Ca' Foscari
University
of Venice

Master's Degree
in
Economia e Finanza

Final Thesis

A multivariate analysis to discover the relation between cryptocurrencies and CO_2 emissions

Supervisor
Prof. Emanuele Aliverti

Graduand
Silvia Bellinzona
989300

Academic Year
2021/2022

Abstract

This study aims to analyze the relationship between cryptocurrencies and carbon dioxide emissions. The research was conducted providing an initial background on the context of cryptocurrencies and emissions, then moving on to the theoretical explanation of univariate models and later to the part concerning the multivariate model: Vector Autoregressive process. Finally, the data, results, and related conclusions are presented.

INDEX

Introduction

It is common to speak of cryptocurrencies as a new generation of digital currencies complementary to legal currencies whose development was made possible by information systems such as blockchains.

The idea of creating an independent system of complementary currencies developed long before the realization of computer networks in order to meet specific needs of operators that could no longer be met by legal tender currencies.

Essentially, the creation of “something else” is firmly anchored precisely in the intrinsic limitations of official currencies.

Monetary theory defines money through the functions it is capable of performing. According to Hicks (1971), money is what money does. More precisely, currency is a good that can perform three basic functions:

- medium of exchange: it can be exchanged for other goods and services;
- measure of value: it measures the market value of other goods and services through the determination of exchange relationships;
- fund of value: it is a means of preserving wealth over time.

Because it is able to perform these three functions this “asset” has the transaction motive, the precautionary motive and the speculative motive at the basis of its demand.

Legal tender currency is that to which the law assigns the characteristics of being compulsorily accepted in exchanges with power to extinguish obligations, it has the value of the amount printed on it, and it is issued exclusively by a central entity to which the management of monetary policies is entrusted (European Central Bank, 2015).

The prerequisites for the creation of a complementary currency occur when the characteristics and functions of legal money do not meet, except in part, the particular needs that people develop in a particular economic environment and in a particular historical period.

The concept of money complementarity refers to the relationship with legal tender; complementarity is constituted and finds its *raison d'être* in the very existence of legal currencies.

There is no doubt that the advent of Internet technology has profoundly impacted people's lives, cultural patterns, social trends, as well as politics and economics. “*The Web* - these are the words of Tim Berners Lee, the young researcher at Cern in Geneva who is considered the inventor of the World Wide Web - *is more a social creation than a technical one*".

These changes have generated several needs, including in the area of economic transactions, that were not felt during previous centuries and to which new technologies have made it possible to respond. The world of complementary currencies could not remain immune to this evolution by finding the ideal tools to build an alternative economic and monetary system: virtual currencies were born.

Virtual/digital currency, as defined by EU Directive 2018/843, is «*a digital representation of value that is not issued or guaranteed by a central bank or a public authority, is not necessarily attached to a legally established currency and does not possess a legal status of currency or money, but is accepted by natural or legal persons as a means of exchange and which can be transferred, stored and traded electronically*».

From this definition, its peculiar characteristics emerge: first, the absence of central control and lack of legal recognition, the dematerialization, that is the absence of a physical entity to act as the value-bearing instrument, and finally, the voluntary element to which its circulation is linked.

They are thus intended to efficiently perform the function of a medium of exchange (lower transaction costs and greater confidentiality) while the ability to act as a unit of account and fund of value are affected not only by the lack of legal recognition but also by the currency's excessive divisibility and extreme volatility. Precisely because of the fact that they only partially fulfill the functions proper to money it would be preferable according to some to refer to them as "virtual currencies" (V. Carlini, 2018).

Following a useful classification made by the European Central Bank, it is possible to distinguish virtual complementary currencies according to a scheme that considers their degree of convertibility.

We can thus have a closed virtual currency, for which there is no interaction with legal money, a virtual currency in which interaction occurs only partially at the entry of the process with the exclusive objective of generating virtual money, and finally to an open virtual currency in which interdependence with the monetary world occurs both at the moment of creation of the virtual currency and at the moment when the virtual currency is reconverted into legal money. In the latter case, the relevance of the amount of virtual currency impacts the traditional system and competes with currencies managed by monetary authorities.

Relegating virtual currencies, therefore, to an exclusively monetary phenomenon would be reductive: they are an expression of a different way of understanding economic relations that represents a true philosophy, a kind of technological utopia in which individuals cooperate outside of imposed and unshared regulation, giving rise to a horizontal community having as its fundamental principle trust.

The cultural atmosphere in which the first virtual currencies developed was the result of two causes of "*people's revolt*" which both shared the common characteristic of being antiestablishment.

The first is the impact that Internet had on communication: if on one hand it allowed people to access to any information, communicate with everyone and express their opinions to everyone, on the other hand it entailed a colossal digital profiling of people thus conceived as a sum of data to be collected. The network is thus no longer seen as a means of redistribution of power but as a system of control and registration in the service of a new economy based on the collection and exploitation of personal data in which everyone's privacy and freedom to decide what to make public and what not are severely compromised.

In the early 1990s, Erich Hoghes[1] published the manifesto of the Cypherpunk movement, in which he theorized the use of cryptography with the goal of protecting and enhancing each individual's privacy. «*We the Cypherpunks are dedicated to building anonymous systems. We are defending our privacy with cryptography, with anonymous mail forwarding systems, with digital signatures, and with electronic money*» (Hughes, E. 1993).

Setting the technical foundation of the Cypherpunk movement was David Lee Chaum, a cryptographer and computer scientist who in 1985 published an article titled « *Security without identification: transaction systems to make big brother obsolete* ».

Chaum developed a revolutionary idea, based on the concept of an entirely electronic currency (DigiCash) that can work through cryptography and it can be spent anonymously. It is the prototype of cryptocurrencies.

This first attempt at a complementary virtual currency was followed by others but failed to develop stable systems.

[1] He is an American mathematician, computer programmer, and a cypherpunk. More importantly, he is credited with being one of the movement's founders.

The prerequisites for the emergence of the most widespread cryptocurrency were only realized with the 2008 financial crisis, which triggered yet another reason for rebellion by shattering a certain covenant between the elites and the people (A. Baricco, 2019).
The struggles and dramas experienced by millions in the aftermath of the subprime crisis are proof not only of the establishment's inability to foresee crises but also that rules and market control systems have served little purpose except to secure the elites by bouncing sacrifices on people.
In 2008 Satoshi Nakamoto published a revolutionary paper "*Bitcoin: a peer-to-peer electronic cash system*" in which he described a protocol for the creation of a new digital currency with the intention of offering people an alternative to the traditional services offered by banks and a solution to the problems that the system regulated by Central Banks had proven unable to cope with. In Nakamoto's *white paper*, the elimination of intermediaries and the subtraction of the power of control and regulation from a third-party authority results first and foremost in the use of *peer-to-peer* networks, networks that are not organized hierarchically into *server and client* but by equivalent nodes, capable of functioning by ensuring the security of transactions through the use of blockchain technology.
Simplifying, the model for creating and using Bitcoin, and cryptocurrencies more generally, is based on three basic elements: asymmetric cryptography, the blockchain, and the mining process.
Asymmetric cryptography involves splitting a code into two keys: a public key and a private key, mathematically linked to each other. The data (date, amount, traders…) of each transaction are transformed, using a cryptographic hash function, into a code that contains the public key while the private key is known only to the trader. This allows all transactions to be made public in a ledger without sharing personal data that are instead accessible only with the use of the private key. The ledger in which all transactions are stored, in an unchangeable manner, constitutes a data base distributed among all nodes whose updating with new transactions is subject to a validation process that is completed once the agreement of the majority of the nodes in the network has been obtained.
This process uses blockchain technology in which data are collected in blocks that are inseparably and indelibly linked together in succession to form a chain so that it is always possible to trace back to previous blocks of data. A powerful and growing algorithm automatically governs the entire system. In the blockchain model, trust in the manager is therefore untied from any subjective aspect but is objectively placed in the system.

The mining process relates to both the creation and the exchange of Bitcoin: the production system is designed in a way that exactly mirrors the exchange system through the establishment of blocks, decryption power and validation speed. Anyone who wants to increase the amount of cryptocurrency in circulation must become a miner spending resources, or it would be better to say energy, to increase the amount of gold in circulation. How? Each time a trader requests a transaction, or a group of transactions, it is destined to make up a completed block which initiates a validation process that first engages miners.

For this essential first validation, the miners are asked to solve a problem through a rather complex algorithm that requires having computing power obtainable only through the simultaneous deployment of a large number of machines to ensure that the problem is solved in a short period of time. As soon as the solution is found by the fastest miner, it is communicated to the other miners (this is the procedure that goes by the name of *proof of work*) and goes to form a new block in the chain that can then be sent to the network confirmation process. The "deminer" receives as a reward for the work done (*fee*) equal to a certain amount of cryptocurrency thus contributing to increasing the amount of "currency" in circulation.

A little more than a decade after the birth of Bitcoin, we can say that cryptocurrencies represent one of the most interesting and innovative examples of technology development and diffusion, capable of operating a profound change in the economic system. There is no doubt that this is neither a transitory nor an insignificant phenomenon with which the Monetary Authorities will have to deal because the absence of regulation (today limited to the tax and anti-money laundering sectors) can only lead to the decomposition of the world monetary and financial system, opening up scenarios far removed from the utopia of an egalitarian technological community.

The spread of cryptocurrencies, however, has highlighted not only their opportunities and innovative aspects but also a number of critical issues that were only partly predictable by their creators.

Firstly, the difficulties in accessing the virtual currency market and the concentration of trading platforms have led to the attribution to communities using cryptocurrencies of an elitist character far removed from the egalitarian and anti-systemic ideals of the early days; secondly, anonymity, the absence of intermediaries and state controls, as well as de-localized access have fostered degenerations that lead many to consider cryptocurrencies as ideal tools for money laundering and illicit transactions.

Furthermore, cryptocurrency transactions have gradually seen the speculative nature of cryptocurrencies prevail, thus distorting the monetary function for which they were designed. What the technologies available to the creators of cryptocurrencies could not take into account was the individual psychological factor in the decision-making processes of users who in fact prioritized investment attracted by the volatility of these currencies while helping to fuel it.
Upon closer analysis of the phenomenon, we could say that rather than a "degenerative" phenomenon the speculative function is inherent in the very nature of cryptocurrency in that «*as long as it remains a side and secondary means of payment its transfer for solving purposes is never independent of an objective and predominant speculative function. Cryptocurrency is and remains, at this stage, nothing more than a form of investment, while the monetary component (beyond any hypocritical ideological excuses) is but a kind of cover, a legitimation pretext to conceal functionality of a different nature*» (Girino, E. 2018).
However, one of the most controversial aspects in the spread of cryptocurrencies, toward which analysts' attention has been focused, is their environmental impact. This is the most interesting aspect for the purposes of the analyses conducted in this study.
The growing social awareness of the value of sustainable development has led governments, public agencies, and businesses to place ecology at the center of their strategies and has paved the way for serious reflection on the sustainability of technological development more generally and, in the specific case, on the processes of cryptocurrency production and use. That a digital currency pollutes it is not immediately intuitive, but the illusion that the dematerialized is also environmentally friendly, almost by definition, has long since fallen.
We have seen that the revolutionary aspect of cryptocurrencies, starting with Bitcoin, is the use of blockchain technology, which allows transactions to be secured (avoiding double spending) through a validation system that disregards a third-party guarantor authority.
In order to generate and validate a Bitcoin transaction, it is necessary to solve a complex mathematical "problem": the computational power required for this process far exceeds the capabilities of a single computer or a single operator and therefore requires the simultaneous use of a large number of machines (concentrated in mining farms) also coordinated among several operators (mining pools) that synergically and continuously

work in order to ensure the solution in the shortest possible time of the computational puzzles and then share the reward in cryptocurrency. All this requires a lot of energy.
What the "alternative currency" system wants to test is not so much the solving ability of the miners, but to measure their computing power because this makes the whole system increasingly efficient, attractive and secure.
In 2019, a study[2] by the Technical University of Munich in collaboration with the Massachusetts Institute of Technology (MIT) coordinated by Christian Stoll verified that while in January 2011 a miner with a 2GH/s GPU could expect to de-mining two chain blocks per day, in 2018, due to the increasing difficulty of the operations to be solved, the same miner, with the same power, had an expectation of solving one block every 472.339 years.
The exponential growth in demand for computing power, accompanied by the explosion in cryptocurrency transactions recorded after the first few years of trial, simultaneously became exponential growth in energy consumption. Stoll's study, starting from the observation that most mining farms are concentrated in countries that rely on coal-generated energy, developed a technical economic model for determining the energy consumption required by the Bitcoin network, arriving at quantifying its carbon footprint as more than 22 million tons of CO_2 (the reference is to 2019) released into the atmosphere each year, and the number was expected to grow rapidly.
In May 2021, a study published in the Rivista Banca d'Italia[3] compared the carbon footprint of the instant payments platform TIPS with the cryptocurrency system: it turned out that in 2019 TIPS carbon footprint had been almost 40.000 times smaller than that of Bitcoin. According to this study, the energy expenditure and the high environmental cost of the cryptocurrency system was nothing more than the price of trust: «*The huge discrepancy in the carbon footprints of TIPS and Bitcoin stems from the fact that the latter uses a large amount of energy to generate trust and consensus among participants in the Bitcoin network, while in the case of TIPS this trust is provided by the Eurosystem*» (Rivista Banca d'Italia, 2021).
If we have to consider, on one hand, that much of this criticism against the cryptocurrency system comes precisely from the banking world and the financial establishment with respect to which they clearly compete, on the other hand, it is undeniable that the goal of

[2] *The Carbon Footprint of Bitcoin*, Joule n.3, 17th July 2019.
[3] *The carbon footprint of the Target Instant Payment Settlement (TIPS) system: a comparative analysis with Bitcoin and other infrastructures,* Rivista Banca d'Italia n. 5/21, 20 May 2021

reducing CO_2 emissions to which all states have committed themselves to varying degrees, has focused the attention on a sector that, while only responsible for a small part of global emissions, has energy-intensive characteristics that are worth focusing on for the adoption of containment and regulatory measures.

CO_2 emissions

When we talk about climate change, we relate to the fact that the climate in the last years has changed a lot. Global warming, indeed, has always been present, in our and in the history of the Planet, but the one that we are all witnessing is anomalous due to man and his activities.

