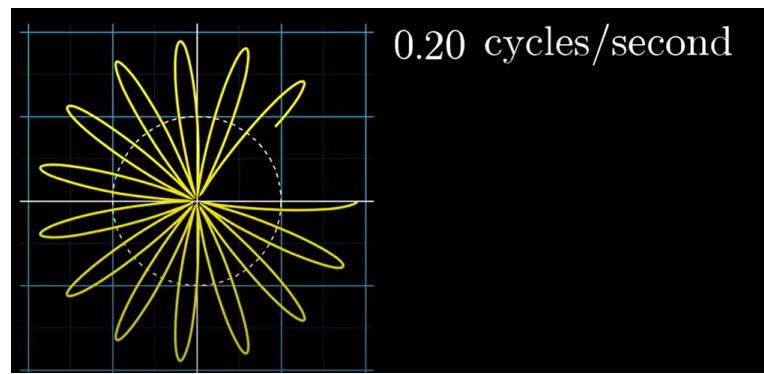


Figure 9: Wrapping the pure signal with 3 beats per second up around a circle slower

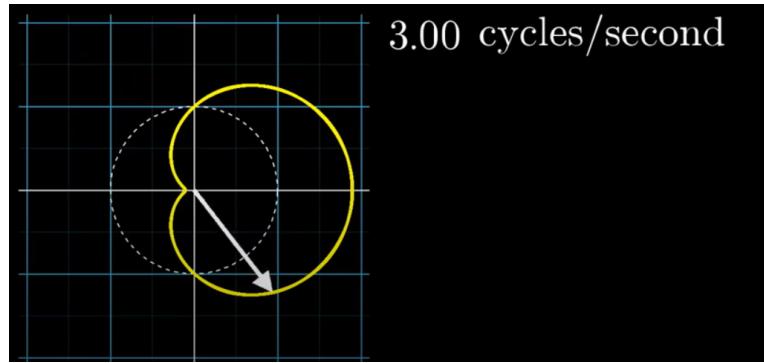


Figure 10: The winding frequency matches the frequency of our signal (three beats per second)

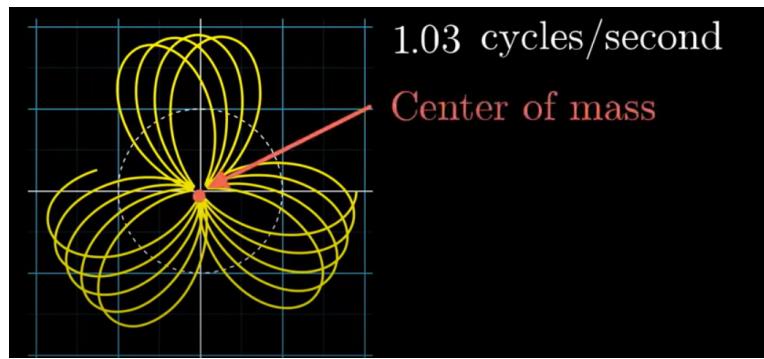


Figure 11: The center of mass

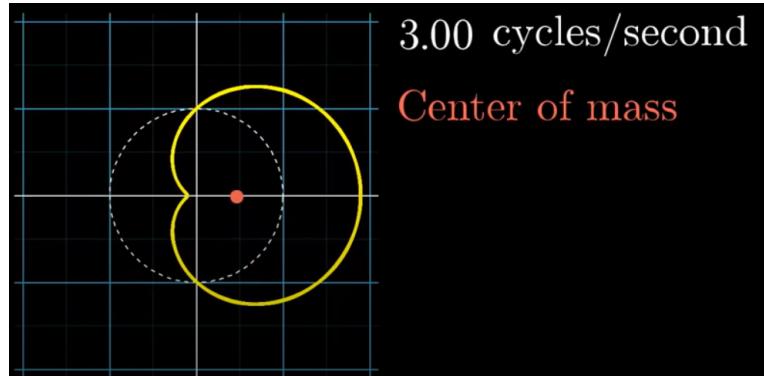


Figure 12: The center of mass when the winding frequency matches the frequency of our signal

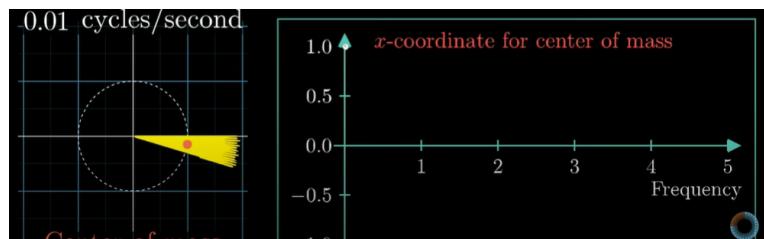


Figure 13: Graph of the center of mass as function of the winding frequency

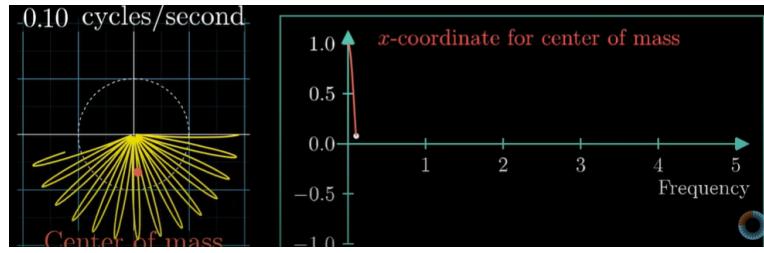


Figure 14: Graph of the center of mass as function of the winding frequency

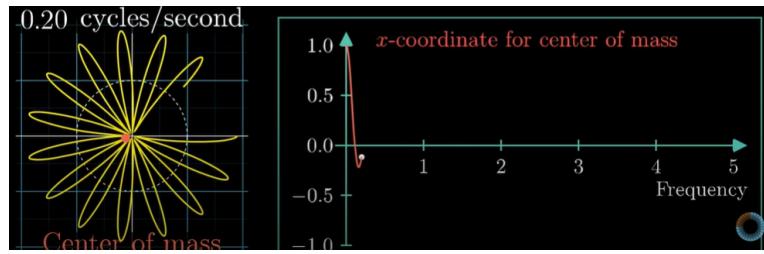


Figure 15: Graph of the center of mass as function of the winding frequency

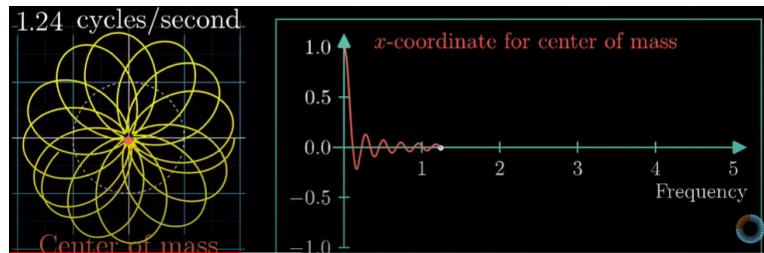


Figure 16: Graph of the center of mass as function of the winding frequency

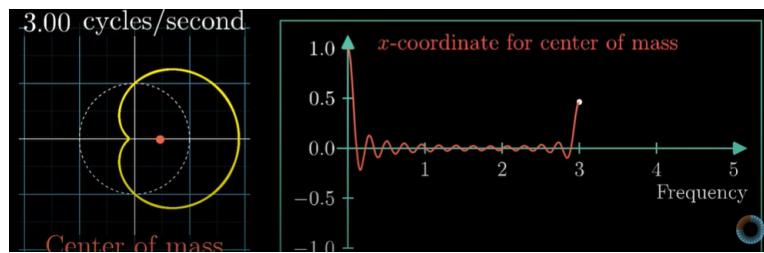


Figure 17: Graph of the center of mass as function of the winding frequency

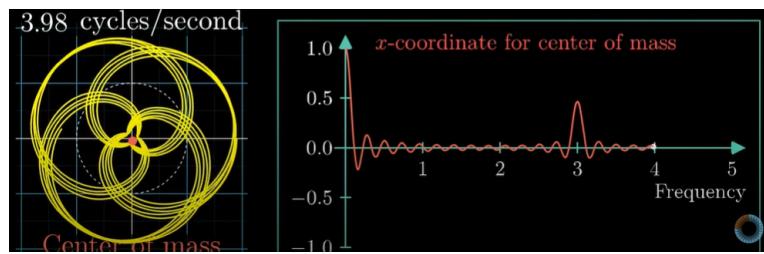


Figure 18: Graph of the center of mass as function of the winding frequency

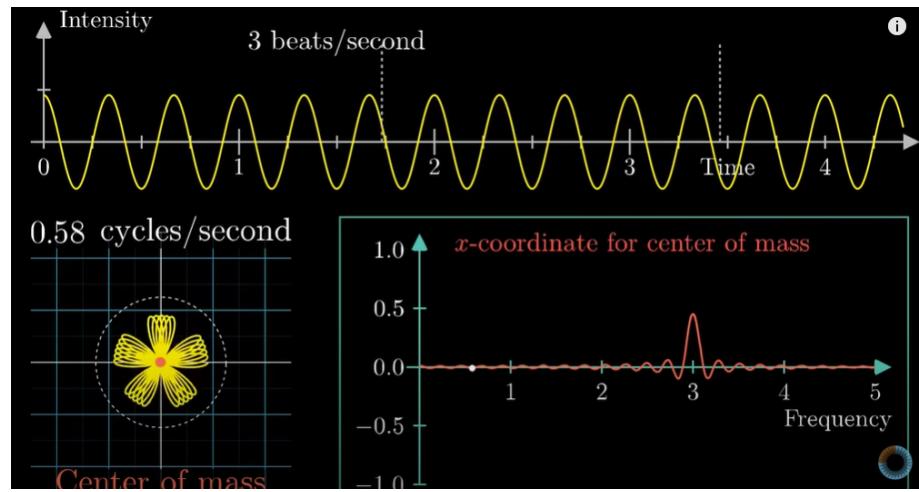


Figure 19: The original cosine wave when centered around zero

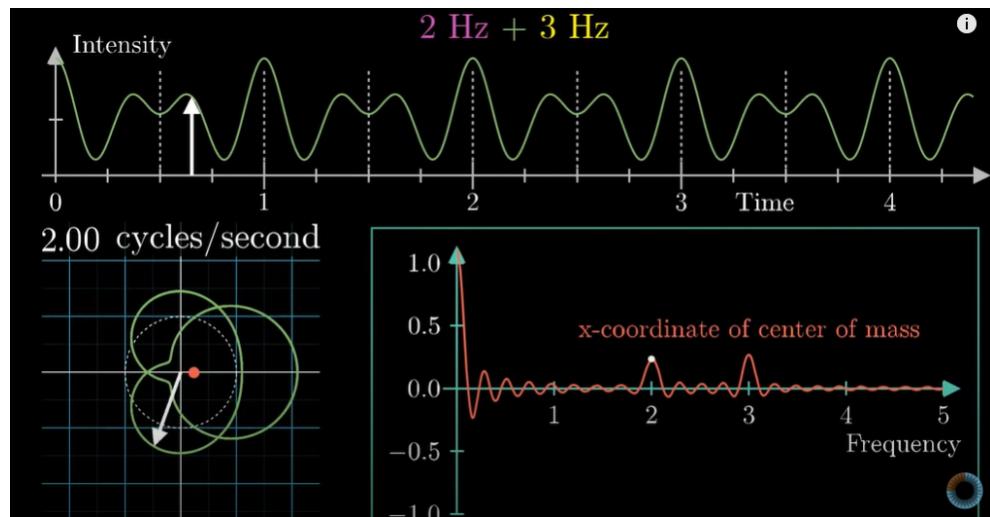


Figure 20: Decomposing the pure signals of an accumulated signal

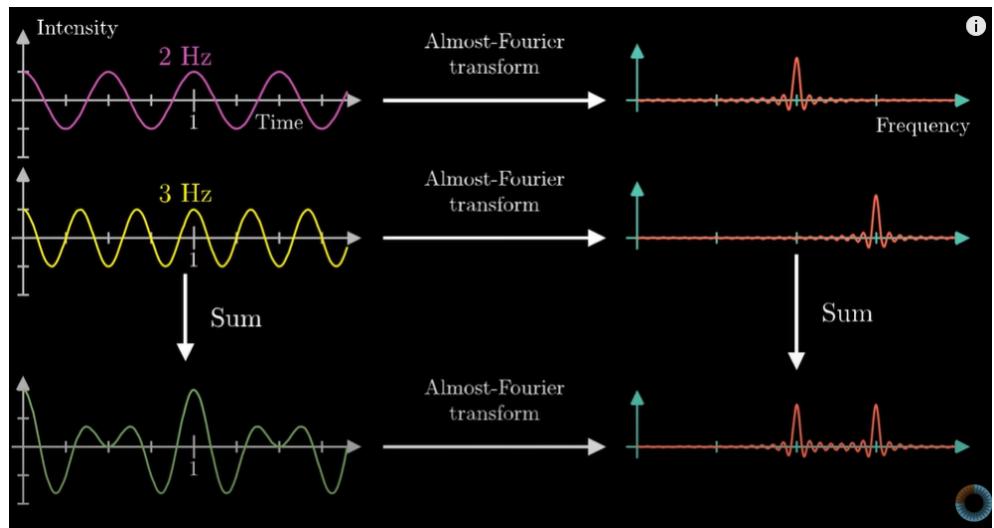


Figure 21: Decomposing the pure signals of an accumulated signal

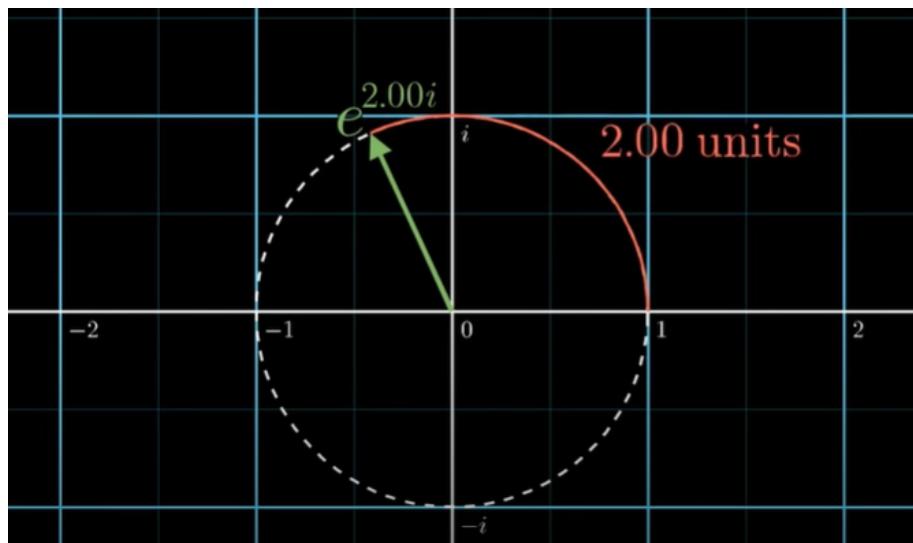


Figure 22: Euler's Formula on the Complex Plane

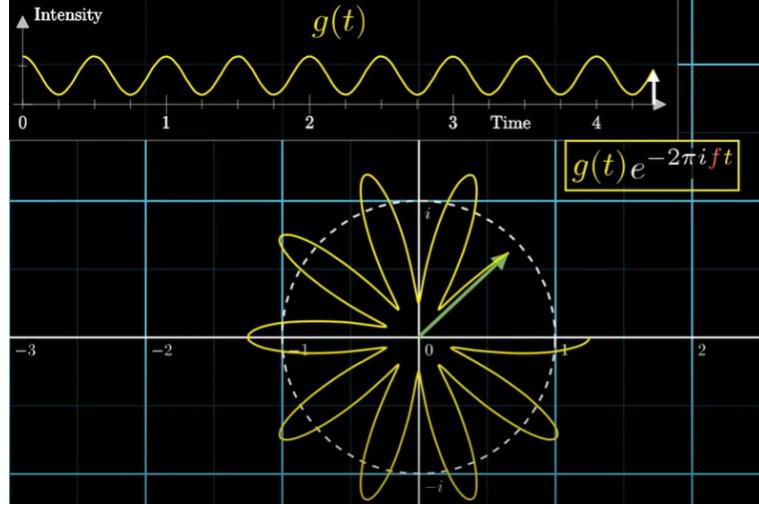


Figure 23: Using Euler's Formula to get the winding version of the original signal

## 6.5 Fourier Transform - Formal Definition

Given a (non-stochastic) sequence  $x_t$ ,  $t = \dots, -1, 0, 1, \dots$ , we can define the **Fourier Transform** of  $x_t$  as:

$$\tilde{x}(\omega) = \sum_{t=-\infty}^{\infty} x_t e^{-it\omega}$$

, in which

- $\omega \in (-\pi, \pi]$ :
  - This is to put an explicit boundary to the frequency: we are limited to have something between 0 and 1 cycle, which in this case is  $2\pi$  radians.
  - Why? When we observe economic data for any unit of time that means that each point changes every time  $t$  changes.
  - That is, we cannot have a movement between  $t=1$  and  $t=2$  since by construction there is nothing being measured there.
  - For example, when observing GDP for  $t$  in months: when  $t$  goes from Dec2000 to Jan 2001, there is no change in the value of GDP between those two units of time because it is not being measured.
  - However if we go to the example of the music tone, we have that the air pressure reaches its max 3 times between  $t=1$  second and  $t=2$  seconds.
  - Thus, we have 3 beats per second and eventually 3 cycles per second. That cannot happen with economic data. Let's imagine  $t=1, 2, 3, 4$ . Then if GDP is 1000, 2000, 3000, 2000, we have that it reaches its max every 3 months, so eventually we will have  $1/3$  cycles per month (or  $2\pi/3$  radians per unit of time).
  - Now, if GDP is 1000, 2000, 1000, 2000, we have that it reaches its max every 2 months, so eventually we will have  $1/2$  cycles per month (or  $2\pi/2$  radians per unit of time).
  - Thus, by construction when looking at economic data for any frequency of time, we will not have something like 1, 2, 3, ... cycles per second.

- As much we can have  $1/2$  cycles per unit of time ( $2\pi/2$  radians per unit of time). Thus, having  $\omega \in (-\pi, \pi]$  makes a lot of sense.
- $\omega$  is the winding frequency but in a different measure:

$$\omega = 2\pi f$$

**It is radians per unit of time**

- Any given, fixed value to  $\omega$  says how quickly  $e^{-it}$  circumnavigates the unit circle. For instance,  $\omega = 2\pi$  means that a full rotation of the unit circle in 1 unit of  $t$ , that is 1 cycle in one unit of time. This is the same as  $f = 1$ , so we get  $\omega = 2\pi$ , which means 1 cycle per unit of time.
- We can read it as  $\omega$  radians per unit of time, which does not imply a full rotation or cycle. It is a full rotation when  $\omega = 2\pi$ .
- So if the unit of  $t$  is a quarter:
  - $\omega = \pi$ :
    - \* One pi radian per unit of time or half cycle per unit of time.
    - \* This is like in the previous subsection when using  $f = 1/2$ , so we obtain the same exponent:  $\pi t i$ .
    - \* That is 1 cycle every 2 units of time, which is equivalent to 0.5 cycle every 1 unit of time.
    - \* If the true  $g(t)$  is such that  $x_t = g(t) = \cos(\omega t)$ , then we have that it has 1 beat every two quarters.
    - \* This means alternations every unit of time: in  $t$  we are on the valley, in  $t + 1$ , we are on the peak, in  $t + 2$  we are in the valley again.
  - $\omega = \pi/2$ :
    - \* half pi radians per unit of time.
    - \* That is, 0.25 cycle per unit of time.
    - \* This is like in the previous subsection using  $f = 1/4$ , so we obtain the same exponent:  $\frac{\pi}{2} t i$ .
    - \* That is 1 cycle every 4 units of time, which is equivalent to 0.25 cycle every 1 unit of time.
    - \* If the true  $g(t)$  is  $x_t = \cos(\omega t)$ , we have that it has 1 beat every 4 quarters.
    - \* So we have annually recurring events.
- Recall, by Euler's formula:

$$e^{i\omega} = \cos(\omega) + i \sin(\omega)$$

$$e^{-i\omega} = \cos(\omega) - i \sin(\omega)$$

- Using the exponential form of complex number we can express:

$$z = e^{-i\omega}$$

, with the property of being on the unit circle

$$|z| = 1$$

- Thus, the Fourier transform can also be written as:

$$\tilde{x}(\omega) = \sum_{t=-\infty}^{\infty} x_t z^t$$

, with  $z = e^{-i\omega}$  over  $\omega \in (-\pi, \pi]$

## 6.6 Fourier Inversion

The Fourier transform can be inverted to give the original sequence, term by term. When  $\tilde{x}$  is the Fourier transform of sequence

$$x_t = \{ \dots, x_{-1}, x_0, x_1, \dots \}$$

then

$$x_j = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \tilde{x}(\omega) e^{i\omega j} d\omega$$

for every integer  $j$ ,  $j = \dots, -1, 0, 1, \dots$

### 6.6.1 Proof

- This follows from writing out the right side

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{x}(\omega) e^{i\omega j} d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_t x_t e^{-i\omega(t-j)} d\omega = \frac{1}{2\pi} (2\pi) x_j = x_j$$

- after noticing that for all  $j \neq t$ , we have that:

$$\int_{-\pi}^{\pi} e^{-i\omega(t-j)} d\omega = 0$$

## 6.7 Alternative Conventions for formal definitions

### 6.7.1 Putting the factor of $\frac{1}{2\pi}$ in the Fourier transform instead of in its inverse

- This is how Uhlig defines it on its notes (if you are interested in understanding sign identification).
- The Fourier Transform:

$$\bar{x}(\omega) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} x_t e^{-it\omega}$$

