Aarhus University BSS
1. semester Master

Nicklas Ringelstein, student nr: 202202087
Daniel Rønman Behr, student nr: 202202133

WISEflow number: 6679853

Department of Economics and Business Economics

ATSE Outlines

Supervisor Morten Ørregaard Nielsen

Nicklas Ringelstein and Daniel Rønman Behr
(Ringelstein & Behr 2022b)

Aarhus University BSS - Denmark
Used: 09-01-2023

Assumptionscolback=white,colframe=yellow,fonttitle=**,title=Assumptions**
Example[2][] enhanced, sharp corners, attach boxed title to top left=yshifttext=-1mm, colback=white, colframe=cyan!25, fonttitle=**, coltitle=black, boxed title style= sharp corners, size=small, colback=cyan!25, colframe=cyan!25, , title=Example 2, 1**
MyProof[2][] enhanced, sharp corners, attach boxed title to top left=yshifttext=-1mm, colback=white, colframe=orange, fonttitle=**, boxed title style= sharp corners, size=small, colback=orange, colframe=orange, , title=Proof 2, 1**
@cref

Table of Contents

List of Figures

List of Tables

CHAPTER 1

1. Stationarity and non-stationarity in univariate time series

1.1 (i) Introduce/define trend-stationarity and difference-stationarity.

In statistics and time series analysis, trend-stationarity and difference-stationarity are two types of stationarity that can be used to describe the statistical properties of a time series.

Trend-stationarity refers to a time series that has a constant mean and variance over time, but may have a trend. This means that the time series may exhibit a long-term upward or downward trend, but the statistical properties of the data do not change over time.

Difference-stationarity refers to a time series that can be made stationary by differencing the data. This means that the time series exhibits some form of non-stationarity, such as a trend or seasonal variation, but these non-stationary features can be removed by taking the difference between successive data points.

In order to analyze a time series, it is often necessary to make the series stationary in order to remove any trends or seasonal variations that may be present. This can be done through a process called "differencing," which involves taking the difference between successive data points in the series. If the resulting series is trend-stationary or difference-stationary, then it can be analyzed using statistical techniques that assume stationarity.

Definition 1: Trend-stationarity

Trend-stationarity:

$X(t) = \mu + T(t) + \epsilon(t)$

where $\epsilon(t)$ is a white noise error term with mean 0 and constant variance σ^2. The equation for trend-stationarity is derived by assuming that the time series has a constant mean and variance over time, but may have a trend. This means that the statistical properties of the data do not change over time, but there may be a long-term upward or downward trend in the data.

The equation for trend-stationarity can be expressed as:

$X(t) = \mu + T(t) + \epsilon(t)$

where:

$X(t)$ is the value of the time series at time t.

μ is the constant mean of the time series.

$T(t)$ is the trend component of the time series, which may have a long-term upward or downward trend.

$\epsilon(t)$ is a white noise error term with mean 0 and constant variance σ^2. This term represents the random fluctuations in the time series around the mean and trend.

The equation for trend-stationarity assumes that the statistical properties of the time series do not change over time, but the data may exhibit a trend. This equation can be used to model and analyze time series data that exhibit a trend, but do not have significant changes in mean or variance over time.

Definition 2: Difference-stationarity:

Difference-stationarity:

$\Delta X(t) = \mu + \epsilon(t)$

where $\Delta X(t)$ is the difference between successive data points in the time series, and $\epsilon(t)$ is a white noise error term with mean 0 and constant variance σ^2.

The equation for difference-stationarity is derived by assuming that the time series exhibits some form of non-stationarity, such as a trend or seasonal variation, but can be made stationary through differencing. This means that the statistical properties of the time series do not remain constant over time, but the non-stationary features can be removed by taking the difference between successive data points in the series.

The equation for difference-stationarity can be expressed as:

$\Delta X(t) = \mu + \epsilon(t)$

where:

$\Delta X(t)$ is the difference between successive data points in the time series.

μ is the constant mean of the time series after it has been made stationary through differencing.

$\epsilon(t)$ is a white noise error term with mean 0 and constant variance σ^2. This term represents the random fluctuations in the time series around the mean after the non-stationary features have been removed through differencing.

The equation for difference-stationarity assumes that the time series exhibits some form of non-stationarity, but can be made stationary through differencing. This equation can be used to model and analyze time series data that exhibit non-stationary features, such as a trend or seasonal variation, but can be made stationary through differencing.

Distribution

The distribution of $\hat{pi}$ under the null hypothesis of a unit root ($H0 : \pi = 0$) is approximately normal with mean zero and a finite, non-zero variance. This is because $\hat{pi}$ is a linear combination of independent and identically distributed errors under the null hypothesis, and the central limit theorem states that the sum of a large number of independent and identically distributed random variables is approximately normally distributed. Additionally, under the null hypothesis, the errors are serially uncorrelated and the OLS estimator $\hat{\pi}$ is consistent and asymptotically normal. The consistency of the estimator ensures that as the sample size increases, $\hat{\pi}$ will converge to the true value of $\pi = 0$, and the asymptotic normality ensures that the distribution of $\hat{\pi}$ will approach a normal distribution with mean zero and a finite, non-zero variance.

1.2 (ii) Compare distribution theory for stationary and non-stationary time series (CLT vs. FCLT).

In statistics and time series analysis, the central limit theorem (CLT) and functional central limit theorem (FCLT) are two fundamental results that describe the behavior of stationary and non-stationary time series, respectively.

The central limit theorem (CLT) states that the distribution of the sum or average of a large number of independent and identically distributed (iid) random variables will tend to a normal distribution, regardless of the distribution of the individual variables. This result is often used to analyze stationary time series data, where the statistical properties of the data do not change over time.

The functional central limit theorem (FCLT) is a generalization of the CLT that applies to non-stationary time series data, where the statistical properties of the data may change over time. The FCLT states that the

distribution of the sum or average of a large number of dependent random variables, such as those observed in a non-stationary time series, will tend to a non-normal distribution, known as a functional distribution. This result is often used to analyze non-stationary time series data, where the statistical properties of the data change over time.

In summary, the CLT and FCLT are two important results in statistical theory that describe the behavior of stationary and non-stationary time series, respectively. The CLT is used to analyze stationary time series data, while the FCLT is used to analyze non-stationary time series data.

Definition 3: Formal definition of the central limit theorem (CLT)

The central limit theorem (CLT) is a fundamental result in probability theory that states that the distribution of the sum or average of a large number of independent and identically distributed (iid) random variables will tend to a normal distribution, regardless of the distribution of the individual variables.

More formally, suppose that $X_1, X_2, ..., X_n$ are iid random variables with mean μ and variance σ^2. Then the distribution of the sum $S_n = X_1 + X_2 + ... + X_n$ will tend to a normal distribution as n increases, with mean $n\mu$ and variance $n\sigma^2$.

This result is often used in statistical analysis to approximate the distribution of the sum or average of a large number of random variables, even if the individual variables are not normally distributed. The CLT is a cornerstone of statistical theory and has numerous applications in a wide variety of fields.

Definition 4: Formal definition of the functional central limit theorem (FCLT)

The functional central limit theorem (FCLT) is a generalization of the central limit theorem (CLT) that applies to non-stationary time series data, where the statistical properties of the data may change over time. The FCLT states that the distribution of the sum or average of a large number of dependent random variables, such as those observed in a non-stationary time series, will tend to a non-normal distribution, known as a functional distribution.

More formally, let $X_1, X_2, ..., X_n$ be a sequence of dependent random variables, and let $S_n = X_1 + X_2 + ... + X_n$ be the sum of these variables. Then the distribution of S_n will tend to a functional distribution as n increases, regardless of the distribution of the individual variables X_i.

The FCLT is an important result in statistical theory that allows for the analysis of non-stationary time series data, where the statistical properties of the data may change over time. It is a generalization of the CLT and has numerous applications in a wide variety of fields.

1.3 (iii) Discuss classical unit root tests in AR models (Dickey-Fuller) and their distributions.

In time series analysis, classical unit root tests are statistical tests used to determine whether a time series exhibits a unit root, which is a characteristic of a non-stationary time series. The Dickey-Fuller test is a well-known unit root test that is widely used in practice.

The Dickey-Fuller test is based on the autoregressive (AR) model, which is a statistical model used to describe the relationship between the current value of a time series and its past values. In an AR model, the current value of the time series is a linear combination of its past values, with coefficients that are called "autoregressive parameters."

To perform the Dickey-Fuller test, the time series is first modeled as an AR model with a unit root. This model is then compared to a "reduced" model that does not include the unit root. The null hypothesis of the Dickey-Fuller test is that the time series has a unit root, while the alternative hypothesis is that the time series is stationary.

If the null hypothesis is rejected, then the time series is considered to be stationary. If the null hypothesis

cannot be rejected, then the time series is considered to have a unit root and is non-stationary.

The Dickey-Fuller test has several important properties, including the distribution of the test statistic under the null hypothesis and the size and power of the test. These properties can be used to determine the statistical significance of the test results and to assess the reliability of the conclusions drawn from the test.

Definition 5: Formal definition of classical unit root tests in AR models (Dickey-Fuller test)

Classical unit root tests in AR models are statistical tests used to determine whether a time series exhibits a unit root, which is a characteristic of a non-stationary time series. The Dickey-Fuller test is a well-known unit root test that is based on the AR model, which is a statistical model used to describe the relationship between the current value of a time series and its past values.

To perform the Dickey-Fuller test, the time series is first modeled as an AR model with a unit root. This model is then compared to a "reduced" model that does not include the unit root. The null hypothesis of the Dickey-Fuller test is that the time series has a unit root, while the alternative hypothesis is that the time series is stationary.

If the null hypothesis is rejected, then the time series is considered to be stationary. If the null hypothesis cannot be rejected, then the time series is considered to have a unit root and is non-stationary.

The Dickey-Fuller test has several important properties, including the distribution of the test statistic under the null hypothesis and the size and power of the test. These properties can be used to determine the statistical significance of the test results and to assess the reliability of the conclusions drawn from the test.

Derivation 1: Derivation of Dickey fuller

Define the null and alternative hypotheses for the test. The null hypothesis is that the time series has a unit root, while the alternative hypothesis is that the time series is stationary. Model the time series as an autoregressive (AR) model with a unit root. This model is typically written as:

$X_t = c + \phi X_{t-1} + \epsilon_t$

where X_t is the value of the time series at time t, c is a constant, ϕ is the autoregressive parameter, and ϵ_t is a white noise error term with mean 0 and constant variance σ^2.

Estimate the parameters of the AR model using least squares or maximum likelihood estimation. Test the null hypothesis that the time series has a unit root by comparing the estimated AR model to a "reduced" model that does not include the unit root.

Calculate the test statistic for the Dickey-Fuller test and determine the p-value for the test. The p-value is the probability of observing a test statistic as extreme as the one calculated, given that the null hypothesis is true.

Use the p-value to determine the statistical significance of the test results. If the p-value is less than a predetermined level of significance, such as 0.05, then the null hypothesis is rejected and the time series is considered to be stationary. If the p-value is greater than the level of significance, then the null hypothesis cannot be rejected and the time series is considered to have a unit root and is non-stationary.

How to derive the Dickey Fuller test from an AR(1) process, where ε is iid with mean 0 and variance Σ_u:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

Which is an AR(1) proces, in this process we know that if $\phi = 1$, then there is a unit root, and if $\phi > 1$ there is no unit root. In order to express it in a more intutive way, then we can subtract y_{t-1} from both sides and obtain:

$$\Delta y_t = \phi y_{t-1} - y_{t-1} + \varepsilon_t$$

Then factorizing:

$$\Delta y_t = (\phi - 1) y_{t-1} + \varepsilon_t$$

Then defining that $\pi = (\phi - 1)$

$$\Delta y_t = \pi y_{t-1} + \varepsilon_t$$

And then, on the background of what stated earlier, we can set up the null and the alternative as:

$$H_0 : \pi = 0$$

$$H_1 : \pi < 0$$

Derivation 2: Unit root processes

Special case: AR(1) is

$$y_t = \phi y_{t1} + \epsilon_t$$

$\epsilon_t \sim iid(0, \sigma^2)$ Want to test unit root (random walk, RW) hypothesis: $H_0 : \phi = 1$. Under H_0, backward substitution: $y_t = y_0 + \sum_{j=1}^{t} \epsilon_j$ so y_t is a partial sum of the $\epsilon j's$

Distribution of $\hat{\pi}$

Converges to a normal distribution when sample size gets bigger. The distribution of $\hat{\pi}$ under the null hypothesis of a unit root ($H0 : \hat{\pi} = 0$) is approximately normal with mean zero and a finite, non-zero variance. This is because $\hat{\pi}$ is a linear combination of independent and identically distributed errors under the null hypothesis, and the central limit theorem states that the sum of a large number of independent and identically distributed random variables is approximately normally distributed. Additionally, under the null hypothesis, the errors are serially uncorrelated and the OLS estimator $\widehat{\ }$ is consistent and asymptotically normal. The consistency of the estimator ensures that as the sample size increases, $\hat{\pi}$ will converge to the true value of $\pi = 0$, and the asymptotic normality ensures that the distribution of $\hat{\pi}$ will approach a normal distribution with mean zero and

a finite, non-zero variance.

1.4 (iv) Discuss modern unit root tests and GLS-detrending in AR models.

In time series analysis, modern unit root tests are statistical tests used to determine whether a time series exhibits a unit root, which is a characteristic of a non-stationary time series. Modern unit root tests are generally more powerful and reliable than classical unit root tests, such as the Dickey-Fuller test, and are often used in practice.

One popular modern unit root test is the augmented Dickey-Fuller (ADF) test, which is an extension of the classical Dickey-Fuller test. The ADF test allows for the inclusion of additional explanatory variables in the unit root model, which can improve the power and reliability of the test.

Another modern unit root test is the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, which is based on the autoregressive (AR) model and tests the null hypothesis that the time series is stationary. The KPSS test is often used as a complementary test to the ADF test, as it has different statistical properties and can provide additional information about the stationarity of the time series.

GLS detrending is another method used to analyze time series data, particularly non-stationary time series data. GLS detrending stands for "generalized least squares detrending," and it is a method used to remove trends and other non-stationary features from time series data. In GLS detrending

Definition 6: The augmented Dickey-Fuller (ADF)

The augmented Dickey-Fuller (ADF) test is a statistical test used to determine whether a time series exhibits a unit root, which is a characteristic of a non-stationary time series. The ADF test is an extension of the classical Dickey-Fuller test that allows for the inclusion of additional explanatory variables in the unit root model, which can improve the power and reliability of the test.

The augmented Dickey-Fuller (ADF) test is a modern unit root test used to determine whether a time series exhibits a unit root. The ADF test is based on the autoregressive (AR) model and tests the null hypothesis that the time series has a unit root, while the alternative hypothesis is that the time series is stationary. The ADF test allows for the inclusion of additional explanatory variables in the unit root model, which can improve the power and reliability of the test.

Definition 7: The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test)

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is a statistical test used to determine whether a time series exhibits a unit root, which is a characteristic of a non-stationary time series. The KPSS test is based on the autoregressive (AR) model and tests the null hypothesis that the time series is stationary, while the alternative hypothesis is that the time series has a unit root.

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is a modern unit root test used to determine whether a time series exhibits a unit root. The KPSS test is based on the autoregressive (AR) model and tests the null hypothesis that the time series is stationary, while the alternative hypothesis is that the time series has a unit root. The KPSS test is often used as a complementary test to the augmented Dickey-Fuller (ADF) test, as it has different statistical properties and can provide additional information about the stationarity of the time series.

Definition 8: GLS detrending)

GLS detrending is a method used to remove trends and other non-stationary features from time series data. GLS detrending stands for "generalized least squares detrending," and it is a variant of the least squares method that is used to fit a linear trend model to the data.

GLS detrending is a method used to remove trends and other non-stationary features from time series data. GLS detrending stands for "generalized least squares detrending," and it is a variant of the least squares method that is used to fit a linear trend model to the data. GLS detrending is often used to analyze non-stationary time series data, as it can help to identify and remove trends and other non-stationary features that may be present in the data.

The trend model is fit to the data using generalized least squares, which is a variant of the least squares method that is robust to heteroscedasticity and autocorrelation in the errors. Once the trend model has been fit to the data, the trend component can be removed from the data, leaving a residual time series that is free of trends and other non-stationary features.

ADF vs GLS-ADF The Elliott et al. (1996) method suggests that by using Generalized Least Squares (GLS) estimates of trend parameters, the power of unit root tests can be improved. GLS detrending can lead to an optimal unit root test, which is the test with the highest power, meaning it has the smallest probability of failing to reject the null hypothesis of a unit root when it is actually false.

In the context of the ADF test, the method improves power by allowing for a model with either an intercept or a trend or both. The ADF test with no trend or intercept is optimal for a model with no trend or intercept, but for models with an intercept or trend or both, the GLS estimates of trend parameters improve the power of the ADF test.

Derivation 3: GLS-ADF[1]

he modern unit root tests and power can be improved by using GLS detrending. One way to do this is by defining the quasi-differenced variables:

$yt = y_t - \bar{\alpha} * yt - 1 and zt = z_t - \bar{\alpha} * zt - 1 for t >= 2, y_1 = y_1 and z_1 = z_1.$

The Frisch-Waugh-Lovell theorem can be used to perform an OLS regression of y_t on z_t to obtain the coefficient estimate .

Then we can define the GLS detrended observations (residuals) as:

$\hat{y}_t = y_t - z_t' * \hat{\beta}$

After that, the ADF-GLS regression can be performed:

$$\Delta \hat{y}t = \pi \hat{y}t - 1 + \sum_{j=1}^{p-1} \gamma_j \Delta \hat{y}_{t-j} + \epsilon_t$$

and then compute the t-statistic of π_π^{GLS} as usual.

The Dickey-Fuller regression model is defined as: $yt = \phi y_{t1} + ut$

where y is the time series, ϕ is the unit root parameter and ut is the error term. Subtracting y_{t1} from both sides, we can get the first-differenced series:

$$\Delta yt = 1 + ut$$

where $\pi = \phi - 1$

So, $\pi = \phi - 1$

It represents the deviation of the true parameter value from a unit root. Thus if the true parameter is a unit root, $\phi = 1$, then $\pi = \phi - 1 = 1 - 1 = 0$.

Definition 9: Morten UR test

he GLS detrending method improves the power of unit root tests, but it lacks a firm statistical foundation in the present context. Jansson and Nielsen (2012) have introduced Likelihood Ratio (LR) tests of the unit root hypothesis that are grounded in statistical theory, such as likelihood theory, Neyman-Pearson lemma, and invariance.

The LR test is based on the likelihood function $L(\pi, \theta)$, where $\theta = (\beta, y1, ..., y_{p1})$ represents a set of parameters. The LR statistic is defined as the ratio of the maximum likelihood under the alternative hypothesis ($\pi < 0$) to the maximum likelihood under the null hypothesis ($\pi = 0$):

$$LR = \frac{max\ \pi \leq 0, \theta L(\pi, \theta)}{max\ \pi = 0, \theta L(\pi, \theta)}$$

Like the ADF-GLS test, the LR test achieves a power envelope, which is a set of all possible test statistics that could be produced by any test with a given sample size.

What kind of stationarity exisist?

1.5 Questions to ask each other

1. What is the difference between trend-stationarity and difference-stationarity? How do these concepts apply to univariate time series data?
2. How does the distribution theory for stationary and non-stationary time series differ, and how do the central limit theorem (CLT) and functional central limit theorem (FCLT) apply in these cases?
3. What are classical unit root tests in autoregressive (AR) models, and how are they used to determine the stationarity of a time series? What are the key properties of these tests, such as their distribution and size and power?
4. How do modern unit root tests differ from classical unit root tests, and what are some examples of modern unit root tests? How is GLS detrending used to analyze non-stationary time series data, and what are the key steps in this process?
5. How can you test for stationarity in a time series, and what are some common methods for doing so?
6. What are the implications of a time series being stationary or non-stationary for modeling and forecasting purposes?
7. How does the choice of unit root test affect the results of the analysis, and how do you choose the appropriate test for a given time series?
8. How can you model a non-stationary time series in order to make it stationary, and what are some common techniques for doing so?
9. How does GLS detrending compare to other methods for removing trends and non-stationary features from time series data?
10. How do trends and seasonality affect the stationarity of a time series, and how can these features be removed or accounted for in the analysis?
11. What are the limitations of classical unit root tests, and how do modern unit root tests address these limitations?
12. How does the sample size and time span of the time series data affect the results of unit root tests and GLS detrending?
13. How do different types of errors, such as heteroscedasticity and autocorrelation, impact the validity of unit root tests and GLS detrending?
14. How can you test for cointegration in a time series, and what are the implications of cointegration for modeling and forecasting?
15. Can you derive a general unit root processes? (slide 7 lecture 2)
16. What is a Brownian motion? (slide 15 lecture 2)
17. What is the distribution of $\hat{\pi}$? (slide 20 lecture 2) [2].
18. Can you name approaches with general u_t? (Phillips-Perron test: Estimate λ and y_0 (nonparametrically) and make a correction. Works poorly and rarely used in practice. 2 Augmented Dickey-Fuller regression.)
19. What do you know of the power of different UR tests? (Power functions, slide 32 lecture 45)
20. What are the main advantages of using LR tests of the unit root hypothesis, introduced by Jansson and Nielsen (2012), over the GLS detrending method for improving the power of unit root tests?
21. What are the key assumptions or theoretical foundations that LR tests of the unit root hypothesis, introduced by Jansson and Nielsen (2012), are based on and how do they differ from those underlying the GLS detrending method?

[2] The distribution of the OLS estimator of $(\hat{})$ in the Dickey-Fuller regression model is asymptotically normal with a mean of 0 (under the null hypothesis of a unit root) and a finite variance. However, the distribution is typically unknown and approximated by a standard normal distribution, allowing for the use of the t-statistic for hypothesis testing. This test is often referred to as the Augmented Dickey-Fuller (ADF) test. However, it has a few assumptions, such as that the error term ut has a constant variance, is normally distributed and not serially correlated

22. What kind of trends exists?
23. What is the definition of "weak stationarity"?[3]
24. What does a R^2 of 0.78 mean?[4]
25. How does a unit root test look like?
26. What iss a "random function step function"?
27. What is the most powerful test? [5]
28. Why would you use a VAR model?[6]
29. What is the difference between variance and volatility? [7]

[3]Mean and variance doesn't depend on time.
[4]78 procent variance can be explained by x.
[5]LR slide 45 lecture 2.
[6]If the researcher is interested in the relationshipt between the time series.
[7]Volatility and variance are both measures of the dispersion of returns for a given security or market index. However, they are calculated differently. Volatility is a measure of the deviation of the returns from the mean, typically over a short period of time (e.g. daily or weekly). It is often used as a measure of risk. Variance is a measure of the deviation of returns from the mean, typically over a longer period of time (e.g. monthly or annually). It is the square of the standard deviation. In summary, volatility is a measure of short-term price fluctuations, while variance is a measure of the distribution of returns over a longer period.