The main cause of climate change is the burning of fossil fuels (oil, coal, natural gas, etc.) which release greenhouse gases into the atmosphere. There are also other human activities, such as agriculture and deforestation, which contribute to their spread. The problem is that these gases retain heat in the atmosphere: the one that is known as greenhouse effect.

The greenhouse effect is necessary to retain some of the solar irradiation that otherwise would be lost to space. Without this natural effect the average temperature of our planet, which is currently +15 °C, would be -18 °C which would make the planet hostile to most forms of life. What human activities are causing is an excessive rise in the concentration levels of these gases with the effect of accelerating global warming and causing uncontrollable climate change.

Despite the numerous agreements establishing rules to steam the risk, the level of carbon dioxide (CO_2) in the atmosphere continues to grow and reached another record in 2022. [4]

[4]As measured by the National Oceanic and Atmospheric Administration (NOAA), in May 2022, CO_2 emissions reached 420,99 ppm (parts per million), last year it was around 419,13 ppm. The observatory, where were made the observations, is located on Mauna Loa, in Hawaii, a privileged place that allows you to measure concentrations in the upper atmosphere, away from local pollution sources (source: Global Monitoring Laboratory).

We can see how this data are so concerning because before the Industrial Revolution, CO_2 levels were around 280 ppm, and these concentrations remained almost constant over the previous 6,000 years (Iconaclima, 2022)

The University of Colorado[5] analyzed the CO_2 emissions of more than 29.000 fossil fuels power plants of 221 countries, ending up with the conclusion that only 5% of the global power plants represents almost three-fourths of the global emission of CO_2, coming from power production.
«*The maps below reveal, as one might suspect, that the plants that did the most absolute damage to the atmosphere were fired by coal (indicated in blue). Most of these plants were clustered in the United States, Europe, India, and East Asia*» (Don Grant *et al* 2021).

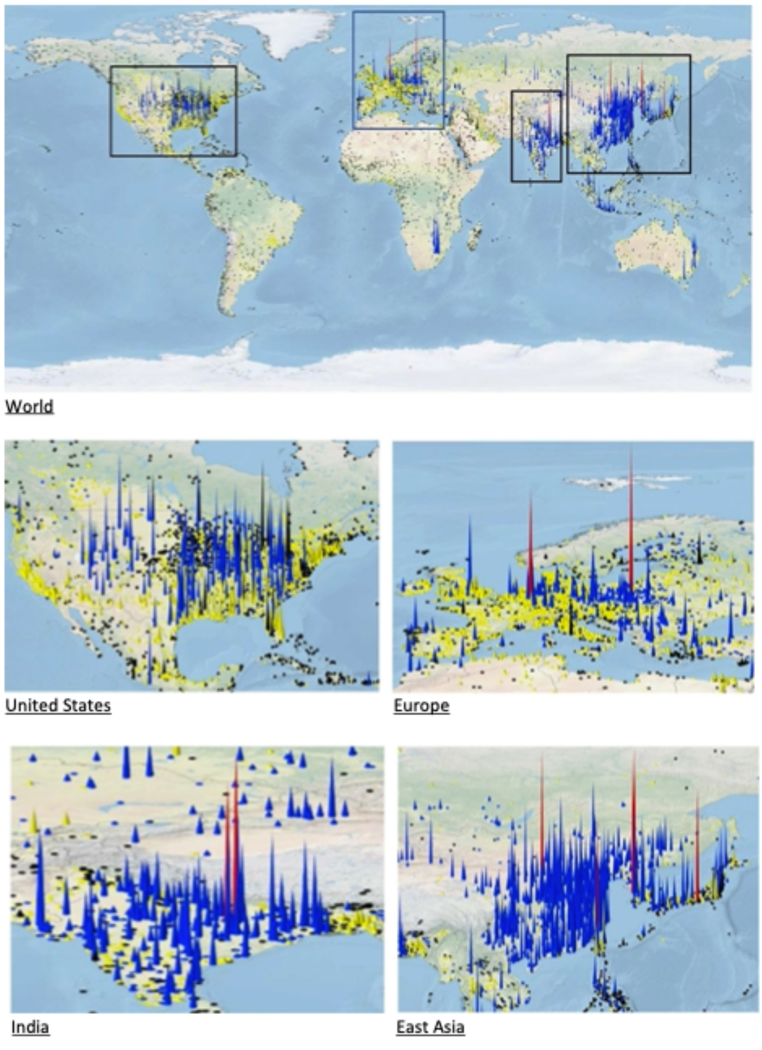


Figure 1: Maps of fossil-fueled power plants' CO_2 emissions. Taller spikes indicate that plants emit CO_2 at higher levels. Colors signify plants' primary fuels (blue = coal, yellow = natural gas, black = oil). Plants with red spikes are the world's ten biggest polluters (all of which rely primarily on coal).
2018
Source: Don Grant *et al* 2021 *Environ. Res. Lett.* **16** 094022

[5]The study was carried out by the CU Boulder (Professor Don Grant and his team).

The investigation highlighted some countries, as reported above, that the team of Don Grant referred to as “super polluters”.

The carbon footprint of cryptocurrencies

The electricity that is consumed in the mining process of Bitcoin (by far the most important cryptocurrency) has become a topic of heated debate in recent years.
Investigating the environmental impact of cryptocurrencies and their contribution to accelerating global warming, first requires distinguishing between energy consumption required by the process and the carbon footprint generated. It is abundantly clear that the consumption of energy generated by renewable sources such as solar or wind power has a very different environmental impact than that generated by sources such as coal, oil and gas.
As we saw, at the base of this connection there is the concept of cryptocurrency mining: «*mining is the process that Bitcoin and several other cryptocurrencies use to generate new coins and verify new transactions*»[6].
As a way to perform these functions, computers, which are connected to the network, have to solve complex mathematical calculus, and for this reason are necessary very powerful computers that require large amounts of energy. But this is not the only point, because it has to be taken into account that all these machines require, obviously, to be always turned on and, more importantly, it is necessary the presence of internal fans and air conditioning to cool the hardware of computers.
Since for this validation work only the fastest miner is rewarded with commissions and allocation of new Bitcoins, the business logic dictates that a natural competition is triggered among them, which results first and foremost in investments in increasingly high-performance technologies. Thus, there has been a shift from the use of the CPU to the GPU, to more powerful graphics cards for gaming use and finally to dedicated ASIC hardware that can calculate billions of Hash per second.
Competition among miners then became competition among microchip manufacturers between those who produce not only the fastest but also the one that consumes the least power. The rapid physical and technological obsolescence of machines (the life cycle of

[6] Coinbase, *What is mining?*

an ASIC machine is estimated at about two years) exposes miners to an economic chase in order to increase efficiency and contain energy costs often with negative outcomes: the Swedish miner KNC Miner went out of business in 2016.

The estimation of the direct energy consumption of Bitcoins (or other cryptocurrencies using blockchain technology) is by no means simple. For this reason, several theoretical mathematical models have been developed which lead to estimate electricity consumption (or it would be better to say a range of consumption) from the limited data available and often with very different results.

For Bitcoin, the value of the Hashing power (the working power of all the miners within the network) is available in real time: by dividing this data by the Hashing power of a single typical machine (considering a typical machine or a mix of machines), it is possible to estimate how many machines are working and to trace back through the data sheets to the power globally consumed by all the miners.

Other models such as, for example, the CBECI (Cambridge Bitcoin Electricity Consumption Index) developed by M. Bevand of the University of Cambridge in 2017 and the BECI (Bitcoin Energy Consumption Index) developed by A. De Vries for Digiconomist in 2019 propose the determination of energy consumption of the Bitcoin network by moving from an economic perspective.

Given that miners' revenues and costs are interrelated and that energy costs account for the largest share of the latter, the two models, although using different techno-economic approaches, arrive at determining through a series of approximations, the total electricity consumption, starting from miners' profitability. The higher the total value of miners' rewards, the more energy-consuming machines that can be used.

The transition from estimating energy consumption to quantifying its carbon footprint, on the contrary, requires knowing where this energy comes from; the location of miners, in this sense, is the key ingredient in understanding whether this energy is "dirty or clean".

The mining sector is now becoming more centralized, with more than 80% of all the cryptocurrencies extracted in specific countries: China, Russia, United States and Canada. An analysis by Rystad Energy[7] shows that until 2021 China's Bitcoin production was 65% of the global total. By 2020, China was producing 63% of its energy from coal which

[7] Rystad Energy, *Bit late for bitcoin: How China's crackdown is reducing more emissions than whole countries emit*. (2021, July 14)

led to the conclusion that 40% of Bitcoins mined globally were fueled exclusively by burning coal.

The estimated carbon emissions from total energy production in China were about 5.200Mt[8] for a total energy production of 7.815 terawatt hours (TWh). Considering that Bitcoin mining in China required 86 TWh leads to determine that the carbon emission for this activity in 2020 was 56Mt equal to the emissions of countries such as Portugal or Peru (Rystad Energy, 2021).

Despite cryptocurrency supporters claiming that mining is increasingly relying on renewable energy, there are many nations that have blocked the way for digital currencies and cryptocurrency mining by passing legislation against this industry.

Although the impact of cryptocurrencies was extremely small compared to the total emissions produced (1.1%), China in July 2021 decided for a heavy reduction in Bitcoin mining. Other countries to which cryptocurrency miners have moved after China's ban (Iran and Kazakhstan) have imposed restrictions and bans on digital currency production activities. In Europe in early 2022, Kosovo, under pressure from the energy crisis, banned cryptocurrency mining within its borders while Sweden and Norway have asked the European Union to take a position on this issue.

Justified alarmism or a biased attitude?

This study aims to provide insight into the relationship between cryptocurrency transaction volume and CO_2 emissions.

The investigation will be conducted through the protracted observation and comparison of several variables over time, namely:

a. the prices of major cryptocurrencies;
b. CO_2 emissions recorded in China's Liaoning region where one of the world's largest mining farms resides, in Dalian. Liaoning is one of the country's major industrial provinces and it is a major producer of electricity, much of which is generated by large coal-fired thermal power plants and an increasing share consists of hydropower;
c. CO_2 emissions recorded in the Qinghai region. This is a mountainous province bordering Tibet characterized by limited industrial development. Abundant water resources have been exploited for hydropower generation in large power plants.

[8] Source: ourworldindata.org

All of this will be put in place with an initial univariate analysis of the data we have, mainly following the method described by the twentieth-century statisticians Box and Jenkins, so that we will have a general view of what we are going to approach. Next step will be the analysis related to the multivariate world, specifically the vector autoregressive model, which is going to provide us with a range of information, including the impulse response function, useful to allow us to draw conclusions regarding the relationship linking cryptocurrencies and carbon dioxide emissions.

Univariate stochastic processes

A time series is simply the chronological record of observations of a variable: in the time series the order of the data has a fundamental importance, given by time and the direction of it.

In time series the natural tendency of many phenomena to evolve more or less regularly leads to the idea that the data collected at a given time instant t are more similar to those collected at instant $t-1$ with respect to those collected at distant periods. It can be said, therefore, that the time series has "*self-memory*" known as persistence.

Mathematical models that go by the name of stochastic processes are used to describe the probabilistic law by which a certain phenomenon may evolve over time. From a practical point of view, stochastic process is a form of representation of a quantity that varies over time randomly and with certain characteristics.

The complete determination of a stochastic process is referred to as a trajectory. We can therefore define the time series as the partial realization of a trajectory of the stochastic process.

In mathematics, a stochastic process, denoted by $\{Y_t\}_{-\infty}^{+\infty}$ with $t \in \mathbb{Z}$, is defined as a family of random variables, belonging to the real number set ($Y_t \in \mathbb{R}^d$ $with$ $d \geq 1$), ordered by n-integer parameter t, such that for all $n \in \mathbb{N}$ and all n-ples $t_1, \dots, t_n$ in $\mathbb{Z}$, the joint probability distribution of all random variables (from time t_1 up to time t_n) is well defined.

The methodology that will be followed for identifying stochastic models follows the pattern of the Box-Jenkins method[9], which involves preliminary analysis, model identification, parameter estimation and verification of the model.

We said that the stochastic process is nothing more than a form of representation of a quantity that varies over time randomly and with certain characteristics.

To conduct this "representation", various models of analysis are used, which can be grouped into two macro categories: univariate models and multivariate models of analysis.

[9] Box, G. and Jenkins, G. (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco

The former consider only one variable, and thus the analysis of data from a single time series, in our case for example the time series of CO_2 emissions from the Liaoning region. Multivariate analysis models, on the other hand, involve the joint study of two or more variables, and thus in our investigation, for instance, the time series on cryptocurrency values and the time series in CO_2 emissions of Chinese regions are considered simultaneously.

The latter are the statistical models that are certainly of most interest; however, this does not preclude the fact that univariate analysis has its own importance for the study of the intrinsic characteristics of the single variable under analysis and is therefore functional for the subsequent multivariate analysis.

Stochastic processes' characteristics

In the stochastic process, it is first necessary to check certain characteristics of the behavior of the variable under observation that go by the names of stationarity and ergodicity.

The first essentially refers to the characteristics of the underlying stochastic process that generated the time series; therefore, when the characteristics of the stochastic process change over time we have a nonstationary process. Basically, dealing with nonstationary processes implies dealing with a variable trend, which depends on many unknown parameters. Hence, non-stationarity involves the difficulty of having to estimate too large number of parameters with a limited quantity of observations.

Thus, we can say that if a series is stationary, it is possible to use its past history (by means of an equation with fixed coefficients) to predict its future behavior (this is in the case of the AR models that are soon to be introduced).

In the case, on the other hand, in which the series under consideration is not a stationary series then a transformation must then be made to induce this stationarity.

Thus, stationarity is a fundamental property for time series as it allows parameters to be estimated with high accuracy, given that the process can be modeled with relatively few parameters.

Stationarity should be distinguished into two subsets: strong stationarity and weak, or second-order, stationarity.

Strong stationarity is a condition for which all aspects of the behavior of a stochastic process are unchanged by changes over time. Consider a stochastic process $\{Y_t\}$ and two vectors of random variables defined as follows:

$$\left(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}\right)^T \sim F$$

$$\left(Y_{t_{1+h}}, Y_{t_{2+h}}, \dots, Y_{t_{n+h}}\right)^T \sim F$$

with h representing the shift in time ($h > 0$).