- Inverse Fourier transform:

$$x_j = \int_{-\pi}^{\pi} \tilde{x}(\omega) e^{i\omega j} d\omega$$

### 6.7.2 Splitting the factor of $\frac{1}{2\pi}$ evenly between the Fourier transform and its inverse

- The Fourier Transform:

$$\bar{x}(\omega) = \frac{1}{\sqrt{2\pi}} \sum_{t=-\infty}^{\infty} x_t e^{-it\omega}$$

- Inverse Fourier transform:

$$x_j = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \tilde{x}(\omega) e^{i\omega j} d\omega$$

## 6.8 Lag Operator Calculus and Fourier Transforms

- Consider

$$y_t = h(L)\varepsilon_t$$

- For any form of  $h(L)$ , we have that the **Fourier Transform** of  $y_t$  is given by:

$$\tilde{y}(\omega) = h(e^{-i\omega}) \tilde{\varepsilon}(\omega) = 2\pi \tilde{h}(\omega) \tilde{\varepsilon}(\omega)$$

- So, the Fourier Transform of  $y_t$  is given by the lag polynomial evaluated at the **unit circle** times the Fourier Transform of  $\varepsilon_t$
- How is this possible? In the following lines I will show it

### 6.8.1 Case 1

- Consider

$$h(L) = h_0 + h_1 L$$

So, we have that:

$$y_t = h_0 \varepsilon_t + h_1 \varepsilon_{t-1}$$

- The **Fourier Transform** of  $y_t$  is given by:

$$\tilde{y}(\omega) = \sum_{t=-\infty}^{\infty} y_t e^{-it\omega}$$

$$\tilde{y}(\omega) = \sum_{t=-\infty}^{\infty} (h_0 \varepsilon_t + h_1 \varepsilon_{t-1}) e^{-it\omega}$$

$$\tilde{y}(\omega) = \sum_{t=-\infty}^{\infty} h_0 \varepsilon_t e^{-it\omega} + \sum_{t=-\infty}^{\infty} h_1 \varepsilon_{t-1} e^{-it\omega}$$

**TRICK:** We will try to put the second term such that it is indexed to  $t - 1$

$$\begin{aligned}\tilde{y}(\omega) &= \sum_{t=-\infty}^{\infty} h_0 \varepsilon_t e^{-it\omega} + \sum_{t=-\infty}^{\infty} h_1 \varepsilon_{t-1} e^{-it\omega} e^{-i\omega} e^{i\omega} \\ \tilde{y}(\omega) &= \sum_{t=-\infty}^{\infty} h_0 \varepsilon_t e^{-it\omega} + \sum_{t=-\infty}^{\infty} h_1 e^{-i\omega} \varepsilon_{t-1} e^{-it\omega+i\omega} \\ \tilde{y}(\omega) &= \sum_{t=-\infty}^{\infty} h_0 \varepsilon_t e^{-it\omega} + \sum_{t=-\infty}^{\infty} h_1 e^{-i\omega} \varepsilon_{t-1} e^{-i\omega(t-1)} \\ \tilde{y}(\omega) &= h_0 \sum_{t=-\infty}^{\infty} \varepsilon_t e^{-it\omega} + h_1 e^{-i\omega} \sum_{t=-\infty}^{\infty} \varepsilon_{t-1} e^{-i\omega(t-1)}\end{aligned}$$

Recall that the **Fourier Transform** pf  $\varepsilon$  is given by:

$$\tilde{\varepsilon}(\omega) = \sum_{t=-\infty}^{\infty} \varepsilon_t e^{-it\omega}$$

Given that the definition of a Fourier Transform implies that we are using all values of  $t$  (i.e., it goes from negative infinity to positive infinity), then the we simply change the indexation in the infinite sum from  $t$  to  $t - 1$  and it will still be the **Fourier Transform** of  $y_t$ :

$$\tilde{\varepsilon}(\omega) = \sum_{t=-\infty}^{\infty} \varepsilon_{t-1} e^{-i(t-1)\omega}$$

Therefore, we have that:

$$\tilde{y}(\omega) = h_0 \tilde{\varepsilon}(\omega) + h_1 e^{-i\omega} \tilde{\varepsilon}(\omega)$$

$$\tilde{y}(\omega) = (h_0 + h_1 e^{-i\omega}) \tilde{\varepsilon}(\omega)$$

$$\tilde{y}(\omega) = h(e^{-i\omega}) \tilde{\varepsilon}(\omega)$$

in which,  $h(e^{-i\omega})$  is the polynomial  $h(L)$  evaluated at  $L = e^{-i\omega}$ .

$$h(e^{-i\omega}) = h_0 + h_1 e^{-i\omega}$$

That is, we have that the lag polynomial  $h(L)$  is being evaluated at the **unit circle** given that  $e^{-i\omega}$  is the exponential form of a complex number such that its modulus equals 1:

$$|z| = 1 \quad \text{for } z = e^{-i\omega}$$

### 6.8.2 Case 2: AR(m)

- Consider

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \dots + \rho_m y_{t-m} + \varepsilon_t$$

- We can rewrite it as:

$$(1 - \rho(L))y_t = \varepsilon_t$$

$$\varepsilon_t = (1 - \rho(L))y_t$$

$$\varepsilon_t = h(L)y_t$$

Then, we have that:

$$\tilde{\varepsilon}(\omega) = h(e^{-i\omega}) \tilde{y}(\omega)$$

$$(1 - \rho(e^{-i\omega})) \tilde{y}(\omega) = \tilde{\varepsilon}(\omega)$$

### 6.8.3 Case 3: MA(n)

- Consider

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_n \varepsilon_{t-n}$$

- We can rewrite it as:

$$y_t = \theta(L)\varepsilon$$

- Then, we have that:

$$\tilde{y}(\omega) = \theta(e^{-i\omega}) \tilde{\varepsilon}(\omega)$$

## 6.9 Spectral Density

### 6.9.1 Some Intuition

- From the previous section, we know that a time series  $X_t$  can be expressed as its Fourier Transform  $\hat{x}(\omega)$
- We can construct the autocovariance function between  $X_t$  and  $X_{t+s}$ :

$$\gamma_x(r, r+s) = \text{Cov}(X_r, X_{r+s})$$

- If  $X_t$  is **covariance stationary**, then we have that:

$$\gamma_x(r, r+s) = \gamma_x(0, s) = \gamma_x(s)$$

That is, we can express the autocovariance function of  $X_t$  as a function of only one variable,  $s$ , instead of two variables  $(r, r+s)$ . This is key because it will allow us to generate a new time series.

- Since  $s$  denotes the autocovariance function of  $X_t$  at lag  $s$ , we could generate a new time series  $\gamma_x(s)$  for  $s = 0, 1, 2, 3, 4, 5, 6, \dots$ .
- It is a time series because it is indexed to  $s$ , which is a unit of time (e.g., depending on the unit of time for  $t$ , the lag can be in terms of years, months, quarters, seconds).
- Changing a bit the notation we could have the new time series:

$$\gamma_x(j) \equiv \gamma_j$$

for  $j = 0, 1, 2, 3, 4, 5, 6, \dots$

- Thus, we could also have the Fourier transform of  $\gamma_j$ :

$$\tilde{\gamma}(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-ij\omega}$$

, in which  $\omega \in (-\pi, \pi]$  represents the frequency. This is known as the **Spectral density** of  $X_t$

- This is extremely useful, since from the **Spectral density** of  $X_t$  we are able to find each autocovariance of  $X_t$ ,  $\gamma_j$ , by using the **Fourier Inverse**.

### 6.9.2 Formal Definition

- Let  $x_t \in \mathbb{R}, t = \dots, -1, 0, 1 \dots$  be **covariance stationary** with **mean zero**.
- Let's denote the autocovariance function of  $x_t$  as  $\gamma_j$ 
  - If  $x_t$  is real-valued, we have that it is a symmetric function:

$$\gamma_j = E[x_t x_{t-j}] = \gamma_{-j}$$

- If  $x_t$  is complex-valued, we have that it is a **Hermitian** function:

$$\gamma_j = E[x_t \bar{x}_{t-j}] = \bar{\gamma}_{-j}$$

- Then, the **Spectral Density** of  $x_t$  is given by the **Fourier Transform** of the autocovariance function of  $x_t$ :

$$s_x(\omega) = \tilde{\gamma}(\omega) = \sum_{j=-\infty}^{\infty} \gamma_j e^{-ij\omega}$$

, in which  $\omega \in (-\pi, \pi]$  represents the frequency.

- Since  $\gamma_j$  is symmetric about 0, i.e.,  $\gamma_j = \gamma_{-j}$ , the spectral density is also **symmetric** about 0:

$$s_x(\omega) = s_x(-\omega)$$

For multivariate case:  $s_x(\omega) = s_x(-\omega)'$

### 6.9.3 Approximation of the spectral density by the second moment of the Fourier transform

- Notice that:

$$E[\tilde{x}(\omega)\tilde{x}(\overline{\omega})] = E\left[\sum_{t=-\infty}^{\infty} x_t e^{-it\omega} \sum_{t=-\infty}^{\infty} \bar{x}_t e^{-it\overline{\omega}}\right]$$

Let's suppose we only have 3 observations:  $t = -1, 0, 1$ :

$$\begin{aligned} E[\tilde{x}(\omega)\tilde{x}(\overline{\omega})] &= E\left[\sum_{t=-1}^1 x_t e^{-it\omega} \sum_{t=-1}^1 \bar{x}_t e^{-it\overline{\omega}}\right] \\ &= E\left[\left(x_{-1}e^{-i(-1)\omega} + x_0e^{-i(0)\omega} + x_1e^{-i(1)\omega}\right) \left(\bar{x}_{-1}\overline{e^{-i(-1)\omega}} + \bar{x}_0\overline{e^{-i(0)\omega}} + \bar{x}_1\overline{e^{-i(1)\omega}}\right)\right] \\ &= E[x_{-1}\bar{x}_{-1}e^{-i(-1)\omega}\overline{e^{-i(-1)\omega}} + x_{-1}\bar{x}_0e^{-i(-1)\omega}\overline{e^{-i(0)\omega}} + x_{-1}\bar{x}_1e^{-i(-1)\omega}\overline{e^{-i(1)\omega}} + \\ &\quad x_0\bar{x}_{-1}e^{-i(0)\omega}\overline{e^{-i(-1)\omega}} + x_0\bar{x}_0e^{-i(0)\omega}\overline{e^{-i(0)\omega}} + \\ &\quad x_0\bar{x}_1e^{-i(0)\omega}\overline{e^{-i(1)\omega}} + x_1\bar{x}_{-1}e^{-i(1)\omega}\overline{e^{-i(-1)\omega}} + \\ &\quad x_1\bar{x}_0e^{-i(1)\omega}\overline{e^{-i(0)\omega}} + x_1\bar{x}_1e^{-i(1)\omega}\overline{e^{-i(1)\omega}}] \end{aligned}$$

Since when  $z = x + iy$ , we have that  $\bar{z} = x - iy$ , then when  $z = e^{-i(-1)\omega}$ , we must have that  $\bar{z} = e^{-i(-1)\omega} = e^{i(-1)\omega}$ . Further, recall  $z\bar{z} = |z|^2$ ; for  $z = e^{-i(-1)\omega}$ , we have  $|z| = 1$ . Finally, notice that  $e^{-i(0)\omega} = e^0 = 1$

$$\begin{aligned} &= \gamma_0 + \gamma_{-1}e^{-i(-1)\omega} + \gamma_{-2}e^{i\omega}e^{i\omega} + \\ &\quad \gamma_1e^{-i(1)\omega} + \gamma_0 + \gamma_{-1}e^{-i(-1)\omega} + \gamma_2e^{-i(2)\omega} + \\ &\quad \gamma_1e^{-i(1)\omega} + \gamma_0 \\ &= \gamma_{-2}e^{-i(-2)\omega} + 2\gamma_{-1}e^{-i(-1)\omega} + 3\gamma_0e^{-i(0)\omega} + 2\gamma_1e^{-i(1)\omega} + \gamma_2e^{-i(2)\omega} \end{aligned}$$

Thus, it is approximate:

$$\cong \sum_{j=-2}^2 \gamma_j e^{-ij\omega}$$

Therefore, for a more general case with infinite observations for  $x_t$ , the **Spectral density** of  $x_t$  can be approximated by:

$$\begin{aligned} s_x(\omega) &= \tilde{\gamma}(\omega) = \sum_{j=-\infty}^{\infty} \gamma_j e^{-ij\omega} \\ &\cong E[\tilde{x}(\omega)\tilde{x}(\overline{\omega})] \end{aligned}$$

#### 6.9.4 Finding the autocovariance: Fourier Inverse

- Since the Spectral density is a Fourier Transform, then we can always apply the Fourier Inverse to find each coefficient inside the infinite sum of the Fourier Transform, which in this case is  $\gamma_j$  (i.e., each autocovariance of  $x_t$ )
- The **Fourier Inverse** of the **Spectral density** of  $x_t$  gives us:

$$\gamma_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} s_x(\omega) e^{i\omega j} d\omega$$

, in which  $\gamma_j$  is the autocovariance of  $x_t$  of order  $j$  for every integer  $j$ .

- For  $j = 0$ , we have that

$$\gamma_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} s_x(\omega) d\omega$$

Thus, we can interpret  $s_x$  as spreading out  $x$ 's variability across the range of frequencies  $\omega \in (-\pi, \pi]$ .

#### Example: White Noise

- Suppose

$$x_t = \varepsilon_t$$

, where  $\varepsilon_t$  is a white noise with variance  $\sigma^2$ .

- Thus, we have that for  $x_t$ :

$$\gamma_j = \begin{cases} \sigma^2 & \text{for } j = 0 \\ 0 & \text{otherwise} \end{cases}$$

- Thus, the spectral density of  $x_t$  is given by only one element:

$$s_x(\omega) = \tilde{s}_x(\omega) = \sum_{j=-\infty}^{\infty} \gamma_j e^{-ij\omega} = \gamma_0 = \sigma^2$$

#### Example: MA(1)

- Suppose

$$x_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

, where  $\varepsilon_t$  is a white noise with variance  $\sigma^2$ .

- Thus, we have that for  $x_t$ :

$$\gamma_j = 0 \text{ for } j > 1$$

That is, only  $\gamma_0$ ,  $\gamma_1$  and  $\gamma_{-1}$  exist.

- Thus, the spectral density of  $x_t$  is given by only three elements:

$$s_x(\omega) = \sum_{j=-\infty}^{\infty} \gamma_j e^{-ij\omega}$$

$$s_x(\omega) = \gamma_0 + \gamma_1 e^{-i\omega} + \gamma_{-1} e^{i\omega}$$

- Given that  $\gamma_1 = \gamma_{-1}$ :

$$s_x(\omega) = \gamma_0 + \gamma_1 (e^{-i\omega} + e^{i\omega})$$

- By using Euler's formula:

$$s_x(\omega) = \gamma_0 + \gamma_1 (\cos \omega - i \sin \omega + \cos \omega + i \sin \omega)$$

$$s_x(\omega) = \gamma_0 + \gamma_1 (\cos \omega + \cos \omega)$$

$$s_x(\omega) = \gamma_0 + 2\gamma_1 \cos \omega$$

### Example: Stationary AR(1)

- Suppose

$$x_t = \rho x_{t-1} + \varepsilon_t$$

, where  $\varepsilon_t$  is a white noise with variance  $\sigma^2$  and  $|\rho| < 1$  so that  $x_t$  is stationary.

- Thus, we have that for  $x_t$ :

$$\gamma_j = \left( \frac{\sigma^2}{1 - \rho^2} \right) \rho^{|j|}$$

for all integer  $j$ .

- Thus, the spectral density of  $x_t$  is given by:

$$s_x(\omega) = \sum_{j=-\infty}^{\infty} \gamma_j e^{-ij\omega}$$

$$s_x(\omega) = \sum_{j=-\infty}^{\infty} \rho^{|j|} e^{-i\omega j} \frac{\sigma^2}{1 - \rho^2}$$

$$s_x(\omega) = \frac{\sigma^2}{1 - \rho^2} \sum_{j=-\infty}^{\infty} \rho^{|j|} e^{-i\omega j}$$

- Now, it is a more complicated result. However, we can use a property for the lag operator when finding the spectral density

### 6.9.5 Lag Operator Calculus, Stationarity and the Spectral density

**Theorem 7** Consider

$$y_t = \sum_{j=0}^{\infty} \psi_j L^j x_t = \psi(L) x_t$$

, in which  $x_t$  is a (complex-valued) zero-mean stationary process.

- Then, if  $\psi(z)$  converges on the unit circle or if that  $\left| \sum_{j=0}^{\infty} \psi z^j \right|^2$  converges on the unit circle, we have that  $y_t$  is stationary with its **Spectral density** given by:

$$s_y(\omega) = \left| \sum_{j=0}^{\infty} \psi_j e^{-iwj} \right|^2 s_x(\omega)$$

$$s_y(\omega) = |\psi(e^{-iw})|^2 s_x(\omega)$$

So, the spectral density of  $y_t$  in terms of the spectral density of  $x_t$  involves the **lag polynomial evaluated on the unit circle**, which implies that the **lag polynomial evaluated on the unit circle** must converge. This is why stationarity plays an important role when finding the spectral density of an AR process.

**Proof**

- Notice that from Theorem 3 and Theorem 4, we immediately have that  $y_t$  has a finite second moment.
- Further, given that  $x_t$  is stationary,  $y_t$  is also stationary (Theorem 6).
- Now, let's allow both variables to be possibly complex valued. Then, since, the autocovariance function of  $y_t$  is given by:

$$\gamma_y(h) = E(y_{t+h} \bar{y}_t) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \bar{\psi}_k \gamma_x(h-j+k), \quad h = 0, \pm 1, \dots \quad (9)$$

The Fourier inverse of the spectral density of  $\{x_t\}$  gives us:

$$\gamma_x(h) = \int_{-\pi}^{\pi} s_x(\omega) e^{i\omega h} d\omega \quad (10)$$

Then, we can rewrite (9) as:

$$\begin{aligned} \gamma_y(h) &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \bar{\psi}_k \int_{-\pi}^{\pi} s_x(\omega) e^{i\omega(h-j+k)} d\omega \\ &= \int_{-\pi}^{\pi} s_x(\omega) \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \bar{\psi}_k e^{i\omega(h-j+k)} d\omega \\ &= \int_{-\pi}^{\pi} s_x(\omega) \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j e^{-i\omega j} \bar{\psi}_k e^{i\omega k} e^{i\omega h} d\omega \\ &= \int_{-\pi}^{\pi} s_x(\omega) \sum_{j=0}^{\infty} \psi_j e^{-i\omega j} \sum_{k=0}^{\infty} \bar{\psi}_j e^{i\omega j} e^{i\omega h} d\omega \end{aligned}$$

By rearranging the terms inside the integral, we obtain:

$$\gamma_y(h) = \int_{-\pi}^{\pi} \left| \sum_{j=0}^{\infty} \psi_j e^{-i\omega j} \right|^2 s_x(\omega) e^{i\omega h} d\omega \quad (11)$$

By using the Fourier inverse definition for the Spectral density, it must be that:

$$\gamma_y(h) = \int_{-\pi}^{\pi} s_y(\omega) e^{i\omega h} d\omega \quad (12)$$

Then, the spectral density must be:

$$s_y(\omega) = \left| \sum_{j=0}^{\infty} \psi_j e^{-i\omega j} \right|^2 s_x(\omega) \quad , \quad \omega \in [-\pi, \pi] \quad (13)$$

$$s_y(\omega) = |\psi(e^{-i\omega j})|^2 s_x(\omega) \quad , \quad \omega \in [-\pi, \pi] \quad (14)$$

Recall  $|e^{-i\omega j}| = 1$ . Thus, we are evaluating  $\psi(z)$  on the unit circle:  $|z| = 1$

### Case 1: Stationary AR(1)

- Consider

$$\epsilon_t = (1 - \rho L)y_t$$

- We can express the above as:

$$\epsilon_t = h(L)y_t$$

with  $h(L) = 1 - \rho L$

- Then, we immediately have a relationship between the **spectral densities**:

$$s_\varepsilon(\omega) = h(e^{-i\omega}) h(e^{i\omega}) s_y(\omega)$$

$$s_\varepsilon(\omega) = (1 - \rho e^{-i\omega})(1 - \rho e^{i\omega}) s_y(\omega)$$

$$s_\varepsilon(\omega) = (1 - \rho e^{i\omega} - \rho e^{-i\omega} + \rho e^{-i\omega} \rho e^{i\omega}) s_y(\omega)$$

$$s_\varepsilon(\omega) = (1 - \rho e^{i\omega} - \rho e^{-i\omega} + \rho e^{-i\omega+i\omega}) s_y(\omega)$$

$$s_\varepsilon(\omega) = (1 - \rho e^{i\omega} - \rho e^{-i\omega} + \rho e^0) s_y(\omega)$$

$$s_\varepsilon(\omega) = (1 - \rho e^{i\omega} - \rho e^{-i\omega} + \rho) s_y(\omega)$$

$$s_\varepsilon(\omega) = (1 - \rho(e^{i\omega} + e^{-i\omega}) + \rho)s_y(\omega)$$

- By using Euler's formula:

$$s_\varepsilon(\omega) = (1 - \rho(\cos \omega - i \sin \omega + \cos \omega + i \sin \omega) + \rho)s_y(\omega)$$

$$s_\varepsilon(\omega) = (1 - 2\rho \cos \omega + \rho^2)s_y(\omega)$$

- Given that  $\varepsilon_t$  is a white noise with variance  $\sigma^2$ :

$$\sigma^2 = (1 - 2\rho \cos \omega + \rho^2)s_y(\omega)$$

Thus, can we just simply write:

$$s_y(\omega) = \left( \frac{1}{1 - 2\rho \cos(\omega) + \rho^2} \right) \sigma^2$$

? **No.** In short, the problem is that we cannot simply put  $(1 - 2\rho \cos \omega + \rho)^{-1}$  on the RHS because it could be the case that that polynomial does not converge

- **Why do we need the process to be stationary?:** Well, we are almost done in finding the spectral density of  $y_t$ . However, we need to realize that we have:

$$\sigma^2 = h(e^{-i\omega}) h(e^{i\omega}) s_y(\omega)$$

, in which **the lag polynomial**  $h(L)$  is being evaluated at  $L = e^{-i\omega}$  and  $L = e^{i\omega}$ . Given that  $|e^{-i\omega}| = |e^{i\omega}| = 1$ , we are analyzing the **the lag polynomial on the unit circle**. From Section 2.3, we know that  $h(L)^{-1} = (1 - \rho L)^{-1}$  when evaluated on the unit circle will converge only if  $|\rho| < 1$ , which is the same condition for AR(1) to be stationary.

