Financial Econometrics 1 - M2 FTD

Empirical Applications

Luis Miguel Fonseca
Stéphane Eloundou Mvondo
Natalia Cárdenas Frías

December 12, 2023

Contents

Introduction **3**

1 Series Dynamics **3**
- 1.1 Seasonality 4
- 1.2 Unit root and trends 4
  - 1.2.1 ADF - Test jointly for deterministic and stochastic trend (with drift) 6
  - 1.2.2 ADF - Test jointly for stochastic trend and drift 8
  - 1.2.3 ADF - Test for stochastic trend only 9
- 1.3 Check stationarity of the series in deltas if UR in levels 9
- 1.4 Cyclical component 10

2 Canonical VAR model application **10**

3 Cointegration theory **12**

4 Impulse Response Analysis **12**
- 4.1 Canonical IRF 12
- 4.2 Structural IRF 12

5 Introduce non-linearities **12**
- 5.1 Markov-switching model 12
- 5.2 STR model 12

6 Difference-in-Difference **12**

Appendix A Code - Data Cleaning **13**

Introduction

something, probably describe how all applications make sense one after the other and what is the research question we could have made ourselves when doing the applications, try to give a coherent look to the whole thing.

This document compiles all our applications for the Financial Econometrics course. Each section represents a specific application, but we tried to make them coherent across them around a broad question:

1 Series Dynamics

Note: Depending on each exercise along these applications we might use different series. In this first section, we performed the stationarity and component analysis of all of them to be able to use them rapidly without having to worry about seasonality or the presence of UR. Therefore, this section encompasses more than the 3 series that were asked in the exercise.

In this work, we focus on the US market. We use the following series retreived for the most part from FRED with its Python API (FRED tickers are in square brakets) :

Inflation Expectation [MICH] This data series is made public by the University of Michigan from their Survey of Consumers. The series represent the median expected value of the percent change in prices over teh next year. The series is not seasonally adjusted.

GDP deflator [A191RI1Q225SBEA] As a measure of inflation, we decided to use the implicit price deflator of teh US GDP. Unlike measures like the CPI deflators do not consider baskets of goods and therefore are broader measures of the price changes across all the economy that measure teh ratio of the GDP in value and in volume. It is a measure produced by the US Bureau of Economic Analysis as a quaterly measure in percent change QoQ. The raw series is already seasonally adjusted at an annual rate.

Unemployment rate [UNRATENSA] It represents the share (in percent) of unemployed people over the labor force [1]. The source is the 'Current Population Survey (Household Survey)' of the US Bureau of Labor Statistics

FED func rate

S&P 500 price

[1] The labor force data in this context encompasses people older than 16, living in continental US, who don't reside in institutions like prisons or homes for the aged, and who are not in active duty in the military.

Corporate Debt Returns [BAMLCC0A0CMTRIV] The series represents the total returns of the debt index ICE Bank of America US Corporate Index value. This capitalization-weighted index tracks the performance if teh following debt instruments[2]:

- Of corporate debt issued in USD in the US domestic market with an investment grade rating[3]. It must have a remaining maturity of at least one year, a fixed-coupon and an outstanding amount of at least 250 million dollars.
- Of original zero-coupon bonds ussued simultaneously in the US and the Eurobond market.
- Of 144a securities and pay-in-kind securities, including toggle notes, that can only be traded bu qualified institutional buyers with large portfolios.
- Of callable perpetual securities if they are at least one year from the first call date.
- Of fixed-to-floating rate securities if they are callable within the fixed rate period and are at least one year from the last call prior to the date the bond transitions from a fixed to a floating rate security.

The data's source is ICE Data Indices, and it is a daily series with close prices. It is not seasonally adjusted.