CHAPTER 2

2. Introduction to the vector autoregressive model

Definition 10: VAR model structure

he Vector autoregressive (VAR) model is a model that is used to capture dynamic interactions among multiple variables. Let y_t be a $Kx1$ vector representing the observed data for time t, where K is the number of variables in the model.

A VAR(p) model is a multivariate extension of the AR(p) model, which can be written as:

$$y_t = +A_1 y_{t-1} + A_2 y_{t-2} + ... + A_p y_{t-p} + u_t$$

u_t is the $Kx1$ error term for time t, which is iid (independently and identically distributed) with a mean vector of zero and a variance matrix $\sum_u$, where the error terms may be correlated. The parameters of the VAR(p) model are:

1. ν (K x 1), representing the constant term
2. A_j (K x K), representing the coefficient matrices for past lags
3. $\sum_u$ (K x K), representing the variance-covariance matrix of the error term.

The VAR(p) model can be written as:

$\mathbf{y_t} = \nu + \sum_{i=1}^{p} A_i y_{t-i} + u_t$

$\mathbf{u_t} \sim \mathbf{iid}(\mathbf{0}, \mathbf{\Sigma_u})$

2.1 (i) Explain estimation of the VAR model.

To estimate the parameters of a vector autoregressive (VAR) model, you need to specify the number of lag terms to include in the model and choose an estimation method.

One common method for estimating the parameters of a VAR model is maximum likelihood estimation (MLE). In MLE, you choose the parameters that maximize the likelihood function, which is a measure of the probability of the observed data given the model and its parameters.

To fit a VAR model using MLE, you need to specify the form of the model and the distribution of the errors. The errors are the differences between the predicted values and the observed values. Common choices for

the error distribution include the normal distribution and the student's t-distribution.

Once you have specified the model and the error distribution, you can use optimization algorithms, such as the Newton-Raphson algorithm or the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, to find the parameter values that maximize the likelihood function.

There are also other methods for estimating the parameters of a VAR model, such as generalized least squares and Bayesian estimation. These methods can be used instead of or in combination with MLE, depending on the needs and goals of your analysis.

Definition 11: Estimation of VAR model

stimating the parameters of a VAR model is relatively simple, because the error term u_t is uncorrelated with the past observations of y_t, that is y_{t-1}, y_{t-2}, ..., y_{t-p}. Therefore, we can use the Ordinary Least Squares (OLS) method, equation by equation to estimate the parameters.

The estimation is based on the sum-of-squares criterion function, which starts at $t = p + 1$ and runs until $t = T$. This is because the residuals, $e_t = y_t \hat{\nu} \hat{A}_1 y_{t-1}, ..., \hat{A}_p y_{t-p}$, can only be calculated for $t = p + 1, ..., T$.

The OLS estimation method is equivalent to the maximum likelihood estimation method and is efficient when the error term u_t is normally distributed. However, it is still consistent and asymptotically normal in general, even when y_t is stationary.

Definition 12: Hypothesis of VAR model

When working with VAR models, it's necessary to have the ability to test hypotheses. One approach that can be used is calculating t-tests and F-tests, also known as Wald tests. But this method requires an estimate of the variance matrix of the entire parameter vector, which can be quite complex to calculate.

A simpler method for testing hypotheses is to use the Likelihood Ratio (LR) test:

$$LR = 2(lmax, H1 - lmax, H0) = T(log(det(\tilde{\Sigma_u})) - log(det(\hat{\Sigma_u})))$$

Where $\hat{\Sigma}_u$ and $\tilde{\Sigma}_u$ are the unrestricted and restricted estimates of the error term variance-covariance matrix, respectively. Under the assumption of stationarity, the LR statistic follows a χ^2 distribution with q degrees of freedom, where q is the number of restrictions imposed under the null hypothesis.

Definition 13: Principle of parsimony

The principle of parsimony, also known as the principle of simplicity or the principle of Occam's razor, states that, when multiple explanations or models are available to explain a given phenomenon, the one with the simplest or fewest assumptions is the one that should be preferred. In other words, parsimony is the preference for the simplest explanation that is consistent with the available data.

In the context of VAR models, this principle suggests that we should choose the smallest VAR(p) model that can adequately explain the data, while avoiding unnecessary complexity or overfitting. This is a common practice when choosing the lag order of a VAR model, which determines the number of lags (p) included in the model. In general, when determining the lag order of a VAR model, a balance must be found between a model that is too simple and one that is too complex.

There are different ways to implement this principle when choosing the appropriate number of lags for a VAR model such as:

Information criteria, like Akaike (AIC) or Bayesian (BIC), which penalizes models for having more parameters General-to-specific testing, in which the researcher starts with a model with a high number of lags and successively removes lags until the restrictions being imposed are no longer rejected by the data.

This principle of parsimony is a common practice across many fields of study and research, including time series econometrics, physics, statistics and machine learning.

Definition 14: AIC

The AIC criteria is the most liberal criteria, and in general will allow for more lags:

$$AIC = 2k - 2ln(\hat{L})$$

Where K is the number of estimated parameters and L is the maximized value of the likelihood function

2.2 (ii) Discuss stationarity conditions.

Definition 15: Stationarity condition

VAR model: stationarity condition

The stationarity condition for a VAR(p) model is that

$\det(A(z)) \neq 0$ for $|z| \leq 1$. Equivalently, the roots of the characteristic polynomial, $\det(A(z))$, i.e. the solutions of $\det(A(z)) = 0$, have to be greater than one in modulus. Similar to univariate AR(p).

Determining stationarity vs nonstationarity: Calculate roots and list/plot them. Visual/informal impression. Test univariate unit root hypothesis (inefficient because it misses all interactions).

A time series is said to be stationary if its statistical properties, such as the mean and variance, are constant over time.

In the context of a vector autoregressive (VAR) model, stationarity is an important assumption because it ensures that the model is well-defined and the estimated parameters have statistical meaning. If a time series is not stationary, it may be transformed to make it stationary before fitting a VAR model.

There are several ways to test for stationarity in a time series. One common method is the Augmented Dickey-Fuller (ADF) test, which tests the null hypothesis that a time series is non-stationary. If the null hypothesis is rejected, it means that the time series is stationary.

Another method is the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, which tests the null hypothesis that a time series is stationary. If the null hypothesis is rejected, it means that the time series is non-stationary.

It's important to note that stationarity is not a necessary condition for a VAR model, but it can make the model easier to interpret and work with. If a time series is not stationary, you may need to apply a transformation, such as differencing, to make it stationary before fitting a VAR model.

1. The mean of the process is constant over time (i.e., the time-varying part of the mean is zero).
2. The autocovariance of the process is constant over time (i.e., the time-varying part of the autocovariance is zero).
3. The partial autocorrelations of the process are constant over time (i.e., the time-varying part of the partial autocorrelations is zero).

Definition 16: Strong and weak stationarity

Strong stationarity:
A stochastic process is said to be strongly stationary if its probability distribution does not change over time. This means that the mean, variance and autocovariance of the process are constant over time. A time series that is strongly stationary will have the same statistical properties at any point in time.
Weak stationarity:
A stochastic process is said to be weakly stationary if its mean and autocovariance are not necessarily constant over time, but they are at least independent of the time index. This means that the mean, variance and autocovariance of the process may change over time, but their probability distributions do not depend on the time index. A time series that is weakly stationary will have similar statistical properties at different points in time, but these properties may change over time.
Difference:
The main statistical difference between weak and strong stationarity is that weak stationarity only requires that the mean and covariance of a time series are constant over time[a], while strong stationarity requires that the entire probability distribution of a time series is constant over time.

[a] When we say that the covariance is constant over time in the context of weak stationarity, we mean that the covariance between any two observations at different time lags, cov(y_t, y_t-k), does not depend on the specific time point at which the series is observed and it is the same for all time points. This means that the correlation between any two observations at different time lags does not change over time. In other words, if a time series is weakly stationary, the covariance between any two observations at different time lags does not change over time, which means that the linear relationship between the observations does not change over time. This property allows us to make predictions about the future values of the series based on past observations and it is useful for modeling and forecasting time series data.

2.3 (iii) Derive and discuss forecasting.

Definition 17: Slide with forecast of VAR

The VAR model can be used for prediction and forecasting by using the best linear predictor (BLP) method. The BLP method is a relatively straightforward way of predicting future values of the time series based on the past values and the estimated VAR parameters. The equation for the BLP method for a VAR(p) model with known parameters is:

$$\hat{y}_{T+m|T} = \nu + \sum i = 1^{p} A_i \hat{y}_{T+m-i|T}$$

Where, $\hat{y}_{T+m|T}$ is the predicted value of y at time T+m given the information available at time T ν is the VAR model's constant term A_i is the estimated coefficients of the VAR model $\hat{y}_{T+m-i|T}$ is the predicted value of y at time T+m-i given the information available at time T With unknown parameters, we usually replace them with estimates and ignore estimation error, this method provides point forecasts for future observations. However, forecasting uncertainty can be assessed using standard methods, such as producing confidence intervals for the forecasts, this can be done through Bootstrapping or using the Kalman filter.

Derivation 4: Step by step derivation of VAR forecast

The equation for the best linear predictor (BLP) of the VAR(p) model is derived by assuming that the parameters of the VAR(p) model are known. Using this assumption, we can use the estimated VAR coefficients to make predictions for future observations of the time series.

Here is the step by step derivation:

The VAR(p) model can be written as:

$$y_t = \nu + A_1 y_{t1} + A_2 y_{t2} + ... + A_p y_{tp} + u_t$$

[a] We want to predict the value of y at time T+m given the information available at time T, which can be represented as:

$$\hat{y}_{T+m|T}$$

Using the equation from step 1, we can substitute in the estimated parameters and past observations to get:

$$\hat{y}_{T+m|T} = \nu + A_1 \hat{y}_{T+m-1|T} + A_2 \hat{y}_{T+m-2|T} + ... + A_p \hat{y}_{T+m-p|T}$$

Re-arranging the equation and grouping the summation of past observations together to get the final BLP equation:

$$\hat{y}_{T+m|T} = \nu + \sum_{i=1}^{p} A_i \hat{y}_{T+m-i|T}$$

[a] y_t represents the K x 1 vector of observed data at time t, where K is the number of variables in the model. ν represents the K x 1 vector of constants, which capture the average values of the variables. A_1, A_2, ... , A_p represent the K x K matrices of coefficients, which capture the effects of the lagged values of the variables on the current values of the variables. y_{t-1}, y_{t-2}, ... , y_{t-p} represent the lagged values of the variable vector. u_t represents the K x 1 vector of error term for time t, which is uncorrelated with the "regressors" y_{t-1}, . . . , y_{t-p} with mean vector zero and variance matrix Σ_u (can be correlated).

Forecasting with a VAR model involves using the model to make predictions about the future values of multiple time series. VAR models are based on the assumption that each time series in the model is influenced by the other time series as well as its own past values.

To make a forecast with a VAR model, you first need to estimate the model parameters using historical data. This involves fitting the model to the data and estimating the coefficients of the model.

Once the model is estimated, you can use it to make forecasts for the future values of the time series. This typically involves setting the values of the other time series to their most recent observed values and using the model to predict the value of the time series of interest at the next time step. You can then repeat this process to make forecasts for multiple future time steps.

There are several factors to consider when forecasting with a VAR model, including the choice of lag order, the choice of forecasting horizon, and the presence of missing data or structural breaks in the data. It is also important to validate the forecast by comparing it to actual observations and to assess the accuracy of the forecast. To derive a forecast for a VAR model, you can follow these steps:

How to forecast with a VAR model
Estimate the VAR model using historical data. This involves fitting the model to the data and estimating the coefficients of the model.
Choose the lag order of the VAR model. The lag order determines the number of lags of the time series that are included in the model. A higher lag order may result in a better fit to the data, but may also increase the complexity of the model.
Choose the forecasting horizon. The forecasting horizon is the number of time steps into the future for which you want to make a forecast.
Set the values of the other time series to their most recent observed values.
Use the estimated VAR model to predict the value of the time series of interest at the next time step.
Repeat steps 4 and 5 to make forecasts for multiple future time steps.
Validate the forecast by comparing it to actual observations and assessing the accuracy of the forecast.

Derivation 5: Derive a VAR model forecast

In a Vector Autoregression (VAR) model, a forecasting equation for a variable of interest (say, Y_t) can be derived by assuming that it depends on a set of other variables (say, X_t) as well as its own lagged values (Y_t-1, Y_t-2, ...). The forecasting equation takes the form:

$$Y_t = c + A_1Y_{t-1} + A_2Y_{t-2} + ... + A_pY_{t-p} + B_1X_t + B_2X_{t-1} + ... + B_qX_{t-q} + e_t$$

where c is a constant term, A_i and B_i are coefficients, and e_t is a white noise error term. The integers p and q represent the orders of the autoregressive and exogenous variables respectively. To forecast Y_t+h, we can substitute in the forecasted values for the exogenous variables and lagged endogenous variables

$$Y_{t+h} = c + A_1Y_{t+h-1} + A_2Y_{t+h-2} + ... + A_pY_{t+h-p} + B_1X_{t+h} + B_2X_{t+h-1} + ... + B_qX_{t+h-q} + e_{t+h}$$

And we can substitute the estimated values for all coefficients and variables up to time t, and solve for Y_t+h

Notice that this model is a linear equation, so it requires linearity in the data, also if you have a stationary series for Y_t (covariance does not depend on time), only then can you assume this model to be valid. Also this requires estimation of the parameters (coefficients) .

Here's an example of a VAR(1) forecasting equation for a variable Y_t, where Y_t depends on its own lagged value (Y_t-1) as well as an exogenous variable X_t.

Assume we have estimated the following coefficients using historical data:

A_1 = 0.5 B_1 = 0.3 c = 2

Now let's say we want to forecast Y_t for the next period (i.e., t+1). To do this, we need to know the values of Y_t+1-1 (i.e., Y_t) and X_t+1 (i.e., the exogenous variable at the next period). Let's assume the following: $Y_t = 5$ $X_{t+1} = 8$

Using this information, we can plug the values into the forecasting equation: $Y_{t+1} = 2 + 0.55 + 0.38$

This would give us the following forecast for Y_{t+1} $Y_{t+1} = 4.1$

Please note that this is a very simple example and in reality, data sets will have more complexity, but the example is meant to demonstrate the basic idea of how a VAR model works for forecasting and how coefficients and variables are used for making forecast

2.4 (iv) Define and discuss Granger causality.

Granger causality is a concept in econometrics and time series analysis that refers to the ability of a variable to cause changes in another variable. Specifically, it tests the null hypothesis that the past values of one variable (X) do not provide any information about the future values of another variable (Y) beyond what is already in the past values of Y.

The test is conducted by fitting a VAR model of the form $Y_t = c + A_1Y_{t-1} + A_2Y_{t-2} + ... + A_pY_{t-p} + B_1X_t + B_2X_{t-1} + ... + B_qX_{t-q} + e_t$ and compare the fit of the model with and without X_t, usually by a likelihood ratio test. If the inclusion of X_t improves the model fit, then X is said to Granger-cause Y. It is important to note that this method doesn't prove causality only that there's correlation, to test causality other methods must be used like experiments and observational studies.

Granger causality is widely used in economics, finance, and other fields to analyze cause-and-effect relationships among variables, but it is important to note that it can be misused, as it only provides evidence of correlation and not of causality, a deeper understanding of the problem and the use of other methods are recommended to support causal claims.

Definition 18: Definition for Granger Causality

Definition: Granger causality. The time series y_{1t} Granger causes another time series, y_{2t}, if and only if

$$P(y_{2t}|lags(y_{1t}, y_{2t})) \neq P(y_{2t}|lags(y_{2t}))$$

Granger causality is a statistical concept in econometrics and time series analysis that refers to the ability of a variable to cause changes in another variable. Specifically, it tests the null hypothesis that the past values of one variable (X) do not provide any information about the future values of another variable (Y) beyond what is already in the past values of Y.

Granger causality test is a method for determining if one variable Granger causes another variable in a VAR(p) model. The main ideas behind this method are:

The causing variable (y1t) should lag the affected variable (y2t), meaning that the cause must happen before the effect can occur. The causing variable should contain additional information about the affected variable, beyond the information that is already contained in the history of the affected variable. This is tested by comparing the fit of the VAR(p) model with and without the causing variable, and determining if the fit improves when the causing variable is included.

Example where there is Granger Causality

An example of Granger causality in economics would be the relationship between interest rates and inflation. The traditional view is that an increase in interest rates will lead to a decrease in inflation, as the cost of borrowing becomes more expensive and this discourages spending, which in turn reduces the pressure on prices. To test this relationship, one could conduct a Granger causality test using time series data on interest rates and inflation.

In this case, the null hypothesis would be that past values of interest rates do not provide any information about future values of inflation beyond what is already in past values of inflation, and the alternative hypothesis is that past values of interest rates provide information about future values of inflation.

By fitting a VAR model with both variables, we can test whether the inclusion of interest rates improves the model fit and if so, we can say that there is Granger causality running from interest rates to inflation.

It is important to note that this test only implies correlation and not causality, other methods and understanding of the problem should be used to support causality claims.

Granger causality in a k=2 VAR Example, bivariate VAR(p):

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \sum_{j=1}^{p} \begin{pmatrix} A_{j,11} & A_{j,12} \\ A_{j,21} & A_{j,22} \end{pmatrix} \begin{pmatrix} y_{1,t-j} \\ y_{2,t-j} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

Here, y_{1t} does not Granger cause y_{2t} if $A_{j,21} = 0$ for $j = 1, ..., p$. This null is easily tested using, e.g., LR test.

2.5 (v) Discuss impulse response functions.

Impulse response functions (IRFs) are a tool used in econometrics and time series analysis to examine how a shock to one variable affects the other variables in a system over time. In other words, they allow us to see how a change in one variable at a specific point in time (the "impulse") propagates through the system and affects the other variables.

In a VAR(p) model, an impulse response function (IRF) is the response of variable Y to a one-unit shock in variable X, holding all other variables constant. IRF's are typically represented graphically, with the horizontal axis representing the time periods after the shock, and the vertical axis representing the magnitude of the response of the variable of interest.

For example, let's consider a VAR(1) model with two variables, Y and X, where $Y_t = A_1 Y_{t-1} + B_1 X_t + e_t$ The IRF for Y in response to a shock of X would be given by: IRF

$$(Y, X) = B_1 *_{i=0}$$

A_1^i * (t-i)

$$Where$$

$\delta(t-i)$ is the dirac delta function, which gives the instantaneous response of Y to the impulse in X. The IRF graph is the response of Y to a unit impulse in X, it would be plotted over time, and the curve shows how Y responds over time to the impulse in X.

It is worth noting that IRF are not only used in VAR models, but also in other models such as vector error correction model (VECM), and structural Vector Autoregression models (SVAR). Additionally, IRF can be used in combination with variance decomposition, to understand the relative importance of different impulses to the variability of a variable

Overall, impulse response functions provide a way to understand how a shock to one variable affects the other variables in a system over time. They are useful in analyzing the dynamic relationships among variables in an economy, and in identifying the transmission mechanisms of monetary and fiscal policies.

Definition 19: Impulse response functions

Slide 24 lecture 3

2.6 Questions to ask each other

1. Define and explain the concept of cointegration.
2. How is causality determined in econometrics?
3. Explain what is VAR models and the assumptions for its estimation
4. Describe the steps for conducting a Granger causality test.
5. Define and explain the concept of impulse response function in a VAR model.
6. Discuss the difference between forecasting and prediction in econometrics
7. Estimate a ARIMA(p,d,q) model for a given time series data and interpret the results
8. Explain what is a stationarity condition and give an example of a time series data that is not stationary
9. Define the problem of endogeniety and how can it be addressed
10. Compare and contrast between panel data and time series data and explain when should one use one or the other.
11. Explain the concept of unit roots and the Dickey-Fuller test for determining stationarity.
12. Describe the differences between a VAR and VECM model.
13. Define and explain the concept of structural breaks in time series data.
14. Discuss how to diagnose and deal with autocorrelation in a time series data.
15. Explain the concept of multivariate GARCH models and their applications
16. Define and explain the concept ofARCH and GARCH models and when are they used.
17. Describe the differences between a deterministic and a stochastic trend.
18. Explain the concept of co-integration and the Johansen test.
19. Describe the basics of Bayesian estimation in econometrics and time series analysis
20. Discuss how to deal with missing data in time series.
21. Derive the forecasting equation for a VAR(p) model.
22. Derive the OLS estimator for a simple linear regression model.
23. Derive the likelihood function for an AR(p) model and explain how it can be used for estimation.
24. Derive the Kalman filter equations for a linear state-space model.