- Thus, the **Spectral density** of a stationary AR(1) process is given by:

$$s_y(\omega) = \left( \frac{1}{1 - 2\rho \cos(\omega) + \rho^2} \right) \sigma^2$$

**Case 2:** AR( $m$ )

- Consider

$$\epsilon_t = (1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_m L^m)y_t$$

, in which  $\varepsilon_t$  is a white noise with variance  $\sigma^2$

- Let  $\lambda_i$  be the roots of the polynomial:

$$\tilde{\pi}(z) = z^m - \rho_1 z^{m-1} - \rho_2 z^{m-2} - \cdots - \rho_m$$

Then, by using the Fundamental Theorem of Algebra we can write:

$$(1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_m L^m) = (1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_m L)$$

- Thus, we can rewrite the process as:

$$\epsilon_t = (1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_m L)y_t$$

- Then, we immediately have a relationship between the **spectral densities**:

$$s_\epsilon(\omega) = (1 - \lambda_1 e^{-i\omega})(1 - \lambda_1 e^{i\omega})(1 - \lambda_2 e^{-i\omega})(1 - \lambda_2 e^{i\omega}) \dots (1 - \lambda_m e^{-i\omega})(1 - \lambda_m e^{i\omega})s_y(\omega)$$

- Given that the process is stationary, we can put on the LHS all the inverted lag polynomials evaluated on the unit circle and by using the Euler's formula as in the previous case we have that the spectral density of an AR(m) process is given by:

$$s_y(\omega) = \sigma^2 \prod_{j=1}^m \frac{1}{1 - 2\lambda_j \cos(\omega) + \lambda_j^2}$$

## 7 AR(P)

In this chapter we will use all the above results to analyze the properties of a time series AR(p)

### 7.1 The lag polynomial

- Consider the classic AR(p) process for  $X_t$ :

$$X_t = \rho_1 X_{t-1} + \rho_2 X_{t-2} + \dots + \rho_p X_{t-p} + \varepsilon_t \quad (15)$$

, in which  $\varepsilon_t$  is i.i.d. with mean zero and variance 1

- Thus, we have:

$$(1 - \rho_1 L - \rho_2 L^2 - \dots - \rho_p L^P)X_t = \varepsilon_t$$

- Let's define the **lag polynomial**:

$$\pi(L) = 1 - \rho_1 L - \rho_2 L^2 - \dots - \rho_p L^P$$

- Thus, we can express (15) as:

$$\pi(L)X_t = \varepsilon_t \quad (16)$$

- Now, we would like to analyze the properties of  $X_t$
- Thus, we would like to solve the difference equation for  $X_t$
- To solve means to express  $X_t$  a function only of  $\varepsilon$  and initial values of  $X_t$  (i.e.  $X_0$ )

### 7.2 Solving the difference equation

- One way to solve it from looking at (15) is to use the Iterative method, but that can be really tedious as  $P$  gets larger.
- One straightforward way to do it, from looking (16), is to premultiply  $\pi(L)^{-1}$  in (16).
- We would like  $\pi(L)^{-1}$  to be:
  - **Defined:** in the sense that it exists. Thus, we avoid the values of  $L$  such that  $\pi(L) = 0$

- **Regular/ Analytic / Holomorphic:** if it is regular around zero, the function can be expanded in a Taylor’s Series that converges in the largest open disk (i.e., around zero since it is the largest distance it can be made given the formula in Taylor’s Theorem) that does not contain any singularity so that we can express  $\pi(L)^{-1}$  as a infinite sum that converges to something finite (i.e., a power series that converges to that analytic function)
- So, it is super clear why we need the function  $\pi(L)^{-1}$  to be defined. Otherwise, it does not exists (i.e., 1/0).
- Being well defined might seem to be enough, but it is not!
- If  $\pi(L)^{-1}$  exists, we can express  $X_t$  as function of  $\varepsilon_t$  so we solved the difference equation for  $X_t$ . However, the solution implies a nonlinear function:  $\pi(L)^{-1}$ .
- Recall that we are interested in finding the properties of  $X_t$  (e.g., is it stationary?) and it is impossible to if we want to analyze:

$$\pi(L)^{-1} \varepsilon_t$$

For instance,  $L$  does not mean anything by its own, it needs to be applied to a time series. Can I apply it to  $\varepsilon_t$  using the expression above? ... clearly NO! not even for the AR(1) case:

$$\frac{1}{1 - \rho_1 L} \varepsilon_t$$

- I know how to deal with  $L$  when it is all by itself (to any power) multiplying a time series. Then, the question is can I express  $\pi(L)^{-1}$  in such a way that I can get that?
- Luckily for us the answer is YES!. We can take advantage of **Taylor’s Theorem** so we can express  $\pi(L)^{-1}$  as a power series that converges in the largest open disk to  $\pi(L)^{-1}$ .
- For doing that, we need to check if the  $\pi(L)^{-1}$  is a **regular function**.
- If so, we can express  $\pi(L)^{-1}$  as an infinite sum that converges and we are able to analyze the features of  $X_t$  since we will be able to apply  $L$  to  $\varepsilon_t$ .

## 7.3 Is the inverse of the lag polynomial well defined?

### 7.3.1 The roots of the characteristic polynomial

- If  $\pi(L) = 0$ , we have that  $\pi(L)^{-1} = \infty$  so it is not well defined and  $\pi(L)^{-1}$  cannot be used
- Thus, I need to state that  $L$  is not taking the values of the roots in the polynomial  $\pi(L)$ . Otherwise I cannot invert it
- I know that  $L$  is the lag operator but when analyzing his properties inside a function we treat it as a variable, taking the most general case  $L \in \mathbb{C}$
- We want to know when do we get  $\pi(L) = 0$
- Finding the values for  $L$  that makes  $\pi(L) = 0$  it is called finding the **roots**
- The roots of a polynomial are the values of the variable in the polynomial that makes the polynomial equal to 0
- Thus, we are interested in finding the roots of this polynomial
- To find the roots, we just need to find the roots in the characteristic function of  $\pi(z)$

- Let's denote the roots of the characteristic function  $\pi(z)$  as  $z_i \forall i = 1, 2, 3 \dots, p$ .
- Once the roots are found, I just have to write that for any  $z \in \mathbb{C} \setminus \{z_1, z_2, z_3, \dots, z_p\}$ , the characteristic function  $\pi(z)^{-1}$  is well defined.

### 7.3.2 An alternative characteristic polynomial: The reflected polynomial

The roots in the characteristic polynomial  $\pi(z)$  are equal to the inverse of the roots in its reflected polynomial  $\tilde{\pi}(z)$ .

$$\tilde{\pi}(z) = z^p \pi(z^{-1}) \quad (17)$$

- $\tilde{\pi}(z)$  is known as the **Reflected polynomial** of  $\pi(z)$ . Therefore,  $\pi(z)$  is also the **Reflected polynomial** of  $\tilde{\pi}(z)$ :

$$\pi(z) = z^p \tilde{\pi}(z^{-1}) \quad (18)$$

Example:

$$\pi(z) = 1 - \rho_1 z - \rho_2 z^2$$

$$\pi(z^{-1}) = 1 - \rho_1 z^{-1} - \rho_2 z^{-2}$$

$$\tilde{\pi}(z^1) = z^2(1 - \rho_1 z^{-1} - \rho_2 z^{-2})$$

Thus, (17) holds:

$$\tilde{\pi}(z^1) = z^2 - \rho_1 z^1 - \rho_2$$

Now,

$$\tilde{\pi}(z^{-1}) = z^{-2} - \rho_1 z^{-1} - \rho_2$$

$$\pi(L) = z^2(z^{-2} - \rho_1 z^{-1} - \rho_2)$$

Thus, (18) holds:

$$\pi(z) = 1 - \rho_1 z - \rho_2 z^2$$

actually follows the same strategy to define the characteristic polynomial:

$$- \tilde{\pi}(z) \equiv P(\lambda)$$

- Let's denote  $\lambda_i$  as the roots in the characteristic polynomial  $\tilde{\pi}(z)$
- Given (17), we know that the characteristic polynomials are related. So, are the roots also related?
- Indeed the roots are related. When finding the roots for  $\tilde{\pi}(z)$ , we are finding the values of  $z$  such that:

$$\tilde{\pi}(z) = 0$$

From (17), it is the same as finding the values of  $z$  such that:

$$z^p \pi(z^{-1}) = 0$$

Since those values cannot be zero, then we are finding the values of  $z$  such that:

$$\pi(z^{-1}) = 0$$

Let's denote  $z_i$  as the roots of the characteristic polynomial,  $\pi(z)$ . Given this notation, the roots of  $\pi(z^{-1})$  should be  $z_i^{-1}$  since  $\pi(z^{-1})$  is  $\pi(z)$  using the inverse variable. Therefore, we have that:

$$\lambda_i = z_i^{-1} \quad (19)$$

## 7.4 Is the function a regular one?

- Notice that the lag polynomial is also a function.
- Checking if a complex function is regular follows Section 2.
- However, there is an issue. It gets more complicated to analyze  $\pi(z)^{-1}$  when  $p$  is greater than 1.
- Then, is there any trick to deal with it?
- The answer is YES! We will take advantage of the roots we have found for establishing when  $\pi(z)^{-1}$  is well defined by applying the **Fundamental Theorem of Algebra**

### 7.4.1 Inverting $\pi(L)$

- Notice that we can always use the roots to rewrite any characteristic equation (**Fundamental Theorem of Algebra**: Finding roots of a polynomial is therefore equivalent to polynomial factorization into factors of degree 1.) e.g.:

$$P(x) = x^3 - 2x^2 - x + 2$$

, in this case the three roots of the polynomial  $P(x)$  are calculated such that

$$P(z) = 0$$

so, we have that the roots are:  $z_1 = 2$ ,  $z_2 = 1$  and  $z_3 = -1$ . Thus, we can rewrite the polynomial as follows:

$$x^3 - 2x^2 - x + 2 = (x - z_1)(x - z_2)(x - z_3)$$

$$x^3 - 2x^2 - x + 2 = (x - 2)(x - 1)(x + 1)$$

$$P(x) = (x - 2)(x - 1)(x + 1)$$

- So, for the reflected characteristic polynomial  $\tilde{P}(x)$ , we have:

$$\tilde{P}(x) = (x - \frac{1}{2})(x - 1)(x + 1)$$

- Further, taking into account that  $P(x)$  is also the reflected polynomial of  $\tilde{P}(x)$  :

$$P(x) = x^3 \tilde{P}(x^{-1})$$

We can also express  $P(x)$  in terms of the roots of its reflected characteristic polynomial as:

$$P(x) = \left(1 - \frac{1}{2}x\right)(1-x)(1+x)$$

for further on this, please refer to Section 1.7.

- Going back to the general case, we know that  $\lambda_i$  are the roots of the characteristic polynomial  $\tilde{\pi}(z)$ . Thus,  $\lambda_i$  are the values of  $z$  such that  $\tilde{\pi}(z) = 0$
- We know that  $\lambda_i = z_i^{-1}$  and  $z_i$  are the roots of the characteristic polynomial  $\pi(z)$
- Thus, we can also write:

$$\pi(z) = \left(1 - \frac{1}{z_1}z\right)\left(1 - \frac{1}{z_2}z\right)\left(1 - \frac{1}{z_3}z\right)\dots\left(1 - \frac{1}{z_p}z\right) \quad (20)$$

- Therefore, we can express  $\pi(z)^{-1}$  in a nicer way:

$$\pi(z)^{-1} = \left(\frac{1}{(1-\lambda_1 z)}\right)\left(\frac{1}{(1-\lambda_2 z)}\right)\left(\frac{1}{(1-\lambda_3 z)}\right)\dots\left(\frac{1}{(1-\lambda_p z)}\right) \quad (21)$$

- This is a nicer way because to know if  $\pi(z)^{-1}$  is a regular function, we just have to check if each  $\frac{1}{(1-\lambda_i z)}$  is a regular function: If two functions  $f(z)$  and  $g(z)$  are analytic in a domain  $D$ , then their sum and their product are both analytic in  $D$ . For further on this, check Section 1.10
- Based on our results in Section 2.4, we conclude that the characteristic function  $\pi(z)^{-1}$  is well defined and analytic for all  $z \in \mathbb{C} \setminus \mathbb{L}$ , in which:

$$\mathbb{L} = \left\{ z \in \mathbb{C} : \operatorname{Re}(z) \geq \frac{1}{\lambda_i} \text{ and } \operatorname{Im}(z) = 0 \right\}$$

or

$$\mathbb{L} = \{z \in \mathbb{C} : \operatorname{Re}(z) \geq z_i \text{ and } \operatorname{Im}(z) = 0\}$$

- Thus, from **Taylor's Theorem** we have that  $\pi(z)^{-1}$  can be expressed as a power series with  $z_0 \in \mathbb{C} \setminus \mathbb{L}$ :

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!} (z - z_0)^n$$

where the series converges on any disk  $|z - z_0| < r$  contained in  $\mathbb{C} \setminus \mathbb{L}$ . Since  $0 \in \mathbb{C} \setminus \mathbb{L}$ , we select  $z_0 = 0$  (largest open disk). Thus, the series converges on the largest open disk only if:

$$|\lambda_i z| < 1$$

- We know that we can find the values for the roots  $\lambda_i$ . However, a natural question that emerges at this point is what is  $z$  doing in there. Well, remember that  $z$  is just the argument in the characteristic function and it is originally the variable  $L$ . But it doesn't help too much to know that. I mean what is the intuition of choosing a value for  $L$ . It was supposed to be a trick just to get along with time series. We know for sure that we need  $z$  such that:

$$z \in \mathbb{C} \setminus \mathbb{L}$$

and

$$|\lambda_i z| < 1$$

A set of values for  $z$  that satisfies both conditions are those such that  $|z| = 1$ . That is, when  $\pi(z)^{-1}$  is evaluated on the **unit circle**. We will pick up this because we need it for the **Spectral density**.

#### 7.4.2 Is the characteristic polynomial $\pi(z)^{-1}$ analytic on the unit circle, $|z| = 1$ ?

- So, we have that  $\pi(z)^{-1}$  can be expressed as an infinite sum as long as the value of  $z$  is such that:

$$z \in \mathbb{C} \setminus \mathbb{L}$$

and

$$|\lambda_i z| < 1$$

- We said that we usually pick up  $|z| = 1$ , but why?
- To better understand this, recall that we are interested in analyzing the properties of  $X_t$
- Let's assume that  $\pi(z)^{-1}$  is a regular function that converges on the largest open disk. That is, we have  $z$  and  $\lambda_i$  satisfying the conditions listed above.
- If so, we have for AR(1):

$$\pi(z)^{-1} = \sum_{i=0}^{\infty} (\lambda_1 z)^i$$

Notice that for AR(1),  $\lambda_1 = \rho$ . Then, by using the inverse lag polynomial. That using  $L$  instead of  $z$ , we have:

$$\pi(L)^{-1} = \sum_{i=0}^{\infty} (\rho L)^i$$

Replacing it on the DGP for  $X_t$ :

$$X_t = \sum_{i=0}^{\infty} (\rho L)^i \varepsilon_t$$

By using applying the lag operator to  $\varepsilon_t$ :

$$X_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}$$

- Then, to know the properties of  $X_t$ , we need to study the properties of:

$$\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}$$

As this is an infinite sum, we would like to know if this converges or not. **Be careful!** Having  $\pi(z)^{-1}$  as a regular function is just to be able to express the function as an infinite sum. However, after replacing the infinite sum for  $z = L$  in the DGP of  $X_t$ , we have an infinite sum that involves  $\varepsilon_t$ . So, to fully analyze  $X_t$ , we need to analyze the infinite sum involving  $\varepsilon_t$ .

- Given that this infinite sum involves a random variable,  $\varepsilon_t$ , we know that we are talking about convergence in probability and converges in  $q$ -th moment. (See Section 4).
- Thus we would like to be able somehow to apply Theorem 3
- We said that we needed to assume  $z$  and  $\lambda_i$  to satisfy the conditions for  $\pi(z)^{-1}$  to be regular. A set of values that satisfies both requirements is given by:

$$|z| = 1$$

$$|\lambda_i| < 1$$

- If so, we have that for the AR(1), we have:

$$\sum_{i=0}^{\infty} \rho^i < \infty$$

Given that  $|\lambda_1| < 1$ , we also have that

$$\sum_{i=0}^{\infty} |\rho^i| < \infty$$

Lets' denote:

$$\psi_j = \rho^j$$

Thus, we have one of the two conditions of Theorem 3.

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

The other condition is satisfied because we assumed  $\varepsilon_t$  is i.i.d. with zero mean and variance 1. Thus, it is a particular case of the one in Theorem 3.

- From Theorem 3, we immediately know that converges in the 1st moment and therefore by Markov inequality it converges in probability.
- Further, from Theorem 4, we immediately know that it converges in the second moment as well.
- Thus, we know that  $X_t$  has a finite second moment, which implies a finite first moment. That is, it has a finite variance and a finite mean.
- In short,
  1. We need to be able to express  $\pi(L)^{-1}$  as an infinite sum to analyze  $X_t$ . Otherwise we simply cannot do it.
  2. Thus we need  $z$  and  $|\lambda_i|$  to be such that:

$$z \in \mathbb{C} \setminus \mathbb{L}$$

and

$$|\lambda_i z| < 1$$

3. Assuming those, it is not enough given that now we are dealing with an infinite sum that involve  $\varepsilon_t$
  4. If  $|\lambda_1| < 1$ , and given the properties of  $\varepsilon_t$ , the conditions in Theorem 3 and Theorem 4 are satisfied and we conclude that  $X_t$  has a finite variance
- However, we would be using the same tools and therefore landing to the same conclusions if we just proceed to analyze if the characteristic function  $\pi(z)^{-1}$  is **holomorphic on the unit circle** and converges on the largest open disk.
  - Thus, to know is  $X_t$  converges on the second moment, we just need to check if the roots,  $\lambda_i$ , in the characteristic polynomial  $\tilde{\pi}(z)$  are such that:

$$|\lambda_i| < 1$$

## 7.5 Is the process stationary?

- From the Theorem 6, we have that  $X_t$  is stationary if:

1.

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

2.

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

3.

$$\varepsilon_t \text{ is stationary}$$

- The first requirement is to be able to express  $X_t$  as an infinite sum, which can be achieved if  $|\lambda_i| < 1$
- The second requirement is immediately satisfied if  $|\lambda_i| < 1$  given that  $\psi_j$  is just a multiplication of  $\lambda_i$  (see Section 7.4.1)

- The third requirement is satisfied by definition of  $\varepsilon_t$
- Therefore, we yield to the same conditions as in the previous section: For  $X_t$  to be stationary we only need to check if the roots,  $\lambda_i$ , in the characteristic polynomial  $\tilde{\pi}(z)$  are such that:

$$|\lambda_i| < 1$$

, which is the same as saying:

$$\lambda_i \in \text{the open unit disk}$$

## 7.6 The Spectral Density

- If the process is stationary, we can find the spectral density of  $X_t$  by applying Theorem 8.
- Now, it is crystal clear why we care about the case when the lag polynomial is on the unit circle.

## 7.7 Impulse response Function

- Consider the mean zero (de-means) weakly stationary AR(1) model.

$$y_t = \beta_1 y_{t-1} + \varepsilon_t$$

- We might want to know what we should expect the future value of  $y_{t+k}$  to look like if  $y_t$  were one unit larger holding all  $y_{t-j}$   $j > 0$  fixed.
- This is equivalent to asking how we should expect  $y_{t+k}$  to change given a one unit change in  $\varepsilon_t$ .
- **WHY?:** From the DGP of  $Y_t$  if we hold  $y_{t-1}$  fixed, the only way to increase  $y_t$  is through increases in  $\varepsilon_t$
- The impulse response function is the path that  $y$  follows if it is kicked by a single unit shock  $\varepsilon_t$ , i.e.,  $\varepsilon_{t-j} = 0, \varepsilon_t = 1, \varepsilon_{t+j} = 0$ .
- This function is interesting since it allows us to start thinking about "causes" and "effects"
- For example, you might compute the response of GDP to a shock to GDP and interpret the result as the "effect" on GDP of a shock over time.
- The  $MA(\infty)$  representation is the same thing as the impulse response function.
- Thus, for a stationary AR(p) process, we know that its stationary solution is given by:

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

This is also known as the  $MA(\infty)$  representation or the impulse-response function. From this representation we can calculate the effect of a single shock overtime.