All (monthly[4]) time series de can be decomposed in the following elements:

$$X_t = \underbrace{\alpha}_{\text{drift}} + \underbrace{\beta \times t}_{\text{deterministic trend}} + \underbrace{\gamma T t}_{\text{stochastic trend}} + \underbrace{\Sigma_{i=1}^{11} \rho_i \, \mathbb{1}_{m_i}}_{\text{seasonality}} + \underbrace{c_i}_{\text{cyclical component}}$$

Remark that according to definitions, the deterministic might as well include the drift/constants α. Also, distinguishing the presence of a deterministic and stochastic trend needs to be done jointly and with specific procedures (see Section 1.1 for the implementation of the ADF unit root tests).

Using R's built-in function `stl`[5], we can decompose the time series. The results are plotted in Figure 1.

1.1 Seasonality

From the decompositions made in the previous section, in Figure 1 we can see that R picks up small seasonal variations in each series (the values on the y-axis are rather very small). The estimations of the seasonal coefficients are reported in Table 1, which indeed shows very small coefficients. We nonetheless decided to work with deseasonalized series even if the changes are minimal.

1.2 Unit root and trends

As for any time series analysis, the first analysis to perform is regarding the presence of unit roots in the series that would make them non-stationary. To do so, we perform the Augmented Dickey-Fuller tests that evaluate

[2] Taken from the FRED's website

[3] To be considered to have an investment grade ranking to enter the index, the company and the country must be considered investment grade by the average of the ratings of Fitch, S&P and Moody's.

[4] One should adapt the number of indicators in the seasonal components according to the data frequency.