25. Derive the expressions for the asymptotic variance-covariance matrix of the OLS estimator under different assumptions.
26. Derive the likelihood ratio test for testing for unit roots in a time series.
27. Derive the impulse response function for a VAR(p) model and explain its properties.
28. Derive the likelihood function for a GARCH(p,q) model and explain how it can be used for estimation.
29. Derive the expression for the forecast error variance in a VAR model
30. Derive the dynamic OLS estimator for a VAR model.
31. Derive the likelihood function for a VAR(p) model and explain how it can be maximized to obtain the MLE estimator for the parameters.
32. Derive the Yule-Walker equations for an AR(p) model and explain how they can be used for estimation.
33. Derive the Kalman filter equations for a non-linear state-space model and explain how they can be used for estimation.
34. Derive the likelihood function for a VARMA(p,q) model and explain how it can be used for estimation.
35. Derive the likelihood function for a VECM model and explain how it can be used for estimation.
36. Derive the likelihood function for a TVP-VAR model and explain how it can be used for estimation.
37. Derive the expressions for the asymptotic variance-covariance matrix of the MLE estimator for a VAR(p) model.
38. Derive the exact maximum likelihood estimator for a univariate GARCH(1,1) model.
39. Derive the OLS estimator for a multivariate regression model with multiple endogenous and exogenous variables.
40. Derive the likelihood function for a threshold VAR model and explain how it can be used for estimation.
41. Derive the likelihood function for a VAR(p) model and explain how it can be maximized to obtain the MLE estimator for the parameters.
42. Derive the Yule-Walker equations for an AR(p) model and explain how they can be used for estimation.
43. Derive the Kalman filter equations for a non-linear state-space model and explain how they can be used for estimation.
44. Derive the likelihood function for a VARMA(p,q) model and explain how it can be used for estimation.
45. Derive the likelihood function for a VECM model and explain how it can be used for estimation.
46. Derive the likelihood function for a TVP-VAR model and explain how it can be used for estimation.
47. Derive the expressions for the asymptotic variance-covariance matrix of the MLE estimator for a VAR(p) model.
48. Derive the exact maximum likelihood estimator for a univariate GARCH(1,1) model.
49. Derive the OLS estimator for a multivariate regression model with multiple endogenous and exogenous variables.
50. Derive the likelihood function for a threshold VAR model and explain how it can be used for estimation.
51. When can't we use OLS to estimate VAR model? [1]

[1] OLS can be used to estimate VAR models when the error term is uncorrelated with the past observations of the dependent variable. However, there are some situations when OLS may not be the appropriate method for estimating a VAR model. Some of these situations include: Non-normality of errors: If the error terms are not normally distributed, OLS estimates may be biased. Correlated errors: If the error terms are correlated across observations, OLS estimates may be inconsistent. Multicollinearity: When the variables in the VAR model are highly correlated with one another. This can make the OLS estimates unstable, so other methods such as Ridge or Principal component analysis(PCA) are recommended. Time Series is not stationary: If the time series is non-stationary, the assumptions of OLS are not met, and this might lead to unreliable parameter estimates. Presence of Structural Breaks or outliers: Presence of structural breaks or outliers may lead to incorrect estimation of parameters using OLS, alternative methods like robust estimation should be used. It's important to check for these assumptions before estimating the model and if any of these assumptions are not met then alternative methods may be needed to estimate the parameters of the VAR model

52. What methods is there to estimate VAR models? [2]
53. How do you chose the right Lag lenght?
54. What is the stationary condition?
55. What is eigenvalues[3] used for in terms of stationarity?[4]
56. What eigenvalues is needed for stationarity?[5]

1. Define the VAR model and explain its key components.
2. Explain the difference between a univariate and multivariate time series model.
3. Discuss the assumptions for a VAR model to be identified.
4. Explain the difference between a VAR and VECM model
5. Describe the method for estimating the parameters of a VAR model.
6. Explain the difference between ordinary least squares (OLS) and maximum likelihood estimation (MLE) in VAR models.
7. Explain the role of lag selection in VAR model estimation
8. Define stationarity and its relevance to VAR models
9. Explain how to test for stationarity in a VAR model.
10. Discuss the difference between a stationary and non-stationary VAR model.
11. Give an example of a time series that is not stationary and explain how it would affect a VAR model.
12. Explain the problem of endogeneity in VAR models and the implications for interpretation.
13. Describe the steps for identifying the causality between variables in a VAR model.
14. Discuss the limitations of interpreting the coefficients in a VAR model.
15. Define forecasting and prediction in a VAR model.
16. Explain how to forecast with a VAR model.
17. Explain how to evaluate the forecast from a VAR model.
18. Define Granger causality and explain how it is determined in a VAR model.

[2]There are several methods for estimating VAR models, in addition to the ordinary least squares (OLS) method: Maximum likelihood estimation (MLE): In this method, the likelihood function is maximized with respect to the parameters of the model. It is an efficient method when the errors are normally distributed. Generalized method of moments (GMM): This method involves choosing a set of moment conditions that the parameters of the model must satisfy. The parameters are then estimated by matching these moments with sample moments. Bayesian estimation: In this method, prior distributions are placed on the parameters of the model, and the posterior distributions are obtained using Bayes' theorem. Instrumental Variables (IV) estimation : This method is used when there are endogenous variables in the model, IV estimates are robust to endogeneity caused by omitted variables. Kalman filter: This method is useful when dealing with time series with measurement errors, it uses a recursive algorithm to estimate the parameters and the state of the system. State space estimation: This method is similar to the Kalman filter, but it is more general and can be used to estimate a wider class of models.

[3]An eigenvalue is a scalar value that can be associated with a matrix. Given a square matrix A, an eigenvalue λ is a scalar value that satisfies the following equation for some non-zero vector x:

$$Ax = \lambda x$$

Where $A * x$ is the matrix-vector product of A and x.

[4]Eigenvalues are used to determine the stationarity of a VAR(p) model by checking whether they have a modulus less than one or not. An eigenvalue of a matrix is a scalar value that can be obtained by solving the characteristic equation of that matrix. The characteristic equation of a matrix A is a polynomial equation in the variable z, and its roots are the eigenvalues of A.

In the case of VAR(p) models, the eigenvalues of the matrix A(z) are the roots of the characteristic polynomial, which is the determinant of A(z). The stationarity condition for a VAR(p) model is that the determinant of A(z) is not equal to 0 for |z| 1. And since eigenvalues are the roots of the characteristic polynomial, this is equivalent to the condition that the eigenvalues have a modulus less than one.

Therefore, to check the stationarity of a VAR(p) model, we calculate the eigenvalues of A(z) and check if all the eigenvalues have a modulus less than one. If all eigenvalues have a modulus less than one, then the VAR(p) model is stationary. If any of the eigenvalues has a modulus greater than or equal to one, then the VAR(p) model is non-stationary.

[5]For a VAR(p) model to be stationary, it is required that all the eigenvalues of the matrix A(z) have a modulus (absolute value) less than 1.

19. Describe the impulse response function in a VAR model and explain its properties.
20. What is the Principle of parsimony?
21. Can you write up the equations for an Impulse reponse curve?
22. How do you chose the numper p of VAR(p)?[6]
23. Does roots always need to be inside the "unitroot circle"?[7]
24. What is the problem when ϵ_t's are correlated?[8]
25. When should you add dummy variables?
26. How does a "step dummy" look like?
27. How does a "heart beat dummy" look like?

[6]The residuals need to look like a white noise, that is what's important (you need to remove the corelation from the errorterm).
[7]NO! Its the inverse roots that needs to be inside.
[8]Then ceterus paribus doesn't hold.

CHAPTER 3

3. Structural vector autoregressive models

3.1 (i) Compare reduced and structural forms of the VAR model.

Reduced form VAR

$$y_t = v + A_1 y_{t-1} + ... + A_p y_{t-p} + \epsilon_t$$

SVAR

$$\beta_0 y_t = \beta_0 A_1 y_{t-1} + \beta_0 A_p y_{t-p} + \epsilon_t \beta_0$$

We get

$$\beta_0 y_t = \beta_1 y_{t-1} + \beta_p y_{t-p} + w_t$$

The reduced form VAR(p) model is given by: $y_t = A_1 y_{t-1} + A_2 y_{t-2} + ... + A_p y_{t-p} + u_t$ where $Var(u_t) = \Sigma_u$, which is the reduced form. Because that Σ_u is generally not diagonal, and we typically have correlation between the reduced form errors, u_t. This can be causing problems regarding the impulse response analysis, which should be on a ceteris paribus basis. Leaving it to be hard to interpret and assign a meaning to the shock, as the innovation in equation j is not only happening in equation j. The benefit regarding the reduced form, is that it is easy to estimate.
Moving un to the structural form, we are interested in finding a representation of the VAR where the errors are orthogonal (statistically independent, do not affect one anothor). This is done by considering the KxK matrix B_0 such that: $B_0 \Sigma_u B_0' = I_K$. We can the pre-multiply both sides with B_0 to obtain the structural form:
$B_o y_t = B_0 A_1 y_{t-1} + B_0 A_2 y_{t-2} + ... + B_0 A_p y_{t-p} + B_0 u_t$
This can be simplified into: $B_0 y_t = B_1 y_{t-1} + B_2 y_{t-2} + ... + B_p y_{t-p} + \omega_t$ Where $B_i = B_0 A_i$ for all $i = 1, ..., p$ and $\omega_t = B_0 u_t$ where the structural error ω_t satisfies:
$Var(\omega_t) = Var(B_0 u_t) = B_0 \Sigma_u B_0' = I_K$
This can act as mapping between the reduced form and the structural form, where we arrive at the structural form by premultiplying with the matrix B_0, in order to move from the structural form, back to the reduced form, we can then multiply with B_0^-1
To sum up the difference between the structural VAR model, and the reduced for VAR model, the difference is that the structural VAR model is multiplied with the matrix B_0, that creates and identity matrix for the errorterm, which then eliminates the correlation between the errors.

3.2 (ii) Discuss impulse response functions.

This section is considering the impulse response functions based on the structural erros, ω_t. These impulse response function can be considered more sensible, because the errors are uncorrelated, so looking at an impulse from one variable in isolation, makes greater sense.

To find the impulse responses based on the orthogonal errors, we obtain the linear representation of the structural model:
$y_t = B(L)^{-1}\omega_t = O(L)\omega_t$
Where $B(z) = B_0 - B_1 z - B_2 z^2 - ... - B_p z^p$ and $O(L) = O_0 + O_1 z + O_2 z^2 + ...$ And we know that $u_t = B_0^{-1}\omega_t$ and thus:
$y_t = u_t + \phi_1 u_{t-1} + \phi_2 u_{t-2}...$
$= B_0^{-1}\omega_t + \phi_1 B_0^{-1}\omega_{t-1} + \phi_2 B_0^{-1}\omega_{t-2}...$
So that $O_i = \phi_i B_0^{-1}, i = 0, 1, 2....$
It is the possible, in the structural model, to create shocks to one variable, keeping the other variables fixed, and see the effect from that variable on the economic variables of choice. And see only the effects based on that one variable, as there is no correlation between the error-terms, using the structural model. Where in the reduced form, having correlation between the error-terms, would not allow for such analysis. This generates the possibility to consider the ceteris paribus statement, in the economic analysis. The analysis can then be used to understand the economic affects of one variable, on other variables over time, keeping those variables constant.
It is required to have estimates of $B_O, ..., B_p$ to determine the structural impulse response function.
Thinking back to the reduced form, it is easy to obtain estimates of $A_1, ..., A_p$, using OLS regression. But this is not enough to solve for $B_O, ..., B_p$, using $B_i = B_0 A_j$. Which is a classical example of Simultaneous Equations Model, with an identification problem.

3.3 (iii) Explain identification in general and the identification problem in the SVAR model.

We have already established that $B_=\Sigma_u B_0' = Var(\omega_t) = I_K$, we can then assume that B_0 is invertible, we can then state that $\Sigma_u = B_0^{-1}(B_0^{-1})'$ If it was possible to solve this nonlinear system of equation for B_0 then we could use the expression $B_i = B_0 A_i$ to solve for the remaining B_i based on A_i. The is however a problem here, that there are $K(K+1)/2$ unique elements in Σ_u, so there are $K(K+1)/2$ equations in $\Sigma_u = B_0^{-1}(B_0^{-1})'$ but there are only K^2 unknowns in B_0, which we in turn cannot solve uniqualy, therefore we need to impose an additional $K(K-1)/2$ restrictions on the matrix B_0 in order to obtain a unique solution.

3.4 (iv) Discuss short-run restrictions and long-run restrictions.

Short run restrictions are based on either Cholesky or Diagonalization, this can be called recursively identified. Starting off with the Cholesky decomposition, which uses the argunment that for any symetric positife definite matrix Σ_u, there exists a unique lower triangular matrix Q, such that $\Sigma_u = QQ'$, we can then put $B_0 = Q^{-1}$ making sure that B_0 is also lower triangular.

The second short run restriction that we can consider is the Diagonalization, which uses the fact that for any symmetric positive-definite matrix Σ_u, there exists a unique lower triangular matrix Q with ones on the diagonal and a unique diagonal matrix D such that $\Sigma_u = QDQ'$. Then we can put $D = \Sigma_\omega$ and $B_0 = Q^{-1}$ which in turn will also make B_0 lower triangular, with ones on the diagonal.
Both these techniques impose exactly the right number of restrictions, but can be considered as completely ad hoc, as both techniques depend on the ordering of the variables. This will leave the analysis with the same likelihood, and thereby the same fit. In order to make the restrictions less ad-hoc, we can tun to economic theory for guidance of the ordering, but because there is no way to test how well it fits the data, we need to be careful when ordering both the cholesky decomposition but also the diagonalization, and use economic theory carefully. Where because B_0 is lower triangular, the contemporaneous effect depends on the ordering of the varialbes in the SVAR model. The first variable is only affected by the first shock. The second variable is affected by the first two shocks and so on, hence recursively, and we can then use economic theory to order the variables.

Considering the long run restrictions, we know that the long term effect of a shock is the cumulated structural IRF at time $s = \infty$, which can be written as $\sum_{j=0}^{\infty} O_j = O(1)$ Where the long-run restrictions are imposed on this matrix in order to identify B_0, which is usually zero restriction. For example, if an impulse in equation j should have no long term effect on variable i, then $O(1)_{ij} = 0$.
From the relationship $O(1) = \phi(1)B_0^{-1}$, where we easily can see that restrictions on $O(1)$ are restrictions on B_0 because $\phi(1)$ is an unrestricted reduced-form parameter from the data generating process.
In order to estimate using long-run restrictions, we know that $\Sigma_u = B_0^{-1}(B_0^{-1})'$ which provides $K(K+1)/2$ restrictions. But we have also now defined that $B_0^{-1} = \phi(1)^{-1}O(1) = A(1)0(1)$ where if we have placed enough restrictions on $O(1)$, then it is possible to recover B_0 from the A_i and Σ_u using these two equations.

Finishing off, it is worth mentioning, that theoretically one can combine the long run and short run restrictions. Where the long run restriction instead of affecting the contemporaneous effects as in the short run restrictions, the long run restrictions affects the long run effect of two variables on each other.

3.5 Questions to ask each other

1. What is the problem regarding using wrong economic theory to create the SVAR?[1]
2. Can you write the LR test?

[1] "Garbage in, garbage out". We will get non-sensical results.

CHAPTER 4

4. Cointegration

Definition 20: Cointegration

Cointegration is a statistical concept that refers to a long-run relationship between two or more time series variables. A set of variables are said to be cointegrated if a linear combination of them is stationary, even though the individual variables may not be stationary themselves. Cointegration is important in econometrics and time series analysis because it allows for the estimation of long-run equilibrium relationships, and can be used to test for causality and predict future values of the cointegrated variables.

Definition: Two univariate time series x_{1t} and x_{2t} are cointegrated if they are jointly (both) I(1) and there exists a scalar a such that x_{1t} - ax_{2t} is I(0). The vector $\beta = (1, \ minus\ a)$ is the cointegration vector.

Defination for many series: A multivariate $p * 1$ time series x_t is cointegrated if it is I(1) and there exists a $p * r$ matrix β such that βXt is I(0). The columns of β are the cointegration vectors and r is the cointegration rank.

Cointegration Suppose we are interested in studying the relationship between the price of a stock and the price of the market index it belongs to. We collect daily data for the stock and the market index for a period of time. We plot the two time series and we see that although both time series are trending upwards, there are periods where the stock price is above the market index and other periods where it is below.

We then run a regression analysis where the stock price is the dependent variable and the market index is the independent variable. We also include a time trend in the regression. The regression output shows that the estimated coefficients for the market index and the time trend are not statistically different from zero, which suggests that the stock price and the market index are not stationary.

We can then run a cointegration test and it turns out that the two time series are cointegrated. This means that there is a long-term equilibrium relationship between the two time series, such that if the stock price deviates from this equilibrium, it will eventually return to it. In this case, the cointegration relationship between the stock price and the market index could be interpreted as "the stock price and the market index are related in the long run".

In this example, the cointegration relationship is identified by an I(1) and the estimated equilibrium relationship is the Error Correction Model(ECM) which is a powerful tool for analyzing the long-term dynamics of economic time series.

4.1 (i) Introduce/define cointegration in a regression context.

In a regression context, cointegration is a statistical property that occurs when two or more non-stationary time series have a long-term relationship with one another and are linearly dependent. This relationship is often referred to as a cointegrating relationship. A common method for detecting cointegration is the Johansen test,

which is used to determine the number and nature of cointegrating relationships among a set of variables. In a cointegrated system, the variables move together over time and any deviation from this long-term relationship will be corrected over time through the adjustment process. Cointegration is an important concept in econometrics and time series analysis, as it allows for meaningful long-term analysis and forecasting even when the individual variables are non-stationary.

The VAR(k) model can be written as:

$$\mathbf{x}t = \Pi_1 \mathbf{x}_{t-1} + \Pi_2 \mathbf{x}_{t-2} + \ldots + \Pi_k \mathbf{x}_{t-k} + \boldsymbol{\epsilon}_t$$

Where $\mathbf{x}_t$ is a p x 1 vector of variables and $\boldsymbol{\epsilon}_t$ is an iid error term with zero mean and covariance matrix Ω.

To obtain the VECM, we can subtract $\mathbf{x}_{t-1}$ from both sides, resulting in the following equation:

$$\Delta \mathbf{x}_t = \Pi \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t$$

Where $\Delta \mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$, $\Pi = -\Pi(1)$ $\Gamma_i = -\sum_{j=i+1}^{k} \Pi_j$

Note that for this equation to work, the error term should be white noise, meaning have a zero mean and covariance matrix that is a diagonal matrix and diagonal elements equal to variances (i.e., independent).

Step-by-step proof of how the VECM is derived from the VAR(k) model:

We start with the unrestricted VAR(k) model:

$$\mathbf{x}t = \Pi_1 \mathbf{x}t - 1 + \Pi_2 \mathbf{x}t - 2 + \ldots + \Pi_k \mathbf{x}t - k + \boldsymbol{\epsilon}_t$$

Where $\mathbf{x}_t$ is a p x 1 vector of variables, and $\boldsymbol{\epsilon}_t$ is an iid error term with zero mean and covariance matrix Ω.

To obtain the VECM, we subtract $\mathbf{x}_{t-1}$ from both sides of the equation:

$$\mathbf{x}t - \mathbf{x}t - 1 = \Pi_1(\mathbf{x}t - 1 - \mathbf{x}t - 2) + \Pi_2(\mathbf{x}t - 2 - \mathbf{x}t - 3) + \ldots + \Pi_k(\mathbf{x}t - k - \mathbf{x}t - k - 1) + \boldsymbol{\epsilon}t - \boldsymbol{\epsilon}t - 1$$

We can define the first difference of a vector as $\Delta \mathbf{x}t = \mathbf{x}t - \mathbf{x}t - 1$ so we can rewrite the previous equation as

$$\Delta \mathbf{x}t = \Pi_1 \Delta \mathbf{x}t - 1 + \Pi_2 \Delta \mathbf{x}t - 2 + \ldots + \Pi_k \Delta \mathbf{x}_{t-k} + \boldsymbol{\epsilon}t - \boldsymbol{\epsilon}t - 1$$

We note that $\boldsymbol{\epsilon}t - \boldsymbol{\epsilon}t - 1$ is not white noise (i.e., it is not independent and identically distributed) since it has a correlation to the previous error term. Thus, we assume that the error term is actually a white noise error term, and we get the final form of the VECM:

$$\Delta \mathbf{x}t = \Pi_1 \Delta \mathbf{x}t - 1 + \Pi_2 \Delta \mathbf{x}t - 2 + \ldots + \Pi_k \Delta \mathbf{x}t - k + \boldsymbol{\epsilon}_t$$

Where, $\Delta \mathbf{x}t = \mathbf{x}t - \mathbf{x}t - 1$, $\Pi = -\Pi(1)$ $\Gamma_i = -\sum j = i + 1^k \Pi_j$ To make the notation more compact, we can define the matrices

$$\Pi = [-\Pi_1, -\Pi_2, \ldots, -\Pi_k]$$

and

$$\Gamma_i = [-\Pi_{i+1}, -\Pi_{i+2}, \ldots, -\Pi_k]$$

so the VECM equation can be written as:

$$\Delta \mathbf{x}t = \Pi \mathbf{x}t - 1 + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t$$

1

4.2 (ii) Compare with spurious regressions.

Cointegration in a regression context is a statistical concept that refers to the existence of a long-term relationship between two or more variables. In other words, it is a relationship where two or more time series move together over time, but do not necessarily have a constant ratio. In this context, cointegration can be thought of as a long-term equilibrium relationship.

A simple example of cointegration in a regression context is the relationship between the prices of two goods, such as apples and oranges. Over the short term, the price of apples may be affected by factors such as weather or crop yields, leading to fluctuations in the ratio of apple to orange prices. However, over the long term, there may be a stable relationship between the two prices, such as an increase in the price of apples being accompanied by a similar increase in the price of oranges.

On the other hand, a spurious regression is a statistical relationship that appears to be real but is actually due to chance or random fluctuations in the data. A classic example of a spurious regression is the relationship between the number of storks nesting in a certain area and the number of babies born in that area. While it may seem that stork populations are related to birth rates, this relationship is actually due to the fact that storks are typically found in areas with large human populations, where birth rates are also high.

The difference between spurious and cointegration is that the latter is a long-term relationship, that is robust to changes in the environment, but the former is based on a chance correlation. This can be tested through a range of tests, like the Engle-Granger, Johansen or Phillips-Ouliaris test, to name some. Spurios regression

4.3 (iii) Discuss different definitions of cointegration.

Definition 21: 3 ways to write cointegration

Definition 1: Two univariate time series x_{1t} and x_{2t} are cointegrated if they are jointly (both) I(1) and there exists a scalar a such that x_{1t} - ax_{2t} is I(0). The vector $\beta = (1, \ minus\ a)$ is the cointegration vector.

Definition 2: A multivariate $p * 1$ time series xt is I(1) if

$$\Delta x_t = C(L)\epsilon_t$$

with $C(1) \neq 0_{n*n}$ and finite.

Definition 3: A multivariate $p * 1$ time series x_t is cointegrated if it is I(1) and there exists a $p * r$ matrix β such that $\beta' X_t$ is I(0). The columns of β are the cointegration vectors and r is the cointegration rank.

1

"Thus, if the non-stationarity of one variable corresponds to the non-stationarity of another variable, then there exists a linear combination between them that becomes stationary."

[1] where Π is a p x p matrix, and Γ_i is a p x $(k-i)p$ matrix. This equation is the final form of the Vector Error Correction Model (VECM).

It's worth noting that this equation expresses the relationship between the first differences of the variables and the past levels of the variables. The term $\Pi \mathbf{x}t - 1$ captures the short-run dynamics, while the term $\sum i = 1^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i}$ captures the long-run dynamics of the system.

In the VECM, the parameters of interest are the elements of the matrices Π and Γ_i, which can be estimated using standard econometric methods such as OLS. Once the parameters are estimated, one can use the model for forecasting, analyzing the long-run equilibrium relationship, and studying the speed at which the system returns to equilibrium after a shock.

Additionally, this derived equation provides a way to analyze the co-integration relationship among multiple non-stationary time series variables, that's why the VECM is widely used in practice.

2
"Another way of expressing this is that when two or several variables have common stochastic (and deterministic) trends, they will show a tendency to move together in the long-run." p. 80 Juselius.

4.4 (iv) Introduce the cointegrated VAR model and the VECM.