For instance, the impact of a one unit one unit increase in  $\varepsilon_{t-k}$  on  $y_t$  is given by:

$$\frac{dy_t}{d\varepsilon_{t-k}} = \psi_k$$

- When we plot the change in  $y_t$  given a one unit shock to  $\varepsilon_{t-k}$  we call this an **impulse response function (IRF)**.

- A stationary process has the property that the effect of a shock does not last forever. That is, if a shock occurs at  $t$ , eventually (as  $t$  goes to infinity)  $\psi_\infty$  will be zero:

$$\frac{dy_{t+k}}{d\varepsilon_t} = \psi_k \rightarrow 0 \quad \text{as } k \rightarrow \infty$$

- We can clearly see that with the stationary AR(1) case, in which:

$$\psi_j = \rho^j$$

and as it is stationary we have  $|\rho| < 1$ . So:

$$\rho^j \rightarrow 0 \quad \text{as } j \rightarrow \infty$$

- Figure 7.1 represents the IRF for a stationary AR(1) process.
- Interpretation? If there is an unexpected increase of one additional dollar traded, how much does dollar volume change  $k$ -periods ahead?

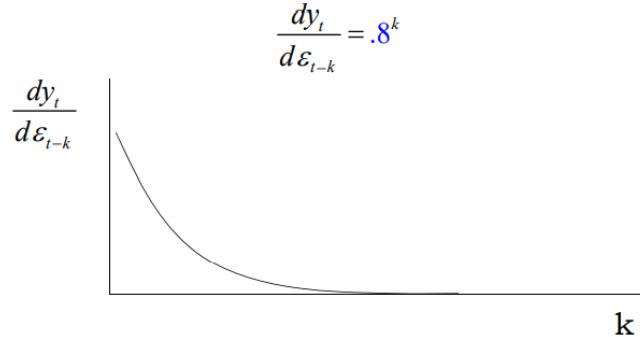


Figure 24: IRF for a stationary AR(1) process

- Having an IRF vanishing over time is a consequence of being a stationary process. However being stationary is a sufficient condition but not a necessary one. That is, we can have a process with IRF vanishing over time that are nonstationary.

## 8 Testing for the presence of Unit Root

### 8.1 Motivation

- As an example, let's focus on the AR(1) case.
- We know that if  $|\rho| = 1$ ,  $y_t$  is nonstationary.
- $|\rho| = 1$  is when the root is on the unit disk (i.e., we have a unit root)
- Thus, we would like to test the hypothesis if the process is stationary or not.
- That is, we would like to know if we have a unit root.
- We would like to apply the usual t-test or Wald test. Can we?

- Sadly, we cannot.
- Consider the simple AR(1) model

$$y_t = \phi y_{t-1} + \varepsilon_t, \text{ where } \varepsilon_t \sim WN(0, \sigma^2)$$

The hypotheses of interest are

$$\begin{aligned} H_0 : \quad & \phi = 1 (\text{ unit root in } \phi(z) = 0) \Rightarrow y_t \sim I(1) \\ H_1 : \quad & |\phi| < 1 \Rightarrow y_t \sim I(0) \end{aligned}$$

The test statistic is

$$t_{\phi=1} = \frac{\hat{\phi} - 1}{SE(\hat{\phi})}$$

where  $\hat{\phi}$  is the estimator of  $\phi$  and  $SE(\hat{\phi})$  is the usual standard error estimate.

The test is a one-sided left tail test. If  $\{y_t\}$  is stationary (i.e.,  $|\phi| < 1$ ) then it can be shown

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} N(0, (1 - \phi^2))$$

or

$$\hat{\phi} \stackrel{A}{\sim} N\left(\phi, \frac{1}{T}(1 - \phi^2)\right)$$

and it follows that  $t_{\phi=1} \stackrel{A}{\sim} N(0, 1)$ . However, under the null hypothesis of nonstationarity the above result gives

$$\hat{\phi} \stackrel{A}{\sim} N(1, 0)$$

which clearly does not make any sense.

## 8.2 The most common unit root tests

- Testing for the presence of a unit root in a time series is now a common starting point that is available as an option in several popular statistical packages.
- For some surveys, see Stock (1994) and Haldrup and Jansson (2006).
- Among the most used statistical tests is the Augmented DickeyFuller (*ADF*) proposed by Said and Dickey (1984) based on Dickey and Fuller (1979).
- Another set of statistics is the family of  $M$  tests that was originally proposed by Stock (1999) and further analyzed by Perron and Ng (1996).
- The  $M$  tests are composed of three statistics:  $MZ_{\hat{\alpha}}, MSB, MZ_{t_{\hat{\alpha}}}$ .
- Elliott, Rothenberg and Stock (1996) proposed a feasible point optimal test ( $P_T^{GLS}$ ) and an  $ADF^{GLS}$  test, both of which are constructed using GLS detrended data in order to increase the power performance of the tests.
- Ng and Perron (2001) used the same strategy applied to the family of  $M$  tests, as well to a feasible optimal point test denominated  $MP_T^{GLS}$ .

### 8.3 The Dickey-Fuller (DF) Statistic

- Based on Dickey and Fuller (1979)
- Let's go back to our beloved AR(1) case:

$$y_t = \rho y_{t-1} + u_t$$

where  $y_t$  is the variable of interest,  $t$  is the time index,  $\rho$  is a coefficient, and  $u_t$  is the error term (assumed to be white noise).

- A unit root is present if  $\rho = 1$ . The model would be non-stationary in this case.
- We can rewrite the model such that:

$$\Delta y_t = (\rho - 1)y_{t-1} + u_t = \delta y_{t-1} + u_t$$

where  $\Delta$  is the first difference operator and  $\delta \equiv \rho - 1$ .

- This model can be estimated, and testing for a unit root is equivalent to testing  $\delta = 0$ .
- It is not possible to use standard t-distribution to provide critical values because under the null,  $y_{t-1}$  is not stationary and therefore CLT fails.
- Under the null this statistic  $t$  has a specific distribution simply known as the **Dickey-Fuller distribution**.
- Therefore to reject the null we just have to compare the the  $t$  statistic againsts the critical values of the Dickey-Fuller distribution.
- It is worth to point out that the critical values depend on the deterministic component of the DGP.
- There are three main versions of the test:

1. Test for a unit root:

$$\Delta y_t = \delta y_{t-1} + u_t$$

2. Test for a unit root with constant:

$$\Delta y_t = a_0 + \delta y_{t-1} + u_t$$

3. Test for a unit root with constant and deterministic time trend:

$$\Delta y_t = a_0 + a_1 t + \delta y_{t-1} + u_t$$

### 8.4 The Augmented Dickey Fuller - ADF

- Based on Said and Dickey (1984)
- The DF test assumes that  $u_t$  is an i.i.d. process and that the DGP for  $y_t$  is an AR(1) process
- However, in general it might be the case that the DGP process for  $y_t$  is AR(P) and that the  $u_t$  is serial correlated.
- If so, we should use the *ADF* equation

$$\Delta y_t = a_0 + a_1 t + \delta y_{t-1} + \beta_1 \Delta y_{t-1} + \cdots + \beta_k \Delta y_{t-k} + \varepsilon_t \quad (22)$$

- For testing about the unit root, we use the same critical values from the DF distribution.
- The  $t$  distribution can be used for testing about  $\beta_i$  for  $i = 1, 1, 2, \dots, k$

- However, a question remains....what is the value of  $k$ ?
- Using Monte Carlo experiments, Schwert (1989) showed that the value of  $k$  has important implications for the size and power of the *ADF* test, in particular when there is strong negative moving average correlation in the residuals.
- Ng and Perron (1995) constitutes the first study dealing with the analysis of lag-length selection using different criteria.
  - They prove that the choice of the data-dependent rule has a bearing on the size and power of the test.
  - Moreover, they show that information-based rules such as the Akaike information criteria (*AIC*) and Bayesian information criteria (*BIC*) tend to select values of  $k$  that are consistently smaller than those chosen through sequential testing for the significance of coefficients on additional lags (*t-sig* method), and that the size distortions associated with the former methods are correspondingly larger.

## 8.5 The Phillips Perron Test

- The DF and ADF forces us to choose the optimal number of lags of the dependent variable or the forecast variable in question.
- Moreover, the functional form of the model must also be specified. Do we include a trend, is there an intercept, do we use both a time trend and an intercept?
- The Dickey Fuller and the Augmented Dickey fuller have been found to have low power (The power of a test is the probability rejecting the null hypothesis when it is false) in some circumstances.
  - Consider a model where  $\phi = 0.95$ .
  - By all accounts, it meets our criteria for a stationary process but the result of the test may indicate non-stationarity especially in data with a low sample size.
- Based on Phillips and Perron (1988)
- The *PP* Test corrects for any serial correlation and heteroscedasticity in the errors by some direct modification to the test statistics.
- This modification is a nonparametric one.
- The *PP* has no need to specify the lag length.

## 8.6 The *ADF – GLS* test

- In their 1996 Econometrica article, Elliott, Rothenberg and Stock (ERS) proposed an efficient test, modifying the Dickey-Fuller test statistic using a generalized least squares (GLS) rationale.
- They demonstrate that this modified test has the best overall performance in terms of small-sample size and power, conclusively dominating the ordinary Dickey-Fuller test.
- In particular, Elliott et al. find that their "DF-GLS" test "has substantially improved power when an unknown mean or trend is present." (1996, p.813)
- Let  $z_t = (1, t)$ .
- For the time series  $y_t$ , regress  $[y_1, (1 - \alpha L)y_2, \dots, (1 - \alpha L)y_T]$  on  $[z_1, (1 - \alpha L)z_2, \dots, (1 - \alpha L)z_T]$  yielding

$$\tilde{\beta}_{GLS}$$

where  $\alpha = 1 + \bar{c}/T$ , and  $\bar{c} = -13.5$  for the detrended statistic.

- Detrended  $\tilde{y}_t = y_t - z'_t \tilde{\beta}_{GLS}$  is then employed in the *ADF* equation, with no intercept nor time trend.

- The  $t$ -statistic on  $\tilde{y}_{t-1}$  is the DF-GLS statistic.
- For the demeaned case, the  $t$  is omitted from  $z_t$ , and  $\bar{c} = -7.0$ .
- To reject the null we compare the  $t$  statistic with the critical values of the  $DF-GLS$  distribution that depends on the deterministic component.
- ERS (1996) show that the choice of  $k$  has a considerable effect on the size (size of a test is the probability of incorrectly rejecting the null hypothesis if it is true) of  $ADF^{GLS}$ .
- In order to select  $k$ , they try  $AIC$ ,  $BIC$  and sequential likelihood ratio statistics. Finally, they use the  $BIC$  to select  $k$  by setting the lower bound at 3, because with zero as the lower bound larger size distortions result.
- Ng and Perron (2001) consider a class of Modified Information Criteria ( $MIC$ ) with a penalty factor that is sample dependent.
- In the Monte-Carlo experiments, they find that the  $MIC$  yields huge size improvements on the  $ADF^{GLS}$ .
- They also show that both the use of the  $MIC$  and allowing for GLS data detrending in the  $M$  test results in a class of  $M^{GLS}$  tests that have desirable size and power properties.
- In conclusion, the  $MIC$  (in particular the  $MAIC$  version) is a superior rule for selecting lag length.

## 8.7 The $M$ statistics

- Perron and NG (1996) showed that for a  $AR(1)$  and  $MA(1)$  with negative coefficient close to  $-1$ , the  $PP$  test exhibits strong size distortions (it did not reject the null when the null was false).
- Thus, they proposed the  $M$  tests that was originally proposed by Stock (1999).
- The  $M$  statistics are composed of three statistics:  $MZ_{\hat{\alpha}}$ ,  $MSB$ ,  $MZ_{t_{\hat{\alpha}}}$ .
- They performed better than the  $PP$  statistic.
- While the power gains of the  $DF^{GLS}$  from using  $GLS$  detrended data are impressive, simulations also show that the test exhibits strong size distortions when dealing with an  $MA(1)$  process with a negative coefficient.
- Since the power gains from the  $DF^{GLS}$  over the  $DF$  come from the use of  $GLS$  detrended data, it is natural to consider the  $M$  tests under  $GLS$  detrending.
- Ng and Perron (2001) analasymptotic properties of the  $M^{GLS}$  tests.
- Ng and Perron (2001) extend the  $M$  tests developed in Perron and Ng (1996) to allow for  $GLS$  detrending of the data.
- They also show that both the use of the  $MIC$  and allowing for  $GLS$  data detrending in the  $M$  test results in a class of  $M^{GLS}$  tests that have desirable size and power properties.
- In conclusion, the  $MIC$  (in particular the  $MAIC$  version) is a superior rule for selecting lag length.

# 9 VAR ( $p$ )

## 9.1 Motivation

- So far we have focused mostly on models where  $y$  depends on past observations of  $y$ .
- More generally we might want to consider models for more than one variable.

- If we only care about forecasting one series but want to use information from another series we can estimate an ARMA model and include additional explanatory variables.
- For example if  $y_t$  is the series of interest, but we think  $x_t$  might be useful we can estimate models like

$$y_t = \beta_0 + \beta_1 y_{t-1} + \gamma x_{t-1} + \varepsilon_t$$

- This model can be fit by least squares. Our dependent variable is  $y_t$  and the independent variables are  $y_{t-1}$  and  $x_{t-1}$
- Once the model is fit, the one-step ahead forecast is given by:

$$E(y_{t+1} | F_t) = \beta_0 + \beta_1 E(y_t | F_t) + \gamma E(x_t | F_t) = \beta_0 + \beta_1 y_t + \gamma x_t$$

- Just like the simple AR model, the one step ahead forecast variance is  $\sigma_\varepsilon^2$ .
- A **joint model** for  $x_t$  and  $y_t$  is required if we are interested in multiple step ahead forecasts, or if we are interested in feedback effects from one process to the other.
- For example, if we want to 2 step ahead forecast for  $y_t$ , we are looking for

$$E(y_{t+2} | F_t) = \beta_0 + \beta_1 E(y_{t+1} | F_t) + \gamma E(x_{t+1} | F_t)$$

Then, the obvious question is what do we use for:

$$E(x_{t+1} | F_t)$$

since  $x_{t+1}$  is not known at  $t$

- Answer: We need a model for  $x$  as well
- Before proceeding to the next section, it is important to review Appendix A.

## 9.2 Companion Form Representation of an AR(p) Model

- Can we convert an AR(p) back into an “AR(1)” type model?
- Consider the AR(p) model:

$$y_t = \sum_{j=1}^p \beta_j y_{t-j} + \varepsilon_t$$

- Next we define the new vectors and matrix:

$$\xi_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}_{p \times 1} \quad v_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{p \times 1} \quad F = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_{p-1} & \beta_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & & \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}_{p \times p}$$

- Then we can write the AR(p) model as the following first order model:

$$\xi_t = F \xi_{t-1} + v_t \tag{23}$$

- This is known as the **Stacked VAR(1)** or the **companion form representation**

### 9.2.1 Stationarity of the Stacked VAR(1)

- From section 7.5, we know that the AR(p) is stationary if the roots of the characteristic polynomial,  $\lambda_i$ , :

$$\tilde{\pi}(z) = z^p \pi(z^{-1})$$

are inside the unit circle:

$$|\lambda_i| < 1$$

or if the roots of the characteristic polynomial,  $z_i$ , :

$$\pi(z)$$

are outside the unit circle:

$$|z_i| > 1$$

- From the stacked form, we have that  $F$  is an **square matrix** and thus, we could find its **eigenvalues**:

$$|F - \lambda I| = 0$$

, which is the same as  $|\lambda I - F| = 0$

- We know that  $|F - \lambda I|$  is a polynomial of degree  $p$ . Given the composition of  $F$ , we have that:

$$|F - \lambda I| = \tilde{\pi}(z)$$

, so to know if the stacked VAR(1) is stationary is the same as knowing if the eigenvalues of  $F$  are inside the unit circle:

$$\lambda_i \in \text{the open unit disk}$$

### 9.3 The VAR(1) model with 2 variables

- Suppose that we have 2 variables that we observe at time period  $t$  and we consider the joint model:

$$\begin{aligned} x_t &= \beta_0^x + \beta_1^x x_{t-1} + \beta_2^x y_{t-1} + u_t^x \\ y_t &= \beta_0^y + \beta_1^y x_{t-1} + \beta_2^y y_{t-1} + u_t^y \end{aligned} \tag{24}$$

- Each equation is like an AR(1) model with one other explanatory variable.
- Each equation depends on its own lag and the lag of the other variable.
- We also now have two errors, one for each equation:  $u_t^x$  and  $u_t^y$
- Since  $x$  depends on  $y$  and  $y$  depends on  $x$ , a more thorough understanding of dynamics and forecasting requires us to jointly consider  $x$  and  $y$  in the system of equations.

### 9.3.1 Matrix Notation

- We can write the model in matrix notation:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \beta_0^x \\ \beta_0^y \end{bmatrix} + \begin{bmatrix} \beta_1^x & \beta_2^x \\ \beta_1^y & \beta_2^y \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} u_t^x \\ u_t^y \end{bmatrix} \quad (25)$$

- By defining the following vectors and matrices we end up with a very simple form for the VAR(1) model. Let:

$$\mathbf{y}_t = \begin{bmatrix} x_t \\ y_t \end{bmatrix}, \mathbf{v}_t = \begin{bmatrix} u_t^x \\ u_t^y \end{bmatrix}, \boldsymbol{\beta}_0 = \begin{bmatrix} \beta_0^x \\ \beta_0^y \end{bmatrix}, \text{ and } \boldsymbol{\beta}_1 = \begin{bmatrix} \beta_1^x & \beta_2^x \\ \beta_1^y & \beta_2^y \end{bmatrix}$$

Then, the model in matrix from is given by:

$$\mathbf{y}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{y}_{t-1} + \mathbf{v}_t \quad (26)$$

### 9.3.2 Assumptions on the errors

- The errors are white noise and uncorrelated with lags of the other errors.
- $u_t^x$  is uncorrelated with  $u_{t-j}^x$  and  $u_{t-j}^y$  for  $j \neq 0$ .
- $u_t^y$  is uncorrelated with  $u_{t-j}^x$  and  $u_{t-j}^y$  for  $j \neq 0$ .
- However there may be contemporaneously correlated. If so, we call them **reduced shocks** since they come from a more compelling model that allows for a contemporaneous relationship between  $x_t$  and  $y_t$ . It might be the case that the system is such that there is contemporaneous relationship:

$$\begin{aligned} x_t &= \beta_0^x + \alpha^x y_t + \beta_1^x x_{t-1} + \beta_2^x y_{t-1} + \varepsilon_t^x \\ y_t &= \beta_0^y + \alpha^y x_t + \beta_1^y x_{t-1} + \beta_2^y y_{t-1} + \varepsilon_t^y \end{aligned} \quad (27)$$

Let's, define:

$$\begin{aligned} \boldsymbol{\alpha} &= \begin{bmatrix} 1 & \alpha^x \\ \alpha^y & 1 \end{bmatrix} \\ \boldsymbol{\varepsilon}_t &= \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^y \end{bmatrix} \end{aligned}$$

Thus,

$$\mathbf{y}_t = \boldsymbol{\alpha}^{-1} \boldsymbol{\beta}_0 + \boldsymbol{\alpha}^{-1} \boldsymbol{\beta}_1 \mathbf{y}_{t-1} + \boldsymbol{\alpha}^{-1} \boldsymbol{\varepsilon}_t \quad (28)$$

so,

$$\mathbf{v}_t = \boldsymbol{\alpha}^{-1} \boldsymbol{\varepsilon}_t$$

If we assume that  $\boldsymbol{\varepsilon}_t$  are **structural shocks** (i.e. the very first and purest shocks in the economy that are not related to each other for any  $t$ ), we have the the **reduced shocks** are such that there are contemporaneously related:

$$\sigma_{u_x u_y} = E[u_t^x u_t^y] - E[u_t^x]E[u_t^y] = E[(\varepsilon_t^x + \alpha^x \varepsilon_t^y)(\alpha^y \varepsilon_t^x + \varepsilon_t^y)] = \alpha^x \sigma_{\varepsilon_x}^2 + \alpha^y \sigma_{\varepsilon_y}^2 \neq 0$$

- $\Omega$  denotes the variance covariance matrix of the reduced error:

$$\Omega = \begin{bmatrix} \sigma_{u_x}^2 & \sigma_{u_x u_y} \\ \sigma_{u_x u_y} & \sigma_{u_y}^2 \end{bmatrix}$$

### 9.3.3 Solving the system of equations: Inverting the matrix

- First, it is worth to point out the it can also be solved by iterating.
- W.O.L.O.G. let's assume the intercept is zero
- Now we have:

$$\mathbf{y}_t = \boldsymbol{\beta}_1 \mathbf{y}_{t-1} + \mathbf{v}_t$$

- Let's use the lag operator:

$$\mathbf{y}_t = \boldsymbol{\beta}_1 L \mathbf{y}_t + \mathbf{v}_t$$

$$(I - \boldsymbol{\beta}_1 L) \mathbf{y}_t = \mathbf{v}_t$$

- So, we would like to be able to express the solution as:

$$\mathbf{y}_t = (I - \boldsymbol{\beta}_1 L)^{-1} \mathbf{v}_t$$

- To do so, first we need to check if the inverse exist. That is, if  $(I - \boldsymbol{\beta}_1 L)$  is a nonsingular matrix. So, we need to check if:

1.  $(I - \boldsymbol{\beta}_1 L)$  is a square matrix
2.  $|(I - \boldsymbol{\beta}_1 L)| \neq 0$  so we have:

$$(I - \boldsymbol{\beta}_1 L)^{-1} = \frac{1}{|(I - \boldsymbol{\beta}_1 L)|} C'$$

, in which  $C'$  is the transpose of the cofactor matrix

Condition 1 is satisfied given that  $\boldsymbol{\beta}_1$  is an square matrix.