[5] Seasonal Decomposition of Time Series by Loess

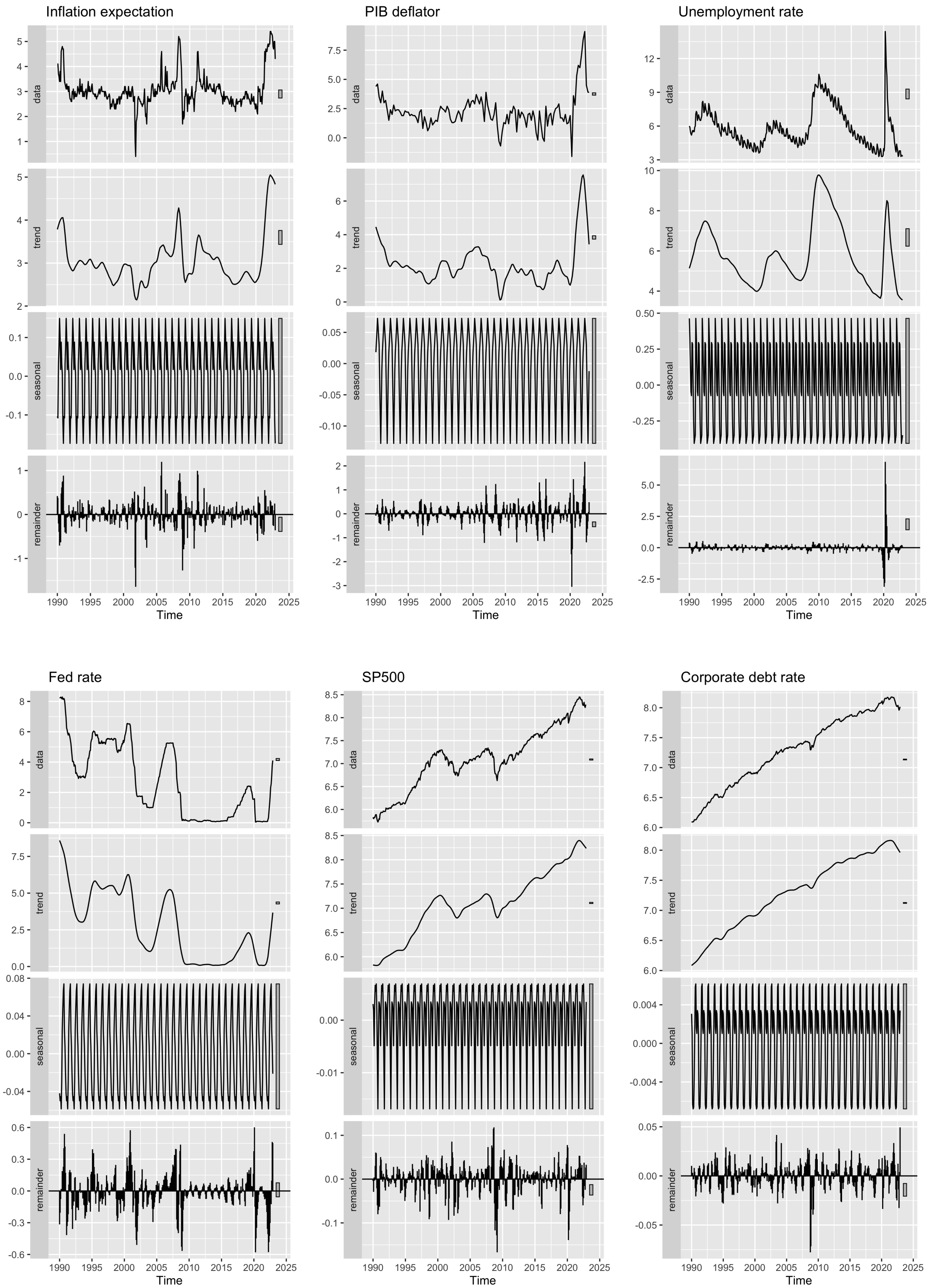

Figure 1: Time series decomposition

Table 1: Estimation of the seasonality of each series

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
infl_e	-0.104	-0.108	0.031	0.078	0.149	0.1	0.017	0.088	0.031	0.006	-0.116	-0.17
deflator	0.018	0.035	0.053	0.072	0.06	0.045	0.03	-0.024	-0.078	-0.128	-0.072	-0.01
unempl	0.464	0.327	0.132	-0.045	-0.074	0.296	0.284	0.007	-0.257	-0.406	-0.379	-0.34
rate	-0.042	-0.05	-0.046	-0.059	-0.03	0.025	0.048	0.067	0.074	0.025	0.008	-0.02
splong	0.003	0.002	-0.005	0.003	0.007	0.006	0.007	0.002	-0.007	-0.017	-0.003	0.003
corp_debt	0.003	0.001	-0.006	-0.007	-0.007	-0.004	0.001	0.006	0.006	0.002	0.001	0.003

the presence of a stochastic trend (a unit root), a deterministic trend, and an intercept or drift. Importantly, this test requires estimating three equations/specifications because it requires investigating the joint presence of both types of trends and drift, for them to discard elements one by one. For all specifications, the main null hypothesis H0 is that the series exhibits a UR. The inference with this test is non-standard and requires to use of corrected critical values to assess significance with the t-statistics.

1.2.1 ADF - Test jointly for deterministic and stochastic trend (with drift)

We first run the following specification to the ADF test to *jointly* investigate the presence of a stochastic and a determinist trend for each series $(X_t)_t$:

$$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \sum_{i=1,2,..} \rho_i \Delta X_{t-i} + \varepsilon_t \tag{1}$$

As per usual, the ADF test assumes H0: $\gamma = 0$ i.e. a unit root exists and the series is non-stationary. We use R's built-in function `ur.df` with `type='trend'` to get this estimation. This function gives us (i) a regression table per series and (ii) a summary table with the following test statistics:

- tau3 refers to the t-statistic associated to H0: $\gamma = 0$ i.e the presence of a UR. H1 refers to the absence of said UR.
- phi2 refers to the F-statistic associated to H0: $\alpha = \beta = \gamma = 0$ [6]
- phi3 is also an F-statistic, now associated to H0: $\beta = \gamma = 0$

Remark that the critical value in both tables can be a little different. This is because they are sensitive to the number of observations in each series. In Table 2, the critical values correspond to those provided directly by R and are associated with $N = 500$, while in Table 3 we give the values for $N = 250$. Since we have 396 data points per series we prefer to refer to the higher critical values but it does not change the analysis done.

