Universität Zürich[UZH]

Abschlussarbeit

zur Erlangung des

Master of Advanced Studies in Real Estate

Application of a Nowcasting model to the Swiss real estate market

A holistic approach to trends evaluation

Name : Anderloni Fabio

Submittend to: Dr. Constantinescu Mihnea

Submission date: 02.09.2019

CONTENTS

`

`

Glossary of terms and abbreviations

2S	Two Step procedure
BFS	Swiss Federal Statistical Office (Bundesamt für Statistik)
CRAN	Comprehensive R Archive Network
ECB	European Central Bank
EM	Expectation-Maximization procedure
FA	Factor Analyses
FED	Federal Reserve
FGV	Getulio Vargas Foundation
FUW	Finanz und Wirtschaft datawarehouse
GDP	Gross Domestic Product
IBRE	Brazilian Institute of Economics
KF	Kalman Filter
KOF	Swiss Economic Institute (Konjunktur Forschungsstelle)
NMEC	Center for Statistical and Computational Methods
OECD	Organization for Economic Co-operation and Development
PCA	Principal Components Analysis
SECO	State Secretariat for Economic Affairs
SNB	Swiss National Bank
VAR	Vector Auto-regression
WMO	World Meteorological Organization,
W&P	Wuest&Partner

List of illustrations

`

List of tables

Executive Summary

Real estate is both an important asset class for investors and a key component of the national economy. A deeper understanding of the correlation between macro-economic fundamentals and real estate prices can provide important insight to investors and property owners. The relevance of this correlation, meanwhile, gains in value when it can interpret the linkage between a broad number of economic indicators and leverages the most recent data streams.

The following analysis applies a nowcasting methodology to the Swiss real estate market to provide a short-run prediction of how property prices change. The model analyzes a large data set containing more than 110 variables describing how real estate economy evolves over time. The collected time-series illustrate with different frequencies (yearly, quarterly, monthly, daily) the fluctuation of economy, finance, demography and property market over time.

The applied statistical learning method leverages sophisticated algorithms: it automatically collects the most recent data available at each new calculation, it assesses the current state of the real estate economy based on the evaluation of past co-movements of all variables and finally it estimates the nowcast of the property's price.

We applied the nowcasting methodology to three sets of housing and apartment sale price variables to evaluate the behavior of the algorithm. These variables include: the Wuest&Partner property sale price index (quarterly), the Wuest&Partner asking price index (quarterly), and the REIDA median apartment transaction price (aggregated each half-year).

The nowcast provides a monthly estimate and is able to fill the gap between the last and the future publication. The results show the model performs well in predicting the trends in prices but lacks precision in the estimation of the levels.

1. Introduction

Real estate is both an important asset class for investors and a key component of the national economy. The estimated total value of Swiss real estate at the end of 2017 was about CHF 2’819 billion, with more than CHF 1’771 billion in residential properties[1]. By comparison, the capitalization value of the Swiss stock market was CHF 1’145 billion[2] and the GDP was CHF 669 billion[3] in the same year. Real estate values and the macro-economy influence each other through a bilateral causation relationship. An understanding of the factors affecting the dynamic of fluctuations in property prices is no less important than understanding the dynamics of other asset classes, such as bonds, stocks and currencies. Some of these influencing factors appear to have a strong relation with macroeconomics fundamentals. A qualitative overview of the linkage between the macroeconomic fundamentals and real estate indicators are described below.[4]

1.1 Problem definition

Estimating property value evolution is critically important to real estate investors. Property values are estimated during the process of sale, purchase, or during a portfolio evaluation. Their appraisal is also a control value for indirect real estate assets. Property estimation is by definition an informed guess and it can be understood as a short term forecast. Yet this forecast is affected by many uncertainties caused by a wide range of factors that directly and indirectly affect price fluctuations. This analysis studies the relation between economic fundamentals and real estate price fluctuations with the aim of estimating in real-time the price evolution based on the observed changes in the economy.