- For condition 2, notice that  $|(I - \boldsymbol{\beta}_1 L)|$  is a second order polynomial on  $L$ :

$$|(I - \boldsymbol{\beta}_1 L)| = \left| \begin{bmatrix} 1 - \beta_1^x L & \beta_2^x L \\ \beta_1^y L & 1 - \beta_2^y L \end{bmatrix} \right|$$

$$|(I - \boldsymbol{\beta}_1 L)| = (1 - \beta_1^x L)(1 - \beta_2^y L) - \beta_2^x L \beta_1^y L$$

$$|(I - \boldsymbol{\beta}_1 L)| = 1 - \beta_2^y L - \beta_1^x L + \beta_1^x L \beta_2^y L - \beta_2^x L \beta_1^y L^2$$

$$|(I - \boldsymbol{\beta}_1 L)| = 1 - (\beta_2^y + \beta_1^x)L - (\beta_2^x \beta_1^y - \beta_1^x \beta_2^y)L^2$$

Thus, we have that:

$$|(I - \boldsymbol{\beta}_1 L)| = \pi(L) = 1 - \pi_1 L - \pi_2 L^2 \quad (29)$$

- Thus, condition 2 is the same as checking that  $\pi(L)^{-1}$  exists.
- Further, we are in a similar situation as is the usual AR(p) case. That is, we would like:
  1.  $\pi(L)^{-1} < \infty$
  2.  $(1 - \pi_1 L - \pi_2 L^2)^{-1}$  has no meaning in Time Series (i.e. How do I multiply it to the error term?). So, we would like  $\pi(z)^{-1}$  to be a regular function so we can use Taylor's Theorem around 0 and express it as an infinite sum.
  3. Check if the power series converges when  $|z| = 1$  (So we know that the infinite sum time the error term (i.e. the RHS) converges in the second moment)
- We know that we can find the roots of the characteristic polynomial  $\pi(z)$ ,  $z_i$ , and express it as:

$$\pi(z) = \left(1 - \frac{1}{z_1}z\right) \left(1 - \frac{1}{z_2}z\right)$$

- Further, we know that, the characteristic function  $(1 - \rho z)^{-1}$  is regular (analytic) for all  $z \in \mathbb{C} \setminus \mathbb{L}$ , in which

$$\mathbb{L} = \{z \in \mathbb{C} : \operatorname{Re}(z) \geq z_i \text{ and } \operatorname{Im}(z) = 0\}$$

Thus, from Taylor's Theorem we have that  $\pi(z)^{-1}$  can be expressed as a power series with  $z_0 \in \mathbb{C} \setminus \mathbb{L}$ .

- The Taylor series expansion around zero is valid for  $|z_i^{-1}z| < 1$
- For  $|z| = 1$ , the Taylor series expansion around zero is valid if

$$|z_i| > 1$$

Otherwise, the infinite sum does not converge to the function  $(1 - \rho z)^{-1}$

- Thus, we need the roots of the characteristic polynomial  $|I - \beta_1 z|$  to be outside the unit circle.
- If so, we have that the solution for the system is given by:

$$\mathbf{y}_t = (I - \beta_1 L)^{-1} \mathbf{v}_t$$

$$\mathbf{y}_t = \frac{1}{|(I - \beta_1 L)|} C' \mathbf{v}_t$$

$$\mathbf{y}_t = \frac{1}{(1 - \pi_1 L - \pi_2 L^2)} C' \mathbf{v}_t$$

$$\mathbf{y}_t = \left( \left(1 - \frac{1}{z_1}L\right) \left(1 - \frac{1}{z_2}L\right) \right)^{-1} C' \mathbf{v}_t$$

$$\mathbf{y}_t = \left(1 - \frac{1}{z_1}L\right)^{-1} \left(1 - \frac{1}{z_2}L\right)^{-1} C' \mathbf{v}_t$$

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \left( \frac{1}{z_1} \right)^j L^j \sum_{j=0}^{\infty} \left( \frac{1}{z_2} \right)^j L^j C' \mathbf{v}_t$$

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \psi_j L^j C' \mathbf{v}_t$$

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \psi_j L^j \begin{bmatrix} 1 - \beta_2^y L & -\beta_2^x L \\ -\beta_1^y L & 1 - \beta_1^x L \end{bmatrix} \mathbf{v}_t$$

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \begin{bmatrix} c_{1j} & c_{2j} \\ c_{3j} & c_{4j} \end{bmatrix} L^j \mathbf{v}_t$$

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{C}_j L^j \mathbf{v}_t$$

- So, we have an infinite MA representation for each variable with two errors. Using Theorem 6, we know that the solution does not only converges in the second moment but it is also a **stationary solution**.
- Therefore, if the roots of the characteristic polynomial  $|I - \boldsymbol{\beta}_1 z|$  are outside the unit circle, the stationary solution of the system is given by:

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{C}_j L^j \mathbf{v}_t \quad (30)$$

#### 9.3.4 The roots of the characteristic polynomial and the eigenvalues

- Notice that the characteristic polynomial  $|Iz - \boldsymbol{\beta}_1|$  is the **reflected polynomial** of  $|I - \boldsymbol{\beta}_1 z|$ :

$$|I - \boldsymbol{\beta}_1 z| = \pi(z)$$

$$|Iz - \boldsymbol{\beta}_1| = \tilde{\pi}(z)$$

$$\tilde{\pi}(z) = z^p \pi(z^{-1})$$

- Thus, the roots of  $\tilde{\pi}(z)$  are the inverse of the roots of the reflected polynomial  $\tilde{\pi}(z)$ :

$$\frac{1}{z_i} = \lambda_i$$

- Recall that the roots of the (reflected) characteristic polynomial  $|I\lambda - \boldsymbol{\beta}_1|$  are the **eigenvalues** of the matrix  $\boldsymbol{\beta}_1$ .
- Thus, we have that talking about the roots of the characteristic polynomial  $|I - \boldsymbol{\beta}_1 z|$  is the same as talking about the inverse of the eigenvalues of  $\boldsymbol{\beta}_1$ .

### 9.3.5 Stationary Solution

- If the roots of the characteristic polynomial  $|I - \beta_1 z|$  are outside the unit circle:

$$|z_i| > 1$$

- or if the eigenvalues of  $\beta_1$  are inside the unit circle (i.e. the roots of the (reflected) characteristic polynomial  $|I\lambda - \beta_1|$ ):

$$|\lambda_i| < 1$$

Then, the VAR(1) has the following stationary solution:

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{C}_j L^j \mathbf{v}_t \quad (31)$$

### 9.4 VAR(1) with $n$ variables

- The requirement for the stationary solution is the same as in section 9.3
- The difference is the number of roots
- The characteristic polynomial is of degree  $n + 1$
- Thus, we have  $n + 1$  roots
- In general, the number of roots depends on the number of variables and on the lag length.

### 9.5 The VAR(2) model

$$\begin{aligned} x_t &= \beta_1^x x_{t-1} + \beta_2^x y_{t-1} + \beta_3^x x_{t-2} + \beta_4^x y_{t-2} + u_t^x \\ y_t &= \beta_1^y x_{t-1} + \beta_2^y y_{t-1} + \beta_3^y x_{t-2} + \beta_4^y y_{t-2} + u_t^y \end{aligned} \quad (32)$$

#### 9.5.1 Matrix Notation

- We can write the model in matrix notation:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \beta_1^x & \beta_2^x \\ \beta_1^y & \beta_2^y \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \beta_3^x & \beta_4^x \\ \beta_3^y & \beta_4^y \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} u_t^x \\ u_t^y \end{bmatrix} \quad (33)$$

- By defining the following vectors and matrices we end up with a very simple form for the VAR(2) model. Let:

$$\mathbf{y}_t = \begin{bmatrix} x_t \\ y_t \end{bmatrix}, \mathbf{v}_t = \begin{bmatrix} u_t^x \\ u_t^y \end{bmatrix}, \boldsymbol{\beta}_1 = \begin{bmatrix} \beta_1^x & \beta_2^x \\ \beta_1^y & \beta_2^y \end{bmatrix}, \text{ and } \boldsymbol{\beta}_2 = \begin{bmatrix} \beta_3^x & \beta_4^x \\ \beta_3^y & \beta_4^y \end{bmatrix}$$

Then, the model in matrix from is given by:

$$\mathbf{y}_t = \boldsymbol{\beta}_1 \mathbf{y}_{t-1} + \boldsymbol{\beta}_2 \mathbf{y}_{t-2} + \mathbf{v}_t \quad (34)$$

### 9.5.2 Solving the system of equations: Inverting the matrix

- Let's use the lag operator:

$$\mathbf{y}_t = \beta_1 L \mathbf{y}_t + \beta_2 L^2 \mathbf{y}_t + \mathbf{v}_t$$

$$(I - \beta_1 L - \beta_2 L^2) \mathbf{y}_t = \mathbf{v}_t$$

- Let's define:

$$A(L) = (I - \beta_1 L - \beta_2 L^2)$$

- So, we would like to be able to express the solution as:

$$\mathbf{y}_t = A(L)^{-1} \mathbf{v}_t$$

- To do so, first we need to check if the inverse exist. That is, if  $A(L)$  is a nonsingular matrix. So, we need to check if:

1.  $A(L)$  is a square matrix
2.  $|(I - \beta_1 L - \beta_2 L^2)| \neq 0$  so we have:

$$(I - \beta_1 L - \beta_2 L^2)^{-1} = \frac{1}{|(I - \beta_1 L - \beta_2 L^2)|} C'$$

, in which  $C'$  is the transpose of the cofactor matrix

Condition 1 is satisfied given that  $\beta_1$  and  $\beta_2$  are square matrices.

- For condition 2, notice that  $|(I - \beta_1 L - \beta_2 L^2)|$  is a fourth order polynomial on  $L$ :

$$|(I - \beta_1 L - \beta_2 L^2)| = \left| \begin{bmatrix} 1 - \beta_1^x L - \beta_3^x L^2 & \beta_2^x L - \beta_4^x L^2 \\ \beta_1^y L - \beta_3^y L^2 & 1 - \beta_2^y L - \beta_4^y L^2 \end{bmatrix} \right|$$

Thus, we have that:

$$|(I - \beta_1 L - \beta_2 L^2)| = \pi(L) = 1 - \pi_1 L - \pi_2 L^2 - \pi_3 L^4 - \pi_4 L^6 \quad (35)$$

- Thus, condition 2 is the same as checking that  $\pi(L)^{-1}$  exists.
- Further, we are in a similar situation as is the usual AR(p) case. That is, we would like:
  1.  $\pi(L)^{-1} < \infty$  (i.e. it exists)
  2.  $\pi(L)^{-1}$  has no meaning in Time Series (i.e. How do I multiply it to the error term?). So, we would like  $\pi(z)^{-1}$  to be a regular function so we can use Taylor's Theorem around 0 and express it as an infinite sum.
  3. Check if the power series converges when  $|z| = 1$  (So we know that the infinite sum time the error term (i.e. the RHS) converges in the second moment)

- We know that we can find the roots of the characteristic polynomial  $\pi(z)$ ,  $z_i$ , and express it as:

$$\pi(z) = \left(1 - \frac{1}{z_1}z\right) \left(1 - \frac{1}{z_2}z\right) \left(1 - \frac{1}{z_3}z\right) \left(1 - \frac{1}{z_4}z\right)$$

- Further, we know that, the characteristic function  $(1 - \rho)^{-1}$  is regular (analytic) for all  $z \in \mathbb{C} \setminus \mathbb{L}$ , in which

$$\mathbb{L} = \{z \in \mathbb{C} : \operatorname{Re}(z) \geq z_i \text{ and } \operatorname{Im}(z) = 0\}$$

Thus, from Taylor's Theorem we have that  $\pi(z)^{-1}$  can be expressed as a power series with  $z_0 \in \mathbb{C} \setminus \mathbb{L}$ .

- The Taylor series expansion around zero is valid for  $|z_i^{-1}z| < 1$
- For  $|z| = 1$ , the Taylor series expansion around zero is valid if

$$|z_i| > 1$$

Otherwise, the infinite sum does not converge to the function  $(1 - \rho z)^{-1}$

- Thus, we need the roots of the characteristic polynomial  $|I - \beta_1 z|$  to be outside the unit circle.
- If so, we have that the solution for the system is given by:

$$\mathbf{y}_t = |A(L)|^{-1} C' \mathbf{v}_t$$

$$\mathbf{y}_t = \pi(L)^{-1} C' \mathbf{v}_t$$

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \left(\frac{1}{z_1}\right)^j L^j \sum_{j=0}^{\infty} \left(\frac{1}{z_2}\right)^j L^j \left(\frac{1}{z_3}\right)^j L^j \left(\frac{1}{z_4}\right)^j L^j C' \mathbf{v}_t$$

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \psi_j L^j C' \mathbf{v}_t$$

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \psi_j L^j \begin{bmatrix} 1 - \beta_2^y L - \beta_4^y L^2 & -\beta_2^x L + \beta_4^x L^2 \\ -\beta_1^y L + \beta_3^y L^2 & 1 - \beta_1^x L - \beta_3^x L^2 \end{bmatrix} \mathbf{v}_t$$

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \begin{bmatrix} c_{1j} & c_{2j} \\ c_{3j} & c_{4j} \end{bmatrix} L^j \mathbf{v}_t$$

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{C}_j L^j \mathbf{v}_t$$

- So, we have an infinite MA representation for each variable with two errors. Using Theorem 6, we know that the solution does not only converges in the second moment but it is also a **stationary solution**.

- Therefore, if the roots of the characteristic polynomial  $|A(z)|$  are outside the unit circle, the stationary solution of the system is given by:

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{C}_j L^j \mathbf{v}_t \quad (36)$$

### 9.5.3 Stationary Solution

As we have seen in section 8.5.2,

- If the roots of the characteristic polynomial  $|I - \beta_1 z - \beta_2 z|$  are outside the unit circle:

$$|z_i| > 1$$

- or if the roots of the (reflected) characteristic polynomial  $(|I\lambda^2 - \beta_1\lambda - \beta_2|)$  are inside the unit circle:

$$|\lambda_i| < 1$$

Then, the VAR(2) has the following stationary solution:

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{C}_j L^j \mathbf{v}_t \quad (37)$$

## 9.6 The VAR(p) model

The model in matrix form is given by:

$$\mathbf{y}_t = \beta_1 \mathbf{y}_{t-1} + \beta_2 \mathbf{y}_{t-2} + \beta_3 \mathbf{y}_{t-3} + \cdots + \beta_p \mathbf{y}_{t-p} + \mathbf{v}_t \quad (38)$$

Let's define

$$A(L) = (I - \beta_1 L - \beta_2 L^2 - \beta_3 L^3 - \cdots - \beta_p L^p)$$

So, the model can be rewritten as:

$$A(L)\mathbf{y}_t = \mathbf{v}_t$$

When solving the system, the difference is in the order of the characteristic polynomial  $|A(z)|$ . It is of order  $(p+n) \times (p+n)$ .

### 9.6.1 Stationary Solution

As we have seen in section 8.5.2,

- If the roots of the characteristic polynomial  $|A(z)|$  are outside the unit circle:

$$|z_i| > 1$$

- or if the roots of the (reflected) characteristic polynomial  $z^p|A(z)|$  are inside the unit circle:

$$|\lambda_i| < 1$$

Then, the VAR(P) has the following stationary solution:

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{C}_j L^j \mathbf{v}_t \quad (39)$$

, in which  $\mathbf{C}_j$  is an  $n \times n$  matrix.

## 10 Structural Vector Autoregressions

### 10.1 Motivation

- A classic question in empirical macroeconomics:
  - what is the effect of a policy intervention (interest rate increase, fiscal stimulus) on macroeconomic aggregates of interest - output, inflation, etc?
- Let  $Y_t$  be a vector of macro time series.
- Let  $\varepsilon_t^r$  denote an unanticipated monetary policy intervention.
- We want to know the dynamic causal effect of  $\varepsilon_t^r$  on  $Y_t$  :

$$\frac{\partial Y_{t+h}}{\partial \varepsilon_t^r}, h = 1, 2, 3, \dots$$

where the partial derivative holds all other interventions constant.

- In macro, this dynamic causal effect is called the **impulse response function (IRF)** of  $Y_t$  to the "shock" (unexpected intervention)  $\varepsilon_t^r$ .
- The challenge is to estimate

$$\left\{ \frac{\partial Y_{t+h}}{\partial \varepsilon_t^r} \right\}$$

from observational macro data.

- Two conceptual approaches to estimating dynamic causal effects (IRF):
  - Structural model (Cowles Commission): DSGE or SVAR
  - Quasi-Experiments

### 10.2 The Reduced VAR

- Consider the Reduced form VAR( $p$ ) :

$$Y_t = A_1 Y_{t-1} + \dots + A_p Y_{t-p} + u_t$$

or

$$A(L)Y_t = u_t$$

where

$$A(L) = I - A_1 L - A_2 L^2 - \dots - A_p L^p$$

where  $A_i$  are the coefficients from the (population) regression of  $Y_t$  on  $Y_{t-1}, \dots, Y_{t-p}$ .

- If  $u_t$  were the shocks, then we could compute the structural IRF using the MA representation of the VAR,

$$Y_t = A(L)^{-1} u_t$$

- But in general  $u_t$  is affected by multiple shocks: in any given quarter, GDP changes unexpectedly for a variety of reasons.
- Is there a way to identify the structural shocks?
- For that we need to find the relationship between the reduced VAR and structural VAR.

### 10.3 The Structural VAR

- Consider a bivariate first-order VAR model:

$$\begin{aligned} y_t &= b_{10} - b_{12}x_t + \gamma_{11}y_{t-1} + \gamma_{12}x_{t-1} + \varepsilon_{yt} \\ x_t &= b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}x_{t-1} + \varepsilon_{xt} \end{aligned}$$

- The error terms (structural shocks)  $\varepsilon_{yt}$  and  $\varepsilon_{xt}$  are white noise innovations with standard deviations  $\sigma_y$  and  $\sigma_x$  and a zero covariance.
- From this representation we would be able to find the IRF for the structural shocks!
- However, we have an issue:
  - The two variables  $y$  and  $x$  are endogenous
  - Note that shock  $\varepsilon_{yt}$  affects  $y$  directly and  $x$  indirectly.
- It is worth to point out that here we there are **10** parameters to estimate.

### 10.4 From a SVAR to an RVAR

- The structural VAR is not a reduced form.
- In a reduced form representation  $y$  and  $x$  are just functions of lagged  $y$  and  $x$ .
- To solve for a reduced form write the structural VAR in matrix form as:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix}$$

$$RY_t = \Gamma_0 + \Gamma_1 Y_{t-1} + \varepsilon_t$$

- Premultiplication by  $R^{-1}$  allow us to obtain a standard VAR(1) :

$$\begin{aligned} RY_t &= \Gamma_0 + \Gamma_1 Y_{t-1} + \varepsilon_t \\ Y_t &= R^{-1}\Gamma_0 + R^{-1}\Gamma_1 Y_{t-1} + R^{-1}\varepsilon_t \\ Y_t &= \Phi_0 + \Phi_1 Y_{t-1} + u_t \end{aligned}$$

- This is the reduced form.
- We no longer have the endogeneity issue
- We can estimate the parameters by OLS equation by equation (ofc, we need to check for stationarity first)
- It is worth to point out we are estimating 9 parameters in the reduced VAR.

## 10.5 The Identification Problem

- Remember that we started with a structural VAR model, and jumped into the reduced form or standard VAR for estimation purposes.
- Is it possible to recover the parameters in the structural VAR from the estimated parameters in the standard VAR? No!!
- There are 10 parameters in the bivariate structural VAR(1) and only 9 estimated parameters in the standard VAR(1).
- The VAR is **underidentified**.

## 10.6 Reduced form to structure

- The Reduced VAR

$$\begin{aligned} A(L)Y_t &= u_t \\ Y_t &= A(L)^{-1}u_t = C(L)u_t \\ A(L) &= I - A_1L - A_2L^2 - \dots - A_pL^p \\ E[u_tu_t'] &= \Sigma_u \text{ (unrestricted)} \end{aligned}$$

- The Structural VAR

$$\begin{aligned} B(L)Y_t &= \varepsilon_t \\ Y_t &= B(L)^{-1}\varepsilon_t = D(L)\varepsilon_t \\ B(L) &= B_0 - B_1L - B_2L^2 - \dots - B_pL^p \\ E[\varepsilon_t\varepsilon_t'] &= \Sigma_\varepsilon = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_k^2 \end{pmatrix} \end{aligned}$$

- Because  $\varepsilon_t = Ru_t$ ,

$$RA(L)Y_t = Ru_t = \varepsilon_t.$$

- Letting  $RA(L) = B(L)$ , this delivers the structural VAR,

$$B(L)Y_t = \varepsilon_t,$$

- The MA representation of the SVAR delivers the structural IRFs:

$$\begin{aligned} Y_t &= D(L)\varepsilon_t \\ D(L) &= B(L)^{-1} = A(L)^{-1}R^{-1} \end{aligned}$$

- Impulse response:  $\frac{\partial Y_{t+h}}{\partial \varepsilon_t} = D_h$

- Therefore, we have that:

$$\begin{aligned} Ru_t &= \varepsilon_t \\ B(L) &= RA(L) \quad (B_0 = R) \\ C(L) &= A(L)^{-1} \\ D(L) &= C(L)R^{-1} \end{aligned}$$

### 10.6.1 A note about $B_0$

- We know that  $B_0$  is the first element in the polynomial  $B(L)$ .
- Since we know that  $B_0$  captures the contemporaneous relationship between the variables, we are very tempted to say that the diagonal of  $B_0$  is full of 1s.
- However, if we do that, we are already making a normalization, which is not bad at all but you have to take that into account.
- In general, we could have that  $B_0$  is full of parameters.
- For example, for the VAR with 2 variables:

$$B_0 = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

- Recall that  $B_0 = R$

### 10.6.2 Identification of $R$

- In population, we can know  $A(L)$ .
- If we can identify  $R$ , we can obtain the SVAR coefficients,

$$B(L) = RA(L)$$

- The question here is how can we identify  $R$ :
  - Identification by Short Run Restrictions: Sims (1980)
  - Identification by Long Run Restrictions: Blanchard and Quah (1989)
  - Identification from Heteroskedasticity: Rigobon (2003)
  - Identification by Sign Restrictions: Uhlig (2005)
  - Identification by External Instruments: Stock (2007), Stock and Watson (2012); Mertens and Ravn (2013); Gertler and P. Karadi (2014); for IV in VAR (not full method) see Hamilton (2003), Kilian (2009).
- The answer lies on the following equations:

$$Ru_t = \varepsilon_t$$

$$E[u_t u_t'] = \Sigma_u$$

$$E[\varepsilon_t \varepsilon_t'] = \Sigma_\varepsilon$$

$$\Sigma_\varepsilon = R \Sigma_u R'$$

- Notice that  $\Sigma_u$  is identified, but the not identified parameters here are  $\Sigma_\varepsilon$  and  $R$ .