Let us examine each series' results, summarized in the following tables:

Inflation expectation We find $t_\gamma = -4.490 < -3.43$ we reject H0 ie we can't say that the series has a UR. Note that the F-statistic corresponding to the nullity of all the coefficients of interes (phi2) leads us to reject H0:

[6] As with all F test, the alternative hypothesis is that at least one of these coefficients in non-null. Since this is general, we don't explicitly signal the H1 hereafter.

Table 2: ADF test - 1st regression with drift, deterministic trend and stochastic trend

	gdp	dpi	infl_e	deflator	rate	splong	corp_debt	CV 1pct	CV 5pct	CV 10pct
tau3	-2.150	-2.694	-4.650	-5.197	-1.709	-1.975	-1.867	-3.980	-3.420	-3.130
phi2	14.496	7.880	7.375	9.097	2.015	3.875	10.326	6.150	4.710	4.050
phi3	2.388	3.849	11.061	13.644	2.917	1.997	4.247	8.340	6.300	5.360

Table 3: ADF test - 1st regression t statistics

	gdp	dpi	infl_e	deflator	rate	splong	corp_debt
alpha	2.197	2.714	4.004	2.593	0.698	2.101	1.987
gamma	-2.150	-2.694	-4.650	-5.197	-1.709	-1.975	-1.867
beta	2.094	2.577	1.265	1.166	-0.181	1.726	1.524
rho	16.326	-11.089	-0.255	14.255	15.719	4.099	6.824

Notes: With N=396, critical values at 5%: alpha = 3.09 ; gamma= -3.43 ; beta = 2.79

$\alpha = \beta = \gamma = 0$ (same for phi3), leading us to believe that the series has either a drift and/or a deterministic trend. We therefore compare the t-statistics associated with α and β to the standard interest threshold (the critical values below Table 3 are conditional on having a UR). Since $|t_\alpha| > 1.96$ and $|t_\beta| > 1.96$, we fail to reject the absence of a drift and of a deterministic trend. We conclude that the series is *stationnary with a constant and a deterministic trend*.

GDP deflator With $t_\gamma = -5.166 < -3.43$, as before we can reject H0 indicating that the series is stationary in levels. Since $|t_\alpha| = 2.584 > 1.96$, we reject the nullity of the drift. Finally, since $|t_\beta| = 1.154 < 1.96$ we cannot reject the absence of a deterministic trend. We conclude that the series is *stationary with a constant*

Unemployment rate Since $t_\gamma = -3.346 > -3.43$, we fail to reject H0. We then test the significance of β with respect to the non-standard threshold 2.79 (bilateral test). Once again, the values of the F-statiss of the other two tests phi2 and phi3, lead us to believe that the series has a constant and/or a deterministic trend (we fail to reject both phi2 and phi3). The ADF methodology leads us to use the second specification for this series and discard the existence of a deterministic trend.

Fed fund rate $t_\gamma = -1.672 > -3.43$, we are in the same situation as the previous series where we cannot reject the existence of a UR. Because we cannot reject phi2 nor phi3 ($F_{phi2} = 2.022 < 4.71,\ F_{phi3} = 2.927 < 6.3$) we need to continue testing this series with the other specifications as we don't reject the existence of a stochastic trend and we cannnot reject the nulity of the trend coefficient ($|t_\beta| = 0.127 < 2.79$).

S&P500 Without many surprises for price series, $t_\gamma = -1.949 > -3.43$ and we cannot reject the existence of a UR. Moreover, $F_{phi2} = 3.926 < 4.71$ and $F_{phi3} = 1.953 < 6.3$ (non-rejection of the null for both tests) leads to conclude that at least one of these coefficients are non-null (note that $t_\beta = 1.696 < 2.79$ and thus we cannot reject the nullity of β). We continue testing this series with the second specification of the test.

Corporate debt Finally, this series has $t_\gamma = -2.588 > -3.43$, $F_{phi2} = 5.6 > 4.71$ and $F_{phi3} = 3.351 < 6.3$. While we fail to reject the presence of a UR and that both γ and β are jointly null (β is not significant as $|t_\beta| = 2.560 < 2.79$), we get that at least one coefficient is significantly different from zero, hinting to the constant. We need to continue to test this series too with the second specification of the ADF test.