Thus, the two vectors of random variables are identically distributed, although this does not mean that they take the same values, but it simply means that the joint probability distribution of the first random vector is the same as that of the second random vector, with the only difference being a shift in time (h). For this reason, strong stationarity is a very difficult hypothesis to observe, because it requires all aspects of the behavior to be constant over time.

On the other hand, weak stationarity, or second-order stationarity, requires the mean, variance and covariance of the process to remain unchanged with respect to temporal changes. While as for the correlation between two observations, the latter depends only on the lag, that is, the temporal distance between the two observations. In other terms:

Expectation	$\mu = E(Y_t)$
Autocovariance function	$\gamma_j = cov\left(Y_j, Y_{t-j}\right)$
Variance	$\sigma^2 = cov(Y_t, Y_t) = var(Y_t) = \gamma_0$
Autocorrelation	$\rho_j = \dfrac{cov\left(Y_t, Y_{t-j}\right)}{\sqrt{var(Y_t)var(Y_{t-j})}} = \dfrac{\gamma_j}{\sqrt{\gamma_0\gamma_0}} = \dfrac{\gamma_j}{\gamma_0}$

Table 1: Functions which characterize the stochastic process

A strongly stationary stochastic process is also (always) stationary of second order; the reverse is not true.
An ergodic stochastic process is defined as one in which the sample mean, autocovariance and autocorrelation are all consistent estimators. Thus, as the time series becomes longer the estimates produced by these estimators become increasingly precise with respect to the parameter of interest. This is the property whereby, while collecting more and more observations, we continue to learn something new about the process.
With some technicalities, it is possible to state that a weak stationarity process is ergodic for the mean if

$$\bar{y} = \frac{\sum_{t=1}^{T} y_t}{T}$$

converges in probability to $E(Y_t)$ as $T \to \infty$. A process will be ergodic for the mean provided that the autocovariance γ_j goes to zero sufficiently quickly as j increases.
Similarly, a second order stationary process is said to be ergodic for second moments if

$$\frac{\sum_{t=j+1}^{T} (Y_t - \mu)(Y_{t-j} - \mu)}{T - j} \xrightarrow{p} \gamma_j$$

for all j (James D. Hamilton, 1994).
On the other hand, a nonergodic process is one that has such emphasized persistence characteristics that one segment of the process, however long, is not sufficient to say anything meaningful about its distributional characteristics. In other words, if the process is non-ergodic, it does not matter how many observations we collect, as there is no additional information we are collecting, since everything is known from the beginning (from the initial value taken by our process).
Therefore, what we must try to have it is a process that is characterized by weak stationarity and ergodicity, to possess a wide range of information that can be exploited.
Some methods for detecting the presence of stationarity will be presented later.

Introduction to ARMA processes

One of the models belonging to the univariate class is the Autoregressive Moving Average (ARMA) model, also referred to as the Box-Jenkins model.
The ARMA is nothing more than a model that allows us to define our historical series as consisting of its values at previous times and random variations (unpredictable shocks).
It is thus a tool that permits to analyze and predict future values and consists essentially of two parts, an autoregressive part of order p and a moving average part of order q.

Moving Average stochastic processes of order $\boldsymbol{q}$

The first univariate model we introduce is the moving average process. A moving average process asserts that the relationship between the present value and the present and past error terms is linear. Once more, it is assumed that the error terms are normally distributed and mutually independent, exactly like white noise.
This model has the same form as Wold's representation theorem[10], so it needs no restriction on the characteristic of stationarity, given the fact that this is always given.
A moving average process of order q has indeed the equivalent form of the Wold, but this time the summation does not run up to infinity, it stops at lag q.
Said so, it is a sequence of random variables which is written as

$$Y_t = \mu + \sum_{j=0}^{q} \theta_q \varepsilon_{t-q} + \varepsilon_t$$

[10] Wold theorem states that any zero-mean covariance stationary process $\{Y_t\}$ can be represented in the form

$$Y_t = \mu + \sum_{j=0}^{\infty} \zeta_j \varepsilon_{t-j}$$

where ε_t is a white noise[10], ζ_j are constant numbers and $\zeta_0 = 1$, and $\sum_{j=0}^{\infty} \left|\zeta_j\right|^2 < \infty$.
The result is very powerful since holds for any covariance stationary process.

where ε_t is a white noise with zero mean and constant variance (σ^2).
The following tables summarizes the properties of the model.

Expectation	$\mu = E(Y_t)$
Autocovariance	$\gamma_j = cov(Y_t, Y_{t-j}) = \begin{cases} \sigma^2 \sum_{i=0}^{q} \vartheta_j^2 & j = 0 \\ \sigma^2 \sum_{k=0}^{q-j} \vartheta_k \vartheta_{k+j} & 1 \le j \le q \\ 0 & j > q \end{cases}$
Autocorrelation	$\rho_j = corr(Y_t, Y_{t-j}) = \frac{\gamma_j}{\gamma_0}$

Table 2: Functions of $MA(q)$ process

The expectation and the variance are constant in time, and the autocovariance function depends on the temporal lag.
By the autocovariance function we can easily find out the autocorrelation function. If the autocovariance function is identically equal to zero, for $j > q$, the autocorrelation function will share the same properties.
An important property of the process is the invertibility, which represents the constraint for the moving average process, and it is not determined by stationarity reasons. The purpose is to have a one-one relationship between the autocorrelation function (or the autocovariance function) and the parametrization of this stochastic process. In other words, for any value of the moving average coefficient there is one and only one autocorrelation function corresponding to those values, and vice versa. The condition, called invertibility, is guaranteed only when the roots of the moving average polynomial are not less than one in absolute value:

$$|z| > 1 \; and \; z = \frac{1}{\theta} \to |\theta| < 1$$

where the moving average polynomial is:

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$$

$$\theta(z) = 0$$

Hence, conditions for stationarity and invertibility contemplate the roots of the characteristic equation to lie strictly outside the unit circle in the complex plane.

Autoregressive stochastic processes of order $\boldsymbol{p}$

Another important class of processes are the Autoregressive processes. These are models that provide an easier representation of a persistent series as compared to moving average ones.

The main idea in this set of models lies in the fact that the variable of interest, at time t, is a function of its past values, plus an error, a white noise. Moreover, it is called autoregressive because, it is a regression of the variable against itself. In other words, each variable is modeled as a function of the past values, that is the predictors are nothing but the lags (time delayed value) of the series.

The process has the following representation:

$$Y_t - \mu = \varphi_1(Y_{t-1} - \mu) + \varphi_2(Y_{t-2} - \mu) + \cdots + \varphi_p\left(Y_{t-p} - \mu\right) + \varepsilon_t$$

with $\varepsilon_t \sim WN(0;\, \sigma^2)$, but it can also be represented, exploiting the Wold representation, as an $MA(\infty)$ stochastic process:

$$Y_t = \mu + \sum_{i=0}^{\infty} \varphi^j \varepsilon_{t-j}$$

where:

- $\zeta_0 = \varphi^0 = 1$
- $\sum_{i=0}^{\infty} |\varphi|^{2j} < \infty$ this condition is guaranteed if and only if $|\varphi| < 1$.

So, this condition guarantees that the AR process is second order stationary and ergodic.

The stationarity and ergodicity are properties which characterizes AR processes and they are satisfied, in the case of this specific model, if the autoregressive polynomial has roots that in absolute value are higher than one, that is

$$|z| > 1 \; and \; z = \frac{1}{\phi} \rightarrow \; |\phi_p| < 1$$

where the autoregressive polynomial is:

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$$

$$\phi(z) = 0$$

Then, with the assumption that the expectation is equal to zero, the properties of the model are the one reported below.

Expectation	$E(Y_t) = 0$
Autocovariance	$\gamma_j = \begin{cases} \dfrac{\sigma^2}{1 - \varphi^2} & j = 0 \\ \varphi^j \gamma_0 & j > 0 \end{cases}$
Autocorrelation	$\rho_j = \begin{cases} 1 & j = 0 \\ \dfrac{\gamma_j}{\gamma_0} = \dfrac{\varphi^j \gamma_0}{\gamma_0} = \varphi^j & j > 0 \end{cases}$

Table 3: Functions of $AR(p)$

Autoregressive Moving Average stochastic process of order $\boldsymbol{p, q}$

We can thus arrive at the definition of the ARMA model, as a combination of the two stochastic processes just described.
The autoregressive moving-average model $ARMA(p, q)$ is thus a generalization of the stochastic processes AR and MA, and is defined as below:

$$A(L)(Y_t - \mu) = C(L)\varepsilon_t$$

where

$$A(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$$
$$C(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$$

and p is the order of the polynomial $A(L)$ and q is the order of the polynomial $C(L)$.
For this reason, the AR and MA processes are special cases ($q = 0$ e $p = 0$, respectively).
An additional and more extensive notation of the model is:

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p\left(Y_{t-p} - \mu\right) + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$$

The Moving Average Autoregressive model is, as previously mentioned, a combination of two polynomials, and this union determines some important characteristics, already discussed, of the process: stationarity, ergodicity and invertibility.
If the AR and the MA polynomial share some common roots then there will always exists a stationary and invertible ARMA process with order p' and q', with $p' < p$ and $q' < q$.
Moreover, since it is desirable to have a parsimonious model, the choice will always fall in the model with smaller autoregressive and moving average order. This is due to the fact that the autocorrelation function these models gives rise is the same that generates the two processes with lower order.

Estimate of ARMA models' parameters

Another key aspect for our subsequent practical analysis is that of model parameter estimation.

If the stochastic process parameters are not known, and they never are, we need to use estimates. The most basic technique is that of Maximum Likelihood.

First things first the Likelihood is the sample's density function, computed in the correspondent point of the observed sample. This function depends on a vector ψ of unknown parameters, which will determine the shape of it; we call this new function as follows

$$L(\psi)$$

Maximizing this function, we are going to obtain the estimate of the maximum likelihood. When we observe a realization of a stochastic process, the likelihood function is nothing more than the joint density function of the observed part of the process, namely the marginal density function of the random vector, calculated in the observed values.

In the case of an ARMA process it will depend on the vector of parameters

$$\psi = (\mu; \phi_1, \dots, \phi_p; \theta_1, \dots \theta_q; \sigma^2)$$

If we assume that the process is Gaussian, the likelihood function is nothing but the density function of a normal multivariate:

$$L(\psi) = f(x; \psi) = \frac{1}{\sqrt{(2\pi)^T \Sigma}} e^{-\frac{(x-k)'(x-k)}{2\Sigma}}$$

where x is the vector of the T observations; k and Σ are its first and second moment, whose depends on ψ.

To be more specific, let's assume that we observe X_t, where t ranges from 1 to n. The X_t generating the data are independent random variables that are all identically distributed, and they all have a common probability density function, which we will refer to as f.

For each observation (x_t), we can compute the value of the probability density function of the corresponding variable, which will depend on some parameter values ψ, that we want to estimate. At this point we define the joint Probability Density Function of the n random variables, giving rise to the n observations in the sample. The joint PDF of these n random variables is equal to the product of the individual PDF, exploiting the total probability Law, as reported below

$$L(\psi) = f(x;\psi) = f(x_1, \dots, x_n;\psi) = \prod_{t=1}^{n} f(x_t;\psi)$$

Therefore, the joint PDF computed on the values of the n observations and for any fix value of the parameter vector, it is equal to the product of the n probability density functions of each one of the X_t, computed in the corresponding values of the observations and for a fix value of the parameter, which we assume known.
Next, by $\hat{\psi}$ we actually denote a function that we simply represent as follows

$$\hat{\psi} = t(x_1, x_2, \dots, x_n)$$

It can be considered as such if at each extracted sample, it assigns a value to the vector ψ that maximizes the likelihood function (MLE). In notation

$$\max L(x,\psi) = L(x,\hat{\psi}) \quad \rightarrow \quad \hat{\psi} = \arg\max L(x,\psi)$$

However, in order to calculate this estimator (MLE) we use the log-likelihood function, which is obtained through the application of the natural logarithm, so we have:

$$\ell(x,\psi) = \ln L(x,\psi) = \sum_{i=1}^{n} \ln f(x_i,\psi)$$

The function

$$L(\psi) = \prod_{t=1}^{n} f(x_t;\psi)$$

is a sort of link between what we have observed on the data and what is the probabilistic model that we assume for the random variables generating those observations (x_t).
Therefore, the Likelihood function is a very useful tool that allows us to compare parameters values in relative terms, meaning that it is coherent with both the model we have assumed and the data, or not.
Consequently, if our purpose is to find an estimate of ψ we might maximize the Likelihood, or the Log-Likelihood, which produces the same results. The logarithm is a monotonically increasing function, therefore, taking the logarithm of the likelihood function and maximizing it, or maximizing directly the likelihood function, gives us the same optimizer, the value of ψ that maximizes the likelihood is the same value that maximizes the log-likelihood. But working with the log-likelihood simplifies things a lot and has important inferential applications.
To continue, in order to maximize the log-likelihood function we have the following:

$$s(x,\psi) = \frac{\partial \ln f(x_i,\psi)}{\partial \psi}$$

where $s(x,\psi)$ is the Score function, which in the case of a sample with n units it is written as follows

$$s(x,\psi) = \sum_{i=1}^{n} s(x_i,\psi) = \sum_{i=1}^{n} \frac{\partial \ln f(x_i,\psi)}{\partial \psi}$$

Thus, the MLE estimate is that value of the vector ψ at which the score is equal to zero, and the likelihood and log-likelihood functions reach the maximum.
That said, when we estimate the unknown parameter ψ, we need to specify some kind of models for the process, for the variables generating the observations.
The estimate we get by maximizing the log-likelihood equation depends on the values that we have observed, of all time series. Clearly, if the time series is generated by a stochastic process the estimate that we get is a transformation of something which is determined by chance. Therefore, the value of the estimates that we get of the unknown

parameter is the determination of something that is random, which is called the estimator of the unknown parameter vector ψ.