1.2 Property prices are subject to fundamentals fluctuations

Macroeconomic fundamentals are quantitative information that describe the evolution of the economy at large. They include measures of supply, demand, inflation,

---

[1] This is measured as aggregated building insurance value excluding land price. The calculation is based on the methodology of the study “Die volkswirtschaftliche Bedeutung der Immobilienwirtschaft der Schweiz“ pom+, Rütter Soceco, 2011.
[2] Source: SIX Swiss Exchange, Swiss Market Index®, full market capitalization (as of 29.12.2017).
[3] Source: Swiss National Bank (Nominal value as of 21.09.2018).
[4] An extensive overview of the academic research about the relationship between macroeconomic and real estate is described in the chapter 5.1.
`

financial accounts, employment, monetary policy and international trade. Generally when economic activity increases, the demand of realty spaces increases as well. Illustration 1 below shows that during the last 20 years, changes in the price of Swiss apartments appear to be positively correlated with movements in the Swiss GDP. One hypothesis to explain this apparent correlation might be that economic growth leads to an increase of the work force and a consequent growth in housing demand. With a relatively inelastic supply of apartments, this change in demand would push the price higher.

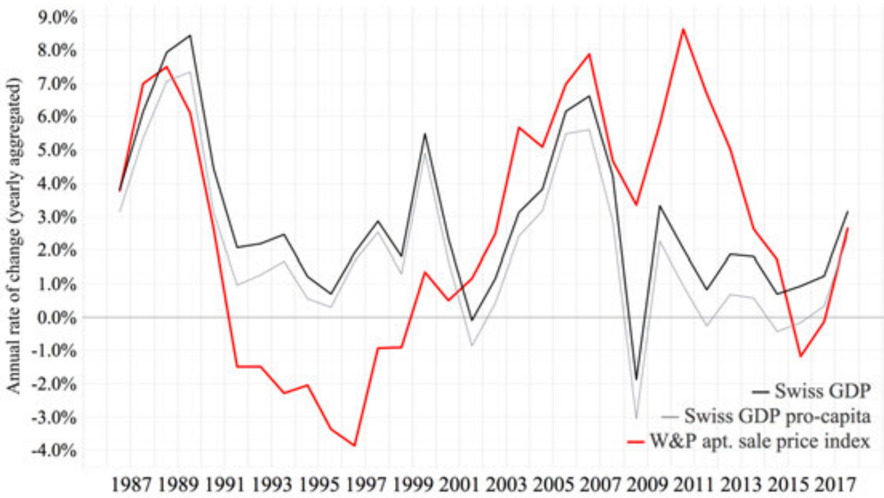


Illustration 1: Aggregate annual rate of change in Swiss GDP and Swiss GDP per-capita versus the apartments sale price index

Source: SNB, Wuest&Partner

Another hypothesis to explain the positive correlation could be that economic growth can lead to a rise of real wages and a broader distribution of wealth across the population, allowing more existing households to buy instead of rent or to separate into multiple households instead than living in a multi-generational domestic environment. If the increase in GDP per-capita reflected a greater dispersion of prosperity across the middle-income classes, more people could then have the freedom to choose to buy. Even in this case, if the supply of housing is relatively inelastic, this increase in demand for housing would push their price higher.

An investment in real estate is commonly associated with borrowing a part of the necessary financial capital. The decision to buy or to rent, or whether to invest in real estate or in other investment vehicles, is strongly influenced by market interest rates.

When the interest rates are high, this increases the price of borrowing capital and makes alternatives to real estate purchases more attractive.

Illustration 2 below shows that Swiss interest rates have declined steadily during the last 20 years. The decreased cost of Swiss capital appears to have encouraged Swiss households to borrow greater sums of money from banks to purchase ever more real estate. Household finance accounts data published by the Swiss National Bank (SNB) reveal that total Swiss household net-worth has been steadily growing during the last 20 years. Today it is approaching the CHF 4'000 billion threshold[5].