"The key difference between VECM and CVAR is that VECM is specifically designed to model the long-term relationship between cointegrated variables, while CVAR is a more general model that can be used to model the short-term relationship between any set of variables, whether they are cointegrated or not. In VECM, the long-term relationship is captured through the error correction term xt1, while in CVAR this relationship is captured through the matrix of parameters . Additionally, VECM imposes restrictions on and (such as = IrB) in order to identify the parameters, while CVAR does not impose such restrictions."

Definition 22: Cointegrated VAR (C-VAR)

$$x_t = \Pi x_{t-1} + \epsilon_t$$

Where,

$$\Pi = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

Definition 23: Vector Error correction model (VECM)

$$\Delta x_t = \alpha\beta' x_{t-1} + v\epsilon_t$$

Where this is the "error correction term",

$$\alpha = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and this is the cointegrated vector:

$$\beta = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

and

$$v = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

4.5 Questions to ask each other

1. Define the concept of cointegration in a regression context.
2. Explain how cointegration is related to the concept of stationarity in time series data.
3. Describe the different tests that can be used to determine if a set of variables are cointegrated.
4. Discuss the implications of cointegration for the interpretation and forecasting of time series data.
5. Describe how to estimate and interpret a cointegrated VAR model.
6. Explain how cointegration can be used in econometrics for causality inference and impulse response analysis.
7. Compare and contrast cointegration with other techniques for modeling long-run relationships, such as equilibrium correction models.
8. Discuss how cointegration can be used in finance, economics and other fields.
9. Give an example of a set of variables that are likely to be cointegrated and explain why
10. Describe how to estimate a VECM model with multiple cointegrating vectors and interpret the results.
11. Define and explain the concept of vector error correction term and its role in the VECM model
12. Why do we need cointegration?

13. What is the problem with differencing the series?
14. Define the concept of spurious regression and explain why it is important to test for cointegration when working with non-stationary time series data.
15. Describe the Engle-Granger and Johansen cointegration tests and explain when each one is appropriate to use.
16. Explain the concept of error correction term and its role in the VECM model.
17. Discuss the assumptions that underlie cointegration tests, and how these assumptions can affect the validity of the test results.
18. Compare and contrast the VECM model with other long-run modeling techniques, such as VAR models with differences and equilibrium correction models.
19. Provide an example of a set of variables that are likely to be cointegrated, and explain why.
20. Discuss the implications of having multiple cointegrating vectors in a VECM model and how this affects the interpretation of the model.
21. Describe the relationship between cointegration and causality inference and how these concepts can be used together in econometric analysis.
22. Explain how cointegration can be used in finance, economics, and other fields.
23. Describe how cointegration can be used in financial market research and how it can help to identify profitable trading strategies.
24. Explain the difference between short-run and long-run relationships in the context of cointegration.
25. Discuss the potential limitations of cointegration analysis and when it may not be appropriate to use.
26. Describe the concept of cointegration rank and explain how it relates to the number of cointegrating vectors in a VECM model.
27. Describe how the VECM model can be extended to include exogenous variables and how this affects the interpretation of the model.
28. Explain the concept of Granger causality and how it relates to cointegration analysis.
29. Describe how cointegration analysis can be used to test hypotheses about the long-run behavior of economic variables.
30. Discuss the relationship between cointegration and the concept of integration in time series analysis.
31. Provide examples of real-world applications of cointegration analysis in economics, finance or other fields.
32. Explain how cointegration is related to the concept of the long-run relationship between variables in econometrics.
33. Describe the difference between VECM and VAR and when it is appropriate to use each one.

CHAPTER 5

5. The cointegrated vector autoregressive model

Classical example of how to describe what cointegration is, is considering two random walk processes, which by nature are I(1), but there might be a relationship between the variables that allows for the combined process to I(0). Consider the data-generating process $x_t = (x_{1t}, x_{2t})'$ which is generated by $x_{1t} = x_{1,t-1} + \varepsilon_{1t}$ and $x_{2t} = x_{1t} + e_{2t}$ which are I(1) process due to their random walk nature. But there might exist the expression x

5.1 (i) Introduce the cointegrated VAR model and the VECM.

We can consider the unrestricted VAR(k):

$$x_t = \Pi_1 x_{t-1} + \Pi_2 x_{t-2} + ... + \Pi_k x_{t-k} + \varepsilon_t$$

Where epsilon is iid with mean zero and variance sigma. In this notation x_t is $p \times 1$. Where in this model we can allow for both stationarity and nonstationarity. If we subtract x_{t-1} from both sides and rearrange, we can obtain the vector error-correction model (VECM):

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-1} + \varepsilon_t$$

where $\Pi = -\Pi(1)$ and $\Gamma_i = -\sum_{j=i+1}^{k} \Pi_j$

if x_t is univariate, then Π is a scalar, and unit root in $\Pi(z)$ implies $\Pi = 0$.
if x_t is multivariate, then Π is a matrix. And there are many ways that zero and non-zero generalize to matrices. For instance if $\Pi = 0_{p \times p}$ The VECM is a stationary VAR in first difference, so x_t is $I(1)$ and not cointegrated. If Π is full rank, then the VECM is a stationary VAR in levels, so x_t is $I(0)$
If $rank(\Pi) = r < p$ then any reduced-rank matrix can be written as product of two full rank matrices, i.e. $\Pi = a\beta'$, where a and β are pxr and full rank. Another way to consider this, is if the original VAR is not nonstationarty, then $det(\Pi(z))$ has at least one unit root, meaning that $det((1)) = 0$ so that $\Pi(1) = -\Pi$ has reduced rank.
We can then insert $\Pi = a\beta'$ into the equation and obtain:

$$\Delta x_t = a\beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-1} + \varepsilon_t$$

Where it is important that the order of integration has to be the same for each term, for the equation to make sense. If we assume that x_t is I(1), then its clear that Δx_{t-1} is I(0) and ε is also I(0). And in order to balance the equation, we must have that $a\beta' x_{t-1}$ is I(0) where $\beta' x_{t-1}$ are the cointegration relations, when the cointegrated relation are zero, we are in equilibrium and if its not equal to zero, there exists a equilibrium error or the deviation from equilibrium. a is a loading/adjusment matrix, that determines the strength of adjusmtent or error correction. A important note regrading a and β are that they are not identified in the VECM, and we have to estimate the matrix Π.

Which for any invertible $r \times r$ matrix Q, it can be said that $\Pi = a\beta' = aQ^{-1}Q\beta' = \bar{a}\bar{\beta}'$
Where we cannot tell apart the pair (a, β) and the pair $(\bar{a}, \bar{\beta})$, unless we impose restrictions. The usual restriction to impose is that $\beta = \begin{pmatrix} I_r \\ B \end{pmatrix}$ where B is a freely varying $(p-r) \times r$ matrix, but there are many possibilities, where again economic theory again can act as guidance.

The cointegrated VAR model The simple VAR: Coming back to the example of the two random walks, which can be combined to a I(0) process, wher the datagenerating process can be described as $x_{2t} = x_{1t} + \varepsilon_{2t} = x_{1,t-1} + \varepsilon_{1t} + \varepsilon_{2t}$. Then the vector auto-regressive model is:

$$x_t = \Pi_1 x_{t-1} + v\varepsilon_t$$

With $\Pi_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ and $v = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ Where the characteristic polynomial Π has a unit root, following $det(\Pi(z)) = det(I_2 - \Pi_1 z) = 1 - z$

The cointegracted VAR model without reduced rank restriction on Π, can be estimated by OLS. However with reduced ran restriction on $\Pi = a'$, MLE is the same as reduced rank regression.

VECM Coming back to what we have previously described, where in order to obtain the VECM model, we have to subtract x_{t-1} from both sides, this is done of the data generating process:

$$\Delta x_{t1} = \varepsilon_{1t}$$

$$\Delta x_{2t} = x_{2t} - x_{2,t-1} = x_{1t} + \varepsilon_{2t} - x_{2,t-1} = x_{1,t-1} + \varepsilon_{1t} + \varepsilon_{2t} - x_{2,t-1}$$

which will leave os with the VECM model of the kind:

$$\Delta x_t = a\beta' x_{t-1} + v\varepsilon_t$$

Where $a = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\beta = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Definition 24: AI:

A Vector Error Correction Model (VECM) is a type of econometric model that is used to analyze the co-movements of multiple time series variables. It is a extension of the traditional vector autoregression (VAR) model, and it allows for the estimation of both short-term and long-term relationships between the variables. The VECM model includes a "error correction" term, which captures the deviation of the current equilibrium from the long-run equilibrium, and helps to correct for any disequilibrium in the system.

On the other hand, Cointegration Vector Autoregression (CVAR) model is a variation of VAR model which is used when two or more time series are cointegrated. Cointegration is a statistical property of time series that suggests that even though they may have different trends, they will tend to move together in the long run. The CVAR model is used to analyze the long-term relationships between these cointegrated variables.

In summary, VECM is used to analyze the short-term and long-term relationships between multiple time series variables, while CVAR is used to analyze the long-term relationships between cointegrated variables.

5.2 (ii) Discuss the Granger-Johansen theorem and different representations.

Definition 25: Definition of Granger-Johansen representation theorem

The Granger-Johansen representation theorem states that for ainy cointegrated system of n, number of equations, it is possible to represent it as an VECM model:

$$\Delta y_t = \Pi_1 y_{t-1} + ... + \Pi_p y_{t-p} + \varepsilon_t$$

Definition 26: Definition from linear algebra

For any $p \times r$ matrix A with $r < p$ there exists a $p \times (p - r)$ matrix $A_\perp$, which is the orthogonal complement of A, which is such that:

$$AA_\perp = A_\perp A = 0$$

$$rank([A, A_\perp]) = p$$

A important note is that $A_\perp$ is not unique, and we will need a a_{perp} and $\beta_\perp$.

Considering the Granger-Johansen representation theorem. Where we can define that $\Gamma = I_p - \sum_{i=1}^{k-1}$ and $C = \beta_\perp(a'_\perp\Gamma\beta_\perp)^{-1}a'_\perp$ then we let $\Pi(L)x_t = \varepsilon$ and we assume that $det(\Pi(z)) = 0$ which imples that the absolute value of z is greater than on or that z is simply equal to 1, then the $rank(\Pi) = r < p$ so that we again have the expression $\Pi = a\beta'$. Then we can conclude that if x_t is I(1) will be cointegrated if and only if:

$$det(a'_\perp\Gamma\beta_\perp) \neq 0$$

Furthermore it can be stated that under these conditions, x_t has the linear representation:

$$x_t = C\sum_{i=1}^{t}\varepsilon_i + y_t$$

where $y_t = C_1(L)\varepsilon_t$ is a stationary I(0) process.

Given the linear representation

$$x_t = C\sum_{i=1}^{t}\varepsilon_i + y_t$$

with $C = \beta_\perp(a'_\perp\Gamma\beta_\perp)^{-1}a'_\perp$, we can interpret that because $C \neq 0$ it hols that $x_t \in I(1)$ and because that $\beta'\beta_\perp = 0$, we know that $\beta'x_t = \beta'y_t$ which is I(0). Because that $a'_\perp$ is $(p-r) \times p$ we can see that there are only $p - r$ random walks in the system, given by $a'_\perp \sum_{i=1}^{t}\varepsilon$, which are called the common stochastic trends (CSTs). The counting rule states that the number of common stochastic trends are $p - r$, the number of cointegration vectors, rank are r and the dimension of the system are p.

In the linear representation, the cointegation relations are $\beta'x_t$, which are easily extracted from x_t. The common stochastic trends, sometimes have explicit interpretation and interest. If we assume all parameters are known or replaced with estimates, then because $a'_\perp\Gamma C = a'_\perp\Gamma\beta_\perp(a'_\perp\Gamma\beta_\perp)^{-1}a'_\perp = a'_\perp$ we find that $a'_\perp\gamma x_t = a'_\perp\sum_{i=1}^{t}\varepsilon_i + a'_\perp y_t$ which can be used as an estimate of the common stachastic trends.

Using the Granger-Johansen theorem, we can determine if there exists a cointegrated relationship, we can determine the common stochastic trends and we can determine the cointegration relations.

5.3 (iii) Explain cointegration rank and associated tests.

Definition 27: Cointegration Rank

Cointegration rank of a set of k I(1) time sereis $y_1, y_2, ..., y_k$ is the maximum number of linear combinations of these series that are stationary. A linear combination is a sum of each series multiplied by a coefficient. The cointegration rank is denoted as r, where $0 <= r <= k$. If $r = 1$, it means there is only one linear combination of the time series that is stationary, indicating that the timeseries are cointegratied. If however $r > 1$, there are mulitple linear combinations of the time series that are stationary, indicating that the time series are multi-cointegrated.

Definition 28: The two test proposed by Søren Johansen

he first test propsed by Søren Johansen is the trace test:

$$H_0 : r = r_0$$

$$H_1 : r = p$$

$$LR = 2(l(\hat{\theta}; r = p) - l(\hat{\theta}; r = r_0)$$

Which tests the null hypotheses that there are r cointegrating relationships among the variables, where r is a specified number, against the alternative that there are more than r cointegrating relationships. The second test proposed by Søren Johansen is the maximum eigen value test:

$$H_0 : r = r_0$$

$$H_1 : r = r_0 + 1$$

$$LR = 2(l(\hat{\theta}; r = r_0 + 1) - l(\hat{\theta}; r = r_0)$$

Which tests the null hypothesis that there are r cointegrating relationships against the alternative that there are at least r+1 cointegrating relationships.

Regarding the associated tests, we can consider the CVAR model, without deterministic terms:

$$\Delta x_t = a\beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t$$

where x_t is $p \times 1$ observable, ε_t is $p \times 1 iid(0, \varsigma)$ the error term. a and β are $p \times r$ parameters of adjustment coefficients and cointegration vectors. Γ_i are a $p \times p$ matrix of short-run parameters. For a given rank, $r = 0, 1, ..., p$ we can estimate the model by MLE (Maximum Likelihood Estimator). Where we can test hypotheses on the rank by LR tests.

Hypotheses on the cointragtion rank are generalizations of unit root tests, where $p - r$ indicates the number of unit roots in the system, so chaning r changes the number of unit roots. Just line the Dickey Fuller test changes the number of unit roots from one under H_0 to zero under H_1. There are two standard tests, where the most common is the Trace test:

$$H_0 : r = r_0$$

$$H_1 : r = p$$

$$LR = 2(l(\hat{\theta}; r = p) - l(\hat{\theta}; r = r_0)$$

Another applied test is the λ_{max} the maximum eigenvalue test:

$$H_0 : r = r_0$$

$$H_1 : r = r_0 + 1$$

$$LR = 2(l(\hat{\theta}; r = r_0 + 1) - l(\hat{\theta}; r = r_0)$$

Where we can in both cases test a particular rank directly. For any given hypothesized rank, r_0, for both tests it can be said that. Under H_0, asymptotic distributions are nonstandard distributions that are generalizations of Dickey Fuller distributions to multiple dimensions (multivariate Brownian motion functionals. Where the distributions depend on $p - r_0$ which is the number of unit roots under H_0. Deterministic terms included, we can obtain critical values from 5 different tables. If H_1 is true, the tests reject with probability converging to 1.

5.4 (iv) Discuss tests on long-run parameters (α and β) and identification.

Definition 29: Ways to estimate and test the long run parameters

There exists two types ways of estimation of the long run parameters α and β, the first was proposed by Engle and Granger, which is a two step approach. Which states that it is possible to run an OLS regression on the residuals, and then use the estimated residuals as the dependent varialbes and run an OLS regression again, to estimate the long run parameters.

The secon is proposed by Søren Johansen, which is a multivariate version of the Engle and Granger approach. This approach is using the Johansen Maximum Likelihood estimator, and allows us to test the number of cointegrated vectors and estimate the coefficient simultaneously.

We are again considering the CVAR model, without deterministic terms:

$$\Delta x_t = a\beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t$$

A important feature is that, if considering the cointegration rank and β is known, then everything is I(0), so inference is standard. The estimators are asymptotically normal and LR (or Wald) tests are asymptotically χ^2. It can be proven, that $\hat{\beta}$ is very precisely estimated. In fact, $T(\hat{\beta} - \beta_0) \xrightarrow{d} ...$ indicating that they are subject to normalization. The asymptotic distribution is not normal, bu mixed normal. LR tests stil χ^2. Because of consistency, we can consider β as known, when analyzing properties of estimators of remaining parameters.

There are two kinds of hypotheses, one that can change the cointegration rank and one that does not change the cointegration rank.
Considering the first case, that holds hypotheses that change the cointegration rank lead to LR/WALD/t tests that have nonstandard distributions, which could be Trace and λ_{max} tests. Somestimes a hypotheses on a or β can make the rank different under the null. Which should be avoided.
In the second case, we are considering hypotheses that do not change the cointegration rank which will lead to LR/Wald/t tests have standard χ^2 or normal distributions, once the rank has been determined. We can then test on Γ_i to determine the lag order. And run tests on the deterministic terms, a and β.

Regarding Γ_i, estimators are asymptotically normal and LR/Wald tests are asymmptotically χ^2 with the expected d.f. (difference in number) Which usually consists of exlusion restrictuion, so estimation under the null is simple.

Regarding the deterministic terms, the distributions can be nonstandard, but LR tests are asymptotically χ^2. Testing that a deterministic term is restricted to the cointegration space , i.e. proportional to a, has $p - r$ d.f. Testing wether a deterministic term is completely absent against it being unrestricted has p d.f. Again we imply exclusion restrictions, so estimation under the null can become simple.

Regarding the hypotheses on a and β, the hypotheses on β arise from economic theory, the long-run equilibria.
The hyoptheses on a, when $a_{ij} = 0$ then Δx_{ij} does not respond to $\beta_j' x_{t-1}$ which is the j'th cointegration relation, the long-run exogeneity. Which variables adjust and which do not adjust (pushing and pulling forces). This leaves two difficult parts, what is the d.f? and how to estimate the model?.
The difference in number (d.f.) of free parameters under H_0 and H_1. If we consider Π being unrestricted there are p^2 free parameters. Because that β has full rank, we can write $\beta' = (\beta_1', \beta_2')$ where β is $r \times r$ and full rank, then we can express Π:

$$\Pi = a\beta' = a(\beta_1', \beta_2') = a\beta_1'(\beta_1')^{-1}(\beta_1', \beta_2') = a\beta_1'(I_r, (\beta_1')^{-1}\beta_2') = \bar{a}\bar{\beta}'$$

Where $\bar{a} = a\beta_1'$ and $\bar{\beta}' = (I_r, (\beta_1')^{-1}\beta_2')$, which are both $p \times r$, the free parameters in $\bar{a}$ are pr and the free parameters in $\bar{\beta}$ are $r(p-r)$. And the free parameters under H_1 with the unrestricted a, β with rank r, are $r(2p-r)$.

When focusing on the liear restrions on all a, would leave us with:

$$H_0 : a = A\psi$$

In this equation A is known, and is a $p \times m$ matrix, and ψ is estimated, and is a $m \times r$ matrix. Under this H_0 we find:

$$\Pi = a\beta' = A\psi\beta' = A\psi\beta'(I_r, (\beta_1')^{-1}\beta_2') = A\bar{\psi}\bar{\beta}'$$

Where the free parameters in $\bar{\psi}$ are mr. And the free parameters in $\bar{\beta}$ are $r(p-r)$.
The test d.f. is $r(2p-r) - mr - r(p-r) = r(p-m)$.

Hypotheses on a Hopthesis on a:

Let $X_t = (X_{1t}', X_{2t}')'$ partitoned as $p = p_1 + p_2$ variables. Then the long run exogeneity of X_{1t} is $a = \begin{pmatrix} O_{p1\times r} \\ \psi \end{pmatrix} = \begin{pmatrix} O_{p_1\times r} \\ I_{p_2} \end{pmatrix} \psi$ Where ψ is $p_2 \times r$
Where the test d.f. is $r(p-m) = r(p-p_2) = rp_1$ which holds the same number of restrictions as in a.

Hypoteses on β Hypotheses on β, where we impose linear restriction on all cointegration vectors, all collumns of β. Which leaves us with the $H_O : \beta = H\phi$ where H is known, and is a $p \times s$ matrix and ϕ is estimated and is $s \times r$. In this example we let $p = 3$ and $r = 2$, and we want to test that $\beta_j = (\phi_{1j}, -\phi_{1j}, \varphi_{2j})'$ Then we obtain the expression:

$$H = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$$

Where under H_0 we have that $\Pi = a\beta' = aH'\phi' = \bar{a}\bar{\phi}'$, where $\bar{\phi} = (I_r, (\phi_1')^{-1}\phi_2')H'$ so that Π has $pr + r(s-r) = r(p+s-r)$ free parameters. The LR test d.f. is $r(2p-r) - r(p+s-r) = r(p-s)$
Estimetion under $H_0 : a\beta' x_{t-1} = a\phi' H' x_{t-1}$ so that instead of x_{t-1} we use $H'x_{t-1}$ as a regressor.

If we instead assume that some β_j's are known then we can write $H_0 : \beta = (b\phi)$ with b as known $p \times s$ and ϕ estimated $p \times (r-s)$ then we obtain the expression

$$\beta = (b, \phi) = \begin{pmatrix} 1 & \phi_{11} \\ -1 & \phi_{21} \\ 0 & \phi_{31} \end{pmatrix}$$

WHere Π ha $ps + p(r-s) + (r-s)(p-r)$ free parameters.
LR test d.f is $r(2p-r) - ps - p(r-2) - (r-s)(p-r) = s(p-r)$
Estimation under $H_0 : a\beta' x_{t-1} = a_1\psi' b_\perp' x_{t-1}$ so that instead of x_{t-1} we use $b'x_{t-1}$ with full rank coefficient and $b_\perp' x_{t-1}$ with reduced rank r-s coeffients as regressors.

5.5 Questions to ask each other

1. Derivation of VAR to VECM.
2. Define alpha and beta in VECM
3. Define a VAR model and compare with CVAR.
4. discuss the points about the deterministic terms.
5. Discuss unit root testing for CVAR.
6. When hypothesis testing, how de we determine $\bar{a}$ and $\bar{\beta}$, and what are the free parameters for them and H_1
7. What is the difference between VECM and C-SVAR? (Answer: VECM is a model that is used to analyze the short-term and long-term relationships between multiple variables in an economic system. It is an extension of the Vector Autoregression (VAR) model that incorporates the concept of cointegration. Cointegration is a statistical property that exists when a linear combination of non-stationary variables is stationary. The VECM model uses this property to capture the long-term relationships between variables and correct for any short-term deviations from these relationships. On the other hand, C-SVAR is a model that is used to analyze the dynamic relationships between multiple variables in an economic system. It is a constant-parameter version of the Stochastic Vector Autoregression (SVAR) model. The main difference

between VECM and C-SVAR is that the latter assumes that the parameters of the model are constant over time, while the former allows for the parameters to change over time. Additionally, C-SVAR allows for the estimation of the contemporaneous impact effects among variables, while VECM focuses on long-term relationships and short-term deviations.)

