Analysis of single time series: Part II

Daniele Bianchi[1]

whitesphd.com

[1]School of Economics and Finance
Queen Mary, University of London

Summary

This week are going to cover forecasting and non-stationarity in time series as well as non-linear models and filtering. All these aspects are fundamental in financial markets since:

- Forecasting is of primary importance for risk and asset management
- Non-stationarity is particularly important in particular for macroeconomic time series.
- Non-linearity matters
- Filters allow to capture dynamic features in financial markets.

Contents

Forecasting

Loss function

Definition (Loss Function)

A loss function is a function of the observed data y_{t+k} and the time-t constructed $\hat{y}_{t+k|k}$, $L\left(y_t, \hat{y}_{t+k|k}\right)$, that has the three following properties:

Property 1: The loss of any forecast is non-negative, so $L\left(y_t, \hat{y}_{t+k|k}\right) \geq 0$

Property 2: There exists a point, y^*_{t+k}, known as the optimal forecast, where the loss function takes the value 0. That is $L\left(y_t, y^*_{t+k}\right) = 0$

Property 3: The loss is non-decreasing away from y^*_{t+k}. That is, forecasts that are more distance from the optimal prediction have increasing values of the loss function.

Mean Square Error

The most common loss function is the Mean Square Error (MSE) which chooses the forecast to minimize

$$E\left[L\left(y_{t+k}, \hat{y}_{t+k|t}\right)\right] = E\left[\left(y_{t+k} - \hat{y}_{t+k|t}\right)^2\right],$$

where $\hat{y}_{t+k|t}$ is the time-t forecast of y_{t+k}.

Notice that this is just the optimal projection problem and the optimal forecast is the conditional mean, $y^*_{t+k|t} = E\left[y_{t+k}\right]$.

It is simple to verify that this loss function satisfies the properties of a loss function, i.e,

- Property 1 holds by inspection.
- Property 2 occurs when $y_{t+k} = \hat{y}^*_{t+k|t}$.
- Property 3 follows from the quadratic form of the MSE.

Forecasting from ARMA models

Fortunately, forecasting from ARMA models is an easy exercise. For instance, let us consider the AR(1) process,

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t,$$

Since the optimal forecast is the conditional mean, all that is needed is to compute $E_t\left[y_{t+k}\right]$ for any k. When $k = 1$,

$$y_{t+1} = \phi_0 + \phi_1 y_t + \epsilon_{t+1},$$

so the conditional expectation is

$$\begin{aligned} E_t\left[y_{t+1}\right] &= E_t\left[\phi_0 + \phi_1 y_t + \epsilon_{t+1}\right], \\ &= \phi_0 + E_t\left[y_t\right] + E_t\left[\epsilon_{t+1}\right], \\ &= \phi_0 + \phi_1 y_t + 0, \\ &= \phi_0 + \phi_1 y_t \end{aligned}$$

which follows since y_t is the time-t information set $\mathcal{F}_t$ and $E_t\left[\epsilon_{t+1}\right] = 0$ by assumption.

Forecasting from ARMA models

Next, consider forecasts from an MA(2)

$$y_t = \phi_0 + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_t,$$

The one-step ahead forecast is given by

$$\begin{aligned} E_t\left[y_{t+1}\right] &= E_t\left[\phi_0 + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_{t+1}\right], \\ &= \phi_0 + \theta_1 E_t\left[\epsilon_t\right] + \theta_2 E_t\left[\epsilon_{t-1}\right] + E_t\left[\epsilon_{t+1}\right], \\ &= \phi_0 + \theta_1 \epsilon_t + \theta_2 \epsilon_{t-1} + 0, \end{aligned}$$

which follows since ϵ_t and ϵ_{t-1} is the time-t information set $\mathcal{F}_t$ and $E_t\left[\epsilon_{t+1}\right] = 0$ by assumption.