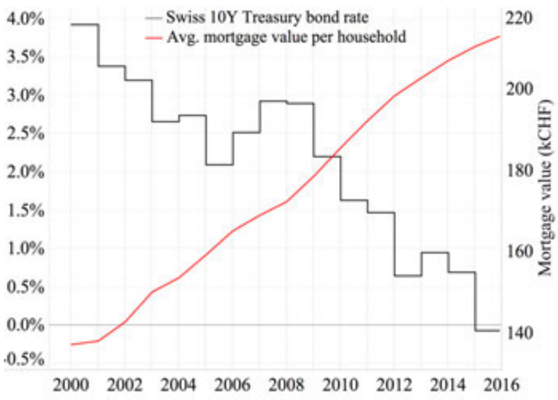


Illustration 2: Swiss 10Y T-bond versus average mortgage value per household

Source: SNB

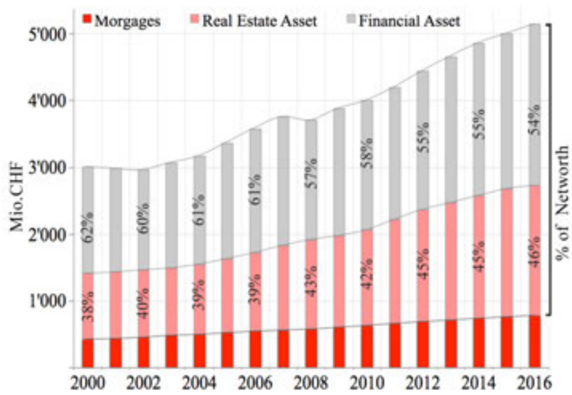


Illustration 3: Household's financial account

Source: SNB

Illustration 3 above portrays the value of the main categories of household finance accounts since the year 2000. It appears that the appreciation in home prices has played an important role in the growth of household net worth. This vehicle has provided a more stable annual contribution to the overall annual growth rate relative to the contribution of other financial assets. In fact, while financial assets have been exhibiting fluctuations (the data reveal slight declines in the value in the years 2002 and 2008) the value of real estate has consistently grown at a regular pace.

Further, the ratio of real estate asset value to the value of stocks and bonds has been increasing during this period. Real estate has shifted from 38% of the total asset value owned by the households in year 2000 to 46% in year 2016. Meanwhile, the ratio of mortgage values to property value has decreased. This relatively slow growth of household exposure to mortgage loans could be interpreted as a decline of demand for

[5] Net-worth = Financial Asset + Real Estate Asset – Liabilities
`

privately owned houses and a rise in leased apartments. This trend could be to the result of a greater need for geographic flexibility, possibly driven by the fluid movements of the labor force. Another hypothesis could be the fact that rising house prices have made private home ownership less affordable. As real estate prices continue to climb, more people choose to rent rather than to own housing leading to a decrease in demand of private mortgages.

Examining the balance sheet of credit institutions in Illustration 4 below reveals that the proportion of private homeowner financing is decreasing in favor of non-private borrowers (companies and institutional investors). When a growing population finds more difficult to obtain a self-owned house, the rising demand of rental apartments could encourage institutional investors to invest in housing as commercial properties increasing their debt exposure.