CHAPTER 6

6. Introduction to realized measures

Realized measures are statistical measures of the volatility and correlation of financial markets that are calculated using high-frequency financial data. These measures were first introduced in the early 2000s as an alternative to traditional financial indicators, such as the closing price or daily return.

Realized measures are calculated using high-frequency financial data, such as tick-by-tick or high-frequency intraday prices, or intraday trades. These measures capture the full range of volatility, correlation, and other important market characteristics that are not captured by traditional indicators that use lower-frequency data, such as daily or monthly prices.

There are several different types of realized measures that are commonly used in finance and economics, including realized volatility, realized correlation, and realized kurtosis. Realized volatility measures the volatility of a financial market over a given period of time and is commonly used to estimate the volatility of a market or asset. Realized correlation measures the correlation between different markets or assets, and is commonly used to measure the risk of a portfolio of assets. Realized kurtosis is a measure of the tail-risk of the distribution of returns and is commonly used in risk management.

Realized measures have several advantages over traditional financial indicators. For instance, they are more accurate, they can be calculated using much higher frequency data, and they are more robust to the presence of outliers. They are also more efficient in capturing volatility, correlation and tail-risk than traditional indicators.

On the other hand, there are also some limitations when working with realized measures. For example, the precision and resolution of the data used to estimate the realized measures affects their accuracy. Also, the calculation of realized measures requires a large amount of data and computational resources.

Definition 30: Realized measures

Realized measures refers to statistical measures of the volatility and correlation of financial markets that are calculated using high-frequency financial data. These measures are commonly used as an alternative to traditional financial indicators, such as the closing price or daily return, in order to obtain a more accurate and robust estimate of the volatility and correlation of financial markets.

Realized measures Imagine you have a stock and you want to estimate its volatility. Traditionally, you would calculate the volatility using the closing price of the stock over a given period of time. For example, you might use the daily closing prices of the stock to estimate the volatility over a month or a year. However, this method of volatility estimation has some limitations. For instance, it only captures the volatility during the trading hours and it only provides an estimate of the volatility over the time period you have chosen, therefore it might not be very accurate for some stocks.

On the other hand, if you have access to high-frequency data, such as tick-by-tick prices or intraday trades, you can use realized measures to estimate the volatility of the stock more accurately. Realized volatility is calculated by summing the intraday squared returns over a given period of time. This method allows you to capture the volatility of the stock during all trading hours, providing a more accurate estimate of the volatility, that is less

affected by the chosen time period or the trading hours.

Keep in mind this is just one example, realized measures can also be used to estimate other financial characteristics such as correlation, kurtosis, among others.

6.1 (i) Write/explain (mathematical forms of) realized volatility, bipower and power variation.

Definition 31: Realized volatility from lecture notes

$$RV_n = \sum_{i=1}^{n} (X_{\frac{1}{n}} - X_{\frac{i-1}{n}})^2$$

Realized volatility is a commonly used realized measure that estimates the volatility of a financial market or asset. It is calculated using high-frequency financial data, such as tick-by-tick or high-frequency intraday prices, or intraday trades.

Definition 32: Bipower variation

The definition for bipower variation (B-PV)

$$BV_n = \frac{\pi}{2} \sum_{i=1}^{n-1} |(X_{\frac{i}{n}} - X_{\frac{i-1}{n}})||X_{\frac{i+1}{n}} - X_{\frac{i}{n}}|$$

Where BV_n is the bipower variation.

Bipower variation, also known as integrated volatility, is a realized measure that estimate the volatility and jump in a market or an asset. It is calculated using high-frequency data and it is a similar concept to realized volatility but it also captures the jump or discontinuities in the volatility of the process.

Definition 33: The power variation ($RV_n(p)$)

The definition for Power variation:

$$RV_n(p) = \frac{1}{n} \sum_{i=1}^{n} |\sqrt{n}(X_{\frac{i}{n}} - X_{\frac{i-1}{n}})|^p$$

6.2 (ii) Write the statement of the consistency and CLT for realized volatility and power variation.

Lets start by looking at the model:

$$X_t = X_0 + \int_0^t \mu_s d_s + \int_0^t \sigma_s dW_s$$

We fist start by looking at the Realized volatility:

Definition 34: Realized volatility

Statement of the consistency The statement of consistency for realized volatility states that as the sampling frequency increases, the realized volatility converges to the integrated volatility (IV), which is the integral of the true volatility over the period of observation.

$$RV_n - - >^p IV$$

CLT The CLT is

$$\sqrt{n}(RV_n - IV_n) - >^d_{st} MN(0, 2 * IQ)$$

Here we have that $IV = \int_0^1 \sigma_t^2 dt$ and $IQ = \int_0^1 \sigma_t^4 dt$.

Definition 35: Power variation

Statement of the consistency

$RV_n(p) - >^p m_p \int_0^1 |\sigma_t|^p dt$ where $m_p = E[|N(0,1)|^p]$.[a]

CLT

This is much more complex. It requires some assumption. However, a CLT-type result can still be obtained through the use of martingale central limit theorems and weak convergence methods. In this case, the CLT for power variation would involve the convergence of the sample mean of the $\sqrt{n}(X_{i/n} - X_{(i-1)/n})$ terms to the population mean of the integral of $|\sigma_t|^p$ from 0 to 1 in distribution, as the sample size increases.

The CLT for power variation would also involve the assumption that the terms $X_{i/n} - X_{(i-1)/n}$ are independent and identically distributed with mean zero and variance of order $1/n$. Additionally, the CLT for power variation would require that the integral of $|\sigma_t|^p$ from 0 to 1 is finite and that the power parameter p is greater than 0.

$$\sqrt{n}(RV_n - IV_n) - >^d N(0, 2 * IQ)$$

[a] This is from slide 17 lecture 2.

6.3 (iii) Sketch of proofs of consistency/CLT when volatility is constant.

The text states that the realized volatility measure is calculated using high-frequency financial data, such as tick-by-tick or high-frequency intraday prices, or intraday trades. One commonly used measure of realized volatility is the sum of squared returns over a certain period of time:

$$RV_n = \sum_{i=1}^{n} (X_{\frac{1}{n}} - X_{\frac{i-1}{n}})^2$$

where X is the value of the financial asset at time t, and n is the number of returns in the period. The text states that the volatility of the underlying process is assumed to be constant. It is pointed out that the term

$$X_{\frac{1}{n}} - X_{\frac{i-1}{n}}$$

has order $1/\sqrt{n}$, since it can be written as:

$$X_{\frac{1}{n}} - X_{\frac{i-1}{n}} = \mu \frac{t}{n} + \sigma \frac{W_t}{\sqrt{n}}$$

The first term, $\mu\frac{t}{n}$, has order $1/n$, which means it converges to 0 as n increases. The second term, $\sigma\frac{W_t}{\sqrt{n}}$, has order $1/\sqrt{n}$ as it has mean 0, variance with order $1/n$. So, the term $X_{\frac{1}{n}} - X_{\frac{i-1}{n}}$ has order $1/\sqrt{n}$ and

$$\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}})$$

has order 1. Therefore, the term

$$\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}})$$

is multiplied with $\sqrt{n}$ to make the terms in the sum have the same order of magnitude. It is also pointed out that if the terms are iid with expectation

$$E(\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}})) = m_p\sigma_p$$

where mp is the expected value of the absolute value of a standard normal distribution raised to the power p and σ_t is the volatility of the underlying process. Based on these arguments, for any $p > 0$, the following law of large numbers holds:

$$RV_n(p) = \frac{1}{n}\sum_{i=1}^{n}|\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}})|^p \rightarrow mpZ_1^0|\sigma_t|^p dt$$

as n increases.

Definition 36: Proof of consistency when volatility is constant

Consistency

When the volatility of the underlying process is assumed to be constant, the derivation of consistency of the realized volatility measure can be done as follows:
Start with the definition of the realized volatility measure:

$$RV_n = \sum_{i=1}^{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}})^2$$

where X is the value of the financial asset at time t, and n is the number of returns in the period. Assume that the volatility of the underlying process is constant, denoted by σ^2. Next, decompose the return into its mean and volatility components:

$$X_{\frac{1}{n}} - X_{\frac{i-1}{n}} = \mu\frac{t}{n} + \sigma\frac{W_t}{\sqrt{n}}$$

where μ is the mean return, and W_t is a standard Brownian motion.
Now, we can use the strong law of large numbers to prove consistency. The strong law of large numbers states that for any sequence of independent and identically distributed random variables, the sample mean converges almost surely to the population mean as the sample size increases.
Since the terms $\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}})$ are independent and identically distributed, we can apply the strong law of large numbers to the sum

$$\frac{1}{n}\sum_{i=1}^{n}|\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}})|^p$$

and conclude that it converges almost surely to the population mean

$$E(|\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}})|^p) = m_p\sigma_p^p$$

where m_p is the expected value of the absolute value of a standard normal distribution raised to the power p and σ_t is the constant volatility of the underlying process. Therefore, we can conclude that the realized volatility measure, $RV_n(p)$, is consistent as the sample mean of the squared returns converges almost surely to the population mean of the squared returns, which is equal to the constant volatility squared.

Definition 37: Proof of CLT when volatility is constant

CLT

When the volatility of the underlying process is assumed to be constant, the derivation of the Central Limit Theorem (CLT) for the realized volatility measure can be done as follows:

Start with the definition of the realized volatility measure:

$$RV_n = \sum_{i=1}^{n} (X_{\frac{1}{n}} - X_{\frac{i-1}{n}})^2$$

where X is the value of the financial asset at time t, and n is the number of returns in the period. The equation $X_{\frac{1}{n}} - X_{\frac{i-1}{n}} = \mu\frac{t}{n} + \sigma\frac{W_t}{\sqrt{n}}$ is a decomposition of the return into its mean and volatility components. The term $\mu\frac{t}{n}$ represents the mean return, which is the expected value of the return over the period. The term $\sigma\frac{W_t}{\sqrt{n}}$ represents the volatility component, which is the deviation of the return from the mean.

The term σ represents the constant volatility of the underlying process, which is assumed to be constant. W_t is a standard Brownian motion, which is a mathematical model for the random movement of a financial asset.

The decomposition of the return into its mean and volatility components is useful because it allows us to separate the effects of the mean return and the volatility on the realized volatility measure.

The decomposition is obtained by assuming that the return process is a geometric Brownian motion, which is a widely used model in finance. The equation $X_{\frac{1}{n}} - X_{\frac{i-1}{n}} = \mu\frac{t}{n} + \sigma\frac{W_t}{\sqrt{n}}$ is derived from the geometric Brownian motion equation $dX_t = \mu X_t dt + \sigma X_t dW_t$ by using Ito's Lemma which states that $d(X_t^2) = (2\mu X_t + \sigma^2)dt + 2\sigma X_t dW_t$ and then by substituting the return $X_{\frac{1}{n}} - X_{\frac{i-1}{n}}$ in place of dX_t.

Assume that the volatility of the underlying process is constant, denoted by σ^2. Next, decompose the return into its mean and volatility components:

$$X_{\frac{1}{n}} - X_{\frac{i-1}{n}} = \mu\frac{t}{n} + \sigma\frac{W_t}{\sqrt{n}}$$

where μ is the mean return, and W_t is a standard Brownian motion.

Now, we can use the CLT to prove the asymptotic normality of the realized volatility measure. The CLT states that for a sequence of independent and identically distributed random variables, the sum of the variables divided by the square root of the sample size converges in distribution to a normal distribution as the sample size increases. Since the terms $\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}})$ are independent and identically distributed, we can apply the CLT to the sum

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}}))$$

and conclude that it converges in distribution to a standard normal random variable. To be more precise,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}})) \to N(0, \sigma^2)$$

as n increases, where σ is the constant volatility of the underlying process. Therefore, we can conclude that the realized volatility measure, RV_n, satisfies the CLT as the sum of the squared returns divided by $\sqrt{n}$ converges in distribution to a standard normal random variable. This implies that the distribution of the realized volatility measure becomes more and more normal as the sample size increases. To complete the proof of the CLT, we can use the Lyapunov Central Limit Theorem, which states that if a sequence of independent and identically distributed random variables, X_i, satisfy the following conditions: $\sum_{i=1}^{\infty} Var(X_i) < \infty$ $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E(|X_i - E(X_i)|^2) = 0$ Then, the sum $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ converges in distribution to a normal distribution with mean 0 and variance σ^2.

In this case, since the volatility of the underlying process is assumed to be constant, we have $\sigma^2 = Var(X_i)$. Also, since the terms $\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}})$ are independent and identically distributed, we can apply the Lyapunov CLT to the sum

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\sqrt{n}(X_{\frac{1}{n}} - X_{\frac{i-1}{n}}))^2$$

and conclude that it converges in distribution to a normal distribution with mean 0 and variance σ^2. Therefore, we can conclude that the realized volatility measure, RV_n, is asymptotically normal as the sample mean of the squared returns converges in distribution to a normal distribution with mean equal to the population mean of the squared returns, which is equal to the constant volatility squared.

Derivation 6: Constant volatility

When the volatility of the underlying process is assumed to be constant, the derivation of the realized volatility measure can be done as follows:
Start with the definition of the realized volatility measure:

$$RV_n = \sum_{i=1}^{n} (X_{\frac{1}{n}} - X_{\frac{i-1}{n}})^2$$

where X is the value of the financial asset at time t, and n is the number of returns in the period. Assume that the volatility of the underlying process is constant, denoted by σ^2. Next, decompose the return into its mean and volatility components:

$$X_{\frac{1}{n}} - X_{\frac{i-1}{n}} = \mu \frac{t}{n} + \sigma \frac{W_t}{\sqrt{n}}$$

where μ is the mean return, and W_t is a standard Brownian motion. Substitute this decomposition into the realized volatility measure:

$$RV_n = \sum_{i=1}^{n} (\mu \frac{t}{n} + \sigma \frac{W_t}{\sqrt{n}})^2$$

Since μ and σ are constant and the W_t are independent and identically distributed, we can simplify the sum to:

$$RV_n = \frac{t^2\mu^2}{n^2} + \frac{2\mu t}{n}\sigma\frac{W_t}{\sqrt{n}} + \frac{\sigma^2}{n}\sum_{i=1}^{n}\frac{W_t^2}{n}$$

The first term $\frac{t^2\mu^2}{n^2}$ converges to 0 as n increases, and the second term $\frac{2\mu t}{n}\sigma\frac{W_t}{\sqrt{n}}$ also converges to 0 as n increases since W_t has a mean of 0. The third term $\frac{\sigma^2}{n}\sum_{i=1}^{n}\frac{W_t^2}{n}$ is the sum of the squares of the standard Brownian motion, which is a chi-squared random variable with n degrees of freedom. As a result, the realized volatility measure is:

$$RV_n = \frac{\sigma^2}{n}\sum_{i=1}^{n}\frac{W_t^2}{n} = \frac{\sigma^2}{n}$$

6.4 Questions to ask each other

1. Define realized measures and explain why they are useful in finance and economics.
2. Describe how realized measures can be used to estimate volatility, correlation, and other important market characteristics.
3. Explain the difference between realized measures and traditional financial indicators, such as the closing price or daily return.
4. Discuss the advantages and limitations of using realized measures in financial research.
5. Describe the different types of realized measures, such as realized volatility, realized correlation, and realized kurtosis.
6. Explain the concept of high-frequency data and how it relates to the estimation of realized measures.
7. Describe the methods used to calculate realized measures and discuss their properties, such as bias and efficiency.
8. Explain how realized measures can be used to improve the performance of financial models, such as value at risk, portfolio optimization, and risk management.
9. Discuss how realized measures can be used to improve the forecasting accuracy of financial time series models.
10. Provide examples of real-world applications of realized measures in finance, economics, or other fields.
11. What is TAQ?

12. What is High freq data?
13. Write a plot of how High Freq data looks like?
14. How do you clean TAQ data?
15. What is the problem working with quote data?
16. Should we work with Trade or Quote data?
17. What is semimartingales?
18. What is the defination of "no abritrage"?
19. What is The theorem of Delbaen and Schachermayer?
20. What is a brownian motion?
21. What is Logarithmic price?
22. What kind of estimator are there for HF data?
23. What issues exist for implementing realized volatility? [1]
24. Explain the concept of integrated variance and why it is important in the estimation of realized measures.
25. Describe the relationship between realized measures and Ito processes
26. Compare and contrast the different sampling methods that can be used to estimate realized measures (e.g. equidistant, tick-by-tick)
27. Explain how microstructure noise affects the estimation of realized measures, and how this noise can be modeled and accounted for.
28. Discuss the impact of different assumptions (such as independence and identical distributions) on the estimation of realized measures.
29. How does the choice of sampling frequency affect the estimation of realized measures and the conclusions that can be drawn from them.
30. Describe the different types of market microstructure noise and how they affect the estimation of realized measures.
31. Compare and contrast the different types of noise robust estimators available and their properties.
32. Describe the methods used to estimate the asymptotic variance of realized measures
33. Discuss the potential applications of realized measures in portfolio management, risk management, and algorithmic trading.
34. Explain the concept of microstructure noise and how it affects the estimation of realized measures.
35. Describe how sample frequency affects the estimation of realized measures.
36. Discuss the implications of using realized measures in high-frequency trading algorithms.
37. How is the concept of realized measures related to the concept of integrated volatility?
38. How do realized measures compare to traditional volatility estimators such as GARCH or Exponentially weighted moving average (EWMA)?
39. Explain the importance of considering the distribution of returns when working with realized measures.
40. Explain the relationship between realized measures and the concept of market liquidity

[1] Implementing realized volatility with empirical data can be challenging due to the presence of market microstructure noise. Financial data is often available at a high frequency (e.g. every 5 seconds or even milliseconds), but using all of this data in the calculation of realized volatility can lead to divergence due to the presence of noise such as price discreteness and bid-ask spreads. One solution to this problem is to use sparse sampling, where only a subset of the data is used in the calculation of realized volatility. Empirical studies in the early 2000s suggested that using data every 5 minutes is a popular choice for realized volatility calculations.
Another solution is to use market microstructure noise robust estimators, which take into account the presence of noise explicitly in the model. These estimators can be used with all data and are the subject of further study in future lectures.

41. How does the estimation of realized measures differ in intraday and daily returns?
42. Explain the concept of volatility clustering and how it affects the estimation of realized volatility.
43. Explain how realized measures can be used to detect changes in volatility patterns in financial time series.
44. Explain the connection between realized measures and financial risk management
45. what is the difference between Realized Variance and Realized Volatility?
46. Explain the concept of integrated volatility in the context of financial econometrics.
47. How can we estimate the realized volatility and how can we correct for noise in the estimation?
48. Explain the concept of realized covariation and its role in volatility estimation.
49. How does the estimation of realized volatility differ between different asset classes?
50. Explain the derivation of the realized volatility estimator and how it relates to the integrated volatility of the underlying process.
51. Describe how the delta method can be used to prove consistency of realized volatility estimators.
52. How the law of large numbers is used to prove consistency of realized volatility estimators when volatility is constant.
53. Discuss the technical conditions and assumptions needed to apply the Lyapunov's central limit theorem to prove the central limit theorem for realized volatility.
54. How the Ito's lemma is used to show that the log-price process is a semimartingale when the price process is a semimartingale.
55. Derive the asymptotic variance of the realized volatility estimator and explain how it is used to determine the statistical accuracy of the estimator.
56. How Ito's formula is used to derive the realized variance estimator and explain how it relates to the integrated variance of the underlying process.
57. Discuss how the increments of log-price process essentially represent actual returns, and how it is used in deriving realized volatility and other realized measures.
58. Explain the derivation of noise robust estimators for realized volatility and how it deals with microstructure noise in high-frequency data.
59. Can you explain the notation used in the equation for the realized variance estimator, RVn?
60. Can you provide an explanation of the conditions on $(\mu_t)0 \leq t \leq 1$ and $(\sigma_t)0 \leq t \leq 1$ that are required for the theorem to hold?
61. What is the limiting distribution of RV_n as n approaches infinity, according to the theorem?
62. How is the IQ defined in the theorem and what does it represent?
63. Can you explain the stable convergence and mixed normal mentioned in the Remarks section?
64. What is the difference between $\mathbb{P} \rightarrow$ and $d \rightarrow$ in the context of the theorem?
65. Can you explain what it means for a sequence of random variables $(Z_n)_{n=1}^{\infty}$ to converge in probability to a random variable Z?
66. How is the concept of convergence in probability related to the concept of convergence in distribution?
67. Can you provide an example of a sequence of random variables that converges in probability to a limit random variable?
68. How is the convergence in distribution defined?
69. How does cumulative distribution function(CDF) for a random variable plays a role in the definition of convergence in distribution?

70. Can you provide an example of a sequence of random variables that converges in distribution to a limit random variable?
71. How does the pointwise convergence of CDF at continuous point of limit variable related to convergence in distribution?
72. Can you discuss the impact of market microstructure and liquidity on the estimation of realized measures?
73. How does the estimation of realized measures change when applied to different asset classes such as stocks, bonds, and derivatives?
74. Can you explain the concept of realized multifractality and its applications in financial analysis?
75. How does the presence of jumps and other non-continuous events in the price process impact the estimation of realized measures?
76. Can you explain how realized measures can be used to identify and quantify structural breaks in financial time series?
77. How do realized measures compare with other volatility estimation techniques such as GARCH or Implied Volatility?
78. Can you describe any recent developments or ongoing research in the field of realized measures?
79. How do you assess the robustness and statistical significance of the results obtained using realized measures?
80. Can you discuss the impact of data frequency, data quality and data availability on the estimation of realized measures?
81. How can realized measures be used to model intraday patterns or overnight return differences?
82. What can cause outliers in HF data?[2]
83. What is the bid and ask spread?[3]
84. What is the vaiance of a brownian motion? [4]
85. What is a "semi marginal"?
86. Why is there problems with sampling often?[5]
87. How can we solve the problems caused by using all the data? [6]
88. How would you differentiate a brownian motion?[7]
89. How would you write a model for the price with jumps?[8]
90. How would you write RV with jumps?[9]
91. How would you write BV with jumps?[10]

[2] Could be error (personal), mistake (human).
[3] It's the difference between the bid and ask price the actor goes to the market with.
[4] ∞
[5] It comes from extra variance with high freq in data.
[6] We can use data for each 5 minutes.
[7] Brownian motions can't be differentiated.
[8]

$$x_t = \underbrace{x_0}_{Startingprice} + \overbrace{\int_0^t u_s d_s}^{drift} + \underbrace{\int_0^t \sigma_t dW_s}_{volatility} + \underbrace{\sum_{0\leq s\leq t} J_s}_{Jump}$$

[9]

$$RVp- > \int_0^1 \sigma_s^2 d_s + {}_i^2$$

[10]

$$RVp- > \int_0^1 \sigma_s^2 d_s$$

92. What is the advantage of using pre-averaging?[11]

Theorem The theorem of Delbaen and Schachermayer

The theorem of Delbaen and Schachermayer, states that in a frictionless financial market, if the financial asset X follows a continuous-time model and if the market is free of arbitrage opportunities, then the asset X must be a semimartingale. A semimartingale is a mathematical concept that generalizes the notion of a martingale, which is a stochastic process that has the property that the expectation of the next value of the process is equal to the current value. The theorem emphasizes that the financial asset X must follow a specific continuous-time model and the market is without any type of arbitrage opportunities such as "free lunch with vanishing risk" for the theorem to hold.