In practice, the one step ahead forecast would be given by

$$E_t\left[y_{t+1}\right] = \hat{\phi}_0 + \hat{\theta}_1 \hat{\epsilon}_t + \hat{\theta}_2 \hat{\epsilon}_{t-1},$$

Forecasting from ARMA models

Finally, consider the one-step ahead forecast from an ARMA(2,2) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_t,$$

The one-step ahead forecast is given by

$$\begin{aligned} E_t\left[y_{t+1}\right] &= E_t\left[\phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_t\right], \\ &= \phi_0 + E_t\left[\phi_1 y_t\right] + E_t\left[\phi_2 y_{t-1}\right] + E_t\left[\theta_1 \epsilon_t\right] + E_t\left[\theta_2 \epsilon_{t-1}\right] + E_t\left[\epsilon_{t+1}\right], \\ &= \phi_0 + \phi_1 y_t + \phi_2 y_{t-1} + \theta_1 \epsilon_t + \theta_2 \epsilon_{t-1}, \end{aligned}$$

Notice that in practice, the parameters and errors will be replaced by their estimates, i.e., $\hat{\phi}_1$ and $\hat{\epsilon}_t$.

Non-stationary time series

Seasonality and diurnality

Seasonality and diurnality are pervasive in financial and economic time series.

While many data series have been *seasonally adjusted* to remove seasonalities, particularly US macroeconomic series, there are many time-series where no seasonally adjusted version is available.

Ignoring seasonalities is detrimental to the precision of parameters and forecasting.

Definition (Seasonality)

Data are said to be seasonal if they exhibit a non-constant deterministic pattern with an annual frequency.

Definition (Diurnality)

Data which exhibit intra-daily deterministic effects are said to be diurnal.

Seasonality and diurnality

Seasonal data are non-stationary, although seasonally adjusted de-trended data may be stationary.

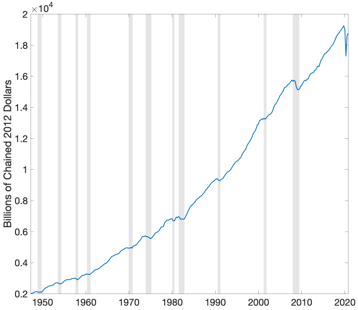
Seasonality is common in macroeconomic time series, diurnality is pervasive in ultra-high frequency data.

Seasonality is, technically, a form of non-stationarity since the mean of a process exhibits explicit dependence on t through the seasonal component.

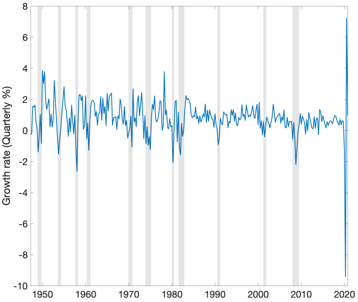
However, a slight change in time scale, where the seasonal pattern is directly modelled along with any non-seasonal dynamics produces a residual series which is stationary.

Seasonality: Macroeconomic example

Real US Gross Domestic Product



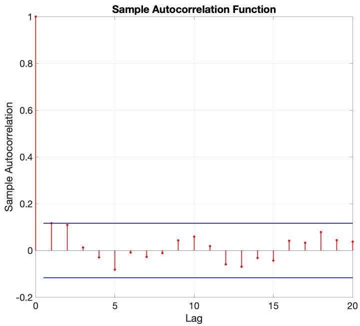
(a) GDP



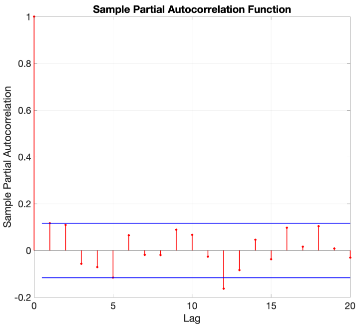
(b) GDP Growth rate

Seasonality: Macroeconomic example

Real US Gross Domestic Product



(c) Autocorrelation Function



(d) PACF

Deterministic trends

The simplest form of nonstationarity is a deterministic trend.