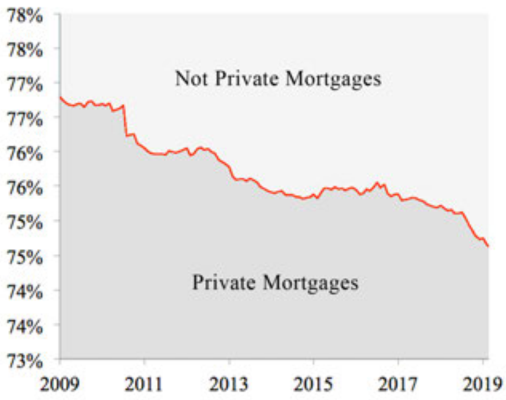


Illustration 4: Private and non-private mortgagor proportion

Source: SNB

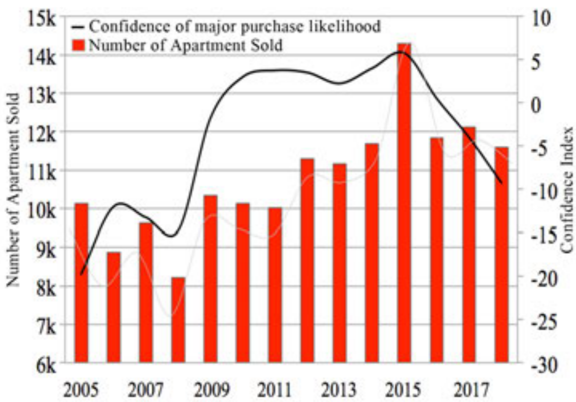


Illustration 5: Number of apartments transaction versus confidence

Source: SECO, W&P

Along with the conventional macroeconomic fundamentals, there are also many "soft variables" that describe behavioral aspects of the housing market, such as consumer sentiment and confidence. One interesting variable based on surveys conducted by the State Secretariat for Economic Affairs (SECO) is the *major purchase likelihood*. This metric reflects how consumers perceive the likelihood of making large expenditure (such as buying a house) in the short run. Illustration 5 above reveals that the number of properties purchased over time appears to be positively correlated with the consumer likelihood of making a large investment.

Demographics can also play a pivotal role in the definition of the real estate market. Not only is the demand for real estate spaces influenced by total population, but the distribution of age classes across the population appears to matter as well. Indeed, population classes can have specific space requirements and preferences. For example, an increase of the population aged 10 years or less may not impact the number of apartment rooms demanded, as small children can often live together in the same room as their siblings or their parents. In a similar way, an increase of the population between 50 and 60 years of age might also fail to increase the growth of apartments demanded. This component of the population has likely been occupying a private home already.

However, examining changes in the proportion of the population between 20 and 29 years of age can reveal a strong correlation with the private construction expenditures, as represented in Illustration 6 below. One hypothesis might be that as young adults begin an independent life away from their parent's household, they increase the demand for new apartments. This would support greater private investment in new construction projects.

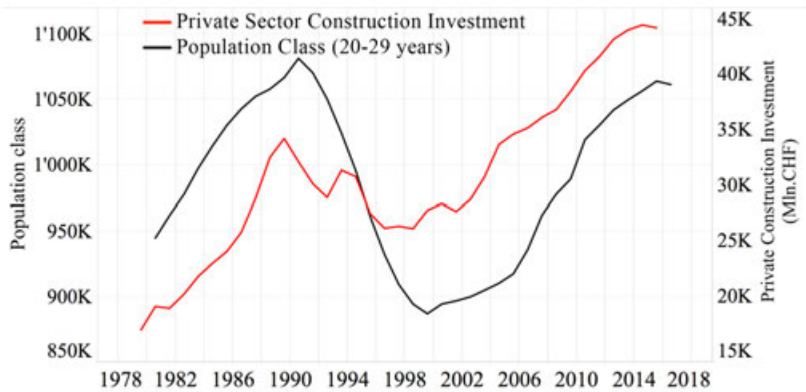


Illustration 6: Young population class versus private construction expenditure

Source: SNB, BFS

As the demand for apartments grows, the supply of housing reacts by increasing new construction projects or renovating existing spaces. The value of construction activities is also influenced by inflation, which in turn is related to the money supply. The thickness of the grey bars in Illustration 7 below portrays the growth rate of M2 money supply for each year. The length of the grey bars reveals the rate of inflation in percentage terms. The red line reveals the annual rate of change in housing construction prices. There appears to be a positive correlation between inflation and construction prices, as inflation impacts the prices of construction materials and machinery. Inflation appears to at least partly drive higher construction costs, raising the prices of new or renovated buildings.

Despite the consistently growing Swiss population, the rented apartment vacancy rate is characterized by fluctuation. One hypothesis for explaining this fluctuation might be that potential tenants persistently have difficulty finding the desired apartments in the market (a mismatch of the demand and supply of apartments related to the size, style, available amenities, etc.). Another hypothesis might be related to the asking price and affordability of housing. Illustration 8 below shows that when vacancy rates decline, the real price of rental housing tends to increase. Likewise, when the vacancy rate raises, the real price of rental housing falls.