How to choose the order of processes

Obtained the parameter estimates, we want now to select the order of the process, consequently it is essential to have a method that can be used to select the optimal one.
There are different criteria in order to select the optimal order for a stochastic process, but the one that we are going to consider is the Ljung-Box Test[11]. This test might allow us to assess whether an observed time series is generated by a white noise or not. it is done by checking if there exists autocorrelation in the time series.
The Ljung-Box test is computed according to the following relationship:

$$Q_{LB} = n(n+2)\sum_{j=1}^{h}\frac{\hat{\rho}_j{}^2}{n-j} \sim \chi_h^2 \quad under\ H_0$$

where n represents the length of the time series, and where we have the sum of the squares of the estimate of all the first j autocorrelation coefficient.
We want to assess whether the first h autocorrelation coefficient is only identically equal to zero.
In general, this test is defined as:

- null hypothesis (H_0): model does not exhibit lack of fit;
- alternative hypothesis (H_1): model exhibit lack of fit.

It can be shown that asymptotically for sufficiently long time series (provided the stochastic process stationary) this test statistic is distributed as a chi-square (χ_h^2) with h degrees of freedom. Accordingly, we will be induced to reject the null hypothesis whenever this test statistic takes sufficiently high values. The decision of rejecting or not the null hypothesis is based firstly on the significance level (α)[12], and secondly by a tool, called P-Value ($\hat{\alpha}$), which is a statistic that works as follows:

[11] This test belongs to a set of asymptotic tests.
[12] The significance level alpha is the probability that our tets has rejected H_0, when H_0 is true (Type I Error).

- $\hat{\alpha} < \alpha$ we reject H_o
- $\hat{\alpha} > \alpha$ we do not reject H_0.

On the basis of these considerations, we will then go on to draw a number of conclusions about the results we are going to obtain in the next part of the analysis.

Identification of the model through the Autocorrelation and the Partial Autocorrelation Function

Simultaneous with the step of choosing the process order, another type of investigation will be carried out; this one concerns the identification of the model through the autocorrelation and partial autocorrelation functions.

In the most common sense, correlation in general represents any relationship, not necessarily linear, between two variables. In statistics, in more detail however, it describes the extent to which two signals have common properties, hence resemble each other, as a function of their mutual lag in time. Rather, when we speak of an autocorrelation function (ACF), we mean the measure of when a signal resembles, thus correlates, with itself lagged by a given time frame. On the other hand, the Partial Autocorrelation Function (PACF) is a measure of the strength of a linear relationship between observations in a time series, say between y_t and y_{t-k}, after eliminating the effects of intermediate linear relationships, or lags in between $(1, 2, \dots, k-1)$.

That said, previously treated processes can be recognized, according to their characteristics, by means of the autocorrelation function and the partial autocorrelation function. The time series that are nothing more than the realization of these processes, should reflect, in their variations, the properties of these models (taking into account the limitation of the range considered). Box and Jenkins, starting from these assumptions, have developed an approach that, using statistical tools, allows to model the time series by identifying the ARMA model that best manages to adapt to the phenomenon in question.

Through the ACF and PACF it is possible, sometimes, to recognize the process we are interested in. Therefore, looking at the ACF and at the PACF, there are a series of rules through which we can identify the three types of Univariate models previously described.

The $AR(p)$ process is characterized by an ACF that gradually decays and a PACF which is truncated at lag p. The $MA(q)$ model is, instead, defined by an ACF that is truncated at lag q and a PACF which gradually decays (geometric decay). To conclude, the $ARMA(p, q)$ stochastic process has both the ACF and PACF described by a geometric decay (they decay exponentially fast).
In this way every stochastic process is going to have its own peculiarities.

Unit roots test: Augmented Dickey-Fuller Test and KPSS Test

As was already anticipated in the explanation regarding stationarity, there are methods that allow us to verify the stationary condition of a series.
There are a set of tests to control for the presence or absence of unit roots that are more rigorous than the simple graphic analysis of the series.
The first we are going to consider is the Augmented Dickey-Fuller test (ADF) which was created by Dickey and Fuller in 1981.
In theory, if we have an $ARMA\ (p, q)$ model this process is second order stationary if only if:

$$1 - \phi_1 L - \cdots - \phi_p L^p = 0 \ \ with\ |z| > 1$$

or it is different from zero for any z, which in absolute value is not greater than one, that is the roots of the equation must be greater than one in absolute value.
Therefore, the aim of this test is to check weather this condition holds or not.
More in detail, suppose that exists:

$$Y_t = \phi Y_{t-1} + u_t$$

where ϕ represents any parameter and u_t is not necessarily a white noise but it is assumed to be a stationary process. If we take the first differences, both from the right and from the left, we get the following:

$$Y_t - Y_{t-1} = \phi Y_{t-1} - Y_{t-1} + u_t$$

$$Y_t - Y_{t-1} = (\phi - 1)Y_{t-1} + u_t$$

Then, we call $\pi = \phi - 1$.

$$\nabla Y_t = \pi Y_{t-1} + u_t$$

If there exists a unit root, $\phi = 1$, it means that the new parameter π is equal to zero, so that we are in a non-stationary case. If, instead, there are no unit roots, we get that $|\phi| < 1$, meaning that π is less than zero, we have a second order stationarity. Therefore, this is exactly the hypothesis system that we want to assess, in which the null hypothesis is equivalent to the non-stationarity of Y_t and the alternative one defines the second order stationarity of the same process.

$$\begin{cases} H_0: \pi = 0 \quad \phi = 1 \\ H_1: \pi < 0 \quad |\phi| < 1 \end{cases}$$

The KPSS test, named after the authors Kwiatkowski, Phillips, Schmidt and Shin, is another test to verify the stationarity of a time series. The null and alternative assumptions for the KPSS test are opposite to those for the ADF test, therefore:

- H_0: The process is stationary
- H_1: The series has a unit root, implying the non-stationarity.

Generally, in most practical cases, it is always better to apply both tests, so that we can ensure that the series is truly stationary.

Analysis of residuals

Something that was included in the previously mentioned Box-Jenkins method is the analysis of residuals. This gives us information about the validity of the model, and it is carried out by analyzing the behavior, indeed, of residuals, through the observation of the presence of some peculiarities. Residuals need to be:

- independent to each other (i.i.d.);
- normally distributed with zero mean and constant variance

It is possible to perform this analysis by way of some tests (for instance the autocorrelation test) that we are going to see in the applied part.

Autoregressive integrated moving average stochastic process of order p, d, q

Since now what we have talked about was stationary processes, nevertheless in the reality things are sometimes different. For the purpose of our future analysis is required to introduce also the following aspects.

Most of the time dealing with nonstationary time series we can talk about ARIMA models, that is autoregressive integrated moving average models. We are thus referring to a particular, but common, category of models always suitable for investigating time series which, however, have special characteristics. The part that differentiates them from the stationary ARMA models, as can be seen, is the integrated process part.

The definition of integrated process states that X_t is an integrated process of order d if it has a stationary and invertible $ARMA$ representation after d differences have been performed on X_t and it is denoted $X_t \sim I(d)$.

For example, if X_t is an integrated process of order 1, then the process:

$$\nabla X_t = X_t - X_{t-1} = \varepsilon_t$$

is stationary.

So, as we have already said, it can be represented through the Autoregressive Integrated Moving Average model (ARIMA) of order p, d, q.

Knowing that an ARMA model is defined as abovementioned, the notation that describes this non-stationary model is the following:

$$X_t \sim ARIMA(p, d, q) \qquad \rightarrow \qquad Y_t = \nabla^d X_t = (1 - L)^d X_t$$

where Y_t is an ARMA model of order p, q.

Seasonal ARIMA (SARMA)

Finally, the presence of any seasonality in the data it must also be taken into account.
Seasonality consists of periodic movements that are repeated more or less regularly throughout the year and from one year to the next.
In order for there to be periodicity, the same trend must be repeated year by year, or rather, the fastest periodicity must be repeated more or less regularly in the slower one.
Seasonality can arise, as it will be in the case of some historical series that we are going to analyze, for example from the succession of the seasons, to the climatic variations due to the Earth's rotation around the sun, etc.
The extension of the Box-Jenkins methodology to seasonal time series can be done by applying an $ARIMA\ (P, D, Q)_S$ model to periodic subsets, that is, a linear stochastic process specific to each "season" which therefore, at least potentially, is different for different periods. The notation has been slightly modified using capital letters to emphasize the fact that the field of application is now seasonal, but the rule of indicating with the P the Autoregressive part, the D for the order of differentiation and Q for the moving average part has remained unchanged.
The seasonal ARIMA model (SARIMA) is written in the following way:

$$ARIMA\ (p, d, q)(P, D, Q)_S$$

Therefore, the non-seasonal terms are simply multiplied by the seasonal ones.

Multivariate stochastic processes

Unlike univariate stochastic processes that consider only one time series, a multivariate stochastic process is a vector whose elements are univariate stochastic processes. This means that the elements of a multivariate model are multiple random variables.
In the multivariate universe the stationarity and ergodicity conditions stay the same, while first and second moment we have what reported below:

$$\begin{cases} E(y_t) = \mu = \begin{bmatrix} E(y_{1t}) \\ E(y_{2t}) \\ \vdots \\ E(y_{nt}) \end{bmatrix} \quad for\ all\ t \\ E[(y_t - E(y_t))(y_{t-k} - E(y_{t-k})'] = \Gamma_k \quad for\ all\ t\ and\ k \end{cases}$$

where if the process y_t has n elements, μ is a $n \times 1$ vector and Γ_k is a matrix $n \times n$ in which:

- For $k = 0$ Γ_k is the variance-covariance matrix of vector y_t
- For $k \neq 0$ the ij elements of Γ_k represents the covariance between the i-th element of y_t and the j-th elements of y_{t-k}.

The autocovariance matrix is defined in such a way that $\Gamma_k = \Gamma'_{-k}$.
In other words, the VAR is considered weakly stationary if the mean of all endogenous variables in the system are the same across time, and if the covariance matrix if y_t and y_{t-k} depends on the time lapsed k.

Vector Autoregressive models

The vector autoregressive model is a model that falls into the category of multivariate stochastic processes, and it is the one we are going to use next to carry out our analysis.
The VAR model was introduced in 1980 by Christopher Sims and its goal is to develop a model that can be economically evaluated, enabling the identification of a long-term relationship involving all the variables taken into consideration.
Sims criticized the use of SEM (Structural Equation modelling)[13] since he thought, conversely to what this approach considers, that variables were all endogenous, in such a way as to give a precise statistical description of the variables under analysis.
From an econometric point of view the VAR estimation is a procedure which can be interpreted as the estimation in the reduced form of a simultaneous equation model.
In the past, VAR models were marked as anti-theoretical models. This is because the VAR model does not include identification restrictions, given that the goal of those who want to estimate the multivariate model is not explain the casual relationships, but only to find a statistically accurate description of the characteristics of persistence of a set of series.
The Vector Autoregression model is a forecasting and structural analysis tool that can be used whenever two or more time series influence each other.
With respect to the univariate models in which it is imposed a unidirectional relationship, here the relation between time series is bidirectional, because the variables influence each other.
A VAR is a linear model in which each given economic variable included in a vector of n indicators is explained by its own past values and the past values of all other $n-1$ variables belonging to that vector. This simple structure provides a systematic way of capturing rich dynamics of data series linked to different time moments.

[13] Structural Equation Modeling is a multivariate statistical analysis technique that allows to verify hypotheses about the influence of a set of variables on others. with expected *a-priori* conditions. SMEs correspond to a family of related procedures aimed at examining the linear relationships between one or more independent variables and one or more dependent variables, which can be measured, that is directly observable, or latent (not directly observable and thus indirectly measured by two or more detectable indicators).

The Vector Autoregressive model of order p [14], which is a generalization of the univariate Autoregressive model of order p, belongs to multivariate stochastic processes (or multivariate linear time series models), more specifically to the VARMA[15] family (extension of the ARMA model presented in the previous chapter).
The VAR models are largely used to investigate aspects of the association among the variables of interest, since they represent the correlation between a set of variables.
The model is defined by two elements: the order of the process, indicated by p, and the number of equations, indicated by k.
Therefore, a vector autoregressive model of order p, and k dimension, can be represented by the following structures:

- using the backshift operator

$$y_t = A_o + \left(A_1 L + \cdots + A_p L^p\right) y_t + \varepsilon_t \quad \rightarrow \quad A(L) y_t = A_o + \varepsilon_t$$

- instead, without the lag operator notation, we have:

$$y_t = A_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t$$

where:

- Y_t is a random vector $(n \times 1)$
- A_0 is an intercept vector $(n \times 1)$
- A_p are fixed coefficient matrices $(n \times n)$, tells us which is the magnitude of each effect.
- ε_t is a white noise vector $(n \times 1)$, white noises processes that may be simultaneously correlated

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} A_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{bmatrix} Y_{t-1} + \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

[14] Where p denotes number of lags of the model, as it was for the univariate AR model.
[15] VARMA processes are complicated to estimate when the polynomial has order higher than zero. Indeed, it is more common to use in empirical applications the VAR model.

The VAR can be also written in a compact form, which is called the companion form. It is a way to write a VAR of any order p comparable to a $VAR(1)$:

$$Y_t = A_0 + \boldsymbol{A} Y_{t-1} + \varepsilon_t$$

If up to now we have spoken in general terms, we are now looking at the simplest case of VAR, the $VAR(1)$ with two variables ($k = 2$), that is written as follows:

$$\begin{cases} y_{1,t} = \alpha_{10} + \beta_{11} y_{1,t-1} + \beta_{12} y_{2,t-1} + \varepsilon_{1,t} \\ y_{2,t} = \alpha_{20} + \beta_{21} y_{1,t-1} + \beta_{22} y_{2,t-1} + \varepsilon_{2,t} \end{cases}$$

In matrix notation:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \alpha_{10} \\ \alpha_{20} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

which is equivalent to the following companion form:

$$Y_t = A + B Y_{t-1} + \varepsilon_t$$

where

- $Y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}$
- $A = \begin{bmatrix} \alpha_{10} \\ \alpha_{20} \end{bmatrix}$
- $B = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}$
- $Y_{t-1} = \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix}$
- $\varepsilon_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$

To signify that we are treating with endogenous variables, variables which influence each other equally, we have changed the notation: $y_{1,t}$ denotes the t-th observation of variable y_1, and $y_{2,t}$ the t-th observation of variable y_2.

VAR stability and stationarity

Establishing the existence of stationarity in this multivariate stochastic process is fundamental, as it was in the case of univariate.