1.2.2 ADF - Test jointly for stochastic trend and drift

The second specification of the test models $\forall (X_t)_t$:

$$\Delta X_t = \alpha + \gamma X_{t-1} + \sum_{i=1,2,..} \rho_i \Delta X_{t-i} + \varepsilon_t \tag{2}$$

The null hypothesis still refers to H0: $\gamma = 0$ the presence of a unit root. We use now `type='trend'` in the `ur.df` function to get this estimation. The output of the test is similar to the previous specification and the same remarks on the critical values apply here. Now the test statistics reported refer to:

- tau2 refers to the t-statistic associated to $\gamma = 0$
- phi1 refers to the F-statistic associated to $\alpha = \gamma = 0$

Table 4: ADF test - 2nd regression with drift and stochastic trend

	gdp	dpi	rate	splong	corp_debt	CV 1pct	CV 5pct	CV 10pct
tau2	-0.621	-1.020	-2.412	-1.005	-2.480	-3.440	-2.870	-2.570
phi1	19.382	8.378	3.014	4.300	14.279	6.470	4.610	3.790

Table 5: ADF test - 2nd regression t statistics

	gdp	dpi	rate	splong	corp_debt
alpha	0.983	1.127	1.508	1.258	2.864
gamma	-0.621	-1.020	-2.412	-1.005	-2.480
rho	16.197	-12.106	15.984	3.977	6.650

Notes: With N=396, critical values at 5%: alpha = 2.53 ; gamma= -2.88

Unemployment rate With $t_\gamma = -3.334 < -2.88$ we reject H0: the series has no unit root. We then chek the F-statistic associated to $H0 : \gamma = \alpha = 0$, i.e. the test *phi1* in the default R test. Given that $F_{phi1} = 5.57 > 4.61$, we reject the null suggesting that the series has a constant which is also supported by the significance of α, $|t_\alpha| = 3.153 > 1.96$. Since γ is still significantly different from zero using the standard threshold -1.64, we finally conclude that the *series is stationary with a constant* (*no deterministic trend*) at 5%.

Fed fund rate Given that $t_\gamma = -2.419 > -2.88$, we cannot reject the null hypothesis. We then check the F-statistic of the joint test *phi1*: $F_{phi1} = 3.032 < 4.61$: we cannot reject the null suggesting that the series has a UR and no drift. Supporting this, we also find that the drift term is not significantly different from zero as $|t_\alpha| = 1.513 < 2.53$. This leads us to use the third specification of the test.

S&P500 Since $t_\gamma = -1.013 > -2.87$, we cannot reject H0. By checking $F_{phi1} = 4.431 < 4.61$ and $|t_\alpha| = 1.270 < 2.53$, we fall in the same case as before where we need to continue testing the series as it seems to have a UR and no drift

Corporate debt For the last series, we find $t_\gamma = -0.383 > -2.87$, leading to a non-rejection of the null hypothesis. Once again we check the other two dtstistics: since $F_{phi1} = 5.051 > 4.61$ we reject the joint nullity of γ and α. We then test the significance of the drift coefficient: $|t_\alpha| = 1.824 < 2.53$ and that γ is not significantly different fom zero. As with the previous series, this result requires us to keep testing the series in the simplest specification of the ADF test.

1.2.3 ADF - Test for stochastic trend only

The last specification of the test keeps only the stochastic trend, $\forall (X_t)_t$:

$$\Delta X_t = \gamma X_{t-1} + \sum_{i=1,2,..} \rho_i \Delta X_{t-i} + \varepsilon_t \tag{3}$$

The null hypothesis still refers to H0: $\gamma = 0$ the presence of a unit root and we use `type='none'`. The output of the test is similar to the previous ones but now there is only one test statistic reported referring to the null (tau1). For this step, we only report the table with the t-statistics as its value for the "gamma" row is identical to the test statistics of tau1.