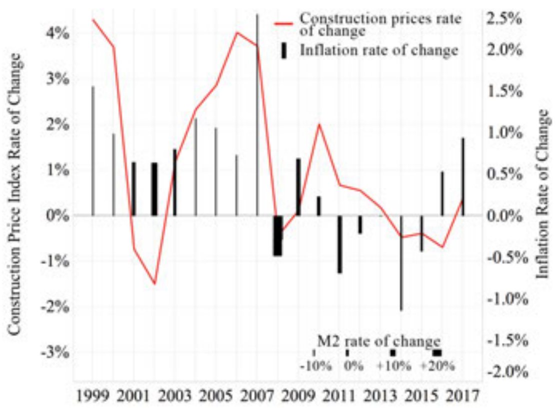


Illustration 7: Inflation and M2 versus construction prices

Source: SNB,

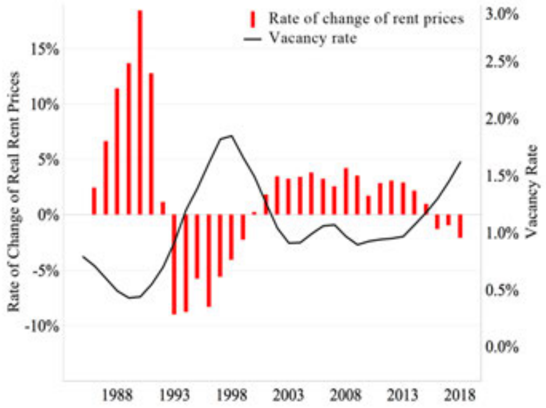


Illustration 8: Asking price versus vacancy

Source: SNB, BFS

These examinations of the data illustrate that a multi-faceted relationship exists between the economic fundamentals of the Swiss economy and the Swiss real estate market, which indicates that such relationships are comprehendible. However, it is difficult to model them in a way that would reflect all of these interwoven and potentially significant quantitative relationships. There are a great number of variables involved, and the relationship between many of them is bilateral. On one hand there is a considerable amount of time-series data that is regularly updated. On the other hand there is not an adequate model to provide a testable method for combining and exploiting these variables in order to understand how their co-movements actually impact real estate prices trends. Moreover, economic data are often released with a certain time delay, which makes it even more complicated to extract useful information for understanding how trends are evolving in a timely manner.

The entanglement between different economic and demographic parameters is fascinating. It is for this very reason the following analysis investigates and tests a method to untie the knots. The nowcasting methodology is used to extract additional information from the complex amount of available data. The goal is to learn how the real estate market is changing in the near future under the influence of a multiplicity of these economic and demographic influences.

1.3 Objectives and limitations of this analysis

The following analysis aims to evaluate how changes in both the Swiss economy and demographics have impacted the evolution of Swiss real estate prices. The analysis is performed with a nowcasting methodology that leverages a large, heterogeneous dataset of observations starting from the year 1970. After the initial statistical learning process, unobserved determinants are distilled from the model and used to nowcast the evolution of real estate prices in the short run. The effort aims to demonstrate how a nowcasting model can expand the real estate investor toolbox and provide additional information for a more accurate evaluation of real estate prices evolution. The data and analysis concern Switzerland as a whole and do not take in consideration regional differences.

The proposed methodology provides a deterministic estimation trough the application of a statistical model without the need for informal judgment. The model observes historical distributions of different variables and outputs estimations based on empirically recognized relationships between them, rather than directed by economic theory. Therefore the proposed method is not a *structural* model, but a *reduced form* econometric model.

1.4 The process of the analysis

The analysis begins with the investigation of the nowcasting method and its actual applications. It continues with the examination of the statistical principles and algorithms that back-up the nowcast estimation. The screening, analysis and classification of about 300 time-series from different sources follows. About 110 variables are selected from the pool of collected data; they are pre-processed and used in the analysis model. A specific algorithm has been developed to perform an update of the dataset with the last available observations as soon as a new estimation is calculated. The results of the calculation show the level of quality of the algorithm's estimated trend in prices during past periods. More importantly, the results offer a short-term estimate of their future evolution, which represents the added value of this method.