Even before discussing about stationarity, an important characteristic that it must be taken into account is the stability condition for the eigenvalues of the matrix $\boldsymbol{A}$. Indicating with $A(z) = I_k - A_1 z - \cdots - A_p z^p$ the characteristic polynomial of the VAR(p) model, then the process $\mathrm{Y_t}$ is stable if the determinant of the $A(z)$ matrix is different from zero for all the z less or equal than one in absolute value.

$$\det(I_n - \boldsymbol{A}z) \neq 0 \quad for\ |z| \leq 1$$

or written as:

$$\det\left(I_n - A_1 z - \cdots - A_p z^p\right) \neq 0 \quad for\ |z| \leq 1$$

Therefore, if the stability condition is verified, the model is also stationary.

Instead, if the model is stationary (first and second moments are time invariant), does not necessarily signify that the model is also stable. An unstable process is not necessarily non-stationary.

As we did for the univariate case, also the VAR can be represented with the Wold Representation. A prerequisite to be written in the way we said, considering the VAR with the lag operator, is that the $A(L)$ polynomial must be invertible, that is if all the np roots of the characteristic equation lie outside the unit circle. If this is the case, we have that it is possible to represent Y_t as the sum of all the past white noise shocks ε_t through a $VMA(\infty)$ representation.

$$Y_t = A_0 + \sum_{i=0}^{\infty} \boldsymbol{\psi}^j\, \varepsilon_{t-j}$$

where ψ is an $n \times n$ coefficients' matrix of lagged values of the error term.

Optimal VAR length

As we did for the univariate case, the choice of the number of p of delays to be included is one of the crucial aspects in specifying a VAR. Many delays results in a very large number of coefficients to be estimated and can cause a sharp decrease in the accuracy of the estimates and a substantial increase in the prediction error.
In other words, we have:

$$n + (n \times n \times p) = n\big(1 + (n \times p)\big)$$

where n is the number of endogenous variable and p is the model order.
Therefore, what reported above represents the total number of parameters to be estimated. Since the VAR of order p has p matrices with dimension $(n \times n)$, each for A_1, A_2, ..., A_p, and everyone represents the coefficients vector linked with the lag on anyone endogenous variable.

A way to determine the lag length of the model is to use the model selection criteria, meaning fitting different model to the same data set. These criteria consist in fitting $VAR(p)$ models with order from 0 to p_{max} and choosing the value p that minimizes some model selection criteria.
A general representation for these model selection criteria is described below:

$$IC(p) = \ln\left|\tilde{\Sigma}(p)\right| + c_T \cdot \phi(n,p)$$

where $\tilde{\Sigma}(p) = \frac{\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t{}'}{T}$ is the residual covariance matrix without a degrees of freedom correction from a VAR(p) model, c_T is a sequence indexed by the sample size T, and $\phi(n,p)$ is a penalization term, penalization because the general rule is to penalize too complicated models, that is with too many parameters.
More specifically we can distinguish three main information criteria, the Multivariate Akaike Information Criterion, the Multivariate Schwarts (or Bayes) Information Criterion, and the Multivariate Hannan-Quinn information criterion, respectively:

$$MAIC(p) = \ln\left|\tilde{\Sigma}(p)\right| + \frac{2}{T} \cdot pn^2$$

$$MBIC(p) = \ln\left|\tilde{\Sigma}(p)\right| + \frac{\ln T}{T} \cdot pn^2$$

$$MHQ(p) = \ln\left|\tilde{\Sigma}(p)\right| + \frac{2\ln(\ln T)}{T} \cdot pn^2$$

The first model selection overestimates the order, which means that the criterion might push to accept or fit models with too many parameters, whereas the other two estimate the order consistently under fairly general conditions if the true order p is less than or equal to p_{max}, that is avoid the risk contemplated in the AIC case.

Estimate of VAR parameters

The vector autoregressive model is a system of equations that represents the entire structure of dynamic correlations between relevant variables. All the variables are assumed to be endogenous, and their simultaneous linkages are "hidden" in the variance/covariance matrix of the error terms, which is typically a non-diagonal matrix.
The two methods to estimate the model's parameters consist in the use, mainly, of the Ordinary Least Square (OLS) and Maximum Likelihood Estimation (MLE) methods, but the one that we are going to apply in our practical analysis is the OLS. However, the maximum likelihood estimator of the VAR parameters asymptotically coincides with the OLS estimator.
The Ordinary Least Square method is a common criterion to estimate models., and it consists in minimizing the sum of squares of the differences between the observed dependent variable in the given dataset and those predicted by the model and does not require any distribution assumption.
Moreover, when the number of variables is considerable, and consequently the number of parameters to be estimated increases significantly, to say the least, we are in the presence of overfitting problems, consequently there is the possibility of resorting to penalty methods such as Lasso, which, for example, performs simultaneously the selection of variables (Shrinkage method) and the estimation of parameters.

Impulse response functions (IRF)

There are several methods for interpreting the estimated model, and one of them is, for example, the impulse response function (IRF).
It can be of relevance to know the response of one variable as a result of an impulse generated by another variable.
Impulse response functions are a graphical representation of an impact of an exogenous shock on VAR variables. Therefore, an IRF is a function that analyzes the output of that system, caused by the input, called an impulse. In general, an impulse response refers to the reaction over time of a dynamic system to some external change. In the economic field, particularly in macroeconomic modeling, impulse response functions describe how the economy reacts over time to exogenous impulses, called "shocks".
Thus, these particular functions summarize the behavior of a variable over time in response to a unit shock occurred in all other variables of the system. This is made possible by the fact that often the shock in one variable directly affects the variable itself but is also transmitted to the other endogenous components of the model according to the "dynamic" mechanism of VAR. If one variable reacts to the change of another, the first cannot be called exogenous for the system (only if one variable is exogenous in the strict sense the responses of this to shock occurred in the other variables will be null).
The assumption that shocks occur on one variable at a time, which is plausible only if the errors are uncorrelated, is one issue with the study of the impulse response functions. However, it is feasible to discover that they are strongly correlated (i.e., they share a component that cannot be assigned to a particular variable) in reality, which makes it so that a shock in one variable is accompanied by other shocks in the system.
One solution to this issue is to assign any shared component's whole effect to the variable that appears first in the VAR system.
When it comes to outside influences, they could also be connected to missing variables, whose impact is reflected in the errors that weren't caught.
If the system contains k variables, then we should assume that a total of k^2 impulse responses will be generated. The shocks might also be either temporary or permanent. Therefore, there is a chance that we would experience a prolonged shock if the path indicated by the IRF is constantly growing or decreasing as t increases or the horizon

travels farther. However, most of the time, we frequently observe brief shocks or departures from the stable state trend values.

Univariate Analysis

The study area, which this research is looking at, regards the timeframe going from January 2019 to December 2021, with daily observations.
The data we are looking at regarding the Bitcoin cryptocurrency was found through the R package crypto2, while the ones concerning CO_2 emissions of the power sector of Liaoning and Qinghai regions, are part of a project, called "Carbon Monitor"[16], led by Tsinghua University, Laboratoire des Sciences du Climat & de l'Environnement, University of California (Irvine), and Chinese Academy of Sciences.

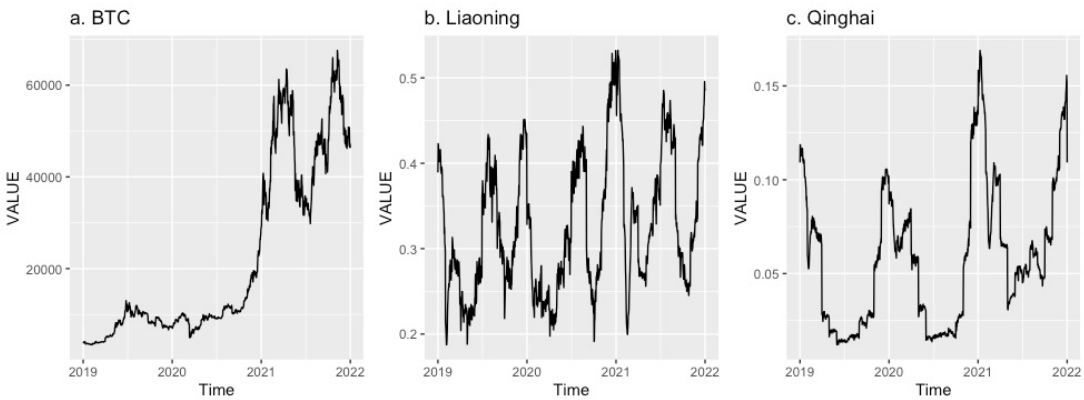


Figure 3: Time series plots: BTC, Liaoning CO_2 emissions and Qinghai CO_2 emissions.

For example, a first graphic analysis of the data set can give an idea of the stationary or non-stationary of the observed series. The time series reported in Figure 3 shows different characteristics. Starting from plot 3a, we can notice two mainly fact, first that there are some trends, secondly the presence of an upward trend. It is possible to say the same for 3b graph. Therefore, it is clear a non-stationary nature of the data, which in the case of cryptocurrencies series is quite normal lately. Lastly, looking at Figure 3c and 3d it is possible to notice a sort of seasonality.
«*[…]Bitcoin accounts for 2/3 of the total energy consumption, and under-studied cryptocurrencies represent the remaining 1/3*» (Gallersdorfer, U., Klaaßen, L., Stoll, C.

[16] Carbon Monitor is a frequently updated daily CO_2 emission dataset, to monitor the variations of CO_2 emissions from fossil fuel combustion and cement production since January 1st 2019 at national level with near-global coverage.

2020), so that we can take into account only the Bitcoin, given its representativeness for our analysis.
Having introduced so, the following plot helps us to have a clearer representation of the time series we consider.

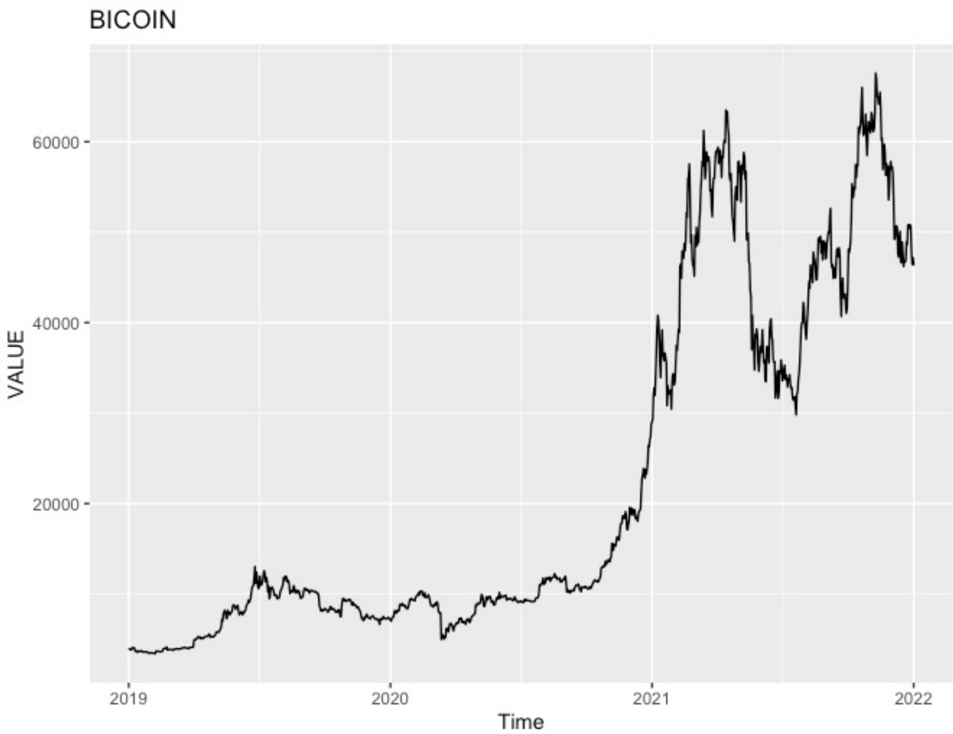


Figure 4: R Autoplot of Bitcoin values ($) of the period
1/01/2019 – 31/12/2021
Source of data R package "crypto2"

From this first graph we can notice the upward trend abovementioned with an obvious increase between the end of 2020 and the beginning of 2021.
Starting from Figure 4, but then more in detail from Table 4, we can observe the way in which the cryptocurrency under analysis went from a value of $3.399, at the beginning of 2019, until reaching its maximum value of $67.567, at the end of 2021.

Statistical measures	**Values ($)**
Min	3.399
1^{st} Quartile	8.206
Median	10.580
Mean	21.973
3^{rd} Quartile	38.202
Max	67.567

Table 4: R summary function which provides a couple of basic statistical measures on the Bitcoin time series

This delta is due to the so-called tokenization that is the phenomenon though which cryptocurrencies are trying to offer to the economic system some added value in terms of decentralization of finance, payment systems, and digitization of real assets.
The field of application is extremely wide (from the banking, financial, recreational, real estate, artistic, etc.) and this explains the growing interest of long-term investors.
Subsequently, looking at the unit root tests (Table 5), since the p-value of the ADF test is greater than 0.05 we fail to reject the null hypothesis. This means that the time series is non-stationary, in other words, there is evidence of a time dependent structure. Indeed, considering also the p-value of the KPSS test, as a confirmation, is less than 0.05 meaning that the null hypothesis of stationarity is rejected.

	Test statistic	**P-value**
ADF	$-2{,}1536$	$0{,}5133$
KPSS	$4{,}457$	$0{,}01 \times 10$[17]

Table 5: R adf.test and kpss.test command of BTC

Confirming the non-stationarity of the process, we funded the model which best fits the data, that is an $ARIMA(2{,}1{,}3)$ with drift, as we can see from Table 6 reported below.

[17] The p-value is smaller than the printed p-value.