Table 6: ADF test - 3rd regression t statistics

	gdp	dpi	rate	splong
gamma	6.148	3.934	-1.935	2.647
rho	16.306	-12.122	15.950	3.976

Notes: With N=396, critical values at 5%: gamma= -1.95

Fed fund rate We find $t_\gamma = -1.94 > -1.95$ thus we cannot reject the existence of a UR (this is a close call but seems adequate since we never rejected the UR and R's built-in `order_integration` also indicates a UR). We conclude that the *series has a unit root with no constant nor time trend*.

S&P500 Similarly, we find $t_\gamma = 2.690 > -1.95$ and we cannot reject H0 and conclude that the *series has a unit root with no constant nor time trend*.

Corporate debt Finally, given that $t_\gamma = 2.592 > -1.95$ do not reject H0 and conclude that the *series has a unit root with no constant nor time trend*.

1.3 Check stationarity of the series in deltas if UR in levels

To conclude that the previous three series are indeed I(1), we need to check that the series of their first differences are stationary (i.e. the series in deltas). We perform the same ADF procedure to test these transformed

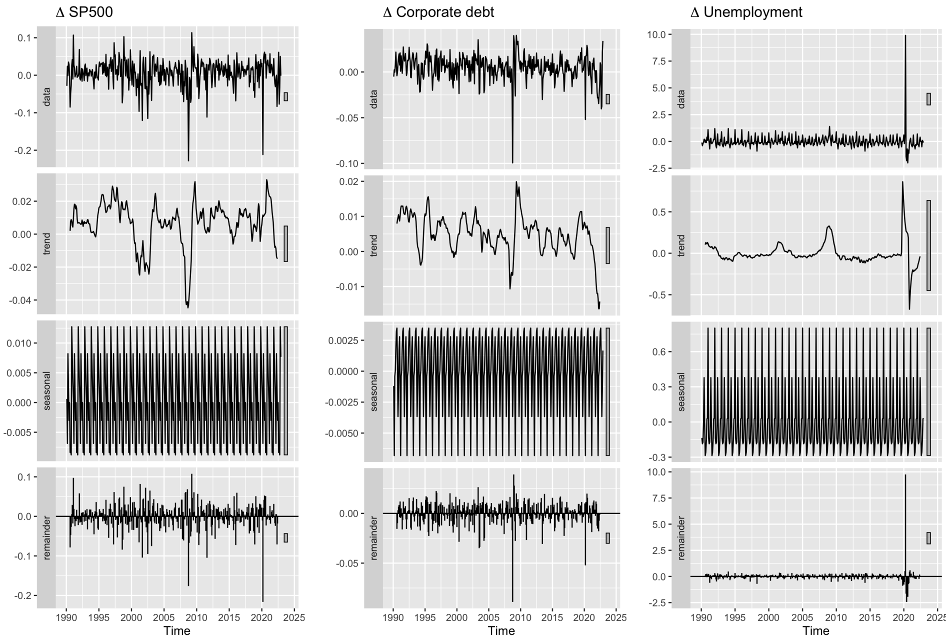

Figure 2: Decomposition of the series in deltas

series.

Table 7: ADF test - 1st regression with drift, deterministic trend and stochastic trend for series in deltas

	d_gdp	d_dpi	d_rate	d_splong	CV 1pct	CV 5pct	CV 10pct
tau3	-10.437	-19.070	-7.107	-13.317	-3.980	-3.420	-3.130
phi2	36.317	121.228	16.895	59.110	6.150	4.710	4.050
phi3	54.473	181.842	25.327	88.665	8.340	6.300	5.360

Table 7 reports the test results in the first specification of the ADF test. We easily see that all the t_γ (the statistic on tau3) are sufficiently negative to reject H0 and conclude that none of the differentiated series has a UR. This is sufficient to conclude that the series on the Fed fund rate, on the returns of the S&P500, and on the return of corporate debt are indeed *integrated of order one*.