Models with deterministic time trends can be decomposed into three components:

$$y_t = \text{deterministic trend} + \text{stationary component} + \text{noise},$$

where y_t would be stationary if the trend were absent.

The two most common forms of time trends are polynomial (linear, quadratic, etc) and exponential.

Processes with polynomial time trends can be expressed

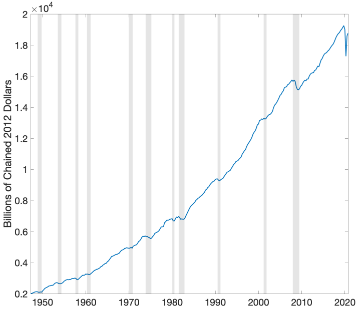
$$y_t = \phi_0 + \delta_1 t + \delta_2 t^2 + \ldots + \delta_s t^2 + \text{stationary component} + \text{noise},$$

and linear time trend models are the most common,

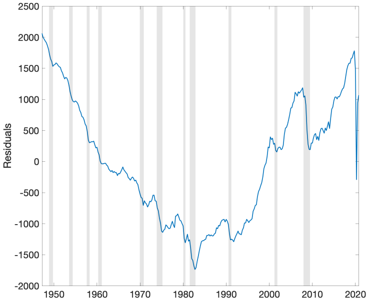
$$y_t = \phi_0 + \delta_1 t + \text{stationary component} + \text{noise},$$

Seasonality: Macroeconomic example

Real US Gross Domestic Product

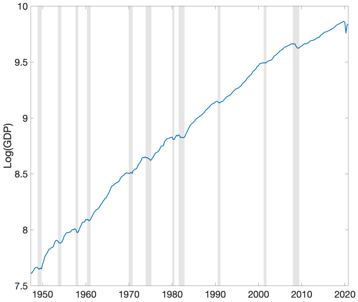


(e) GDP

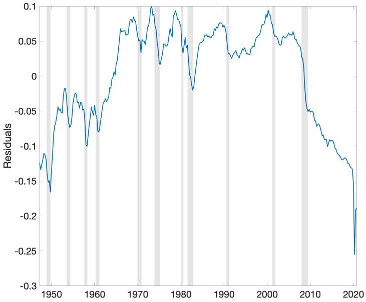


(f) $\hat{\epsilon}$ from $GDP_t = \mu + \delta_1 t + \delta_2 t^2 + \epsilon_t$

Real US Gross Domestic Product



(g) $\ln(GDP)$



(h) $\hat{\epsilon}$ from $ln(GDP)_t = \mu + \delta_1 t + \epsilon_t$

Unit roots

Random Walk

Definition (Random walk)

A stochastic process $\{y_t\}$ is known as a random walk if

$$y_t = y_{t-1} + \epsilon_t,$$

where ϵ_t is a white noise process with the additional property that $E_{t-1}]\epsilon_t] = 0$.

Random Walk

The basic properties of a random walk are simple to derive. First, the conditional expectation:

$$E_t\left[y_{t+k}\right] = y_t, \qquad \text{for any} \qquad k,$$

Second, the variance of a random walk is defined as

$$V\left[y_t\right] = t\sigma^2,$$

Third, the s^{th} autocovariance (γ_s) of a unit root process is given by

$$V\left[\left(y_t - y_0\right)\left(y_{t-s} - y_0\right)\right] = \left(t - s\right)\sigma^2,$$

and the s^{th} autocorrelation is then

$$\rho_s = \frac{t-s}{t},$$

Random Walk plus drift

Definition (Random walk plus drift)

A stochastic process $\{y_t\}$ is known as a random walk plus drift if

$$y_t = \mu + y_{t-1} + \epsilon_t,$$

This can be expressed equivalently as

$$y_t = \delta t + \sum_{i=1}^{t} \epsilon_i + y_0,$$

ans so the random walk plus drift process consists of both a deterministic trend and a random walk.