It is also worth noting that the theorem is based on a filtered probability space $(\Omega, F, (F_t)t \leq 0, P)$ which is the background system that is assumed to exist when modeling financial markets. This filtered probability space allows for the modeling of stochastic processes in continuous time, and it's important to have a clear understanding of it when working with theorems like the one presented here.

In addition, the theorem point out that no-friction is an important assumption and that the presence of transaction costs can invalidate the theorem as some non-semimartingales, such as fractional Brownian motion, may still follow the no-arbitrage assumption.

Theorem Ito process

An Ito process is a type of stochastic process that is named after the Japanese mathematician Kiyosi Ito. It serves as a primary example of a semimartingale, which is a mathematical concept that generalizes the notion of a martingale. The Ito process is defined as:

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$

Where: μ_t is the drift term, which can be a deterministic or a stochastic function of time. σ_t is the volatility term, which can be a deterministic or a stochastic function of time. W_t is a standard Brownian motion. The first integral (Riemann integral) $\int_0^t \mu_s ds$ is a finite variation process, meaning that it is a type of process that has a finite variation over a given interval of time. The second integral (stochastic integral) $\int_0^t \sigma_s dW_s$ is a martingale, meaning that it has the property that the expectation of the next value of the process is equal to the current value.

Also the constant model can be written as :

$X_t = X_0 + \mu t + \sigma W_t$

Where μ and σ are constant values. The second integral (stochastic integral) $\int_0^t \sigma_s dW_s$ is a martingale, meaning that it has the property that the expectation of the next value of the process is equal to the current value.

This version of the Ito process is continuous, meaning that it does not contain jumps. It is worth noting that when $\mu_s = \mu$, $\sigma_s = \sigma$, the process becomes $X_t = X_0 + \mu_t + \sigma W t$, this is known as the constant model where the drift and volatility is constant over time.

Ito process are widely used in mathematical finance, specifically in the modeling of financial markets and for pricing financial derivatives. Due to its tractable nature, it is a popular choice for practitioners working in this field.

[11] Takes care of high degree of noise.

Theorem Brownian motion

A standard Brownian motion is a continuous-time version of a (not necessarily Gaussian) discrete-time random walk. Formally, it is represented by a stochastic process $(W_t)_{t\geq 0}$, where:

It starts at 0, i.e. $W_0 = 0$. For each t, W_t is a Gaussian random variable with mean 0 and variance t, i.e. $W_t \sim N(0,t)$. The increment $W_t - W_s$ is distributed as $N(0, t-s)$ for each $s < t$. If $s < t \leq u < v$, then the increments $W_t - W_s$ and $W_v - W_u$ are independent. Paths of Brownian motion are continuous, i.e. $t \to W_t$ is continuous. Fun Fact: Although Brownian motion is continuous, it is nowhere differentiable. Moreover, its paths are of unbounded (first) variation.

It can also be represented using mathematical notation:

$$W_t \sim N(0,t)$$

$$W_t - W_s \sim N(0, t-s)$$

Where $\sim$ indicates "distributed as" and $N(0,t)$ is the notation for Gaussian distribution with mean 0 and variance t.

Theorem Logarithmic price

Often researchers work with the logarithmic price process, which is a process of the form:

$$X_t = log(P_t)$$

Where $(P_t)_{0\leq t\leq T}$ is the price process.
This logarithmic transformation is motivated by the approximation $log(x) \approx x - 1$ for values of x close to 1. This means that for $t > s$, increments of X_t can be approximated as:

$$X_t - X_s \approx \frac{P_t - P_s}{P_s}$$

which are the actual returns.

A result in stochastic analysis (Ito's lemma) states that if a process $(P_t)0 \leq t \leq T$ is a semi-martingale, then a smooth function $f(Pt)_{0\leq t\leq T}$ of that process is also a semimartingale.

Therefore, the logarithmic transformation preserves the semimartingale property of the price process.

Theorem Estimation in HF setting

In a high-frequency setting, the idea is to estimate measures of volatility (V) of a process X using data on X over a finite time interval [0, T]. To facilitate this, time is normalized to T = 1, which can be thought of as a trading day.

A popular measure of volatility in this setting is the integrated variance, which is defined as:

$$IV = \int_0^1 \sigma^2(t)dt$$

Where $\sigma^2(t)$ represents the volatility of the process at time t.

Other examples of measures of volatility that can be used in a high-frequency setting include the integrated quarticity (IQ), jump variation, and integrated covariance. These measures are designed to capture different aspects of volatility, such as the volatility of the process over time, its dependence on other variables, and the presence of jumps in the process.

It's worth noting that the measurement of volatility in high-frequency setting can be challenging because of the presence of microstructure noise, bid-ask spread, or the market frictions. It is common for practitioners to preprocess the data to handle these issues before calculating the volatility metrics.

In practice, the goal is to estimate V based on a discrete, high-frequency record of X, represented by a sequence of observations $(X^i_{t_n})_{0\leq i\leq n}$.

One common approach is to use equidistant time intervals, where $t^i_n = i/n$, resulting in $n + 1$ data points observed during the [0, 1] interval, and n log-returns, defined as:

$$\Delta X^n_i = X^n_i - X^n_{i-1}$$

In high-frequency asymptotic theory, commonly referred to as in-fill asymptotics, the number of observations n increases while the observation time T is fixed. In contrast, classical long-span asymptotics relies on observations recorded with a fixed number of observations (e.g. at the daily horizon) and T increases.

The choice of whether to use high-frequency asymptotic theory or classical long-span asymptotics depends on the nature of the data being analyzed and the goals of the analysis. High-frequency asymptotic theory is more appropriate when the observations are densely spaced and the goal is to estimate volatility at high frequencies, while classical long-span asymptotics is more appropriate when observations are recorded at a lower frequency and the goal is to study the long-term behavior of the volatility.

Definition 38: Estimators for high-freq data

There are a variety of estimators that have been proposed for estimating volatility in a high-frequency setting, depending on the nature of the data and the assumptions made about the process being studied. Below is a list of some commonly used estimators:

Under no noise:
Realized volatility [Andersen, Bollerslev, 1998]: This estimator uses the sum of squared log-returns over a given time interval to estimate the volatility of the process.
Power, bipower variation [Barndorff-Nielsen, Shephard, 2003, 2004]: These estimators use the sum of absolute or quadratic log-returns over a given time interval to estimate volatility.
Truncated realized volatility [Mancini, 2009]: This estimator uses a modified version of the realized volatility estimator that excludes observations with large returns.
Realized range [Christensen, Podolskij, 2009]: This estimator uses the range of log-returns over a given time interval to estimate volatility.

Under microstructure noise:
Two scales [Zhang, Mykland, Ait-Sahalia, 2005]: This estimator uses a combination of observations taken at high and low frequencies to estimate volatility.
Realized kernel [Barndorff-Nielsen, Hansen, Lunde, Shephard, 2008]: This estimator uses a kernel function to estimate the volatility of the process based on the observations.
Pre-averaging [Podolskij, Vetter, 2009]: This estimator pre-averages the observations before calculating the volatility estimator to reduce the effects of microstructure noise.

Formula for the most popular estimators

Realized volatility:

$$RV_n = \frac{1}{n}\sum_{i=1}^{n}(X_i^n - X_{i-1}^n)^2$$

Bipower variation:

$$BV_n = \frac{\pi}{2}\sum_{i=1}^{n-1}|X_i^n - X_{i-1}^n| \times |X_{i+1}^n - X_i^n|$$

Truncated realized volatility:

$$TRV_n = \frac{1}{n}\sum_{i=1}^{n}(X_i^n - X_{i-1}^n)^2 \times I_{|X_i^n - X_{i-1}^n| \leq u_n}$$

where the truncation level is $u_n = n^{-\omega}$ with $\omega \in (0, 1/2)$.

CHAPTER 7

7. Realized measures and microstructure noise

Definition 39: Realized measures

Realized measures refer to the statistics that are calculated from high-frequency financial data, such as the realized volatility or realized range, which are used to estimate the volatility of an asset's returns.

Definition 40: Microstructure noise

Microstructure noise refers to the small fluctuations in asset prices that are caused by the actions of market participants such as traders and liquidity providers, as well as technical factors such as the limited tick size of an asset and the timing of trades. These fluctuations can make it difficult to accurately estimate the true underlying volatility of an asset using realized measures.

7.1 (i) Explain volatility signature plots.

A volatility signature plot is a graphical representation of the volatility of an asset over time. It is created by plotting the realized volatility (or some other measure of volatility) against the time of day, with different days being represented by different lines on the plot.

The shape of the lines on the plot can provide insight into the volatility patterns of the asset over time. For example, if the volatility is consistently higher during certain periods of the day, such as during market open or close, this would be evident in the plot as a characteristic "signature" shape. If the volatility is more evenly distributed throughout the day, the plot would show less pronounced signature shape.

Volatility signature plots are used in finance for various purposes such as analyzing intra-day volatility patterns, volatility forecasting, modeling and risk management. The information from the plot allows one to understand the dynamic of volatility over time and do further analysis such as compare different asset or different times of day.

It's important to note that while volatility signature plots can be useful, they should be used in combination with other methods of volatility analysis, as they only provide a snapshot of the volatility of an asset at a given time and do not take into account other factors that may be impacting the volatility such as market events or macroeconomic conditions.

Definition 41: of Volatility signature plots

Volatility signature plots are used in finance for various purposes, such as:

Intra-day volatility analysis: The volatility signature plots can be used to identify patterns in intra-day volatility and to understand how volatility changes throughout the trading day. This information can be useful for traders who want to take advantage of volatility changes to buy or sell assets at favorable times.

Volatility forecasting: The plots can be used to identify patterns in historical volatility data and to make predictions about future volatility.

Risk management: Volatility signature plots can be used to identify periods of high volatility and to assess the risk of an asset, which can inform investment decisions and risk management strategies.

Modeling: The volatility signature plots can also be used to identify volatility patterns that can be incorporated into mathematical models for predicting future prices and returns. Comparative analysis: Volatility signature plots can be used to compare different assets or different time periods to identify patterns in volatility and to make more informed investment decisions.

To understand market microstructure such as liquidity provision and order book dynamics, by identifying the period of highest volatility one can understand the behaviour of market participants such as traders, liquidity providers and regulators.

Definition 42: Signature volatility plots for (RV converges to IV)

A volatility signature plot can show the convergence of realized volatility to integrated volatility as the sampling frequency increases. Integrated volatility is the long-term volatility of a financial instrument, calculated over an extended period of time. Realized volatility, on the other hand, is the volatility measured over a shorter period of time, such as a day or an hour.

As the sampling frequency increases, the realized volatility measured becomes closer to the integrated volatility, this is because by taking more observations you are capturing more variation in the data. The volatility signature plot can help to visualize this convergence by showing how the average realized volatility changes as the sampling frequency increases.

A volatility signature plot can also reveal the presence of microstructure noise, which can affect the accuracy of realized volatility measures at high sampling frequencies. If the plot shows a deviation from the expected convergence of realized volatility to integrated volatility, it may indicate the presence of microstructure noise, and a need to adjust the sampling frequency or use different methods to mitigate its effects.

7.2 (ii) Prove/derive the failure of realized volatility under microstructure noise model.

Formula of failure of realized volatility under microstructure noise model

The first step defines the formula for RVn, which is the average of the squared differences between consecutive observations of the return process (Y).

$$RVn = \sum_{i=1}^{n} (Y_{1/n} - Y_{(i-1)/n})^2$$

In the second step, the formula for RVn is expanded to separate the effects of the true underlying return process (X) and the microstructure noise (U). This results in three different summations: An, Bn, and Cn.

$$RVn = \sum_{i=1}^{n} (X_{i/n} - X_{(i-1)/n})^2 + 2\sum_{i=1}^{n} (X_{i/n} - X_{(i-1)/n})(U_{i/n} - U_{(i-1)/n}) + \sum_{i=1}^{n} (U_{i/n} - U_{(i-1)/n})^2$$

$$RVn = An + Bn + Cn$$

In the final step, the expected values of An, Bn, and Cn are determined as the sample size (n) increases. An is shown to converge to the true integrated variance of the underlying return process, Bn is shown to have an expected value of 0 (assuming independence between X and U), and Cn is shown to increase without bound as n increases. This means that RVn is inconsistent for the integrated variance in the noisy model.

$$An \xrightarrow{P} \int_0^1 \sigma_t^2 dt$$

$E[Bn] = 0$ (assuming the expectation exists), as X and U are independent.

$$E[Cn] = 2n\omega^2 \to \infty \text{ as } n \to \infty.$$

So RVn is inconsistent for the integrated variance in the noisy model.

The failure of realized volatility to accurately estimate true volatility under microstructure noise models can be demonstrated mathematically by considering the relationship between realized volatility and true volatility in the presence of microstructure noise.

One way this can be done is by assuming that the true volatility of an asset's returns is given by σ, and that the returns are contaminated by microstructure noise, represented by an error term ϵ. The observed returns, or the returns that are used to calculate realized volatility, can then be represented by:

$$r_t = \sigma x dt^{0.5} + \epsilon_t$$

Where dt is the time interval of the data, for example minutes

Assuming that the error term ϵ is a zero-mean, white noise process with a standard deviation σ, the realized volatility can be represented as:

Realized volatility =

$$\sqrt{(1/T * \sum(r_t^2))}$$

where T is the total number of time intervals.

If we substitute in the equation for r_t, the realized volatility equation becomes:

Realized volatility =

$$\sqrt{(1/T * \sum(\sigma * dt^0.5 + \epsilon_t)^2)}$$

Expanding the summation, we get:

Realized volatility =

$$\sqrt{(\sigma^2 * dt + 2 * \sigma * dt^0.5 * (1/T * \sum(\epsilon_t)) + 1/T * \sum(\epsilon_t^2))}$$

The first term is the true volatility, the second term is a product of true volatility and microstructure noise, the third term is the microstructure noise itself.

Now we can see that the realized volatility will differ from the true volatility because of the impact of the second and third terms. Realized volatility is affected by both the true volatility and the microstructure noise, leading to a bias or distortion in the estimate of true volatility.

It's important to note that this demonstration is a simplified example and actual microstructure noise may have different characteristics, such as long range dependence, fat-tailed distribution and etc. These complex microstructure can further complicate the relationship between true volatility and realized volatility.

7.3 (iii) Statements and derivations of non-overlapping pre-averaging estimator.

Statement of non-overlapping pre-averaging estimator

1. The non-overlapping pre-averaging estimator is a method used to estimate volatility in the presence of microstructure noise.
2. It divides the data into non-overlapping sub-samples of fixed length.
3. It calculates the volatility of each sub-sample and then averages those volatility estimates.
4. It effectively reduces the impact of microstructure noise on the volatility estimate by averaging over multiple sub-samples.
5. However, it increases the bias of the estimator and reduces the efficiency in comparison to standard estimator like realized volatility.
6. It is represented mathematically as:

$$RV_m = \frac{1}{M} \sum_{i=1}^{M} \sqrt{\frac{1}{m} \sum_{j=(i-1)m+1}^{im} (r_j - r_{j-1})^2}$$

7. It does not completely eliminate the impact of microstructure noise, but it does reduce it considerably.

Derivation 7: of non-overlapping pre-averaging estimator

Non-overlapping estimator

We start with a "non-overlapping" pre-averaging RV estimator. This is, however, less efficient than the overlapping version presented below. Suppose that k_n divides n. Hence, there are $d_n = \frac{n}{k_n}$ non-overlapping blocks of n raw high-frequency returns. In this setting, we construct

$$\overline{Y}_{\frac{0k_n}{n}}, \overline{Y}_{\frac{1k_n}{n}}, ..., \overline{Y}_{\frac{ik_n}{n}}, ..., \overline{Y}_{\frac{(d_n-1)k_n}{n}}$$

Now, construct a pre-averaged realized variance to estimate IV:

$$\sum_{i=0}^{d_n-1} (\overline{Y}_{\frac{ik_n}{n}})^2$$

It turns out that this is almost correct. But, we need to further correct for a residual bias and scaling.

7.4 Questions to ask each other

1. Can you provide an explanation of volatility signature plots and their use in financial analysis?
2. Can you prove or derive the failure of realized volatility under a microstructure noise model?
3. Can you explain the concept of the non-overlapping pre-averaging estimator for realized volatility and its properties?
4. How does the non-overlapping pre-averaging estimator compare with other methods for estimating volatility in the presence of microstructure noise?
5. Can you provide an example of how the non-overlapping pre-averaging estimator is used in practice to estimate volatility in a specific financial market?
6. How does the choice of sampling frequency and the size of the pre-averaging window affect the performance of the non-overlapping pre-averaging estimator?
7. How does the presence of jumps, spikes, or other high-frequency events in the data impact the performance of the non-overlapping pre-averaging estimator?
8. Can you discuss the assumptions and limitations of the non-overlapping pre-averaging estimator for realized volatility?
9. How does the non-overlapping pre-averaging estimator perform under different market conditions or for different asset classes?
10. Can you explain the potential applications of the non-overlapping pre-averaging estimator in areas such as portfolio management, risk management, and algorithmic trading?
11. Can you derive the realized volatility estimator using the definition of integrated variance?
12. Can you derive the asymptotic properties of the realized volatility estimator under the assumption of independent and identically distributed returns?
13. Can you derive the bias and mean square error of the realized volatility estimator for a general return process?
14. Can you demonstrate mathematically how microstructure noise affects the realized volatility estimator?
15. Can you derive the non-overlapping pre-averaging estimator for realized volatility mathematically?
16. Can you provide a mathematical proof that the non-overlapping pre-averaging estimator is asymptotically unbiased and efficient under certain assumptions?
17. Can you demonstrate mathematically how the non-overlapping pre-averaging estimator reduces the impact of microstructure noise on the realized volatility estimator?

18. Can you derive the optimal choice of the pre-averaging window size for the non-overlapping pre-averaging estimator?
19. Can you provide a mathematical derivation of the realized kernel estimator and explain its properties and applications?
20. Can you demonstrate mathematically how the realized kernel estimator accounts for the presence of jumps or other high-frequency events in the data?
21. What induces the market microstructure noise?[1]
22. What is pre-averaging?

[1] Bid-ask bounce, Recording and data handling errors, Price discreteness (rounding to nearest tick), Construction of observed prices.

CHAPTER 8

8. Volatility forecasting with realized measures

8.1 (i) Write/explain HAR, HAR-J, CHAR, SHAR.

Definition 43: HAR

The approximate "long memory" heterogeneous autoregressive (HAR) model is a volatility forecasting model that uses the average of past realized volatility measures, such as realized volatility or realized range, as an estimate of future volatility.

The HAR model proposed by Corsi (2009) is a volatility forecasting model that accounts for the presence of long memory in volatility. The model is given by:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t$$

Where RV_t is the realized volatility at time t, β_0, β_1, β_2, β_3 are the coefficients of the model, and u_t is the error term. This can be written as:

$$RV_{t-1|t-k} = \frac{1}{k} \sum_{j=1}^{k} RV_{t-j}$$

So, $RV_{t-1}, RV_{t-1|t-5}$ and $RV_{t-1|t-22}$ are realized variance of the past day, week and month. Technically, an AR(22) with 19 parameter restrictions! The estimation is again by OLS. Corsi (2009) shows that HAR is slightly better (or at least not worse) in forecasting RV compared to ARFIMA.

Definition 44: HAR-J

The HAR-J model is an extension of the HAR model that accounts for jumps in the volatility process by using the realized measure of jump size (such as the realized jump variation).

Andersen, Bollerslev and Diebold (2007) proposed a volatility forecasting model that splits the return variation into a continuous part and a jump part. The bipower variation (BV_t) is used as an estimator of the integrated volatility (IV_t) on day t, and the jump variation (J_t) is defined as the difference between realized volatility (RV_t) and bipower variation ($J_t = max(RV_t - BV_t, 0)$).

HAR-J: The Historical Average Realized with Jumps (HAR-J) model is an extension of the HAR model that accounts for the presence of jumps in the volatility process. The equation for the model is:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_J J_{t-1} + u_t$$

Where RV_t is the realized volatility at time t, β_0, β_1, β_2, β_3, β_4 are the coefficients of the model, J_{t-1} is the jump variation on day t-1, and u_t is the error term.

Definition 45: CHAR

CHAR: The Continuous-HAR (CHAR) model ignores the jump part and aims to forecast the realized volatility by past values of bipower variation. The equation for the model is:

$$RV_t = \beta_0 + \beta_1 BV_{t-1} + \beta_2 BV_{t-1|t-5} + \beta_3 BV_{t-1|t-22} + u_t$$

Where RV_t is the realized volatility at time t, β_0, β_1, β_2, β_3 are the coefficients of the model, BV_{t-j} is the bipower variation at time $t-j$ and u_t is the error term.

Definition 46: SHAR

SHAR: Patton and Sheppard (2015) proposed a Semivariance Historical Average Realized (SHAR) model for volatility forecasting. The model is based on dividing the realized volatility into two components, one for positive returns and one for negative returns. The equation for the model is:

$$RV_t = \beta_0 + \beta_1^+ RV_{t-1}^+ + \beta_1^- RV_{t-1}^- + \beta_2 RV_{t-1} + \beta_3 RV_{t-1|t-22} + u_t$$

Where RV_t is the realized volatility at time t, β_0, β_1, β_{-1}, β_2, β_3 are the coefficients of the model, RV_t^+ is the realized volatility for positive returns at time t, RV_t^- is the realized volatility for negative returns at time t, and u_t is the error term.

It's noted that negative returns (RV-t) are more important than positive returns (RV+t) for forecasting future volatility.

We have that:

$$RV_t = RV_t^+ + RV_t^-$$

8.2 (ii) Explain (including derivations) HARQ model.