1.4 Cyclical component

2 Canonical VAR model application

Table 8: Canonical VAR in levels - Identify order

	1	2	3	4	5	6	7	8	9	10
AIC(n)	-10.01	-10.56	-10.59	-10.73	-10.88	-10.87	-10.93	-11.01	-11.00	-10.98
HQ(n)	-9.96	-10.48	-10.47	-10.57	-10.68	-10.64	-10.66	-10.71	-10.66	-10.60
SC(n)	-9.89	-10.35	-10.29	-10.33	-10.38	-10.29	-10.25	-10.25	-10.14	-10.03
FPE(n)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 9: Level VAR - Estimation

	Dependent variable:		
	deflator	unempl	splong
deflator.l1	1.830***	−0.199	0.026***
	(0.051)	(0.124)	(0.008)
unempl.l1	0.120***	0.956***	0.003
	(0.022)	(0.053)	(0.004)
splong.l1	0.556*	−4.968***	1.174***
	(0.321)	(0.786)	(0.053)
deflator.l2	−0.819***	−0.317	−0.042**
	(0.101)	(0.246)	(0.017)
unempl.l2	−0.117***	−0.131*	−0.004
	(0.031)	(0.075)	(0.005)
splong.l2	−0.553	5.816***	−0.246***
	(0.506)	(1.237)	(0.083)
deflator.l3	−0.894***	1.008***	0.010
	(0.108)	(0.264)	(0.018)
unempl.l3	0.014	0.182**	0.002
	(0.031)	(0.076)	(0.005)
splong.l3	0.149	−0.992	0.137
	(0.521)	(1.275)	(0.085)
deflator.l4	1.526***	−0.716**	0.019
	(0.113)	(0.277)	(0.019)
unempl.l4	−0.013	−0.150**	0.003
	(0.031)	(0.076)	(0.005)
splong.l4	0.428	−0.516	−0.012
	(0.523)	(1.278)	(0.086)
deflator.l5	−0.653***	−0.130	−0.021
	(0.111)	(0.271)	(0.018)
unempl.l5	0.001	0.092	−0.008*
	(0.031)	(0.076)	(0.005)
splong.l5	−0.027	1.417	0.063
	(0.527)	(1.288)	(0.086)
deflator.l6	−0.498***	0.609**	−0.003
	(0.107)	(0.262)	(0.018)
unempl.l6	0.004	−0.105	0.006
	(0.031)	(0.075)	(0.005)
splong.l6	−1.048**	−1.069	−0.302***
	(0.527)	(1.290)	(0.086)
deflator.l7	0.834***	−0.476*	0.019
	(0.103)	(0.252)	(0.017)
unempl.l7	0.081***	0.088	−0.0004
	(0.030)	(0.073)	(0.005)
splong.l7	0.954*	−0.455	0.261***
	(0.526)	(1.287)	(0.086)
deflator.l8	−0.350***	0.153	−0.010
	(0.054)	(0.131)	(0.009)
unempl.l8	−0.075***	0.039	−0.001
	(0.021)	(0.052)	(0.004)
splong.l8	−0.428	0.746	−0.078
	(0.340)	(0.832)	(0.056)
const	−0.268*	0.490	0.015
	(0.149)	(0.365)	(0.024)
Adjusted R^2	0.977	0.913	0.997
Residual Std. Error (df = 363)	0.210	0.515	0.034
F Statistic (df = 24; 363)	684.697***	170.695***	5,592.560***
Note:			*p<0.1; **p<0.05; ***p<0.01

3 Cointegration theory

4 Impulse Response Analysis

4.1 Canonical IRF

4.2 Structural IRF

5 Introduce non-linearities

5.1 Markov-switching model

5.2 STR model

6 Difference-in-Difference

https://www.tidy-finance.org/r/difference-in-differences.html

Appendix A Code - Data Cleaning

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Financial Econometrics - Empirical Applications d
Data Gathering and Data Cleaning

@author: nataliacardenasf
"""

import pandas as pd
import os
import datetime

from fredapi import Fred

os.chdir('/Users/nataliacardenasf/Documents/GitHub/PROJECTS_AP_FE/FinancialEconometrics1')