	Ar1	Ar2	Ma1	Ma2	Ma3	drift
coefficients	$0{,}8357$	$-0{,}9650$	$-0{,}8668$	$0{,}9791$	$-0{,}0023$	$11{,}2880$
s.e.	$0{,}0252$	$0{,}0252$	$0{,}0308$	$0{,}0377$	$0{,}0199$	$12{,}4398$

Table 6: R Auto.arima function results of BTC

Dealing with integrated series, which for definition is a non-stationary one, we have to transform it, by differentiating the series. The following results are obtained:

	Model	ACF test	KPSS test	Ljung-Box test
		P-value		
Bitcoin	$ARIMA(3,0,2)$ $with\ zero\ mean$	$0{,}01$	$0{,}1$	$1{,}11 \times 10^{15}$

Table 7: results of the model and the tests after having differentiated

A lack of fit is shown by the results from table 7, however by looking at the plot, which is the p-values for the Ljung-Box, given by figure 5, uncorrelation for the first 8 lags are pictured. However, after that the behavior starts to present some dependence.
Starting from the end of 2020, looking at the plot of the standardized residuals, periods of higher volatility keep arising, as reflected in the third plot of the p-values.

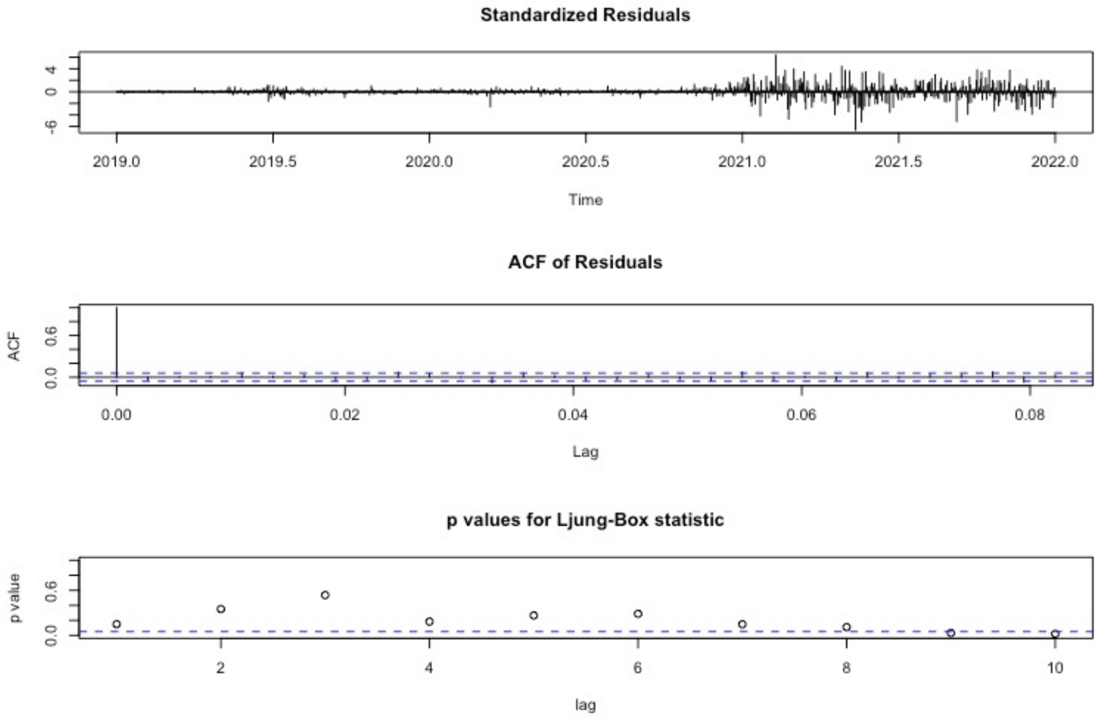


Figure 5: tsdiag function in R of Bitcoin

Furthermore, a noticeable point in this evaluation is how satisfactory the model moves around the mean, despite the non-negligible volatility change in correspondence of the beginning of 2021, which can be solved using some alternative models, such as the GARCH process, representing an approach which estimate the volatility.

Moving forward, for our univariate analysis, we have to consider the CO_2 emissions of the two Chinese regions: Liaoning and Qinghai.

As we have already mentioned before, we have for both the regions a seasonal pattern.

The Augmented Dickey–Fuller (ADF) test for the presence of unit roots and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test to verify the null hypothesis of stationarity were then performed, the results are the one represented in Table 7.

a. LIAONING	Test statistic	P-value
ADF	$-3{,}6452$	$0{,}02833$
KPSS	$0{,}15487$	$0{,}0426$

b. QINGHAI	Test statistic	P-value
ADF	$-2{,}971$	$0{,}1673$
KPSS	$0{,}41485$	$0{,}01$

Table 8: R adf.test and kpss.test command

From the tables reported, it is observable that in relation to the results about Qinghai the time series is clearly non-stationary, actually, we fail to reject the null hypothesis using both tests. On the other side, looking at the first one (8a) the two tests are contrasting. As a result, the models obtained are respectively an $ARIMA(2,1,0)(0,1,0)_{365}$ and an $ARIMA(0,1,1)(0,1,0)_{365}$.

LIAONING	Ar1	Ar2
coefficients	$-0{,}1415$	$-0{,}1419$
s.e.	$0{,}0366$	$0{,}0367$

QINGHAI	Ma1
coefficients	$-0{,}2151$
s.e.	$0{,}0368$

Table 9: R auto.arima function results

Concerning the Liaoning region, we can clearly see the seasonality of the time series from the autocorrelation function reported below.

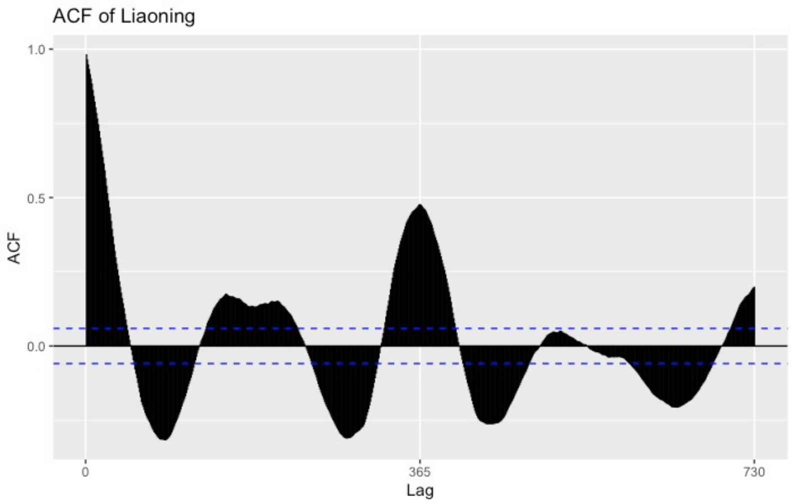


Figure 6: ACF of Liaoning CO_2 emissions time series

This plot represents the need to differentiate the two time series to make them stationary, due to the fact that the seasonal feature characterizing the time series has a strong pattern which we have to eliminate it.
By doing so, the results seasonal pattern released are reported:

Chinese region	**Model**	**ACF test**	**KPSS test**	**Ljung-Box test**
		P-value		
LIAONING	$ARIMA(1,0,0)$ $with\ zero\ mean$	0,01	0,1	$< 2{,}2 \times 10^{16}$
QINGHAI	$ARIMA(2,0,1)$ $with\ zero\ mean$	0,01	0,1	0,000591

Table 10: results of the model and the tests after having differentiated

Evaluating the Ljung-Box Test results we can safely reject the null hypothesis, demonstrating once again, the exhibitions of the model lack of fit.
To support this strong implication figure 7, still glimpse a sort of seasonal pattern.

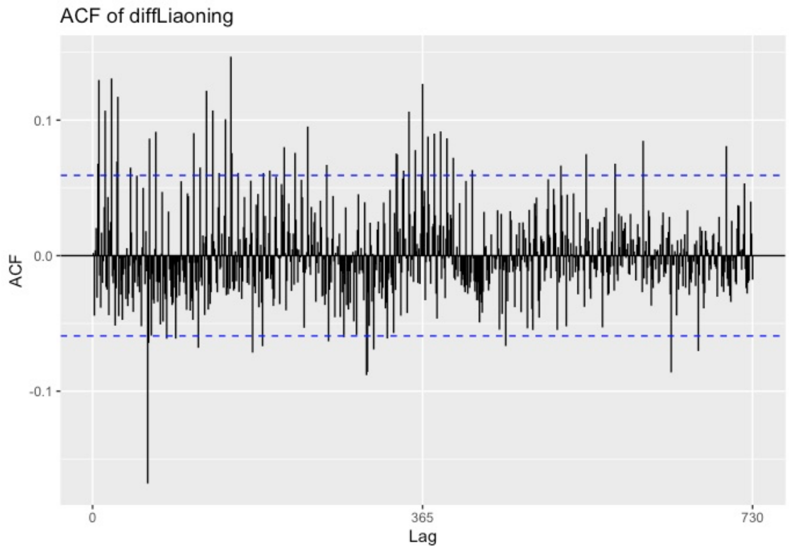


Figure 7: ACF Liaoning plot

Indeed, as the ACF Liaoning plot demonstrate, we can see high values of the estimated autocorrelation coefficient function which fall outside the blue lines, which is a sign of either the dependence presence, or temporal autocorrelation, among the random variables. This is also partially[18] confirmed by the results reported in Figure 8 regarding the Ljung-Box test.

[18] Partially because until lag 6 the behavior of the P-values for Ljung-Box statistic plot works well, due to the reason that the residuals are uncorrelated.

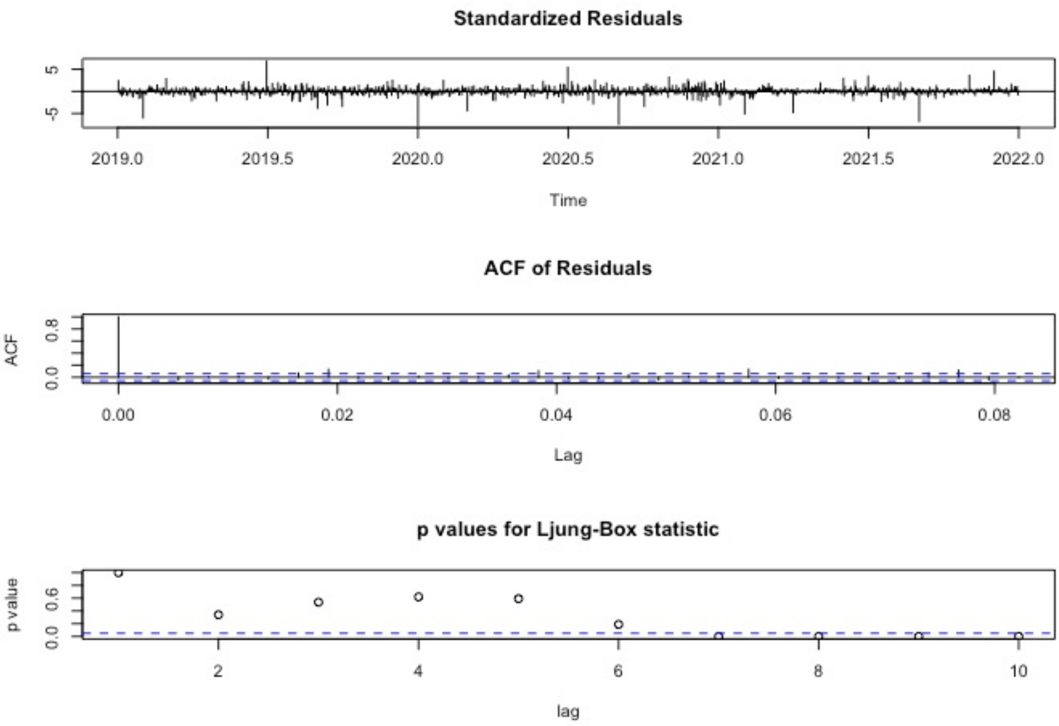


Figure 8: tsdiag function in R of Liaoning CO_2 emissions

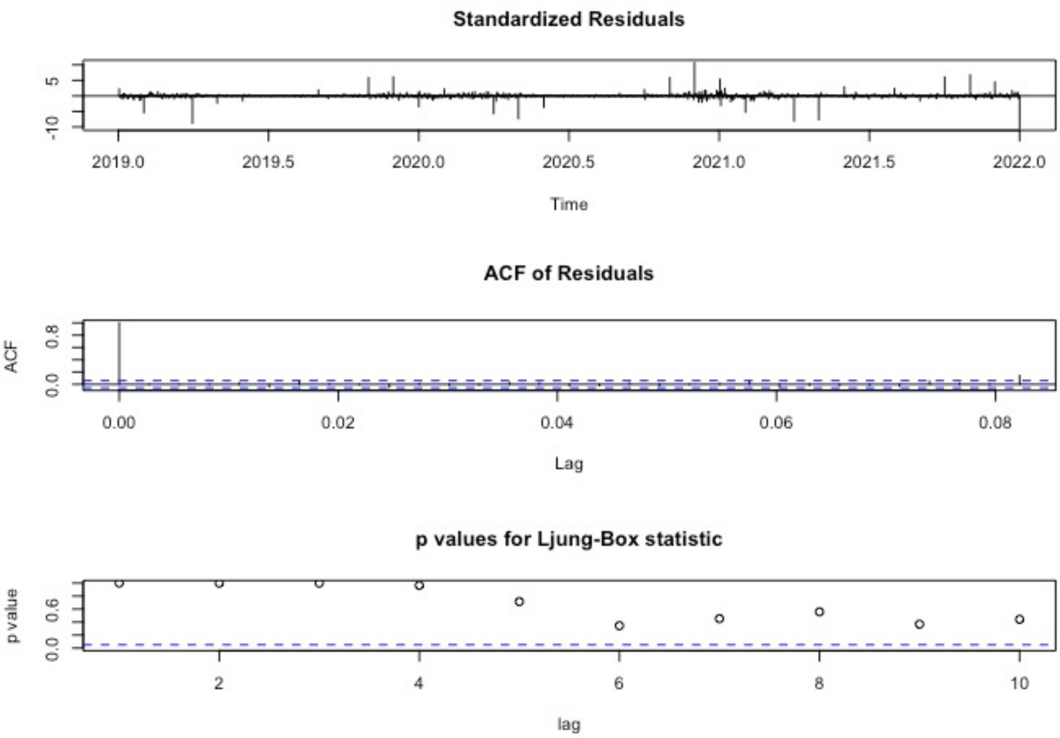


Figure 9: tsdiag function in R of Qinghai CO_2 emission

By contrasting the abovementioned dependence theory, on the other hand, by using the tsdiag function of the Qinghai CO_2 emissions (Figure 9) we found a nicer behavior of the residuals.