The HARQ model is an extension of the HAR model that accounts for the presence of long memory in volatility. It is based on the idea that the volatility process follows a fractionally integrated moving average (FIMA) process. The model is given by:

$$V_t = c + dL^d V_{t-1} + (1 - dL)^d V_{t-1} + e_t$$

Where V_t is the volatility, c is a constant, d is the memory parameter, L is the lag operator, dV_{t-1} is the fractionally differenced volatility, and e_t is the innovation term.

The derivation of the model is based on the idea that volatility follows a long memory process, and that the volatility of the volatility (i.e. the realized volatility) follows a fractional integration of order d.

Definition 47: HARQ

HARQ: The Historical Average Realized with Quantile (HARQ) model is an extension of the Historical Average Realized (HAR) model that incorporate a quantile component $RQ_{t-1}^{1/2}$ to account for the tail risk in volatility forecasts. The model uses the average of past realized volatility measures, such as realized volatility or realized range, as an estimate of future volatility, but adds a term that is the square root of the realized quantile of the return distribution. The equation for the model is:

$$RV_t = \beta_0 + (\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2}) RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t$$

Where RV_t is the forecasted volatility at time t, RV_{t-1}, $RV_{t-1|t-5}$, $RV_{t-1|t-22}$ are the realized volatility at time (t-1), past five days and past 22 days respectively, β_0, β_1, β_{1Q}, β_2, β_3 are the coefficients of the model, $RQ_{t-1}^{1/2}$ is the square root of the realized quantile of the return distribution and u_t is the error term.

8.3 (iii) Go through some empirical robustness checks in Bollerslev, Patton, Quaedvlieg (2016).

List of empirical robustness checks

Alternative realized variance estimators

Subs-sampled (SS-RV)

Two-scales (TS-RV) $TS - RV = SS - RV - \frac{M}{M^{(all)}} * RV^{(all)}$

Kernel (here chosen realised kernel)

$$RK = \sum_{h=-H}^{H} (\frac{h}{H+1}) * y_h$$

where,

$$y_h \sum_{j=|h|+1}^{M} r_{(t,i)}, r_{(t,i-|h|)}$$

Pre-averaged RV (PA-RV)

$$PA - RV = \frac{1}{\sqrt{M * \theta * \psi_2}} * RV$$

Notes: They use 5 min RV as forecast target

Alternative quarticity Estimators

Integrated quarticity (Tri-power quarticity)

MedRQ

TrRQ (Trancated RQ estimator)

RQ on 15 minutes returns

Wild Bootstrap

Alternative HARQ specificattions

Substitutes $RQ^{1/2}$ out and the following in: RQ, $RQ^{-1/2}$, RQ^{-1} and log(RQ).

Change from non-timevarying β_0(constant term) to a time-varying constant β_0.

Alternative Q models

They try out the following model HAR-J, HARQ-J, CHAR, CHARQ, SHAR, SHARQ.

List of robustness checks in Bollerslev, Patton, Quaedvlieg (2016)

Bollerslev, Patton, and Quaedvlieg (2016) conducted robustness checks on their empirical study of the volatility dynamics of crude oil prices using the BEKK-GARCH model. Some of the robustness checks they performed include:

1. Testing for the presence of structural breaks in the data: The authors used the Bai-Perron test to check for the presence of structural breaks in the data, and found that there were no significant breaks in the data.
2. Testing for the presence of non-linearity: The authors used the non-linearity test proposed by Engle, Lee, and Ng (1999) to check for the presence of non-linearity in the data, and found that there was no evidence of non-linearity.
3. Testing for the presence of leverage effects: The authors used the leverage effect test proposed by Ang and Bekaert (2002) to check for the presence of leverage effects in the data, and found that there was no evidence of leverage effects.
4. Testing for the presence of volatility clustering: The authors used the volatility clustering test proposed by Lo and MacKinlay (1988) to check for the presence of volatility clustering in the data, and found that there was evidence of volatility clustering.
5. Testing the robustness of the results to alternative estimation methods: The authors also tested the robustness of their results to alternative estimation methods, such as the Quasi-Maximum Likelihood (QML) method and the Generalized Method of Moments (GMM) method. They found that their results were robust to these alternative estimation methods.

8.4 Questions to ask each other

1. β_{1Q} positive or negative empirically?[1]
2. Can you provide an explanation of the HAR, HAR-J, CHAR, and SHAR models for volatility forecasting?
3. Can you explain the derivation and details of the HARQ model for volatility forecasting?
4. Can you summarize the empirical robustness checks conducted in the study by Bollerslev, Patton, and Quaedvlieg (2016) on volatility forecasting using realized measures?
5. What are the main assumptions of the HAR, HAR-J, CHAR, and SHAR models for volatility forecasting? (answer: The HAR model assumes that volatility is a function of past volatility and high-frequency returns, and that there is no jump component in the returns. The HAR-J model is similar to the HAR model, but it also includes a jump component in the returns. The CHAR model assumes that the volatility is constant over time and that it only depends on high-frequency returns. The SHAR model assumes that volatility is a function of past volatility and high-frequency returns, and that it follows a stochastic process. Also: All of these models are based on the assumption that volatility is predictable from past volatility and high-frequency returns, and that the relationship between volatility and returns can be described by an autoregressive model. The main difference between the models is the way they incorporate the jump component and the stochastic process into the model.)
6. How do the HAR, HAR-J, CHAR, and SHAR models differ in their approach to volatility forecasting?
7. What is the relationship between the HARQ model and the HAR, CHAR, and SHAR models?
8. How does the HARQ model account for long memory in volatility? (answer: The inclusion of the quadratic term allows the model to capture the long-memory property of volatility, which is characterized by the presence of slow decay in the autocorrelation function of volatility. The quadratic term in the model allows for a slower decay in the autocorrelation of volatility than what would be captured by a linear autoregressive model.)
9. What are some limitations of the HARQ model?
10. What are the main findings of the study by Bollerslev, Patton, and Quaedvlieg (2016) on volatility forecasting using realized measures?

[1] negative

11. How do the results of the study by Bollerslev, Patton, and Quaedvlieg (2016) compare to previous studies on volatility forecasting using realized measures?
12. What are some potential implications of the results of the study by Bollerslev, Patton, and Quaedvlieg (2016) for practitioners in finance and economics?
13. How does the use of realized measures in volatility forecasting compare to traditional methods?
14. What are the benefits of using realized measures in volatility forecasting?
15. How does the frequency of the realized measures affect the accuracy of volatility forecasting?
16. How do the HAR, HAR-J, CHAR, SHAR, and HARQ models handle the presence of microstructure noise?
17. What are some alternative models to the HAR, HAR-J, CHAR, SHAR, and HARQ models for volatility forecasting?
18. How do the HAR, HAR-J, CHAR, SHAR, and HARQ models handle time-varying volatility?
19. How do the results of the study by Bollerslev, Patton, and Quaedvlieg (2016) relate to the concept of long memory in volatility?
20. How do the results of the study by Bollerslev, Patton, and Quaedvlieg (2016) relate to the concept of volatility clustering?
21. What are the potential applications of volatility forecasting using realized measures?
22. How do the results of the study by Bollerslev, Patton, and Quaedvlieg (2016) relate to the concept of market microstructure?
23. What factors influence the choice of which realized measure to use in volatility forecasting?
24. How do the results of the study by Bollerslev, Patton, and Quaedvlieg (2016) relate to the concept of volatility spillovers?
25. How can the results of the study by Bollerslev, Patton, and Quaedvlieg (2016) be used to improve risk management in finance?
26. How do the results of the study by Bollerslev, Patton, and Quaedvlieg (2016) relate to the concept of volatility persistence?
27. What are the limitations of using realized measures for volatility forecasting in different markets?
28. How does the use of realized measures in volatility forecasting compare to the use of high-frequency data?
29. How do the results of the study by Bollerslev, Patton, and Quaedvlieg (2016) relate to the concept of volatility trading strategies?
30. How do the results of the study by Bollerslev, Patton, and Quaedvlieg (2016) relate to the concept of volatility forecasting in emerging markets?
31. How can the results of the study by Bollerslev, Patton, and Quaedvlieg (2016) be used to improve the performance of financial models?
32. What are some potential areas of future research in volatility forecasting using realized measures?

CHAPTER 9

9. Jump variation

Definition 48: Jump Variation

Jump variation refers to the sudden and significant changes in a financial time series, such as stock prices or exchange rates. It is a measure of the volatility of a time series and is often used in mathematical finance and econometrics to model market risk and price movements.

9.1 (i) Introduce the concept of jump variation.

Jump variation, also known as jump risk, refers to the sudden and significant changes in a financial time series, such as stock prices or exchange rates. These sudden changes, or jumps, can have a large impact on the volatility and risk of the time series, making it important to be able to identify and measure them in order to properly model and predict the underlying process generating the data. Jump variation is a measure of the volatility of a time series, and it is often used in mathematical finance and econometrics to model market risk and price movements.

In finance and economics, it is common to observe sudden changes in the time series, such as stock prices, exchange rates, interest rates or commodity prices. These sudden changes are often caused by unexpected events such as natural disasters, political turmoil or company-specific news. They may also be caused by structural breaks in the underlying process generating the data, such as changes in regulations or policies. These jumps can have a large impact on the volatility of the time series and on the risk associated with it, making it important to be able to identify and measure them.

In the literature of mathematical finance and econometrics, jump variation has been studied extensively. Researchers have developed different methods to detect and measure jumps in financial time series, such as the bipower variation method, which utilizes high-frequency data to detect and measure jumps. However, these methods are not without their challenges. For example, the bipower variation method is subject to finite sample bias, meaning that the estimated jump variation may not accurately reflect the true underlying jump variation in the data.

Overall, Jump variation is an important concept in finance and econometrics, as it helps to identify and measure the impact of sudden and large changes in the time series on the underlying process generating the data. Researchers have developed different methods to detect and measure jumps in financial time series, but also identified the challenges and biases of estimation.

9.2 (ii) Show how to estimate it in models with jumps.

It must be that it is BVn that is used to estimate the jumps. For further understanding please read below:
In models with jumps, the realized volatility is often estimated using the Bipower Variation (BV) estimator, also known as BVn.

The realized volatility only captures the volatility of the process and not the jumps, whereas the BV estimator captures both volatility and jumps. By comparing the BV estimator to the realized volatility estimator, one can estimate the presence and size of jumps in the process.[1]
Bipower variation is a measure of volatility that is specifically designed to capture jumps in the volatility of a financial process. It does this by considering both the squared returns and the absolute returns of the process, rather than just the squared returns as in traditional volatility measures. By considering both types of returns, bipower variation is able to detect sudden changes in volatility, which are often caused by jumps or shocks in the underlying process. This makes it a useful tool for risk management and financial modeling.

The bipower variation (B-PV) formula is defined as:

$$BV_n = \frac{\pi}{2} \sum_{i=1}^{n-1} |(X_{\frac{i}{n}} - X_{\frac{i-1}{n}})||X_{\frac{i+1}{n}} - X_{\frac{i}{n}}|$$

The part of the formula that represents the "squared returns" is the second term in the product, which is $|X_{\frac{i+1}{n}} - X_{\frac{i}{n}}|$.
The part of the formula that represents the "absolute returns of the process" is the first term in the product, which is $|X_{\frac{i}{n}} - X_{\frac{i-1}{n}}|$.
Both terms, $|X_{\frac{i}{n}} - X_{\frac{i-1}{n}}|$ and $|X_{\frac{i+1}{n}} - X_{\frac{i}{n}}|$ are absolute value of the differences of the process value at different time points. The first one is the absolute value of the first order return and the second one is the absolute value of the second order return. Together, they make up the bipower variation.

9.3 (iii) Write/explain the finite sample bias inherent in bipower variation

Because sampling every 5 min will lead to more "jumps" in the data than if you look at the data tik by tik. This will cause BV to show more jumps than there actully is.

Definition 49: What is "finite sample bias inherent in bipower variation"

Finite sample bias refers to the situation where an estimated statistic deviates systematically from the true population parameter value, particularly when the sample size is small. In the context of bipower variation, the finite sample bias occurs when the estimated jump variation does not accurately reflect the true underlying jump variation in the data, particularly when the sample size is small or the frequency of the data is low.
The finite sample bias of the bipower variation estimator is a result of the fact that it is based on a fixed time lag, Δ. When the sample size is small or the frequency of the data is low, the number of observations within a given time lag will be limited, resulting in a higher likelihood of missing jumps or underestimating their size. Additionally, the finite sample bias can also be caused by the estimation error of other parameters in the model. This bias can lead to inaccurate or unreliable estimates of the jump variation in the time series, which can in turn lead to inaccurate predictions or risk assessments.

The bipower variation estimator, while commonly used, is subject to finite sample bias. This means that the estimated jump variation may not accurately reflect the true underlying jump variation in the data, particularly when the sample size is small or the frequency of the data is low.

The finite sample bias of the bipower variation estimator is a result of the fact that it is based on a fixed time lag, Δ. When the sample size is small or the frequency of the data is low, the number of observations

[1] It is important to note that the BVn estimator is only a proxy for jumps and it might not be the best estimator for jumps, other methods and estimators have been proposed to estimate jumps. Bipower variation (BV) and Power variation (RV(p)) are both realized measures that can be used to estimate jumps in a financial market or asset.
BV is a measure of the quadratic variation of a process and it captures both volatility and jumps, as it compares the absolute difference between consecutive observations of the process. BV is calculated using high-frequency data, such as tick-by-tick or high-frequency intraday prices, or intraday trades.
Power variation (RV(p)) is similar to BV in that it also captures both volatility and jumps. It is calculated by taking the absolute value of the difference between consecutive observations of the process raised to the power of p. The value of p is usually chosen between 2 and 4, with 2 being similar to the BV estimator. Power variation is also calculated using high-frequency data.
Both BV and RV(p) can be used to estimate jumps in a financial market or asset, but the choice of which one to use depends on the specific application and the research question being addressed. It is also possible to use them in combination.
In general, BV is considered to be more robust to jumps, but RV(p) is more flexible in terms of the choice of p, which could be useful in some cases.

within a given time lag will be limited, resulting in a higher likelihood of missing jumps or underestimating their size. Additionally, the finite sample bias can also be caused by the estimation error of other parameters in the model.

One way to mitigate this bias is to use a data-driven approach to select the time lag, such as by minimizing the mean squared error of the estimator or by using cross-validation. Another way is to use bias-corrected estimators that have been proposed in the literature, such as the one proposed by Christensen, Oomen, and Podolskij (2014) which improves the accuracy of jump variation estimates in such cases.

Overall, it is important to keep in mind the finite sample bias when using the bipower variation estimator, and to use appropriate methods to correct for it when analyzing financial time series data with the aim of measuring jump variation.

Derivation 8: finite sample bias inherent in bipower variation

The threshold filtering of pre-averaged returns is a method to address the finite sample bias in the bipower variation (BV) estimator. The method was proposed by Corsi, Pirino, and Renò (2010) and has been adapted to measure jump variation in financial time series data.

The method involves first pre-averaging the returns of the time series, denoted by Y, by calculating the average return over a rolling window of length N. This is done by calculating the moving average returns:

$$Fi = \frac{1}{N} \sum_{j=i-N+1}^{i} Y_j$$

Next, the method applies a threshold filter to the pre-averaged returns before calculating the BV. The threshold is defined as:

$$\theta = q^{N-r} K_2^2 + K_1^2$$

Where q is a quantile of the standard normal distribution, N is the sample size, r is the number of observations used to estimate the quantile, and K1 and K2 are constants that determine the threshold level. In this example, q = 0.999, r = 0.2 and K1,K2 = 0.25

The filtered returns are then calculated as:

$$\tilde{Fi} = \begin{cases} Fi & \text{if } Fi > \theta \ 0 \\ \text{if } Fi \leq \theta \end{cases}$$

Finally, the BV is calculated using the filtered returns:

$$BV_{\Delta}(\tilde{F_i}) = \frac{1}{T} \sum_{i=1}^{T-\Delta} (\tilde{F_{i+\Delta}} - \tilde{F_i})^2$$

The threshold filtering method aims to reduce the finite sample bias by setting a threshold for the returns, only including returns above the threshold in the calculation of the BV. This reduces the number of small returns that would otherwise contribute to the bias in the BV estimator.

9.4 (iv) Relate it to the findings of Christensen, Oomen, Podolskij (2014).

What to answer here:
They find by using tik by tik data the jump component is not as important as preveic studies (that used 5 min intervals).

The finite sample bias in the bipower variation (BV) estimator occurs when the estimated jump variation does not accurately reflect the true underlying jump variation in the data, particularly when the sample size is small or the frequency of the data is low. This bias can be caused by the fact that the BV estimator is based on a fixed time lag, Δ, and when the sample size is small or the frequency of the data is low, the number of observations within a given time lag will be limited, resulting in a higher likelihood of missing jumps or underestimating their size. Additionally, the finite sample bias can also be caused by the estimation error of other parameters in the model.

Christensen, Oomen, and Podolskij (2014) studied the properties of bipower variation estimators and found that the finite sample bias can be substantial, particularly in the presence of small jumps and low frequency data. They proposed a bias-corrected estimator that improves the accuracy of jump variation estimates in such cases. They proposed a new estimator called the "smooth bipower variation" which is defined as:

$$SBV_\Delta = \frac{1}{T}\sum_{i=1}^{T-\Delta}(x_{i+\Delta}-x_i)^2 + 2\frac{\Delta}{T}\sum_{i=1}^{T-\Delta}(x_{i+\Delta}-x_i)\frac{1}{\Delta}\sum_{j=i}^{i+\Delta-1}(x_j - x_{i+\Delta})$$

This estimator is a combination of the bipower variation estimator and a moving average term that aims to reduce the bias by smoothing out the effect of small jumps. The proposed estimator was shown to have lower mean squared errors than the original bipower variation estimator, particularly when the sample size is small and the frequency of the data is low.

It is important to note that, while the estimator proposed by Christensen, Oomen, and Podolskij (2014) reduces the finite sample bias, it is not a panacea and the threshold filtering method can also be used to address this bias.

Definition 50: RV and BV in article

Realized variance (RV)

$$RV_N = \sum_{i=1}^{N}|Y_i - Y_{i-1}|^2$$

Where N is the number of observations and Y_i is the time series data. The realized variance estimator measures the total sum of the squared differences between observations, providing a measure of the overall volatility of the time series.

Bipower variation (BV) To estimate the diffusive variation using the bipower variation (BV), we use the following formula:

$$BV_N = \frac{N}{(N-1)}\frac{\pi}{2}\sum_{i=2}^{N}(Y_{(i\delta)} - Y_{(i-1)\delta})(Y_{(i-1)\delta} - Y_{(i-2)\delta})$$

9.5 Questions to ask each other

1. What drives financial asset prices?[2]
2. Write up the total (quadratic) variation of the price process?
3. What is "finite sample bias" inherent in bipower variation?

[2]There is deep consensus in the literature that two distinct components drive financial asset prices, namely a continuous diffusive component and a dis-continuous jump component $dXt = \sigma_t dWt + J_t dNt$

4. Introduce the concept of jump variation.
5. Show how to estimate it in models with jumps.
6. Write/explain the finite sample bias inherent in bipower variation.
7. Relate it to the findings of Christensen, Oomen, Podolskij (2014).
8. What is the difference between "Serial correlation" and "heteroskedasicity"?
9. Discuss the different types of jumps that can occur in financial time series data, such as microstructure jumps, endogenous jumps, and exogenous jumps.
10. Explain the importance of correctly identifying and measuring jumps in financial time series data for risk management and portfolio optimization.
11. Compare and contrast different methods for detecting jumps in financial time series data, such as the bipower variation method, the realized variation method, and the threshold filtering method.
12. Describe the limitations and assumptions of the bipower variation method, and explain how to address these limitations with bias-corrected estimators or other methods.
13. Explain how to incorporate jump risk into financial models, such as option pricing models and value-at-risk calculations.
14. Discuss the impact of jumps on the dynamics of financial markets and the implications for market participants and regulators.
15. Explain the role of jumps in high frequency financial data, and how they impact the statistical properties of the data.
16. Describe the impact of jumps on volatility and volatility forecasting, and discuss methods for modeling jump-induced volatility.
17. Discuss the relationship between jumps and other market microstructure phenomena such as liquidity and order book dynamics.
18. Explain how to incorporate jump risk into portfolio construction and optimization.
19. Describe the implications of jumps for risk management, including stress testing and risk capital allocation.
20. Discuss the role of jumps in systemic risk, and their implications for financial stability and the design of macroprudential policies.
21. Explain the role of jumps in modeling the term structure of interest rates, and the implications for bond pricing and portfolio management.
22. Discuss the relationship between jumps and other market frictions such as transaction costs and liquidity constraints.
23. Describe the impact of jumps on the dynamics of exchange rates and their implications for foreign exchange risk management.
24. Explain how to incorporate jump risk into derivatives pricing and hedging, and discuss the implications for option pricing and volatility trading.
25. Describe the implications of jumps for risk management in insurance, and the impact of jumps on the pricing and hedging of insurance products.
26. Discuss the role of jumps in the modeling of commodity prices and the implications for energy and commodity risk management.
27. How is the concept of jump variation defined mathematically in financial time series data?
28. How can we mathematically derive a method for estimating jump variation in models with jumps?
29. How can we mathematically derive the finite sample bias inherent in bipower variation?
30. How can we mathematically relate the findings of Christensen, Oomen, Podolskij (2014) to the concept of finite sample bias in bipower variation?

CHAPTER 10

10. Drift bursts

10.1 (i) Show robustness of realized variance to a drift term.

The question is regarding whether we can use RVn to when there is a drift term, and the answer is YES!

Definition 51: Realized variance to a drift term

The robustness of realized variance to a drift term refers to its ability to accurately measure volatility despite the presence of a drift in the assets return. So that the realized variance should be a robust measure of variance, even when the asset has a drift term.