Difference or de-trend?

Detrending removes nonstationarities from deterministically trending series while differencing removes stochastic trends from unit roots.

Question: what happens if the wrong type of detrending is used?

Differencing a stationary series produces another series which is stationary but with largers variance than the de-trended series,

$$y_t = \delta t + \epsilon_t$$
$$\Delta y_t = \delta + \epsilon_t - \epsilon_{t-1}$$

while the properly detrended series would be

$$y_t - \delta t = \epsilon_t,$$

Thus, if ϵ_t is a white noise process, the variance of the differenced series is twice that of the detrended series.

Testing for Unit Roots

Dickey-Fuller tests (DF), and their generalization to augmented Dickey-Fuller tests (ADF) are the standard test for unit roots.

Consider the case of a simple random walk,

$$y_t = y_{t-1} + \epsilon_t,$$

so that

$$\Delta y_t = \epsilon_t,$$

Dickey and Fuller noted that if the null of a unit root were true, then

$$y_t = \phi_1 y_{t-1} + \epsilon_t,$$

can be transformed into

$$\Delta y_t = \gamma y_{t-1} + \epsilon_t,$$

where $\gamma = \phi - 1$.

Dickey-Fuller test

The null hypothesis that can be tested is then

$$H_0 : \gamma = 0, \qquad \text{against} \qquad H_1 : \gamma < 0,$$

This test is equivalent to testing whether $\phi = 1$ in the original model.

$\hat{\gamma}$ can be estimated using a simple regression Δy_t on y_{t-1}, and the t-stat can be computed in the usual way.

If the distribution of $\hat{\gamma}$ were standard normal (under the null), this would be a very simple test.

Unfortunately, it is non-standard since, under the null, y_{t-1} is a unit root.

The solution to this problem is to use the Dickey-Fuller distribution rather than the standard normal.

Dickey-Fuller test

Dickey and Fuller considered three separate specifications for their test,

$$\Delta y_t = \gamma y_{t-1} + \epsilon_t,$$
$$\Delta y_t = \phi_0 + \gamma y_{t-1} + \epsilon_t,$$
$$\Delta y_t = \phi_0 + \delta t + \gamma y_{t-1} + \epsilon_t,$$

which corresponds to a unit root, a unit root with a linear trend, and a unit root with a quadratic trend.

The null and alternative hypotheses are the same: $H_0 : \gamma = 0$, $H_1 : \gamma < 0$ (one-sided alternative).

The null that y_t contains a unit root will be rejected if $\hat{\gamma}$ is sufficiently negative, which is equivalent to $\hat{\phi}$ being significantly less than 1 in the original specification.

Dickey-Fuller test

Unit root testing is further complicated since the inclusion of deterministic regressor(s) affects the asymptotic distribution.

For example, if $T = 200$, the critical values of a Dickey-Fuller distribution are

	No trend	Linear	Quadratic
10%	-1.66	-2.56	-3.99
5%	-1.99	-2.87	-3.42
1%	-2.63	-3.49	-3.13

Dickey-Fuller test

The augmented Dickey-Fuller (ADF) test generalized the DF to allow for short-run dynamics in the differenced dependent variable.

The ADF is a DF regression augmented with lags of the differenced dependent variable to capture short-term fluctuations around the stochastic trend, i.e.,

$$\Delta y_t = \gamma y_{t-1} + \sum_{p=1}^{P} \Delta y_{t-p} + \epsilon_t,$$

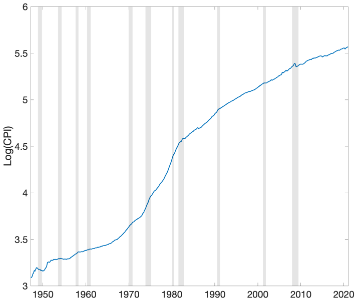
$$\Delta y_t = \phi_0 + \gamma y_{t-1} + \sum_{p=1}^{P} \Delta y_{t-p} + \epsilon_t,$$

$$\Delta y_t = \phi_0 + \delta t + \gamma y_{t-1} + \sum_{p=1}^{P} \Delta y_{t-p} + \epsilon_t,$$