## 10.7 Identification by Short Run Restrictions

- Based on Sims (1980)
- In general, the SVAR is fully identified if

$$R\Sigma_u R' = \Sigma_\varepsilon$$

can be solved for the unknown elements of  $R$  and  $\Sigma_\varepsilon$

- Recall that  $\Sigma_u$  is identified.
- Let's say we have  $k$  variables.
- How many parameters do we have to identify?

- To see that, let's go back to:

$$\Sigma_\varepsilon = R\Sigma_u R'$$

- we can rewrite it as unidentified on the LHS and identified on the RHS:

$$R^{-1}\Sigma_\varepsilon(R')^{-1} = \Sigma_u$$

- Now we can clearly see that we have to identify  $k$  parameters from  $\Sigma_\varepsilon$  since  $\Sigma_\varepsilon$  is a diagonal matrix.
- Further, we also need to identify the parameters inside  $R$ , that is  $k \times k$  parameters (remember that in general  $R$  does not need to have the diagonal full of 1s).
- Therefore, we have  $k + k^2$  parameters to identify.
- How many parameters do we have already identified?
  - The ones on the RHS
  - Recall that  $\Sigma_u$  is not a diagonal matrix because the reduced shocks are correlated.
  - Given the structure of a variance covariance matrix, we do not have  $k \times k$  identified parameters because the lower triangle equals the upper triangle
  - We only have  $k(k + 1)/2$  identified parameters.
- How many parameters on the LHS can be identified?
  - Given the above and that the quality holds:

$$R^{-1}\Sigma_\varepsilon(R')^{-1} = \Sigma_u$$

it must be that we can only identify  $k(k + 1)/2$

- Thus, we have  $k + k(k - 1)/2$  parameters that cannot be identified.
- How can we identify the remaining  $k + k(k - 1)/2$  parameters?
  - First, we impose a normalization condition that helps with the identification of  $k$  parameters. We have two ways to do so:
    1. The diagonal components of  $\mathbf{B}_0$  are 1's
    2. The variance-covariance matrix of structural disturbances is an identity matrix.
  - Now we only have  $k(k - 1)/2$  parameters to be identified.

- How can we identify the remaining  $k(k - 1)/2$  parameters?
  - To fully understand this, you must know about Cholesky decomposition.
  - If not, please refer to Appendix A
  - Letting  $\Phi_0 = B_0^{-1}$ , it follows that
 
$$\Phi_0 \Sigma_\varepsilon \Phi_0' = \Sigma_u$$
  - Let  $\mathbf{P}$  be a lower triangular matrix of the Cholesky decomposition of  $\Sigma_\varepsilon$  so that
 
$$\mathbf{P} \mathbf{P}' = \Sigma_\varepsilon$$

From  $\Phi_0 \Sigma_\varepsilon^{1/2} = \mathbf{P}$ , it follows that

$$\Phi_0 = \mathbf{P} \Sigma_\varepsilon^{-1/2}$$

- Given either of the normalizations, we end up with the condition that  $\Phi_0$  must be lower triangular (it is pretty evident if you normalize such that  $\Sigma_\varepsilon = I$ ).
- If  $\Phi_0$  is lower triangular, then  $R^{-1}$  is lower triangular.
- Since  $R = B_0$ , it follows that  $B_0$  is lower triangular.
- Since  $B_0$  contains the coefficients associated with the contemporaneous relationship in the structural VAR, we are imposing a structure for the Short Run dynamics between the variables.
- If  $B_0$  is lower triangular we are imposing the idea that:
  - $y_{i,t}$  for  $i = 2, 3, 4, \dots, k$  has no contemporaneous effect of  $y_{1,t}$ .
  - The residuals of  $u_{1,t}$  are due to pure shocks to  $y_{1,t}$ .
  - $y_{i,t}$  for  $i = 3, 4, \dots, k$  has no contemporaneous effect of  $y_{2,t}$ .
  - $y_{i,t}$  for  $i = 4, 5, \dots, k$  has no contemporaneous effect of  $y_{3,t}, \dots$  and so on.
  - All the structural shocks have contemporaneous effect on the last variable,  $y_{k,t}$
- This description of identification is via method of moments, however identification can equally be described via IV, e.g. see Blanchard and Watson (1986).

## 10.8 Identification by Long Run Restrictions

- Based on Blanchard and Quah (1989)
- This approach identifies  $R$  by imposing restrictions on the long-run effect of one or more  $\varepsilon$ 's on one or more  $Y$ 's.
- What do we mean here by long-run effect?
  - The long-run effect refers to the long run variance.
  - The long run variance of a variable is its spectral density at zero frequency.
- What is the spectral density?
  - The Spectral Density of  $x_t$  is given by the Fourier Transform of the autocovariance function of  $x_t$  :
$$s_x(\omega) = \tilde{\gamma}(\omega) = \sum_{j=-\infty}^{\infty} \gamma_j e^{-ij\omega}$$

, in which  $\omega \in (-\pi, \pi]$  represents the frequency.

- There is a Theorem that states the following:

**Theorem 8** Consider

$$y_t = \sum_{j=0}^{\infty} \psi_j L^j x_t = \psi(L) x_t$$

, in which  $x_t$  is a (complex-valued) zero-mean stationary process.

- \* Then, if  $\psi(z)$  converges on the unit circle or if that  $\left| \sum_{j=0}^{\infty} \psi_j z^j \right|^2$  converges on the unit circle, we have that  $y_t$  is stationary with its **Spectral density** given by:

$$s_y(\omega) = \left| \sum_{j=0}^{\infty} \psi_j e^{-i\omega j} \right|^2 s_x(\omega)$$

$$s_y(\omega) = |\psi(e^{-i\omega})|^2 s_x(\omega)$$

- Since the Spectral density is a Fourier Transform, then we can always apply the Fourier Inverse to find each coefficient inside the infinite sum of the Fourier Transform, which in this case is  $\gamma_j$  (i.e., each autocovariance of  $x_t$ )
- The **Fourier Inverse** of the **Spectral density** of  $x_t$  gives us:

$$\gamma_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} s_x(\omega) e^{i\omega j} d\omega$$

, in which  $\gamma_j$  is the autocovariance of  $x_t$  of order  $j$  for every integer  $j$ .

- For  $j = 0$ , we have that

$$\gamma_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} s_x(\omega) d\omega$$

- Thus, from an AR(P) representation for  $y_t$  we have that the Spectral Density of  $y_t$  is given by:

$$s_y(\omega) = \left| \sum_{j=0}^{\infty} \psi_j e^{-i\omega j} \right|^2 s_{\varepsilon}(\omega)$$

$$s_y(\omega) = |\psi(e^{-i\omega})|^2 s_{\varepsilon}(\omega)$$

, in which  $\omega \in (-\pi, \pi]$  represents the frequency.

- That was for  $y_t$  being univariate, but how it is if we have that  $Y_t$  being multivariate?
- I will try to answer this intuitively.
- In the univariate case, we know that the Inverse Fourier of the spectral density of  $y_t$  is the autocovariance function of  $y_t$ .
- If we are in the bi-variate case, we should expect that the Inverse Fourier of the Spectral Density of  $Y_t$  gives us the autocovariance function too.

- However, for this case we should get a matrix  $2 \times 2$  as follows:

$$= \begin{bmatrix} \gamma_y(h) & \gamma_{y,x}(h) \\ \gamma_{x,y}(h) & \gamma_x(h) \end{bmatrix}$$

- Thus, it must be the case that the Spectral Density is also a matrix.
- Each element on the main diagonal is the autospectrum (i.e., the Fourier transform of the auto-covariance function of each variable)
- While the elements on its off-diagonals are the cross-spectra between  $y_{i,t}$  and  $y_{j,t}$ ,  $i \neq j = 1, \dots, k$  (i.e., the Fourier transform of the covariance function of  $y_{i,t}$  and  $y_{j,t}$ ,  $i \neq j = 1, \dots, k$ )
- For the  $Y_t$  being multivariate, we have a modified version of the Theorem 8.

**Theorem 9** Consider

$$Y_t = A(L)^{-1} X_t$$

, in which  $X_t$  is a (complex-valued) zero-mean stationary vector.

- \* Then, if  $|A(z)|$  converges on the unit circle, we have that  $Y_t$  is stationary with its **Spectral density** given by:

$$s_y(\omega) = A(e^{-iw})^{-1} s_x(\omega) A(e^{-iw})^{-1'}$$

- Thus, for the VAR(P) for  $Y_t$ , the spectral density is:

$$s_y(\omega) = A(e^{-iw})^{-1} s_u(\omega) A(e^{-iw})^{-1'}$$

Recall we have that

$$Ru_t = \varepsilon_t$$

$$B(L) = RA(L)$$

$$Y_t = B(L)\varepsilon_t$$

Thus, the spectral density can be written also as:

$$s_y(\omega) = B(e^{-iw})^{-1} s_\varepsilon(\omega) B(e^{-iw})^{-1'}$$

- Now, remember we are interested in the long run variance, which is the spectral density at zero frequency:

$$s_y(0) = A(1)^{-1} s_u(0) A(1)^{-1'}$$

$$s_y(0) = B(1)^{-1} s_\varepsilon(0) B(1)^{-1'}$$

- Given that the structural shocks are not correlated and that they are i.i.d. (the autocovariances are zero), we have that:

$$s_\varepsilon(0) = \Sigma_\varepsilon$$

- Therefore, we have that the long run variance of  $Y_t$  is given by:

$$s_y(0) = A(1)^{-1} s_u(0) A(1)^{-1'}$$

$$s_y(0) = B(1)^{-1} \Sigma_\varepsilon B(1)^{-1'}$$

Thus,

$$A(1)^{-1} s_u(0) A(1)^{-1'} = B(1)^{-1} \Sigma_\varepsilon B(1)^{-1'}$$

Using the fact that  $B(L) = RA(L)$ :

$$A(1)^{-1} s_u(0) A(1)^{-1'} = (RA(1))^{-1} \Sigma_\varepsilon (RA(1))^{-1'}$$

Again we need to check the number of parameters that must be identified.

- \* All parameters on the LHS are identified
- \* Recall that the LHS is the long run variance of  $Y_t$ .
- \* Given the structure of a variance covariance matrix, we do not have  $k \times k$  identified parameters because the lower triangle equals the upper triangle.
- \* We only have  $k(k + 1)/2$  identified parameters.
- \* From the RHS, we can see that we have to identify  $k$  parameters from  $\Sigma_\varepsilon$  since  $\Sigma_\varepsilon$  is a diagonal matrix.
- \* Further, we also need to identify the parameters inside  $R$ , that is  $k \times k$  parameters (remember that in general  $R$  does not need to have the diagonal full of 1s).
- \* Therefore, we have  $k + k^2$  parameters to identify.
- \* Given that the above equality holds, we have  $k + k(k - 1)/2$  parameters that cannot be identified.
- \* If we normalize the variance-covariance matrix of structural disturbances (i.e., it is an identity matrix), we only have  $k(k - 1)/2$  parameters to be identified.
- How do we identify the remaining  $k(k - 1)/2$ ?
  - Recall that for IRF, we have that  $Y_t = D(L)\varepsilon_t$
  - Thus,  $D(L) = B(L)^{-1}$
  - We identify the remaining  $k(k - 1)/2$  using the long run neutrality restriction.
  - The main way long restrictions are implemented in practice is by setting  $\Sigma_\varepsilon = I$  and imposing zero restrictions on  $D(1)$ .

- Imposing  $D_{ij}(1) = 0$  says that the effect the long-run effect on the  $i^{\text{th}}$  element of  $Y_t$ , of the  $j^{\text{th}}$  element of  $\varepsilon_t$  is zero
- If  $\Sigma_\varepsilon = I$ , the moment equation above can be rewritten,

$$s_y(0) = D(1)D(1)'$$

where  $D(1) = B(1)^{-1}$ .

- Because  $RA(1) = B(1)$ ,  $R$  is obtained from  $D(1)$  as  $R = A(1)^{-1}B(1)$ , and  $B(L) = RA(L)$  as above.
- If the zero restrictions on  $D(1)$  make  $D(1)$  lower triangular, then  $D(1)$  is the Cholesky factorization of  $s_y(0)$ .

## 10.9 Identification from Heteroskedasticity

- Based on Rigobon (2003)
- Let's assume that:
  1. The variance of the structural shock is such that it has structural break.
  2. The structural shock variance breaks at date  $s$ :  $\Sigma_{\varepsilon,1}$  before,  $\Sigma_{\varepsilon,2}$  after.
  3.  $R$  doesn't change between variance regimes.
- Let's normalize  $R$  to have 1's on the diagonal.
- Thus the unknowns are:
  - in  $R$  we would have had  $k \times k$  unknown parameters, but given the normalization of 1's on the diagonal, we only have :  $(k^2 - k)$  unknown parameters
  - $\Sigma_{\varepsilon,1}$  is a diagonal matrix, so we have  $k$  unknown parameters
  - $\Sigma_{\varepsilon,2}$  is a diagonal matrix, so we have  $k$  unknown parameters
- Recall that  $Ru_t = \varepsilon_t$ , so by looking at the variance, we have that:

$$\text{First period: } R\Sigma_{u,1}R' = \Sigma_{\varepsilon,1}$$

$$\text{Second period: } R\Sigma_{u,2}R' = \Sigma_{\varepsilon,2}$$

We can rewrite the above equations such that the identified parameters are on the LHS and the unidentified ones are on the RHS:

$$\text{First period: } \Sigma_{u,1} = R^{-1}\Sigma_{\varepsilon,1}R'^{-1}$$

$$\text{Second period: } \Sigma_{u,2} = R^{-1}\Sigma_{\varepsilon,2}R'^{-1}$$

Thus, we have:

- For the first period, on the LHS we have  $k(k + 1)/2$  identified parameters
- For the second period, we have  $k^2 - k$   $k(k + 1)/2$  identified parameters.
- On the RHS from both periods we have  $K^2 - k$  unidentified parameters for  $R$ ,  $K$  unidentified parameters for  $\Sigma_{\varepsilon,1}$ , and  $k$  unidentified parameters for  $\Sigma_{\varepsilon,2}$
- In total, we have  $k(k + 1)$  identified parameters, and  $k^2 + k$  unidentified parameters.

- There is a rank condition here too - for example, identification will not be achieved if  $\Sigma_{\varepsilon,1}$  and  $\Sigma_{\varepsilon,2}$  are proportional.
- The break date need not be known as long as it can be estimated consistently
- Different intuition: suppose only one structural shock is homoskedastic. Then find the linear combination without any heteroskedasticity!

## 11 Cointegrated VAR

### 11.1 Motivation

- Many economic variables are not stationary, and we consider the type of non-stationarity that can be removed by differencing.
- Let's think about the reduced VAR in first differences.
- Is it always the right approach if all the variable are  $I(1)$ ?
- What if the variables share a **long-run relationship**?
- That is, a relationship that is stable in the long run.
- Wouldn't it make sense that this stable relationship should be a regressor in the VAR in first differences?
- In the short-run dynamics, the movements of the variables might be guided by a long-run relationship.

### 11.2 I(0) process

- Let in the following  $\varepsilon_t$  be a sequence of independent identically distributed  $p$ -dimensional random variables with mean zero and variance matrix  $\Omega$ .
- A linear process defined by

$$Y_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}$$

is  $I(0)$  if:

1.  $\sum_{j=0}^{\infty} C_j z^j$  is convergent for  $|z| \leq 1$
2.  $\sum_{i=j}^{\infty} C_i \neq 0$

### 11.3 I(d) process

A stochastic process  $X$  is called integrated of order  $d$ ,  $I(d)$ ,  $d = 0, 1, 2, \dots$  if

$$\Delta^d X_t$$

is  $I(0)$ . An important aspect of this definition is that it is enough to have one of the infinite sum of the errors not converging for one element of  $X_t$  thus allowing the component processes to be **integrated of different orders**. Remember that in general we can have  $p$  infinite sums for each element in  $X_t$

## 11.4 Cointegrated process with cointegrating vector $\beta$

- Let  $X_t$  be integrated of order 1.
- We call  $X_t$  cointegrated with cointegrating vector  $\beta \neq 0$  if:

$$\beta' X_t$$

can be made **stationary** by a suitable choice of its initial distribution.

- The **cointegrating rank** is the number of linearly independent cointegrating relations
- The space spanned by the cointegrating relations is the **cointegrating space**.
- Note that  $\beta' X_t$  need not be  $I(0)$
- However, but for AR processes the cointegrating relations we find are in fact  $I(0)$

### 11.4.1 Example

- Consider the following process:

$$\begin{aligned} X_{1t} &= \sum_{i=1}^t \varepsilon_{1i} + \varepsilon_{2t} \\ X_{2t} &= a \sum_{i=1}^t \varepsilon_{1i} + \varepsilon_{3t} \\ X_{3t} &= \varepsilon_{4t} \end{aligned}$$

- Clearly  $X_{3t}$  is  $I(0)$  but the vector process  $X_t = (X_1, X_{2t}, X_{3t})$  is an  $I(1)$  process since the other two elements have infinite sums that do not converge (both have one unit roots)
- It has two cointegrating vectors:

$$\beta_1 = (a, -1, 0)$$

$$\beta_2 = (0, 0, 1)$$

To see this, notice that:

$$\beta_1 X_t = a \sum_{i=1}^t \varepsilon_{1i} + a \varepsilon_{2t} - a \sum_{i=1}^t \varepsilon_{1i} - \varepsilon_{3t} + 0$$

$$\beta_1 X_t = a \varepsilon_{2t} - \varepsilon_{3t}$$

, which is stationary.

## 11.5 Spurious Regression

- For a random walk, unconditional population moments (i.e. which don't depend on time  $t$ ), such as  $E[X]$ , don't exist. (In some loose sense, they are infinite.)
- Similarly, the unconditional Covariance between two independent random walks is zero.

- Let  $\mathbf{Y}_t = (y_{1t}, \dots, y_{nt})'$  denote an  $(n \times 1)$  vector of  $I(1)$  time series that are not cointegrated.

Write

$$\mathbf{Y}_t = (y_{1t}, \mathbf{Y}'_{2t})'$$

and consider regressing of  $y_{1t}$  on  $\mathbf{Y}_{2t}$  giving

$$y_1 = \hat{\beta}'_2 \mathbf{Y}_{2t} + \hat{u}_t$$

- Let's assume  $y_{1t}$  is not cointegrated with  $\mathbf{Y}_{2t}$
- Then,  $\hat{u}_t \sim I(1)$ .
- If so, the above is a spurious regression
- The regression is spurious when we regress one random walk onto another independent random walk.
- It is spurious because the regression will most likely indicate a non-existing relationship:
  - The true value of  $\beta_2$  is zero.
  - The coefficient estimate will not converge toward zero (the true value). Instead, in the limit the coefficient estimate will follow a non-degenerate distribution.
  - The  $t$  value most often is significant.
  - $R^2$  is typically very high.
- Lesson: just because two series move together does not mean they are related!
- Lesson: use extra caution when you run regression using nonstationary variables; be aware of the possibility of spurious regression! Check whether the residual is nonstationary.

## 11.6 Cointegrated VAR(1)

To get some intuition of the Error Correction Model, we will start by analyzing the simple case of a cointegrated VAR(1)

### 11.6.1 Finding the integrated order of the process

- Let the VAR(1) model be:

$$\mathbf{y}_t = \beta_1 \mathbf{y}_{t-1} + \mathbf{v}_t$$

- We know, that if the roots of the characteristic polynomial  $|I - \beta_1 z|$  are outside the unit circle, then the process has a stationary solution:

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \begin{bmatrix} c_{1j} & c_{2j} \\ c_{3j} & c_{4j} \end{bmatrix} L^j \mathbf{v}_t$$

- Thus, the process is  $I(0)$
- Now, let's suppose we have a matrix  $\beta_1$  such that it generates a unit root (e.g.  $c_{1j} \geq 1$  (remember that the unit root is seen directly in having one of the elements in  $\psi_j$  equal to 1), so the infinite sum never converges (i.e. the infinite sum times the error term (the RHS) does not converge in the second moment)).