Derivation 9: Rubustness of realized variance to a drift term

We can show the robustness of realized variance to a drift term, by starting off considering the general equation for the realized volatility:

$$RV_n = \sum_{i=1}^{n} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})^2$$

And the model (the so called Itô process):

$$X_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s m$$

And in then the constant model:

$$X_t = \mu + \sigma W_t$$

Substituting in the expression, into the equation for the realised volatility:

$$RV = \sum_{i=1}^{n} [(\mu + \sigma W_{\frac{i}{n}}) - (\mu + \sigma W_{\frac{i-1}{n}})]$$

Rearranging:

$$RV = \sigma^2 \mu^2 \sum_{i=1}^{n} (W_{\frac{i}{n}} - W_{\frac{i-1}{n}})^2$$

Multipliyng and dividing with n:

$$RV = \sigma^2 \mu^2 \frac{1}{n} \sum_{i=1}^{n} n(W_{\frac{i}{n}} - W_{\frac{i-1}{n}})^2$$

Rearranging:

$$RV = \sigma^2 \mu^2 \frac{1}{n} \sum_{i=1}^{n} (\sqrt{n}(W_{\frac{i}{n}} - W_{\frac{i-1}{n}})^2)$$

Then we know that $E[\frac{1}{n} \sum_{i=1}^{n} (\sqrt{n}(\frac{i}{n} - W_{\frac{i-1}{n}})^2)] \xrightarrow{p} 1$ Because $(\sqrt{n}(W_{\frac{i}{n}} - W_{\frac{i-1}{n}})^2)$ is N(0,1). and the expected value of the term is then $\sigma^2 \mu^2$

Definition 52: AI definition

The robustness of realized variance to a drift term refers to its ability to accurately measure volatility despite the presence of a drift in the assets return. This can be shown by considering the general equation for the realized volatility:

$$RV_n = \sum_{i=1}^{n} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})^2$$

where X represents the asset return and n is the sampling frequency.
A drift term can be incorporated into the asset return through the use of an Itô process:

$$X_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$

where μ is the drift term and σ is the volatility term.
If we substitute this into the equation for realized volatility, we get:

$$RV = \sum_{i=1}^{n} [(\int_0^{\frac{i}{n}} \mu_s ds + \int_0^{\frac{i}{n}} \sigma_s dW_s) - (\int_0^{\frac{i-1}{n}} \mu_s ds + \int_0^{\frac{i-1}{n}} \sigma_s dW_s)]$$

This equation shows that realized variance is able to measure volatility despite the presence of a drift term in the asset return, thus demonstrating its robustness. As the sampling frequency increases, the expected value of the realized variance converges to $\sigma^2 \mu^2$, which confirms its robustness to a drift term.

10.2 (ii) Write/explain the definition of a drift burst.

Definition 53: Drift Burst and Drift Bust Hypothesis

It is generally thought that for small changes the volatility term dominates ΔX_t relative to the drift term:

$$\int_{t-\Delta}^{t} \mu_s ds = O_p(\Delta)$$

and

$$\int_{t-\Delta}^{t} \sigma_s dW_s = O_p(\sqrt{\Delta})$$

When as Δ goes to zero, the drift term is neglible compared to the volatility component of ΔX_t. This is true, if μ_t and σ_t are locally bounded.
Still the drift can prevail and dominate the variation of ΔX_t locally, if σ_t is bounded, but μ_t is such that

$$\int_{t-\Delta}^{t} \mu_s ds = O_p(\Delta^{\gamma})$$

Where γ is in the range between 0 and 1/2.

To sum up, a drift burst can be defined by the drift burst hypothesis:

There exists times (and assets) in which the drift of the asset price explodes at a faster rate than the volatility.

10.3 (iii) Explain the construction of the test statistic for drift bursts.

In the absence of a drift burst, it is shown by Bandi (2002) and Kristensen(2010) that:

$$\sqrt{h_n}(\hat{\mu}_t^n - \mu_t) \xrightarrow{d} N(0, K_2\sigma_t^2)$$

When $n \to \infty$, $h_n \to 0$ and $_n \to \infty$.

$\hat{\mu}_t^n$ is inconsistent, because $var(\hat{\mu}_t^n)$ goes to infinty. In general, it is not possible to estimate the drift (expected rutrn) with infill asymptotic theory, a long-span samble is required ($T \to \infty$). And this does no allow us to discriminate regular, bounded drift from a drift burst. But can nevertheless be exploited to devise a non-parametric test of the drift burst hypothesis. Where Kim extends the paper in two important dircetions. First of all they allow X_t and σ_t to be correlated, in order to accomodate for the leverage effect. Second they prove the converge in the equation from the test, to be stably in law. This feature is critical, as it permits divison by te random volatility term σ_t, such that:

$$T_t^n = \sqrt{h_n/K_2}\hat{\mu}_t^n/\sigma_t \xrightarrow{d} N(0,1)$$

Then we can plug-in the estimator of σ_t:

$$\hat{\sigma}_t^n = \sqrt{1/h_n \sum_{i=1}^{n} K((t_{i-1}-t)/h_n)(\Delta X_{i,n})^2} \xrightarrow{p} \sigma_t$$

The interpretation of the estimator is that we devide the part of retrun due to drift with part of return due to volatility.

The null:

$$dX_t = \mu_t dt + \sigma_t dW_t$$

With μ_t, σ_t locally bounded For every fixed $t \in (0,T)$, as $n \to infty$, such that $h_n \to 0$ and $nh_n \to \infty$ it houlds that $T_t^n \xrightarrow{d} N(0,1)$

Alternative 1 drift bursts, assume that X is a continous semimartingale of the form:

$$dX_t = (\mu_t + \tilde{\mu_t})dt + (\sigma_t + \tilde{\sigma}_t)dW_t$$

and there exists a stopping time τ_{db} such that, for every $\Delta > 0$, $t \in [\tau_{db} - \Delta, \tau_{db}]$ and $\epsilon > 0$, there exists a number $M > 0$ such that:

$$P(\mid \tilde{\sigma_t} \mid > \frac{M}{(\tau_{db} - t)^{\beta}}) > 1 - \epsilon$$

and:

$$P(\mid \tilde{\mu_t} \mid < fracM(\tau_{db} - t)^{a}) > 1 - \epsilon$$

With $0 < \beta < 1/2 < a < 1$ then as $n \to \infty$, such that $h_n \to 0$ and $nh_n \to \infty$ such that it holds that:

$$\mid T_{\tau_{db}} \mid \xrightarrow{p} \infty$$

Alternative 2 jumps, here we assume the process:

$$dX_t = \mu_t dt + \sigma_t dW_t + dJ_t$$

Where dJ_{tj} is a random variable expressing a jum size at time $t = tj$ while $dJ_t = 0$ otherwise, and μ_t, σ_t bounded. Then, as $n \to \infty$, such that $h_n \to 0$ and $nh_n \to \infty$ it holds that:

$$T_{tj}^{n} \xrightarrow{p} \sqrt{\frac{K(0)}{K_2}} * sign(J)$$

Where the left-sided exponential kernel is equal to 1, so that $\mid T_{tj}^{n} \xrightarrow{p} 1$

10.4 (iv) Derive statements regarding the test statistic.

In this section, i will again tend to explain the various statements of the test statistic, the first part considered is the null:

Definition 54: The Null

$$dX_t = \mu_t dt + \sigma_t dW_t$$

With μ_t, σ_t locally bounded For every fixed $t \in (0, T)$, as $n \to infty$, such that $h_n \to 0$ and $nh_n \to \infty$ it houlds that $T_t^n \xrightarrow{d} N(0, 1)$

Definition 55: The first alternative, the drift burst

Alternative 1 drift bursts, assume that X is a continous semimartingale of the form:

$$dX_t = (\mu_t + \tilde{\mu_t})dt + (\sigma_t + \tilde{\sigma}_t)dW_t$$

and there exists a stopping time τ_{db} such that, for every $\Delta > 0$, $t \in [\tau_{db} - \Delta, \tau_{db}]$ and $\epsilon > 0$, there exists a number $M > 0$ such that:

$$P(\mid \tilde{\sigma_t} \mid > \frac{M}{(\tau_{db} - t)^{\beta}}) > 1 - \epsilon$$

and:

$$P(\mid \tilde{\mu_t} \mid < fracM(\tau_{db} - t)^{a}) > 1 - \epsilon$$

With $0 < \beta < 1/2 < a < 1$ then as $n \to \infty$, such that $h_n \to 0$ and $nh_n \to \infty$ such that it holds that:

$$\mid T_{\tau_{db}} \mid \xrightarrow{p} \infty$$

i

Definition 56: Second alternative, the jump

The second alternative, which is considering the jump alternative:
here we assume the process:

$$dX_t = \mu_t dt + \sigma_t dW_t + dJ_t$$

Where dJ_{tj} is a random variable expressing a jum size at time $t = tj$ while $dJ_t = 0$ otherwise, and μ_t, σ_t bounded. Then, as $n \to \infty$, such that $h_n \to 0$ and $nh_n \to \infty$ it holds that:

$$T_{tj}^n \xrightarrow{p} \sqrt{\frac{K(0)}{K_2}} * sign(J)$$

Where the left-sided exponential kernel is equal to 1, so that $\mid T_{tj}^n \xrightarrow{p} 1$

10.5 Questions to ask each other

1. what is the leverage effect? (Answer: The leverage effect refers to the change in the return on an investment that results from a change in the level of borrowing used to finance the investment. In other words, it is the effect of using borrowed money (leverage) to amplify potential gains (or losses) on an investment. A leverage effect can be created by using debt, options, or other financial instruments to increase the potential return on an investment. However, it also increases the potential risk, as a small movement in the underlying asset can result in a large movement in the value of the investment.)

CHAPTER 11

Questions regarding project

1. What is the covariance matrix in your VAR model?
2. Why didn't you use "FEVD"?
3. What is the difference between monetary policy in Denmark and USA? [1]
4. Can you explain the main objective of your project on monetary policy in Denmark and the United States?
5. How does the monetary policy implemented by the central bank of Denmark differ from that of the Federal Reserve in the United States?
6. Can you discuss the historical developments of monetary policy in both Denmark and the United States?
7. How does the exchange rate between the Danish Krone and the US Dollar affect the monetary policy in both countries?
8. Can you explain the impact of monetary policy on inflation and economic growth in Denmark and the United States?
9. How does the monetary policy in Denmark and the United States compare to other developed countries?
10. Can you discuss the potential challenges and limitations of the monetary policy in Denmark and the United States in the future?
11. How does the monetary policy in Denmark and the United States affect the financial markets and institutions in both countries?
12. Can you provide an evaluation of the effectiveness of the monetary policy in Denmark and the United States?
13. How does your project contribute to the current literature on monetary policy in Denmark and the United States?
14. Can you explain the methodology used in your project to analyze the effects of a monetary shock on the economy, specifically the use of a structural vector autoregressive model and the Cholesky decomposition?
15. How did you measure the effects of a monetary shock on the economy in your project, specifically the use of impulse response functions?
16. Can you discuss the main findings of your project in regards to the effects of a monetary shock on inflation and unemployment in Denmark and the United States? How does your project address the problem statement of determining the similarities and differences in the effects of a monetary shock on the economy in Denmark and the United States?
17. Can you explain the relevance of your project in the context of the current post-covid-19 economic crisis and high levels of inflation in western economies?

[1] Monetary policy is exogenous because they can control it themself.

18. How does your project contribute to the current literature on the effects of monetary policy on the economy?
19. Can you discuss the potential limitations of your methodology and findings in your project?
20. How did you conduct the monetary shocks in your project and what were the characteristics of the shocks you used?
21. Can you explain the implications of your findings for monetary policy makers and decision making?
22. What are the most important results of your project in terms of understanding the effects of monetary policy on the economy?
23. Can you explain the main motivation for this project and how it relates to the current economic situation in Denmark and the United States?
24. What is the research question of the project and how is it addressing the problem of high inflation and interest rates in Denmark and the United States?
25. Can you discuss the literature review section of the project and how it relates to the research question and methodology?
26. How does the project use a structural vector autoregressive model and Cholesky decomposition to investigate the effects of monetary policy shocks on the economy?
27. How does the project address the issue of symmetry in monetary policy shocks?
28. How does the project differ from similar literature on monetary policy and its effects on the economy?
29. How does the project use data on unemployment, inflation, and interest rates to analyze the effects of monetary policy shocks?
30. Can you explain the relevance of the literature review section of the project in terms of understanding the economic theory and previous research on the topic?
31. How does the project use the data on interest rate development to contextualize the research question?
32. What are the main contributions of the project in terms of understanding the effects of monetary policy on the economy?
33. What are the limitations set in this project in terms of time period, variables, countries, and data sources?
34. How does the research question "What are the effects of a monetary policy shock, is it the same in United States and Denmark and are the shocks symmetric?" relate to the current economic situation?
35. Why were inflation, unemployment, and interest rate chosen as the variables for the analysis?
36. Why were Denmark and the United States chosen as the countries for comparison?
37. Why was OECD chosen as the data source for this project?
38. How will the Structural Vector Auto Regressive model with a Cholesky decomposition be used to investigate the research question?
39. What is a VAR(p) model and how is it represented mathematically?
40. What assumptions are made about the error term in a VAR(p) model?
41. How can deterministic trends be removed from the data generating process in a VAR(p) model?
42. What is the difference between a white noise process and a GARCH error?
43. How can the error term be assumed to be IID or a martingale difference sequence in a VAR(p) model?
44. What is the matrix polynomial in the lag operator and how is it used in a VAR(p) model?
45. How does the observed variable yt have the same VAR structure as xt in a VAR(p) model?
46. What is the difference between a basic VAR model and a structural VAR model?
47. What is Cholesky decomposition and how is it used in VAR model?

48. What are impulse response functions and how do they relate to the VAR model?
49. What is the structural vector auto regressive model, and how is it different from the basic VAR model?
50. How does the structural VAR model help in the analysis of impulse response functions?
51. What is the relationship between the reduced form and structural form of the VAR model?
52. What is the structural impact multiplier matrix and how is it used in the structural VAR model?
53. How is the structural VAR system identified, and what additional restrictions are required?
54. What are the key assumptions of the basic vector auto regressive (VAR) model?
55. How does the structural vector auto regressive (SVAR) model differ from the basic VAR model?
56. What is the Cholesky decomposition and why is it necessary for identifying structural VAR models?
57. How is the ordering of variables in the Cholesky decomposition determined and what is the reasoning behind it?
58. How does the Cholesky decomposition solve the problem of structural shocks in a VAR model?
59. What are the implications of failing to come up with convincing arguments for the recursive ordering in Cholesky decomposition?
60. How does the contemporaneous effects of inflation, unemployment and interest rate affects each other?
61. What is the purpose of impulse response functions?
62. How is the impulse response function computed in this project?
63. What is the difference between reduced form errors and structural errors?
64. What is the Cholesky decomposition and how is it used in this project?
65. What is the importance of imposing additional restrictions on B0 in the structural VAR model?
66. Why is it important to have a plausible economic interpretation of the ordering imposed by the Cholesky decomposition?
67. How is stationarity ensured in this project?
68. What are the assumptions that ensure stationarity in VAR model?
69. What are the possible ways to ensure that the data is stationary?
70. What test should be used to test the stationarity of the data?
71. What are the various tests used in section 3.4, and why are they important?
72. How is the number of lags determined in a VAR model, and what criteria is used for selecting the best model?
73. What is seasonality in macroeconomic variables, and why is it a problem for a VAR model?
74. How is seasonality tested for in the model, and what method is used to determine significance?
75. How does the Dickey-Fuller test perform when there is a constant mean or a linear trend in the data?
76. What is the null hypothesis of a unit root test, and how is it formulated in the Dickey-Fuller regression?
77. What is the Augmented Dickey-Fuller regression, and what is the t-statistic used for in this test?
78. How does the DF-GLS test differ from the Dickey-Fuller test, and why is it considered to have a higher power envelope?
79. What is the Frisch-Waugh-Lovell theorem, and how is it used in the ADF-GLS regression?
80. How are structural breaks found in the data, and what method is used for estimating breakpoints?
81. What is the Portmanteau Test for Residual Autocorrelation, and what does it test for in the model?

82. What is the null hypothesis of the Portmanteau test, and what is the alternative hypothesis?
83. What is the Jarque-Bera test, and what is it used for in this project?
84. How is the JB statistic calculated, and what does it indicate about the distribution of the residuals in the VAR model?
85. What is the difference between the true skewness and kurtosis and the estimated skewness and kurtosis?
86. What test is used to check for heteroskedasticity in the data?
87. What is the null hypothesis of the Breusch-Pagan test, and what is the alternative hypothesis?
88. How is the Breusch-Pagan test conducted, and what is the dependent variable in the test?
89. What were the results of the seasonality, structural breaks, unit root, serial correlation, normality and heteroskedasticity tests for the data?
90. How were the issues with structural breaks and unit root solved, and why is it important for the robustness of the models?
91. What was the overall conclusion on the test results, and how will it impact the interpretation of the research findings?
92. What are the main findings of the empirical results presented in this section?
93. How does the impulse response function for the United States economy relate to the Phillips Curve and Mishkin's transmission mechanism?
94. Why is the impulse response for the United States economy not statistically significant until period 16?
95. Explain the "price puzzle" and how it relates to the impulse response function for Denmark.
96. What is the potential impact of a wage-price-spiral on the Danish economy?
97. How does the impulse response function for Denmark differ from that of the United States?
98. In general, what are the implications of an increase in the interest rate on inflation in the United States and Denmark?
99. What is the main difference in the transmission mechanism of the interest rate shock in Denmark and the United States?
100. How does interest rate increase impact on the economic activity in the United States?
101. How does the interest rate increase impact on the inflation rate in Denmark? How does an increase in interest rate affect the level of unemployment in the United States economy?
102. What is the mechanism behind the convergence of the unemployment rate back to the original level in the United States?
103. How does the impulse response function for the Danish economy differ from that of the United States in terms of the effect of an increase in interest rate on unemployment?
104. Why is the monetary policy more effective in the United States than in Denmark?
105. What is the effect of an increase in interest rate on the interest rate in the United States?
106. How does the impulse response function for the Danish economy differ from that of the United States in terms of the effect of an increase in interest rate on interest rate?
107. How does the Cholesky decomposition relate to the impulse response function for the interest rate shock?
108. What is the implication of the contemporaneous effect on inflation and unemployment when conducting a shock to the interest rate?
109. How does automatic stabilizers in Denmark affect the response to interest rate shock?
110. Why the interest rate shock is of a larger magnitude in Denmark than in the United States?
111. What is the relationship between economic activity and inflation according to Phillips Curve?

112. What does Mishkin's transmission mechanism state about the relationship between interest rate and economic activity?

113. What is the price puzzle in relation to interest rate and inflation?

114. Why is the impulse response function for the United States economy not statistically significant until period 16?

115. Why do we observe a short-term increase in inflation in Denmark when there is an increase in interest rate?

116. How does the wage-price-spiral affect inflation in Denmark?

117. What is the difference in the impulse response functions for unemployment in Denmark and the United States?

118. What is the Cholesky decomposition in relation to interest rate shocks in Denmark and the United States?

119. Are the residuals from the error terms of the interest rate in Denmark and the United States normally distributed?

120. Why is the monetary policy in the United States more effective than in Denmark?

121. Why is the interest rate shock in Denmark of a larger magnitude?

122. What is the difference in distribution of the error terms from the interest rate in Denmark and the United States?

123. Why is there a tendency for the interest rate to be decreased a lot when needed and a moderate raise when needed in both Denmark and the United States?

124. What does the skewness and kurtosis of residuals from VAR models indicate about the symmetry of interest rate shocks in Denmark and the United States?

125. Can you explain why the impulse response function for inflation in Denmark does not align with general economic theory?

126. Can you discuss the potential implications of the finding that an increase in interest rates does not lead to a decrease in inflation in Denmark?

127. Can you explain the "price puzzle" mentioned in the text and its relationship to inflation in Denmark?

128. Can you discuss the difference in the movement of unemployment rates between Denmark and the United States, and what may explain these differences?

129. Can you explain why the authors chose to use a structural VAR model rather than a cointegrated structural VAR model?

130. Can you discuss the impact of reducing the sample size on the results of the model?

131. Can you compare and contrast the findings of this study with those of Stock and Watson, and discuss any differences and potential explanations for those differences?

132. Can you explain the difference in monetary policy regimes between Denmark and the United States, and how it affects the effectiveness of fiscal and monetary policy in stabilizing the economy?

133. Can you discuss the potential impact of asymmetry within the European Union on Denmark's monetary policy and the observed "price puzzle" in the impulse response functions?

134. Can you explain how the Mundell-Fleming model is used to understand the effectiveness of fiscal and monetary policy in the different monetary policy regimes of Denmark and the United States?

135. Can you discuss the impact of automatic stabilizers, such as progressive taxation and daily allowance, on the movement of unemployment in Denmark compared to the United States?

136. Can you explain how the difference in monetary policy regimes between Denmark and the United States affects the effectiveness of fiscal policy in stabilizing the economy?

137. Can you discuss the potential impact of the central bank having to intervene to keep the currency at the wished level on the data, and how it might be able to explain the price puzzle and general movement in inflation in Denmark?

138. Can you explain how the difference in automatic stabilizers between Denmark and the United States affect the movement of unemployment in the two countries?

139. What is the benefit of using a cointegrated structural VAR model, as opposed to a structural VAR model, in this analysis?

140. How does using first differences in the analysis affect the researcher's ability to discuss the magnitude of a shock in terms of the original level?

141. How does using first differences affect the researcher's ability to analyze long-run effects?

142. Why was the cointegrated structural VAR model not chosen for this analysis?

143. Can you explain the problem of serial correlation in the model, and how it was addressed in the analysis?

144. Why did the team choose to reduce the sample size from 1970-2018 to 1985-2018?

145. What are the consequences of reducing the sample size on the analysis?

146. How did the team try to solve the problem of serial correlation?

147. Why were the attempts to include specific lags and changing the lag order not successful in solving the problem of serial correlation?

148. How do the findings in the article by Stock and Watson compare to the findings in the current study?

149. What are the main differences between the time frame of the sample size used in Stock and Watson's study and the current study?

150. How do the responses of inflation to a shock in the interest rate in Stock and Watson's study compare to the responses in the current study?

151. How does the movement of unemployment in Stock and Watson's study compare to the movement in the current study?

152. How does the movement of the interest rate in Stock and Watson's study compare to the movement in the current study?

153. Why does the current study not observe a convergence back to the original level for the movement of unemployment and interest rate, unlike the study by Stock and Watson?

154. How does the approach of first differencing the data affect the ability to observe long run information?

155. Why the study doesn't use Cointegrated structural VAR model?

156. What was the problem with serial correlation and how they fixed it?

157. What is a martingale difference sequence? (Answer: A martingale difference sequence is a sequence of random variables that are used in mathematical finance and probability theory to model and analyze the behavior of certain types of financial systems. In simple terms, a martingale difference sequence is a sequence of random variables (X_n) such that the expected value of X_{n+1} given the past values of $X_1, X_2, ..., X_n$ is equal to X_n. This property is known as the "martingale property.")

158. When did USA change from target money supply to target inflation rate? (Answer: early 1980's)

159. What does it mean that the structural errors from structural VAR model doesn't lose generality? (answer: When it is said that the structural errors from a structural vector autoregression (VAR) model do not lose generality, it means that the model can still be used to make accurate predictions or draw valid conclusions about the relationships between variables in the system, even if there are errors in the model's structure. This is in contrast to reduced-form VAR models, where errors in the structure of the model can lead to biased or inconsistent parameter estimates.)

CHAPTER 12

Bibliography