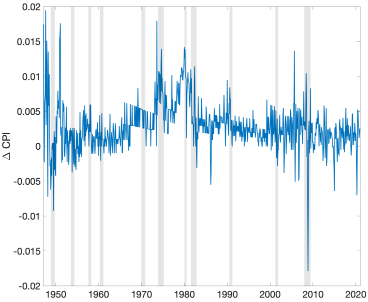
Neither the null and the alternative hypothesis nor the critical values are changed by the inclusion of lagged dependent variables.

Example on Inflation and Default Spread

Inflation and default premium



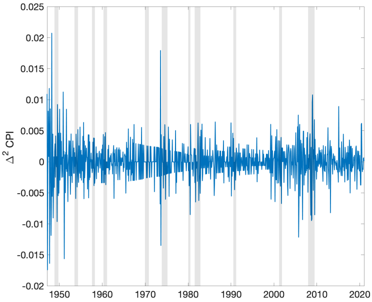
(i) $\ln(CPI)$



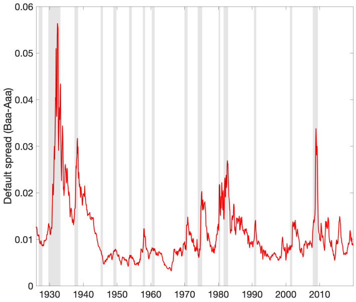
(j) ΔCPI

Example on Inflation and Default Spread

Inflation and default premium



(k) $\Delta^2 CPI$



(l) Default premium

Example on Inflation and Default Spread

Testing unit roots for Inflation and default premium

	$\ln CPI$	$\ln CPI$	$\ln CPI$	$\Delta \ln CPI$	$\Delta \ln CPI$	$\Delta^2 \ln CPI$	Def Sp.	Def Sp.
t-stat	-0.214	-0.934	5.457	-7.161	-4.826	-21.001	-3.432	-1.795
p-val	0.992	0.768	0.999	0.000	0.000	0.000	0.0102	0.069
Deterministic	Linear	Const	None	Const.	None	None	Const.	None
# lags	4	4	4	4	4	4	4	4

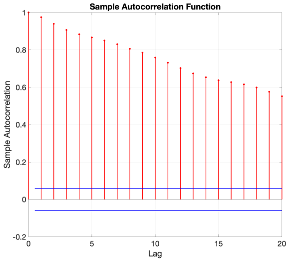
Exercises

Exercises

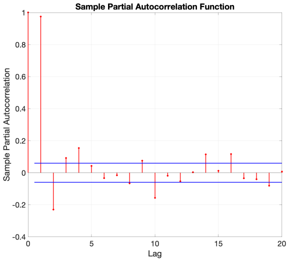
Problem 6.1: Derive the three-step ahead forecast for an AR(1) and an MA(2) model

Problem 6.2: Derive the optimal three-step ahead forecast from the ARMA(1,2) model

Problem 6.3: Justify a reasonable model for the dynamics of the default premium using information from the ACF and PACF



(m) ACF



(n) PACF

Exercises

Problem 6.4: Outline the steps needed to perform a unit root test on a time series of FX rates.

Problem 6.5: Answer the following questions:

- How are the autocorrelations and partial autocorrelations useful in building a model?
- Describe the use of the information criteria.

Problem 6.6: Suppose $y_t = \phi_0 + \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$

- What is $E_t\left[y_{t+1}\right]$?
- What is $E_t\left[y_{t+2}\right]$?
- What is $V_t\left[y_{t+1}\right]$?
- What is $V_t\left[y_{t+2}\right]$?