- Is the process  $I(1)$ ? Yes, since:

$$\mathbf{y}_t = \beta_1 \mathbf{y}_{t-1} + \mathbf{v}_t$$

$$A(L)\mathbf{y}_t = \mathbf{v}_t$$

$$\mathbf{y}_t = \frac{1}{(1 - \pi_1 L - \pi_2 L^2)} C' \mathbf{v}_t$$

$$\mathbf{y}_t = \left(1 - \frac{1}{z_1} L\right)^{-1} \left(1 - \frac{1}{z_2} L\right)^{-1} C' \mathbf{v}_t$$

Let's suppose that  $z_1 = 1$ , so we have one **unit root**. Then we have that:

$$\mathbf{y}_t = (1 - L)^{-1} \left(1 - \frac{1}{z_2} L\right)^{-1} C' \mathbf{v}_t$$

**TRICK:** How do we get rid off the explosive root?

$$(1 - L)\mathbf{y}_t = \left(1 - \frac{1}{z_2} L\right)^{-1} C' \mathbf{v}_t$$

$$\Delta \mathbf{y}_t = \left(1 - \frac{1}{z_2} L\right)^{-1} C' \mathbf{v}_t$$

Given that  $|z_2| > 1$ , the RHS do converges in the second moment. Thus, we have that

$$\Delta \mathbf{y}_t = \sum_{j=0}^{\infty} \tilde{C}_j \mathbf{v}_{t-j}$$

and it is stationary. So  $\Delta \mathbf{y}_t$  is  $I(0)$ .

### 11.6.2 Is it cointegrated?

- Let's assume that there is a vector  $\beta$  such that  $\beta' \mathbf{y}_t$  is stationary, then  $\mathbf{y}_t$  is cointegrated

## 11.7 Rank decomposition

- Before going to the ECM, it is worth to know about **Rank factorization**
- Given an  $m \times n$  matrix  $A$  of rank  $r$ , a rank decomposition or rank factorization of  $A$  is a factorization of  $A$  of the form

$$A = CF$$

, where  $C$  is an  $m \times r$  matrix and  $F$  is an  $r \times n$  matrix.

- The rank of the matrix is the number of columns or rows that are linearly independent

- Every finite-dimensional matrix has a rank decomposition
- The rank of a matrix can be found by finding the **eigenvalues** of the matrix that are nonzero.
- A matrix whose rank is equal to its dimensions is called a full rank matrix.
- When the rank of a matrix is smaller than its dimensions, the matrix is called rank-deficient, singular, or multicollinear.
- Only full rank matrices have an inverse.
- In practice, we can construct one specific rank factorization as follows: we can compute  $B$ , the **reduced row echelon form** of  $A$ . Then  $C$  is obtained by removing from  $A$  all non-pivot columns, and  $F$  by eliminating all zero rows of  $B$ .

## 11.8 The Vector Error Correction model

- The vector error correction model is a alternative way to express the first difference of a cointegrated VAR
- Introduced by Granger (1986)
- The idea comes from an alternative way to write the first difference of an  $I(1)$  process:

$$\mathbf{y}_t - \mathbf{y}_{t-1} = \boldsymbol{\beta}_1 \mathbf{y}_{t-1} - \mathbf{y}_{t-1} + \mathbf{v}_t$$

$$\Delta \mathbf{y}_t = (\boldsymbol{\beta}_1 - I) \mathbf{y}_{t-1} + \mathbf{v}_t$$

Let's define:

$$\Pi = -(I - \boldsymbol{\beta}_1)$$

Then, we have that:

$$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + \mathbf{v}_t \quad (40)$$

- From section 9.4.1, we know that  $\Delta \mathbf{y}_t$  is stationary and by definition  $\mathbf{v}_t$  is stationary.
- We also know that  $\mathbf{y}_{t-1}$  is  $I(1)$
- Thus, for (40) to be true, it must be that either:
  - $\Pi \mathbf{y}_{t-1}$  is stationary
  - or  $\Pi = 0$
- Given the dimensions of  $\boldsymbol{\beta}_1$ ,  $\Pi$  is a  $2 \times 2$  matrix
- We would like to immediately think that  $\Pi \mathbf{y}_{t-1}$  is the cointegrated relationship, but not quite.
- Let's use the **rank decomposition** of  $\Pi$ :

$$\Pi = \alpha \boldsymbol{\beta}'$$

, in which  $\alpha$  is  $2 \times r$  and  $\boldsymbol{\beta}'$  is  $2 \times r$  and  $r$  being the rank of the matrix  $\Pi$ .

- Then, we will say that  $\boldsymbol{\beta}'$  is the matrix that contains the  $r$  cointegrated vectors

- Then, the rank of the matrix  $\Pi$  determines the number of cointegrated vectors
- If the rank of  $\Pi$  is zero, it means that  $\Pi$  must be a zero matrix since the zero matrix is the only matrix whose rank is 0. That implies that there are no cointegrated vectors for  $\mathbf{y}_t$
- If  $\Pi$  is full rank, then we have that it can be inverted so:

$$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + \mathbf{v}_t$$

$$\Pi^{-1} \Delta \mathbf{y}_t = \mathbf{y}_{t-1} + \Pi^{-1} \mathbf{v}_t$$

Given that  $\Pi^{-1}$  is a finite constant matrix, we still have that  $\Pi^{-1} \Delta \mathbf{y}_t$  and  $\Pi^{-1} \mathbf{v}_t$  are stationary. Therefore, the only way in which the above equation can be true is if  $\mathbf{y}_{t-1}$  is stationary, which is a contradiction given that it is  $I(1)$ . Thus:

$$0 \leq r < n$$

- So, the **rank of the matrix**  $\Pi$  is what we should try to find if we are looking for the number of cointegrated vectors
- If  $\mathbf{y}_t$  is  $I(1)$ , we can always express the Error correction model as follows:

$$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + \mathbf{v}_t \tag{41}$$

, in which  $\Pi = \alpha \beta'$ ,  $\alpha$  is  $2 \times r$  that captures the adjustment vectors (to the long-run equilibrium),  $\beta$  is  $2 \times r$  that captures the cointegrated vectors (long-run equilibrium), and  $r$  being the rank of the matrix  $\Pi$ .

## 11.9 Cointegrated VAR(p)

### 11.9.1 VECM representation

- Let's assume the process  $\mathbf{y}_t$  is  $I(1)$  with  $n$  elements
- Thus,  $\Delta \mathbf{y}_t$  is  $I(0)$
- Either if there are or not cointegrating vectors, the VECM representation should be valid as the rank of  $\Pi$  tells us the number of cointegrating relationships
- The idea is similar to the one in section 9.6 but with some extra tricks
- The goal is to get a system of equations in first differences such the there exists an explicit  $\Pi$  matrix such that its rank tells us the number of cointegrating relationships
- From the VAR(p), let's subtract  $\mathbf{y}_{t-1}$  on both sides so we get the first difference of the process on the LHS:

$$\mathbf{y}_t - \mathbf{y}_{t-1} = \beta_1 \mathbf{y}_{t-1} - \mathbf{y}_{t-1} + \beta_2 \mathbf{y}_{t-2} + \beta_3 \mathbf{y}_{t-3} + \cdots + \beta_p \mathbf{y}_{t-p} \mathbf{v}_t$$

$$\Delta \mathbf{y}_t = (\beta_1 - I) \mathbf{y}_{t-1} + \beta_2 \mathbf{y}_{t-2} + \beta_3 \mathbf{y}_{t-3} + \cdots + \beta_p \mathbf{y}_{t-p} + \mathbf{v}_t$$

We would like to argue the same logic as in VAR(1), but it is not possible since the above equation has the variables in levels, which are  $I(1)$ , and we need all the variables in the VECM to be  $I(0)$ .

- **TRICK:** To fix the problem mentioned above, we proceed to use the following trick:

$$\mathbf{y}_{t-1} = \mathbf{y}_t - \Delta \mathbf{y}_t$$

Thus, we proceed to transform all the variables in levels such that they can be expressed in first differences:

$$\Delta \mathbf{y}_t = (\boldsymbol{\beta}_1 - I) \mathbf{y}_{t-1} + \boldsymbol{\beta}_2 (\mathbf{y}_{t-1} - \Delta \mathbf{y}_{t-1}) + \boldsymbol{\beta}_3 (\mathbf{y}_{t-2} - \Delta \mathbf{y}_{t-2}) + \cdots + \boldsymbol{\beta}_p (\mathbf{y}_{t-(p-1)} - \Delta \mathbf{y}_{t-(p-1)}) + \mathbf{v}_t$$

Thus, we have some first differences of the variables, but we still have some variables in levels. Thus, we have to continue the replacing until the only variable in levels is  $\mathbf{y}_{t-1}$ <sup>9</sup>

- By doing so, we end up with the VECM representation of a VAR(p):

$$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + \Gamma_1 \Delta \mathbf{y}_{t-1} + \Gamma_2 \Delta \mathbf{y}_{t-2} + \cdots + \Gamma_{p-1} \Delta \mathbf{y}_{t-(p-1)} + \mathbf{v}_t \quad (42)$$

, in which:

$$\Gamma_j = - \sum_{i=j+1}^p \boldsymbol{\beta}_i, \quad j = 1, \dots, p-1$$

$$\Pi = - (I - \boldsymbol{\beta}_1 - \cdots - \boldsymbol{\beta}_p)$$

- In the VECM representation, we have that the variable on the LHS is  $I(0)$ . Thus, for (42) to be true, we require all the elements on the RHS to be  $I(0)$ , so  $\Pi \mathbf{y}_{t-1}$  must be such that it contains all the cointegrating vectors.
- So, using **rank factorization**, we have that:

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

, in which  $\Pi$  is an  $n \times n$  matrix,  $\alpha$  is  $n \times r$  matrix,  $\beta$  is an  $n \times r$  matrix and  $r$  is the rank of  $\Pi$ . So the rank of  $\Pi$  tells us the number of cointegrating vectors.

- The rank of  $\Pi$  is such that

$$0 \leq r < n$$

since if it is full rank, all the variables in levels would be  $I(0)$  (see explanation in section 9.6). If the rank is zero, then  $\Pi$  is a zero matrix and the variables are not cointegrated.

---

<sup>9</sup>e.g.,  $\mathbf{y}_{t-3} = \mathbf{y}_{t-2} - \Delta \mathbf{y}_{t-2} = \mathbf{y}_{t-1} - \Delta \mathbf{y}_{t-1} - \Delta \mathbf{y}_{t-2}$

## 11.10 The rank of $\Pi$

- To find the rank of  $\Pi$ , we just have to find the number of eigenvalues that are nonzero
- Notice that

$$\Pi = - (I - \boldsymbol{\beta}_1 - \dots - \boldsymbol{\beta}_p)$$

Recall that from our VAR(p) in levels, we have that:

$$A(L) = (I - \boldsymbol{\beta}_1 L - \boldsymbol{\beta}_2 L^2 - \dots - \boldsymbol{\beta}_p L^p)$$

So, there is a relationship between the lag polynomial from the VAR(p) in levels and the  $\Pi$  matrix in the VECM representation:

$$\Pi = -A(1) \tag{43}$$

- So finding the rank of  $\Pi$  is finding the number of eigenvalues of  $A(z)$  for  $z = 1$  that are nonzero:

$$|\Pi - I\lambda| = |-A(1) - I\lambda| \tag{44}$$

- Furthermore, from the VAR(p) analysis in the previous section, we know that the first condition when solving the system is that  $A(L)^{-1}$  exists. For ensuring that, we require the  $|A(L)| \neq 0$ .
- Given the explicit form for  $A(L)$ , we know that  $|A(L)|$  is a polynomial of degree  $(p+n) \times (p+n)$  in  $L$ .
- So, for  $A(L)^{-1}$  to exist we need the characteristic polynomial  $|A(z)|$  to not be equal to zero:

$$|A(z)| = \pi(z) = 1 - \pi_1 z - \pi_2 z^2 - \dots - \pi_{p \times n} z^{p \times n} \neq 0 \tag{45}$$

- Using the **Fundamental Theorem of Algebra**, we know, that we can express the above polynomial in terms of its roots,  $z_i$ :

$$|A(z)| = \pi(z) = \left(1 - \frac{1}{z_1} z\right) \left(1 - \frac{1}{z_2} z\right) \dots \left(1 - \frac{1}{z_{p \times n}} z\right) \neq 0 \tag{46}$$

- We know that Taylor series expansion is valid for each element in the multiplication for  $|z| \leq 1$  if  $|z_i| > 1$
- Now, notice that  $z = 1$  is just one particular case such that  $|z| = 1$ , but it is relevant in this case since even for this trivial case, for the stationary solution to exist we need  $|z_i| > 1$ .
- Now, remember that  $-A(1) = \Pi$ , so:

$$|-A(1)| = |\Pi|$$

and given that  $\Pi$  cannot be full rank (i.e.  $\Pi^{-1}$  does not exist, the matrix is singular) if the variables are  $I(1)$  (see section 9.7.1). So it must be that

$$|\Pi| = 0$$

, which is the same as:

$$|A(1)| = 0$$

, which happens if the characteristic polynomial  $|A(z)|$  evaluated on  $z = 1$  is such that:

$$|A(1)| = \pi(1) = \left(1 - \frac{1}{z_1}\right) \left(1 - \frac{1}{z_2}\right) \dots \left(1 - \frac{1}{z_{p \times n}}\right) = 0 \quad (47)$$

, which happens if at least one of the roots is 1:

$$z_i = 1$$

- Thus, having at least one unit root in the characteristic polynomial  $|A(z)|$  ensures that  $\Pi$  is not a full rank matrix and the VECM representation is valid (i.e., the rank of  $\Pi$  can be zero but it is not full rank so the variables are  $I(1)$ ).
- Finally if at least one root in the characteristic polynomial  $|A(z)|$  is such that

$$|z_i| < 1$$

then, we have that for that root  $z_i$ , the Taylor expansion for  $\left(1 - \frac{1}{z_i}z\right)^{-1}$  is not valid for  $|z| = 1$ . Thus, the characteristic polynomial  $|A(z)|$  cannot be expanded as Taylor Series expansion because we have that the infinite sum for  $|z| = 1$  does not converge (i.e.  $\infty$ ).

- Notice that having  $|z_i| < 1$  for the characteristic polynomial  $|A(z)|$  does not make  $|A(1)| = 0$ , but we would have that:

$$\Pi^{-1} = |\Pi|^{-1} C'$$

$$\Pi^{-1} = |A(1)|^{-1} C'$$

$$\Pi^{-1} = \left(1 - \frac{1}{z_1}\right)^{-1} \left(1 - \frac{1}{z_2}\right)^{-1} \dots \left(1 - \frac{1}{z_{p \times n}}\right)^{-1} C'$$

in which  $C'$  is the transpose of the cofactor matrix. Now, for the root such that  $|z_i| < 1$ , we have that Taylor Series expansion is not valid for  $\left(1 - \frac{1}{z_i}z\right)^{-1}$  for  $|z| = 1$ . Thus,  $|A(z)|^{-1}$  is not convergent for  $|z| = 1$ :

$$|A(z)|^{-1} = \left(1 - \frac{1}{z_1}z\right)^{-1} \left(1 - \frac{1}{z_2}z\right)^{-1} \dots \left(1 - \frac{1}{z_{p \times n}}z\right)^{-1}$$

, in which each element is expanded using Taylor's Theorem so we get an infinite sum for each element. However, for  $|z| = 1$ , the infinite sum linked with the explosive root,  $|z_i| < 1$ , is not convergent (i.e. it is infinity). Thus, for  $|z| = 1$  which includes  $z = 1$ :

$$|A(1)|^{-1} = \infty$$

$$\Pi^{-1} = \infty C' = \infty$$

Thus,  $\Pi^{-1}$  does not exist even though  $|A(1)| \neq 0$ .

- Therefore, if one root in the characteristic polynomial  $|A(z)|$  is such that it is inside the unit circle:

$$z_b \leq 1$$

, we have that  $\Pi^{-1}$  does not exist, so  $\Pi$  is a singular matrix and it does not have full rank:

$$r < n$$

- Recall that  $X$  is  $I(1)$  if it has 1 unit root,  $I(2)$  if it has 2 unit roots. So the above case does not fall into the category of Integrated process because it has an explosive root but not a unit root.

## 12 Estimation of the Cointegrated VAR(P)

### 12.1 Residual-Based Tests for Cointegration

- Let the  $(n \times 1)$  vector  $\mathbf{Y}_t$  be  $I(1)$ .
- Recall,  $\mathbf{Y}_t$  is cointegrated with  $0 < r < n$  cointegrating vectors if there exists an  $(r \times n)$  matrix  $\mathbf{B}'$  such that

$$\mathbf{B}'\mathbf{Y}_t = \begin{pmatrix} \beta'_1 \mathbf{Y}_t \\ \vdots \\ \beta'_r \mathbf{Y}_t \end{pmatrix} = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{rt} \end{pmatrix} \sim I(0)$$

- Cointegration tests cover two situations:
  1. There is at most one cointegrating vector
  2. There are possibly  $0 \leq r < n$  cointegrating vectors.
- Based on Engle and Granger (1987)
- They developed a simple two-step residual-based testing procedure based on regression techniques.
- Engle and Granger's two-step procedure for determining if the  $(n \times 1)$  vector  $\beta$  is a cointegrating vector is as follows:
  1. Form the cointegrating residual  $\beta'\mathbf{Y}_t = u_t$
  2. Perform a unit root test on  $u_t$  to determine if it is  $I(0)$ .
- The null hypothesis in the Engle-Granger procedure is no-cointegration (i.e.,  $u_t$  has a unit root) and the alternative is cointegration.
- There are two cases to consider:
  1. The proposed cointegrating vector  $\beta$  is pre-specified:
    - For example, economic theory may imply specific values for the elements in  $\beta$  such as  $\beta = (1, -1)'$ .

- The hypotheses to be tested are  $H_0 : u_t = \beta' \mathbf{Y}_t \sim I(1)$  (no cointegration)  $H_1 : u_t = \beta' \mathbf{Y}_t \sim I(0)$  (cointegration)
  - Any unit root test statistic may be used to evaluate the above hypotheses.
  - Tests for cointegration using a pre-specified cointegrating vector are generally much more powerful than tests employing an estimated vector.
2. The proposed cointegrating vector is estimated from the data and an estimate of the cointegrating residual  $\hat{\beta}' \mathbf{Y}_t = \hat{u}_t$  is formed:
- Since  $\beta$  is unknown, to use the Engle-Granger procedure it must be first estimated from the data.
  - Before  $\beta$  can be estimated some normalization assumption must be made to uniquely identify it.
  - A common normalization is to specify  $\mathbf{Y}_t = (y_{1t}, \mathbf{Y}'_{2t})'$  where  $\mathbf{Y}_{2t} = (y_{2t}, \dots, y_{nt})'$  is an  $((n-1) \times 1)$  vector and the cointegrating vector is normalized as  $\beta = (1, -\beta'_2)'$ .
  - Engle and Granger propose estimating the normalized cointegrating vector  $\beta_2$  by least squares from the regression

$$y_{1t} = \gamma' \mathbf{D}_t + \beta'_2 \mathbf{Y}_{2t} + u_t$$

$\mathbf{D}_t$  = deterministic terms

and testing the no-cointegration hypothesis with a unit root test using the estimated cointegrating residual

$$\hat{u}_t = y_{1t} - \hat{\gamma}' \mathbf{D}_t - \hat{\beta}_2' \mathbf{Y}_{2t}$$

- Phillips and Ouliaris (1990) show that ADF and PP unit root tests applied to the estimated cointegrating residual do not have the usual Dickey-Fuller distributions under the null hypothesis of no-cointegration.
- Due to the spurious regression phenomenon under the null hypothesis, the distribution of the ADF and PP unit root tests have asymptotic distributions that depend on:
  - \* The deterministic terms in the regression used to estimate  $\beta_2$
  - \* The number of variables,  $n-1$ , in  $\mathbf{Y}_{2t}$
- Hansen (1992):
  - \* The asymptotic distributions of standard cointegration test statistics are shown to depend both upon regressor trends and estimation detrending methods.
  - \* It is suggested that trends be excluded in the levels regression for maximal efficiency.
  - \* Fully modified test statistics are asymptotically chi-square.

## 12.2 Dynamic OLS

- Stock and Watson (1993) suggest adding  $p$  leads and lags of  $\Delta \mathbf{Y}_{2t}$ :

$$\begin{aligned} y_{1t} &= \gamma' \mathbf{D}_t + \beta'_2 \mathbf{Y}_{2t} + \sum_{j=-p}^p \psi'_j \Delta \mathbf{Y}_{2t-j} + u_t \\ &= \gamma' \mathbf{D}_t + \beta'_2 \mathbf{Y}_{2t} + \psi'_0 \Delta \mathbf{Y}_{2t} \\ &\quad + \psi'_p \Delta \mathbf{Y}_{2t+p} + \dots + \psi'_1 \Delta \mathbf{Y}_{2t+1} \\ &\quad + \psi'_{-1} \Delta \mathbf{Y}_{2t-1} + \dots + \psi'_{-p} \Delta \mathbf{Y}_{2t-p} + u_t \end{aligned}$$

- Estimate the augmented regression by least squares.
- The resulting estimator of  $\beta_2$  is called the dynamic OLS estimator and is denoted  $\hat{\beta}_{2,DOLS}$ .
- $\hat{\beta}_{2,DOLS}$  is consistent, asymptotically normally distributed and efficient (equivalent to MLE) under certain assumptions (see Stock and Watson (1993))

### 12.3 Johansen's Methodology for Modeling Cointegration

The basic steps in Johansen's methodology are:

1. Specify and estimate a VAR( $p$ ) model for  $\mathbf{Y}_t$ .
2. Construct likelihood ratio tests for the rank of  $\Pi$  to determine the number of cointegrating vectors.
3. If necessary, impose normalization and identifying restrictions on the cointegrating vectors.
4. Given the normalized cointegrating vectors estimate the resulting cointegrated VECM by maximum likelihood.