### Initialize FRED API
fred = Fred(api_key='23edc2b1b61e17c07b83a97e7abfc02b')

### Import all the data

# S&P 500
sp500 = pd.DataFrame(fred.get_series('SP500')) #daily close, NSA, Index
sp500.columns= ['sp500']

#Inflation expectations from survey UMich
infl_e = pd.DataFrame(fred.get_series('MICH')) #monthly, NSA, median expected in % over next
    12 mo
infl_e.columns= ['infl_e']


#ICE BofA US Corporate Index Total Return Index
corp_debt = pd.DataFrame(fred.get_series('BAMLCC0A0CMTRIV')) #daily, close, NSA, Index
corp_debt.columns= ['corp_debt']


#MP rate
rate = pd.DataFrame(fred.get_series('DFF')) #daily, 7-Day, NSA, %
rate.columns = ['rate']

#Deflator
deflator = pd.DataFrame(fred.get_series('A191RI1Q225SBEA')) #Q, SA Annual Rate
deflator.columns = ['deflator']

#Unemployment
```

```
unempl = pd.DataFrame(fred.get_series('UNRATENSA')) #monthly, NSA, %
unempl.columns=['unempl']

# GDP
gdp = pd.DataFrame(fred.get_series('GDP'))  # quarterly, Billions of Dollars, SA Annual Rate
gdp.columns = ['gdp']

# RPI (Real Personal Income)
rpi = pd.DataFrame(fred.get_series('RPI'))  # monthly, SA rate, deflated
rpi.columns = ['rpi']

# Real Personal Disposable Income
dpi = pd.DataFrame(fred.get_series('DSPIC96'))  # monthly, SA annaul rate, chained 2017 USD
dpi.columns = ['dpi']

# Manufacturing Sector
manufacturing = pd.DataFrame(fred.get_series('MPU9900063'))  # annual, NSA => avoid this
manufacturing.columns = ['manufacturing']

fred.search("MPU9900063").T #this function gives of the info on every series


### Resample into monthly data
sp500 = sp500.resample('1M').mean(numeric_only=True)
infl_e = infl_e.resample('1M').mean(numeric_only=True)
corp_debt = corp_debt.resample('1M').mean(numeric_only=True)
rate = rate.resample('1M').mean(numeric_only=True)
deflator = deflator.resample('1M').mean(numeric_only=True)
unempl = unempl.resample('1M').mean(numeric_only=True)
gdp = gdp.resample('1M').mean(numeric_only=True)
rpi = rpi.resample('1M').mean(numeric_only=True)
dpi = dpi.resample('1M').mean(numeric_only=True)
manufacturing = manufacturing.resample('1M').mean(numeric_only=True)

dta = [infl_e, rate, sp500, corp_debt, deflator, unempl, gdp, rpi, dpi, manufacturing]

### Slice the df to relevant period
#Find common time span
min_date = max([min(i.index) for i in dta])
max_date = min([max(i.index) for i in dta])
print(min_date, max_date)

#Let us work on monthly data for the 1990-2022 period
start = datetime.datetime(1990,1,1)
end= datetime.datetime(2022,12,31)

## SP500 series is too short, I am taking it from Yahoo Finance
import yfinance as yf
```

```python
splong = yf.download('^GSPC', start=start,end=end)['Adj Close'].resample('M').mean(numeric_only=True)
splong = pd.DataFrame(splong)
type(splong)
splong.rename(columns={"Adj Close":'splong'}, inplace=True)

##Get a single DF
dta.append(splong)
for i in range(len(dta)): #we had some indexes at end of month, others at 1st of month: harmonize to 1st each month
    df = dta[i]
    df.index = [pd.datetime(x.year, x.month, 1) for x in df.index.tolist()]
    dta[i] = df.loc[start:end,:]
dta# we're good now
#merge into 1 df, 1 series per column
monthly = pd.concat(dta, axis=1)
#interpolate missing months for deflatior data (Q): uses midpoints ie assumes that each month in the quarter contributes in the same fashion to the increase QoQ
m1 = monthly.interpolate(method ='linear', limit_direction ='forward')


m1.to_csv("DATA/data.csv")
```