Figure 8 and 9 are the conclusive plots of this chapter needed to give a first picture of the type of time series we are going to contemplate in the next part which is the multivariate analysis.

Multivariate Analysis – VAR model

The simultaneous analysis, which is the key aspect of this final investigation, using the already mentioned Vector Autoregressive Model, is built among the CO_2 emissions of the Chinese regions and the value of the principal cryptocurrencies (BTC, ETH, ADA, BNB, USDC, USDT, XRP).

According to what was presented in the theoretical part, models are selected according to information criteria. The results obtained are reported below.

AIC(n)	HQ(n)	SC(n)
9	2	1

Table 11: Liaoning CO_2 emissions and cryptocurrencies
R LagSelection function (Akaike Information Criterion, Hannan-Quinn Criterion, Schwarz Criterion)

AIC(n)	HQ(n)	SC(n)
20	2	1

Table 12: Qinghai CO_2 emissions and cryptocurrencies
R LagSelection function (Akaike Information Criterion, Hannan-Quinn Criterion, Schwarz Criterion)

In both cases (Table 11 and 12), in relation to the order choice, the Schwartz Information Criterion (SC or BIC) was used, as it is more parsimonious in the number of lags to be included in the model, thus in the estimation is possible to avoid burning too many degrees of freedom.

In both cases, the model we are going to estimate is going to be a $VAR\ (1)$.

	Liaoning.l1	BTC.l1	ETH.l1	ADA.l1	USDT.l1	USDC.l1	XRP.l1	BNB.l1	Const
Liaoning	0,0025332	−0,0000002	0,0000044	−0,0051476	0,0631706	0,0024607	−0,0102973	0,0000366	0,0000452
BTC	−5121,0356367	0,1201765	−1,8010470	760,2986705	8377,8898634	−5375,8458271	−2763,9983392	−2,7489072	41,8550764
ETH	−367,5335089	0,0115570	−0,1146706	42,2351969	289,6798432	−129,7459855	−195,8713277	−0,5326797	3,4588384
ADA	−0,1259175	0,0000011	−0,0000338	−0,0055442	0,1393318	−0,1514288	−0,0952541	−0,0002265	0,0013887
USDT	0,0123949	0,0000000	0,0000004	0,0004215	−0,3421565	−0,1143640	−0,0003122	−0,0000039	−0,0000284
USDC	0,0079109	0,0000001	−0,0000001	0,0004403	−0,0042320	−0,3808432	−0,0014130	−0,0000037	−0,0000226
XRP	−0,1867350	0,0000028	−0,0000682	0,0111004	0,0984504	−0,0069390	−0,0424603	0,0000463	0,0005314
BNB	−38,8007291	0,0011270	−0,0024298	4,0967654	−4,6022762	−31,7536919	−28,51557741	−0,1564558	0,5087513

Table 13: Liaoning coefficients matrix

	Qinghai.l1	BTC.l1	ETH.l1	ADA.l1	USDT.l1	USDC.l1	XRP.l1	BNB.l1	Const
Qinghai	−0,0196805	−0,0000001	0,0000028	−0,0006896	0,0189723	0,0192789	−0,0034752	0,0000042	−0,0000123
BTC	−4367,4395704	0,1201995	−1,7592107	760,5463764	8887,3095428	−5804,9817684	−2757,3327980	−2,8175590	41,4694553
ETH	−718,0003206	0,0115181	−0,1118798	42,6914535	315,4173459	−166,6111190	−196,5594505	−0,5335363	3,4473961
ADA	−0,1590818	0,0000011	−0,0000328	−0,0054821	0,1504745	−0,1627557	−0,0952393	−0,0002277	0,0013812
USDT	0,0020761	0,0000000	0,0000003	0,0004301	−0,3436168	−0,1134527	−0,0003528	−0,0000036	−0,0000271
USDC	−0,0018043	0,0000001	−0,0000001	0,0004492	−0,0052478	−0,3803085	−0,0014480	−0,0000035	−0,0000217
XRP	−0,3501764	0,0000028	−0,0000668	0,0113164	0,1119182	−0,0254501	−0,0427677	0,0000457	0,0005250
BNB	−68,2873924	0,0011237	−0,0021312	4,1367903	−1,6841710	−35,5329173	−28,5667582	−0,1566219	0,5072395

Table 14: Qinghai coefficients matrix

The matrices above represent the estimated coefficients of the model. Of particular importance is the structure of these matrices, due to the fact that gives information about the temporal dependence between the time series constituting the multivariate model. Indeed, the values of most coefficients, in both Table 13 and 14, are basically negligible given they are very close to zero. As demonstrated, there are some relevant (negative) dependences, between the data of both regions, concerning the emissions and the firsts two major cryptocurrencies (BTC and ETH).

The size of the coefficient for each independent variable gives the magnitude of the effect that the variable is having on the dependent variable, and the sign of the coefficient (positive or negative) gives the direction of the effect. This signifies that in this case the independent variable of Bitcoin, for example, is expected to decrease 5121 times when the dependent one, the Liaoning emissions, increases by one. Despite of what has been found, this is clearly contradictory to what we are trying to prove, since cryptocurrency transactions would even seem to "reduce" the carbon dioxide emissions of the Chinese region.

To this purpose, the below correlation matrices can also be of greater interest for the analysis, because, in general, it allows us to see the degree of relationship existing between the different variables.

	Liaoning	BTC	ETH	ADA	USDT	USDC	XRP	BNB
Liaoning	1,0000000	−0,037488	−0,0468924	−0,0288041	−0,0184467	0,0006532	−0,03308836	−0,0232434
BTC	−0,037488	1,0000000	0,725753	0,556854	−0,004423	−0,030207	−2757,3327980	−2,8175590
ETH	−0,0468924	0,725753	1,0000000	0,629978	−0,009041	−0,022916	0,597130	0,708966
ADA	−0,0288041	0,556854	0,629978	1,0000000	−0,001224	−0,011055	0,608725	0,574861
USDT	−0,0184467	−0,004423	−0,009041	−0,001224	1,0000000	0,632689	0,003365	−0,006896
USDC	0,0006532	−0,030207	−0,022916	−0,011055	0,632689	1,0000000	−0,0112916	−0,0095548
XRP	−0,03308836	0,584524	0,597130	0,608725	0,003365	−0,0112916	1,0000000	0,584271
BNB	−0,0232434	0,620920	0,708966	0,574861	−0,006896	−0,0095548	0,584271	1,0000000

Table 15: Liaoning correlation matrix

	Qinghai	BTC	ETH	ADA	USDT	USDC	XRP	BNB
Qinghai	1,0000000	−0,013471	−0,006391	−0,004041	−0,018317	−0,028126	−0,032770	−0,014323
BTC	−0,013471	1,0000000	0,723646	0,545296	−0,005611	−0,043030	0,581040	0,598694
ETH	−0,006391	0,723646	1,0000000	0,623723	−0,007947	−0,025754	0,577664	0,693861
ADA	−0,004041	0,545296	0,623723	1,0000000	0,003721	−0,013044	0,595543	0,579548
USDT	−0,018317	−0,005611	−0,007947	0,003721	1,0000000	0,589270	0,006145	−0,002654
USDC	−0,028126	−0,043030	−0,025754	−0,013044	0,589270	1,0000000	−0,019238	−0,007081
XRP	−0,032770	0,581040	0,577664	0,595543	0,006145	−0,019238	1,0000000	0,581337
BNB	−0,014323	0,598694	0,693861	0,579548	−0,002654	−0,007081	0,581337	1,0000000

Table 16: Qinghai correlation matrix

Additionally, observing the correlation matrices in Table 15 and 16, what emerges, is an absence of relationship (correlation coefficient very close to zero) between the emissions of the two regions and the individual cryptocurrencies.

The only significant relationship regards only a few cryptocurrencies, which is not of particular relevance for our purposes.

However, at this point, there is the need to clarify something concerning the market of cryptocurrencies.

As any other financial instrument, the demand-offer intersection determines the value of Bitcoin. However, it's vital to remember that the Bitcoin offer is completely rigid given that the amount of Bitcoin that will be "mined" over time is predetermined in advance and is, therefore, known to all market participants[19].

The number is predetermined means that the number of Bitcoins tends to the 21 million limit, and it took little under 13 years to mine 90% of them, therefore it will probably take around 119 years to mine the remaining 10%. In reality, it is unknown how long it will take as the rate at which blocks are validated is not fixed and the halving that reduces the generation of new BTC occurs every 210.000 blocks added to the blockchain of Bitcoin.

[19] If there are more owners of Bitcoin who want to sell them than there are buyers, the price falls. Conversely, if there are more buyers, the price drops.

On average, it takes about 10 minutes to validate a block, but as hash rate increases this pace accelerates. For example, during this 2021 it often took less than 10 minutes to validate a block.
Therefore, if the current rate is maintained, it will still take nearly 19 years to mine every one of the remaining 2.1 million BTC.
Nevertheless, unlike a traditional currency where a centralized institution (central banks) can shift its supply to try to modify its value, it stands in complete contrast to the fact that the supply is rigid. However, because demand determines the value of Bitcoin, it is not viable to stabilize its value simply adjusting the supply.
Due to the fact that this market has almost always been speculative, it is challenging to apply the fundamental rules determining the supply-demand intersection.
At this point, trying to draw a connection between the value of cryptocurrencies and the carbon dioxide emissions they generate is all but pointless.
The so-called hash rate of the aforementioned cryptocurrencies is a more important factor for the purposes of our research.
Unfortunately, however, due to the lack of data regarding the hash rates, the analysis is going to continue by only taking into account Bitcoin's hash rate values in relation to the emissions in the two Chinese regions.
Having said so, the multivariate analysis, through the use of the Vector Autoregressive model, is carried out by first considering the Bitcoin hash rates together with the CO_2 emissions of the Chinese region of Liaoning and then the hash rates of BTC in relation to the Qinghai region emissions. The reason of this choice lies in the fact that, as already mentioned in the introduction, the Liaoning region is home to one of the largest mining farms in the world; therefore, considering what is the purpose of this thesis, the following association would be of extreme significance. Instead, with regard to the second analysis we want to examine, the relationship between the Bitcoin and the carbon dioxide emissions of a region, that of Qinghai, which has a position that could be assumed relatively marginal with regard to the energy sector, as it is considered one of the most remote regions of China, with one of the least developed economies in the whole country.
First of all, below in Figure 10 we have a representation of the two time series compared. In order to make the comparison visible, it was necessary to change the scale of the Bitcoin hash rate data (dividing by 100.000.000 the actual values) so that the data available to us on emissions were also visible.

Moreover, as can be noticed from the historical series of Bitcoin's hash rate, the drastic drop that is seen in correspondence to the summer months of 2021 refers to the ban imposed by China in June 2021, by which all cryptocurrency-related activities were prohibited. Bans imposed by Chinese authorities concerned not only pollution reasons but mainly illicit and speculative activities, that would disrupt the normal order of the economy and the financial system.

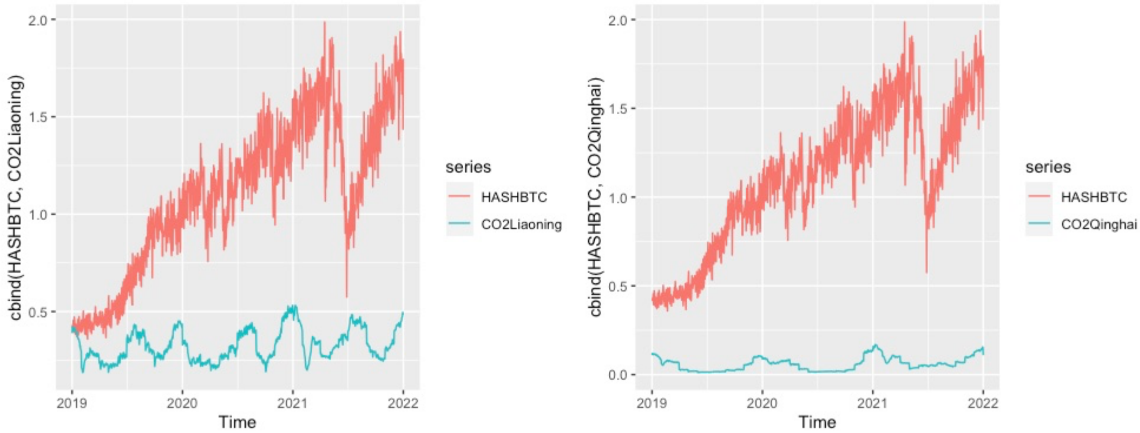


Figure 10: Time series BTC in relation with Liaoning CO_2 emissions and Qinghai CO_2 emissions.

Concerning the choice of the model, given three types of information criterion, the following results are shown:

AIC(n)	**HQ(n)**	**SC(n)**
7	7	4

Table 17: Hash Rate BTC and Liaoning CO2 emissions
R LagSelection function

AIC(n)	**HQ(n)**	**SC(n)**
7	4	4

Table 18: Hash Rate BTC and Qinghai CO_2 emissions
R LagSelection function

Table 17, but also Table 18, suggests that the optimal lag length for our multivariate model is a Vector Autoregressive process of order 4.

	Hash BTC	**CO_2 Liaoning**
Hash BTC	1,0000	0,01333
CO_2 Liaoning	0,01333	1,0000

	Hash BTC	**CO_2 Qinghai**
Hash BTC	1,0000	0,01809
CO_2 Qinghai	0,01809	1,0000

Table 19: Correlation matrices of residuals

From table 19, the two regions under examination, do not present symptoms of correlation with the Hash Rate of Bticoin.
Therefore, from our analysis, once again, the time series considered appear to be almost completely uncorrelated to each other.
It is possible to conclude that even if we have accounted for different kind of data, more significant for the purpose, no evidence is found.

Impulse response functions

In our concluding analysis, regarding the previously discussed impulse response function, before the final results some preconditions are needed.
Primarily, the extreme importance that resides in the order of the variables must be stated, given that the response of one variable to the impulse of another could change if the sequence is altered. On the other hand, the structure of the output plots represents of course the axis origin, but more importantly an equilibrium situation from which a shock will arise.

Arriving at the practical part of the IRF, firstly we obtained is the impact of the Bitcoin has rate on itself.