#### 12.3.1 Likelihood Ratio Tests for the Number of Cointegrating Vectors

- The unrestricted cointegrated VECM is denoted  $H(r)$ .
- The  $I(1)$  model  $H(r)$  can be formulated as the condition that the rank of  $\Pi$  is less than or equal to  $r$ .
- This creates a nested set of models

$$H(0) \subset \cdots \subset H(r) \subset \cdots \subset H(n)$$

$$H(0) = \text{non-cointegrated VAR}$$

$$H(n) = \text{stationary VAR}(p)$$

- This nested formulation is convenient for developing a sequential procedure to test for the number  $r$  of cointegrating relationships.
- Johansen formulates likelihood ratio (LR) statistics for the number of cointegrating relationships as LR statistics for determining the rank of  $\Pi$ .
- Recall, the rank of  $\Pi$  is equal to the number of non-zero eigenvalues of  $\Pi$ .
- Thus, these LR tests are based on the estimated eigenvalues  $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$  of the matrix  $\Pi$ .
- Johansen derived two statistic tests for the number of cointegrating vectors:
  1. Trace Statistic
  2. Maximum Eigenvalue Statistic

#### 12.3.2 Johansen's Trace Statistic

- Johansen's LR statistic tests the nested hypotheses

$$H_0(r) : r = r_0 \text{ vs. } H_1(r_0) : r > r_0$$

- The LR statistic, called the trace statistic, is given by

$$LR_{\text{trace}}(r_0) = -T \sum_{i=r_0+1}^n \ln(1 - \hat{\lambda}_i)$$

- The asymptotic null distribution of  $LR_{\text{trace}}(r_0)$  is not chi-square but instead is a multivariate version of the Dickey-Fuller unit root distribution which depends on the dimension  $n - r_0$  and the specification of the deterministic terms.
- Sequential Procedure for Determining the Number of Cointegrating Vectors:
  - First test  $H_0(r_0 = 0)$  against  $H_1(r_0 > 0)$ .
  - If this null is not rejected then it is concluded that there are no cointegrating vectors among the  $n$  variables in  $\mathbf{Y}_t$ .
  - If  $H_0(r_0 = 0)$  is rejected then it is concluded that there is at least one cointegrating vector and proceed to test  $H_0(r_0 = 1)$  against  $H_1(r_0 > 1)$ .
  - If this null is not rejected then it is concluded that there is only one cointegrating vector.
  - If the  $H_0(r_0 = 1)$  is rejected then it is concluded that there is at least two cointegrating vectors.
  - The sequential procedure is continued until the null is not rejected.

### 12.3.3 Johansen's Maximum Eigenvalue Statistic

- Johansen also derives a LR statistic for the hypotheses

$$H_0(r_0) : r = r_0 \text{ vs. } H_1(r_0) : r_0 = r_0 + 1$$

- The LR statistic, called the maximum eigenvalue statistic, is given by

$$LR_{\max}(r_0) = -T \ln \left( 1 - \hat{\lambda}_{r_0+1} \right)$$

- As with the trace statistic, the asymptotic null distribution of  $LR_{\max}(r_0)$  is not chi-square but instead is a complicated function which depends on the dimension  $n - r_0$  and the specification of the deterministic terms.

### 12.3.4 Estimation of the Cointegrated VECM

- Johansen suggests applying Maximum Likelihood Estimation (MLE) for the Error Correction Model, which is a restricted regression

$$\begin{aligned} \Delta \mathbf{Y}_t = & \Phi \mathbf{D}_t + \alpha \beta' \mathbf{Y}_{t-1} + \Gamma_1 \Delta \mathbf{Y}_{t-1} \\ & + \cdots + \Gamma_{p-1} \Delta \mathbf{Y}_{t-p+1} + \varepsilon_t, \end{aligned}$$

, where  $\mathbf{D}_t$  denotes the deterministic components.

- This regression is restricted in the sense that it imposes the hypothesis the rank of  $\Pi$
- The estimates of  $\beta$  are obtained using a special technique called reduced rank regression.
- Unlike the Engle-Granger Two-Step procedure, Johansen method estimates  $\alpha$  and  $\beta$  simultaneously.
- The MLEs of the remaining parameters are obtained by least squares estimation of

$$\begin{aligned} \Delta \mathbf{Y}_t = & \Phi \mathbf{D}_t + \alpha \hat{\beta}'_{mle} \mathbf{Y}_{t-1} + \Gamma_1 \Delta \mathbf{Y}_{t-1} \\ & + \cdots + \Gamma_{p-1} \Delta \mathbf{Y}_{t-p+1} + \varepsilon_t, \end{aligned}$$

### 12.3.5 Reduced Rank Regression

- The reduced rank regression model is a multivariate regression model with a coefficient matrix with reduced rank.
- The reduced rank regression algorithm is an estimation procedure which estimates the reduced rank regression model.
- It is related to canonical correlations<sup>10</sup> and involves calculating eigenvalues and eigenvectors.
- Reduced rank regression model:
  - We consider the multivariate regression of  $Y$  on  $X$  and  $Z$ :

$$Y_t = \Pi X_t + \Gamma Z_t + \varepsilon_t, \quad t = 1, \dots, T$$

- $Y$  is  $n \times 1$
- $X$  is  $n \times 1$
- $z$  is of dimension  $k$
- The hypothesis that  $\Pi$  has reduced rank less than or equal to  $r$  is expressed as

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

- $\alpha$  is  $n \times r$
- $\beta$  is  $n \times r$
- $r < n$

- Reduced rank regression algorithm:

- In order to describe the algorithm, we introduce the notation for product moments

$$S_{yx} = T^{-1} \sum_{t=1}^T Y_t X'_t$$

$$S_{yx.z} = S_{yx} - S_{yz} S_{zz}^{-1} S_{zx}$$

- The algorithm consists of the following steps:

1. Recall that the Frisch-Waugh theorem shows that one can partial out the parameter  $\Gamma$ . Thus, first, regress  $Y$  and  $X$  on  $Z$  and form residuals ():

$$(Y | Z)_t = Y_t - S_{yz} S_{zz}^{-1} Z_t$$

$$(X | Z)_t = X_t - S_{xz} S_{zz}^{-1} Z_t$$

and product moments:

$$\begin{aligned} S_{yx.z} &= T^{-1} \sum_{t=1}^T (Y | Z)_t (X | Z)'_t \\ &= S_{yx} - S_{yz} S_{zz}^{-1} S_{zx}, \end{aligned}$$

---

<sup>10</sup>In statistics, canonical-correlation analysis (CCA), also called canonical variates analysis, is a way of inferring information from cross-covariance matrices. If we have two vectors  $X = (X_1, \dots, X_n)$  and  $Y = (Y_1, \dots, Y_m)$  of random variables, and there are correlations among the variables, then canonical correlation analysis will find linear combinations of  $X$  and  $Y$  which have maximum correlation with each other.

2. Next, solve the eigenvalue problem

$$|\lambda S_{xx.z} - S_{xy.z} S_{yy.z}^{-1} S_{yx.z}| = 0$$

where  $|.|$  denotes determinant.

3. The ordered eigenvalues are  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$  and the eigenvectors are  $V = (v_1, \dots, v_q)$ , so that

$$S_{xx.z} V \Lambda = S_{xy.z} S_{xx.z}^{-1} S_{xy.z}$$

$\Lambda$  and  $V$  are known as the generalized eigenvalues and eigenvectors of  $S_{xy.z} S_{xx.z}^{-1} S_{xy.z}$  with respect to  $S_{xx.z}$ .

4. Recall that the factorization

$$\hat{\Pi}_{mle} = \hat{\alpha}_{mle} \hat{\beta}'_{mle}$$

is not unique. Thus,  $V$  is normalized so that

$$V' S_{xx.z} V = I_p$$

and

$$V' S_{yx.z} S_{xx.z}^{-1} S_{xy.z} V = \Lambda$$

5. Finally, define the estimators

$$\hat{\beta} = (v_1, \dots, v_r)$$

together with

$$\hat{\alpha} = S_{yx.z} \hat{\beta}$$

$$\hat{\Omega} = S_{yy.z} - S_{yx.z} \hat{\beta} \left( \hat{\beta}' S_{xx.z} \hat{\beta} \right)^{-1} \hat{\beta}' S_{xy.z}$$

Equivalently, once  $\hat{\beta}$  has been determined,  $\hat{\alpha}$  and  $\hat{\Gamma}$  are determined by regression.

- Note the difference between the unrestricted estimate  $\hat{\Pi}_{OLS} = S_{yx.z} S_{xx.z}^{-1}$  and the reduced rank regression estimate  $\hat{\Pi}_{RRR} = S_{yx.z} \hat{\beta} \left( \hat{\beta}' S_{xx.z} \hat{\beta} \right)^{-1} \hat{\beta}'$  of the coefficient matrix to  $X$ .
- For interpretations, it is often convenient to normalize or identify the cointegrating vectors by choosing a specific coordinate system in which to express the variables.
- An arbitrary normalization is to solve for the triangular representation of the cointegrated system (default method in Eviews).
- The resulting normalized cointegrating vector is denoted  $\hat{\beta}_{c,mle}$ .
- The normalization of the MLE for  $\beta$  to  $\hat{\beta}_{c,mle}$  will affect the MLE of  $\alpha$  but not the MLEs of the other parameters in the VECM.
- $\hat{\beta}_{c,mle}$  is super consistent

- Let  $\hat{\beta}_{c,mle}$  denote the MLE of the normalized cointegrating matrix  $\beta_c$ . Johansen (1995) showed that

$$T \left( \text{vec} \left( \hat{\beta}_{c,mle} \right) - \text{vec} \left( \beta_c \right) \right)$$

is asymptotically (mixed) normally distributed

### 12.3.6 Hypothesis Testing about the coefficients

- Asymptotic standard chi-squared inference can be conducted on all parameters of the model using likelihood ratio,  $LR$ , tests.
- For further on this please refer to Johansen et al. (1995)

## 13 Further Research on Cointegration

- Nearly Cointegrated Systems: Ng and Perron (1997).
- Higher Order Cointegrated Systems: Haldrup (1998), Kurita (2013).
- Structural Breaks in Cointegrated Systems: Gregory and Hansen (1996), Maki (2012).
- Time-Varying Cointegration: Bierens and Martins (2010)
- Quantile Cointegration: Kuriyama (2016), Quineche (2022).
- Fractionally Cointegrated Systems: Robinson (2003), Johansen (2008) Quineche (2021).

## References

- Bierens, H. J. and Martins, L. F. (2010). Time-varying cointegration. *Econometric Theory*, 26(5):1453–1490.
- Blanchard, O. J. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review*, 79(4):655–673.
- Blanchard, O. J. and Watson, M. W. (1986). Are business cycles all alike? In *The American business cycle: Continuity and change*, pages 123–180. University of Chicago Press.
- Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*. New York: Springer Verlag, 2 edition.
- Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, pages 427–431.
- Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276.
- Granger, C. W. J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, 48(3):213–228.
- Gregory, A. W. and Hansen, B. E. (1996). Residual-based tests for cointegration in models with regime shifts. *Journal of econometrics*, 70(1):99–126.
- Haldrup, N. (1998). An econometric analysis of i (2) variables. *Journal of Economic Surveys*, 12(5):595–650.
- Hansen, B. E. (1992). Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends. *Journal of econometrics*, 53(1-3):87–121.
- Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory*, 24(3):651–676.

- Johansen, S. et al. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press on Demand.
- Kurita, T. (2013). Modelling time series data of monetary aggregates using i (2) and i (1) cointegration analysis. *Bulletin of Economic Research*, 65(4):372–388.
- Kuriyama, N. (2016). Testing cointegration in quantile regressions with an application to the term structure of interest rates. *Studies in Nonlinear Dynamics & Econometrics*, 20(2):107–121.
- Maki, D. (2012). Tests for cointegration allowing for an unknown number of breaks. *Economic Modelling*, 29(5):2011–2015.
- Ng, S. and Perron, P. (1995). Unit root tests in arma models with data-dependent methods for the selection of the truncation lag. *Journal of the american statistical association*, pages 268–281.
- Ng, S. and Perron, P. (1997). Estimation and inference in nearly unbalanced nearly cointegrated systems. *Journal of Econometrics*, 79(1):53–81.
- Phillips, P. C. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica: journal of the Econometric Society*, pages 165–193.
- Phillips, P. C. and Perron, P. (1988). Testing for a unit root in time series regression. *biometrika*, 75(2):335–346.
- Quineche, R. (2021). Consumption, aggregate wealth and expected stock returns: An fcvar approach. *Journal of Time Series Econometrics*, 13(1):21–42.
- Quineche, R. (2022). Consumption, aggregate wealth and expected stock returns: a quantile cointegration approach. *Studies in Nonlinear Dynamics & Econometrics*, 26(5):i–ii.
- Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics*, 85(4):777–792.
- Robinson, P. M., editor (2003). *Time Series with Long Memory*. Oxford University Press.
- Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607.
- Schwert, G. W. (1989). Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics*, 7(2):147–159.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pages 1–48.
- Stock, J. H. and Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica: Journal of the Econometric Society*, pages 783–820.
- 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.

## Appendix A Some matrix algebra facts

### A.1 Eigenvalues and Eigenvectors

- Eigenvalues are a special set of scalars associated with a linear system of equations (i.e., a matrix equation) that are sometimes also known as characteristic roots, characteristic values (Hoffman and Kunze 1971 ), proper values, or latent roots (Marcus and Minc 1988, p. 144).
- The determination of the eigenvalues and eigenvectors of a system is extremely important in physics and engineering, where it is equivalent to matrix diagonalization and arises in such common applications as

stability analysis, the physics of rotating bodies, and small oscillations of vibrating systems, to name only a few.

- Let  $A$  be an  $n \times n$  matrix (i.e. **square matrix**)
- $x$  be a  $n \times 1$  non-zero vector
- $\lambda$  a scalar.
- Then  $x$  is an **eigenvector** and  $\lambda$  an **eigenvalue** if:

$$Ax = \lambda x \quad (48)$$

- From the above, if it exists, we have that:

$$Ax - \lambda x = 0_n$$

$$(A - I_n \lambda)x = 0_n$$

Thus, we have a trivial solution for  $x$  (i.e. zero vector), if we can isolate the vector  $x$  such that:

$$(A - I_n \lambda)^{-1}(A - I_n \lambda)x = (A - I_n \lambda)^{-1}0_n$$

$$x = (A - I_n \lambda)^{-1}0_n$$

$$x = 0$$

Thus, we need  $(A - I_n \lambda)^{-1}$  to exist.

- Thus, we need the inverse of  $(A - I_n \lambda)$  to exist
- Recall that for a matrix  $B$ , the **inverse** of the matrix  $B$  is defined as:

$$B^{-1} = \frac{1}{|B|} C'$$

, in which,  $C'$  is the transpose of the cofactor matrix, which is a matrix formed by all of the cofactors of the square matrix  $B$ . Thus, we need the determinant of the matrix to not be zero, otherwise the inverse does not exist.

- Thus, equation 8.1 has a nonzero solution for  $x$  if and only if the determinant of the matrix  $(A - I_n \lambda)$  is equal to zero:

$$|A - \lambda I| = 0 \quad (49)$$

- Using Leibniz' rule for the determinant, the left-hand side of Equation (8.2) is a polynomial function of the variable  $\lambda$  and the degree of this polynomial is  $n$ , the order of the matrix  $A$
- Thus, we will find  $n$  values for  $\lambda$

- There is an eigenvector for each eigenvalue. The eigenvectors corresponding to each eigenvalue can be found by solving for the components of  $x$  in the equation

$$(A - I_n \lambda) x = 0_n$$

- If the eigenvalues are unique then the eigenvectors will all be **orthogonal**.

## A.2 Eigenvalues and the characteristic polynomial

- We know that the eigenvalues of  $A$  are values of  $\lambda$  that satisfy the equation:

$$|A - \lambda I| = 0 \quad (50)$$

- Using Leibniz' rule for the determinant, the left-hand side of Equation (8.3) is a polynomial function of the variable  $\lambda$  and the degree of this polynomial is  $n$ , the order of the matrix  $A$
- This polynomial is called the **characteristic polynomial** of  $A$ :

$$|A - \lambda I| = \phi_1 \lambda + \phi_2 \lambda^2 + \dots + \phi_n \lambda^n = \Phi(\lambda) \quad (51)$$

, in which the coefficients in  $\Phi(\lambda)$  depend on the entries of  $A$ .

- The **fundamental theorem of algebra** implies that the characteristic polynomial of an  $n$ -by- $n$  matrix  $A$ , being a polynomial of degree  $n$ , can be factored into the product of  $n$  linear terms:

$$|A - \lambda I| = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda) \quad (52)$$

where each  $\lambda_i$  may be real but in general is a complex number.

- The numbers  $\lambda_1, \lambda_2, \dots, \lambda_n$ , which may not all have distinct values, are **roots of the polynomial** and are the **eigenvalues** of  $A$ .

## A.3 Eigen Decomposition "Matrix Diagonalization"

- The matrix decomposition of a square matrix  $A$  into so-called eigenvalues and eigenvectors is an extremely important one
- This decomposition generally goes under the name **matrix diagonalization**. However, this moniker is less than optimal, since the process being described is really the decomposition of a matrix into a product of three other matrices, only one of which is diagonal, and also because all other standard types of matrix decomposition use the term "decomposition" in their names, e.g., Cholesky decomposition, Hessenberg decomposition, and so on.
- Let  $A$  be an  $n \times n$  and **nonsingular** matrix. That is,

$$|A| \neq 0$$

, so it is **invertible**.

- Assume that **eigenvalues are all unique** (i.e. linearly independent eigenvectors). Define the matrices composed of eigenvectors:

$$P \equiv [ \mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_k ] \\ = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{kk} \end{bmatrix}$$

and eigenvalues

$$D \equiv \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}$$

where  $D$  is a diagonal matrix. Notice that

$$AP = A [ \mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_k ]$$

$$AP = [ A\mathbf{X}_1 \quad A\mathbf{X}_2 \quad \cdots \quad A\mathbf{X}_k ]$$

Recall that for each eigenvalue, there is an eigenvector, so using (48), we can rewrite each element of the above as:

$$= [ \lambda_1 \mathbf{X}_1 \quad \lambda_2 \mathbf{X}_2 \cdots \quad \lambda_k \mathbf{X}_k ] \\ = \begin{bmatrix} \lambda_1 x_{11} & \lambda_2 x_{21} & \cdots & \lambda_k x_{k1} \\ \lambda_1 x_{12} & \lambda_2 x_{22} & \cdots & \lambda_k x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 x_{1k} & \lambda_2 x_{2k} & \cdots & \lambda_k x_{kk} \end{bmatrix}$$

Trick! We can obtain the same matrix by postmultiplying  $P$  by  $D$ :

$$= \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{kk} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}$$

Therefore, by using all eigenvalues and eigenvectors, we can rewrite (48) as:

$$AP = PD$$

Thus, the **Eigen decomposition** is given by:

$$A = PDP^{-1} \tag{53}$$

## A.4 Features of Eigen Decomposition

- Squaring both sides of equation (8.6) gives:

$$\begin{aligned} A^2 &= (PDP^{-1})(PDP^{-1}) \\ &= PD(P^{-1}P)DP^{-1} \\ &= PD^2P^{-1} \end{aligned}$$

- By induction, it follows that for general positive integer powers,

$$A^n = P D^n P^{-1} \quad (54)$$

- The **inverse** of  $A$  is

$$\begin{aligned} A^{-1} &= (P D P^{-1})^{-1} \\ &= P D^{-1} P^{-1} \end{aligned}$$

where the inverse of the diagonal matrix  $D$  is trivially given by

$$D^{-1} = \begin{bmatrix} \lambda_1^{-1} & 0 & \cdots & 0 \\ 0 & \lambda_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k^{-1} \end{bmatrix}$$

## A.5 Cholesky Decomposition

- The Cholesky decomposition of a real symmetric positive-definite matrix  $\mathbf{A}$ , is a decomposition of the form:

$$\mathbf{A} = \mathbf{L}\mathbf{L}'$$

where  $\mathbf{L}$  is a real lower triangular matrix with positive diagonal entries.

- If  $\mathbf{A}$  contains complex numbers,  $\mathbf{A}$  must be a Hermitian positive-definite matrix<sup>11</sup>. If so, the Cholesky decomposition is a decomposition of the form

$$\mathbf{A} = \mathbf{L}\mathbf{L}^*,$$

where  $\mathbf{L}$  is a lower triangular matrix with real and positive diagonal entries, and  $\mathbf{L}^*$  denotes the conjugate transpose of  $\mathbf{L}$ .

- Every Hermitian positive-definite matrix (and thus also every real-valued symmetric positive-definite matrix) has a **unique Cholesky decomposition**.

### A.5.1 Example of a variance-covariance matrix

- As an example let's focus on the 2 variables case.
- Thus, the variance-covariance matrix  $\Sigma_u$  is a  $2 \times 2$  matrix:

$$\Sigma_u = \begin{pmatrix} \sigma_1^2 & \rho\sigma_{1,2} \\ \rho\sigma_{1,2} & \sigma_2^2 \end{pmatrix}$$

which can be rewritten as:

$$\Sigma_u = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

If so, the Cholesky decomposition is given by:

$$\Sigma_u = \mathbf{L}\mathbf{L}'$$

with  $\mathbf{L}$  being a the following lower triangular matrix:

$$\begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix}$$

---

<sup>11</sup>A Hermitian matrix (or self-adjoint matrix) is a complex square matrix that is equal to its own conjugate transpose. We need the conjugate because the covariance of 2 complex variables depends on which variable goes first (for further on this, refer to Section 1.13)