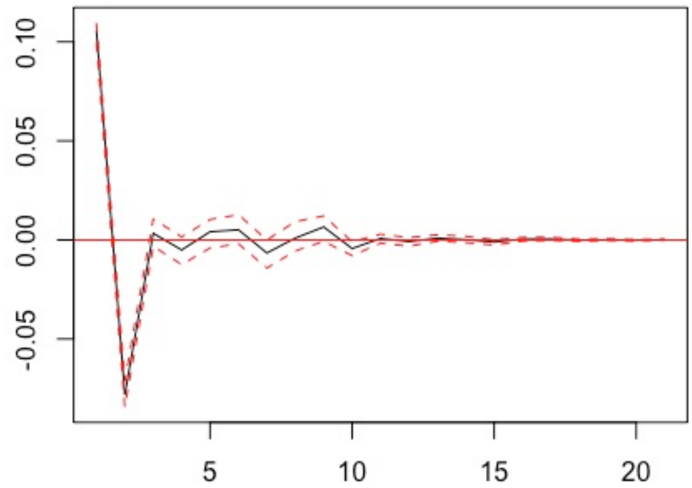
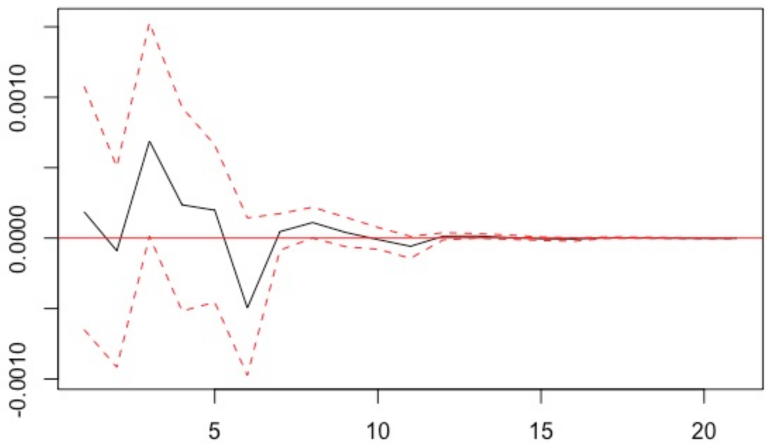
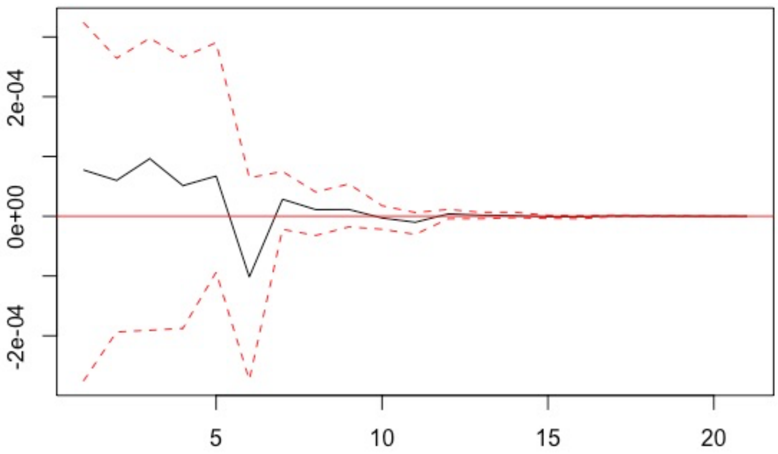


Figure 11: Hash rate BTC’s impulse to Hash rate BTC

As noted, the shock produces, immediately at first lags, a drastic decrease. The implication arising from an impulse in the rate values are given in the following two plots, considering the two Chinese regions.



a.



b.

Figure 12: a. Hash rate BTC’s impulse to Liaoning CO_2 Emissions; b. Hash rate BTC’s impulse to Qinghai CO_2 emissions

As anticipated in previous chapters, the effect of an impulse on a variable is called transitory if the variable returns to its previous equilibrium value (zero) after a few periods. On the other hand, if it does not return to zero and settles on a different equilibrium value, the effect is then said to be permanent. As a result, from the graphs obtained we can state that, in relation to both Figure 12a and 12b, the shock led to effects that were definitely temporary, as a result of the shock the series returned to their steady state.

It is evident from the graphs that a shock to Bitcoin's hash rate would appear to have a smaller effect on Qinghai's emissions than Liaoning's.

Hence, in our case, an input on the Bitcoin hash rate is going to have an effect on the CO_2 emissions after about 6 lags in relation to the Chinese regions. This is because, in general, what is emitted/produced in terms of CO_2 is nothing more than the consequence of the actions of previous periods, as they are effects not necessarily visible right away.

Conclusions

From our analysis what we can conclude is that each model is a misrepresentation of the reality due to the fact that models are based on specific assumptions, which may also be questionable. Bitcoins are criticized under different aspects, in general, cryptocurrencies and their related pollution is nowadays a hot topic which deserves a deeply analysis. Indeed, the new generation expect quick answers about this argument. This is not possible yet when data are not fully available or misrepresented, leading to different or in some cases wrong conclusions.

Even the model presented in this study, while robust in the methodology adopted, it carries limitations and pitfalls that arise precisely from the approximations related to uncertainty and data scarcity.

From the results obtained what emerges is the substantial lack of relationship between cryptocurrencies and CO_2 emissions.

After all, the data previously mentioned with particular reference to China, pointed out that the incidence of the cryptocurrencies system in terms of CO_2 emissions is relatively low compared to emissions from other production and consumption activities, thus destinated to be undetected on a large scale.

Furthermore, to this day the greatest expenditure of energy in the cryptocurrencies framework has been recorded for mining activities, which by the very nature of these coins, will be exhausted in the near future when the quantitative limit imposed by the algorithm of their creation is reached; what will be predominant in the system will be transactions with coins already in circulation.

The complaint about the emissions production of cryptocurrencies has in itself instrumental profiles that can do nothing against a phenomenon that shows no signs of ending. Indeed, it has been seen that restrictive policies put in place by some countries have pushed miners to relocate to so-called "super-polluters" countries with a contrary effect to the one aimed. The circumvention of bans in fact has not left untouched even China where miners have managed to take back their space with "underground" activities that are even more difficult for authorities to detect.

The path undertaken by several countries leading towards greater regulation of the cryptocurrency system is not only related to tax and anti-money laundering aspects but also to the containment of the contribution to global pollution.

Putting together the way in which the blockchain world has moved very rapidly to renewable energy in the past few years and with the fact that today it is estimated that nearly 60% of the Bitcoin mining industry is based on clean energy, still seems not to be enough for cryptocurrencies detractors.
Considering the historical time of energy crisis we are living, shifting cryptocurrencies to use renewable energy would result in a general opinion of unnecessarily waste of precious commodities which could be used to reduce the speculation of the abovementioned crisis.
What cannot be denied in favor of the crypto activities is represented by the birth of efficient and inclusive financial and payments services. Therefore, the need for regulatory intervention that can prevent the risks associated with the use of Distributed Ledger Technologies (DLT) and protect investors, cannot be resolved by nipping in the bud potential beneficial effects in terms of innovation and democratization of finance.
The ideological conflict between regulated market and decentralized system has no reason to be.
Regulatory incursions in a world that was born precisely in open contrast to centralized finance may seem a contradiction under the literally point of view, and it is hard to see why investors who voluntarily took capitalist risk should be protected.
The exploitation of blockchain technology represents a suggested technological solution, which allows the relationships between the players involved to be configured in an innovative way. In this sense, the governor of the Bank of Italy Visco I., in his final considerations about 2021, highlighted the initiatives taken by the central bank to encourage a dialogue with market participants in order to promote the development of technologies capable of bringing greater benefits to the whole community, «*ensuring protection of personal data, security and ease of usage, encouraging innovation and supporting the digital transformation of the economy*» (Visco I., 2022).

Bibliography

Azzalini A., Scarpa B. (2012), *Data Analysis and Data Mining An Introduction*, Oxford Press, New York
Https://Doc.Lagout.Org/Others/Data%20mining/Data%20analysis%20and%20data%20mining_%20an%20introduction%20%5bazzalini%20%26%20scarpa%202012-04-23%5d.Pdf

Baricco A. (2019), *E ora le elites si mettano in gioco,* Repubblica

Box, G. and Jenkins, G. (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco

Carlini V. (2018), *Bitcoin, ecco perché non è una moneta. Il vero valore? La blockchain*, il Sole 24 ore.

Cavicchioli, M. (2015) "Likelihood Ratio Test and Information Criteria for Markov Switching Var Models: An Application to the Italian Macroeconomy", *Italian Economic Journal,* volume 1, pages 315–332.
https://link.springer.com/article/10.1007/s40797-015-0015-6

Claude Sammut, Geoffrey I. Webb (2017), *Encyclopedia of Machine Learning and Data Mining*, Springer, Boston.
Source: https://link.springer.com/referenceworkentry/10.1007/978-1-4899-7687-1_192#citeas

De Vries, A. (2018), "Bitcoin's Growing Energy Problem", *Joule 2*, CellPress

Don Grant *et al (*2021), *Reducing CO2 emissions by targeting the world's hyper-polluting power plants*, Environ. Res. Lett. 16 094022
https://www.almendron.com/tribuna/wp-content/uploads/2021/10/grant-2021-environ-res-lett-16-094022.pdf

Gallersdörfer U., Klaaßen L., Stoll C. (2020), “Energy Consumption of Cryptocurrencies Beyond Bitcoin”, *Joule 4*, CellPress

Girino E. (2018), “Criptovalute: un problema di legalità funzionale*”, Rivista di diritto bancario,* Ottobre/Dicembre 2018

Global Monitoring Laboratory, https://gml.noaa.gov/ccgg/trends/

Granger, Clive W. J. 1969, *Investigating causal relations by econometrics models and cross-spectral methods*. Econometrica, vol. 37.

Hamilton, J.D., Time Series Analysis. (1994), http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/504_02_Hamilton_Time-Series-Analysis.pdf

Hansen, L.P. Sargent, T.J., “Formulating and Estimating Dynamic Linear Rational Expectations Models”, *Journal of Economic Dynamics and Control 2* (1980) https://pages.stern.nyu.edu/~dbackus/GE_asset_pricing/HansenSargent%20JEDC%2080.pdf

Hicks J. (1971)*, Saggio critico di teoria monetaria,* Etas Kompass*,*

Holmes E. E., Scheuerell M. D., and Ward E. J. (2021), *Applied Time Series Analysis for Fisheries and Environmental Sciences* https://atsa-es.github.io/atsa-labs/sec-boxjenkins-aug-dickey-fuller.html

Hughes E. (1993), *A Cypherpunk's Manifesto,*
Source: https://www.activism.net/cypherpunk/manifesto.html

Ruppert D. (2011), *Statistics and Data Analysis for Financial Engineering*, Springer, New York

Stoll C.,Klaaßen L., Gallersdörfer U. (2019), “The Carbon Footprint of Bitcoin”, *Joule n.3*.

Tversky A., Kahneman D.E. (1979), "Prospect Theory: an analysis of decision under risk", *Econometrica*, pp. 263-291

Tiberi P.,(2021), "The carbon footprint of the Target Instant Payment Settlement (TIPS) system: a comparative analysis with Bitcoin and other infrastructures", *Rivista Banca d'Italia* n. 5/21

Yûichirô Kakihara, Stochastic Processes, *Encyclopedia of Physical Science and Technology (Third Edition)*, 2003
https://www.sciencedirect.com/topics/engineering/stationarity

Sitography

Antonelli Fabio, Appunti di processi stocastici,
http://people.disim.univaq.it/~gabriell/didattica/processi.pdf

Banca d'Italia, https://www.bancaditalia.it/pubblicazioni/interventi-governatore/integov2022/cf_2021.pdf

Borsa Italiana, https://www.borsaitaliana.it/notizie/sotto-la-lente/blockchain.htm

Britannica, https://www.britannica.com/place/Liaoning

Britannica, https://www.britannica.com/place/Qinghai

Capterra, https://www.capterra.it/blog/2313/criptovalute-rischi-studio

Money.it, https://www.money.it/mining-criptovalute-cose-e-come-fare

Coinbase, https://www.coinbase.com/it/learn/crypto-basics/what-is-mining

Coinmetrics, https://cmbi-indexes.coinmetrics.io/cmbihash

Criptoinvestire, https://www.criptoinvestire.com/come-funziona-la-crittografia-nelle-blockchain.html

De Santis Alberto, Le serie storiche, http://www.diag.uniroma1.it//~desantis/NOTE/timeseries.pdf

De Vincenzi Tommaso, http://webusers.fis.uniroma3.it/~devincenzi/materiale_didattico/dispense/AA17-18/cap_9_parte1.pdf

Digiconomist, https://digiconomist.net/bitcoin-energy-consumption

ECB, What is money? 24 November 2015 Source: https://www.ecb.europa.eu/ecb/educational/explainers/tell-me-more/html/what_is_money.en.html

Espresso, https://espresso.repubblica.it/mondo/2022/01/18/news/bitcoin_vietati_consumo_energia_alternativa_green-334297695/

Flandoli Franco, http://users.dma.unipi.it/~flandoli/dispense_Istituzioni.pdf

Icona clima, https://www.iconaclima.it/salute-del-pianeta/co2-concentrazioni-mai-cosi-elevate-negli-ultimi-4-milioni-di-anni/

Il Sole 24 ore, https://www.ilsole24ore.com/art/bitcoin-cosi-minatori-cinesi-aggirano-divieti-pechino-seconda-mondo-gli-usa-AE7IhoZB

Koinsquare, https://koinsquare.com/bitcoin-quanta-energia-consuma-realmente/

La Repubblica, https://www.repubblica.it/green-and-blue/2022/01/20/news/bitcoin_divieto_consumo_energetico-334475329/

Natural Gas World, https://www.naturalgasworld.com/east-siberian-producer-eyes-blue-ammonia-prospects-90345

Our World in data, https://ourworldindata.org

Puntoinformatico, https://www.punto-informatico.it/mining-e-staking-cosa-sono-e-come-funzionano/

University of Cambridge, https://ccaf.io/cbeci/index/methodology

University of Colorado, https://www.colorado.edu/asmagazine/2021/09/01/cleaning-worst-hyper-polluting-power-plants-could-have-huge-effect