1.5 Organization of the analysis

The next chapter introduces the nowcasting concept, its implementation in econometrics and explains why nowcasting can be a link between macroeconomic and real estate. Chapter 3 presents an outline of the key statistical learning methods that underpin the nowcasting model. An explanation of the mechanics of the model according to the "Two step" method and its implementation with the "Expectation-Maximization" method are then described in chapter 4. The relationship between economic fundamentals, demographic variables and properties prices is advocated in chapter 5, along with relevant extant academic research. This chapter also lists the selected categories and variables, and explains the applied stationarization processes. Chapter 6 analyzes the results of the nowcast. Chapter 7 summarizes the content of a few interviews with portfolio managers regarding risk assessment and high-frequency data. The conclusions of are then summarized in the chapter 8.

2. Overview about the literature

The term nowcasting refers to a short term forecasting method. This method initially was applied to meteorology, and more recently to analyzing the economy. The UK Meteorological Office created the term "*nowcasting*" in the 1980s as the aggregation of "*now*" with "*forecasting*". Later, the World Meteorological Organization (WMO) defined nowcasting as a form of forecasting based on rapid updates of local details over a period from present to 6 hours ahead (WMO, 2017, p. xi). If a forecast is used to predict the future, a nowcast aims to describe the present and the short-term future focusing on what is known in the real time.

During the last two decades, the application of nowcasting has been extended to econometrics, becoming an established tool for macro-economic analysis that is widely applied by national banks. In this chapter we first explain the meaning of nowcasting compared to forecasting, flash estimates and general statistic reports. Than we provide an overview of the evolution of the nowcasting as an econometrics method applied to macro-economy. Finally we explain its potential link to real estate economics.

2.1 Reports, Flash-estimates, Forecasting and Nowcasting

Statistical data analysis methods change depending on whether the ultimate goal is to describe the past, the present or the future.

Indicators and reports (i.e GDP, Cap-rates, price index) are created by evaluating a complete set of observations within a delimited past period (i.e. quarterly, yearly). Many are published with a lag of several months from the authentic moment of the observation. Therefore, they provide precise information about a time window in the past and do not include any information between the end of the analyzed time-window and "now." This means the recipient of this data is completely blind to the specific events of the recent past.

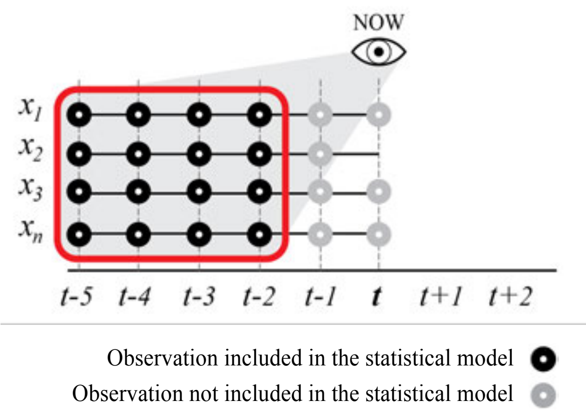


Illustration 9: General statistic diagram

Flash estimates fill this gap by providing an early picture of the considered statistic. As represented in Illustration 10, flash estimates are calculated after the end of the reference period, right up to the current period. This estimate is based on a similar data set of information, even if partially incomplete, with the condition that the amount of available observations is "sufficient to ensure a high degree of compatibility with the first regular computation," as stated by the United Nations (UN) in the *Handbook on Rapid Estimates* (2017, p.35). Flash estimates are a statistical robust picture of the near past and are calculated with a partially incomplete data set. For this reason it is referred to as an "estimate."

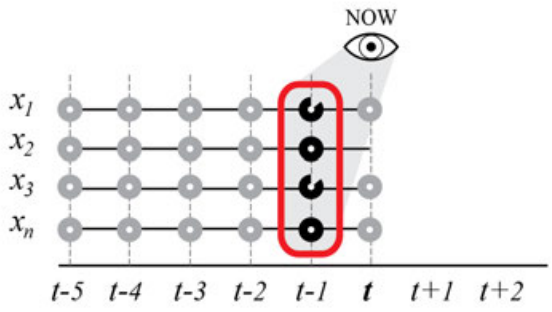


Illustration 10: Flash estimate diagram

Forecasts switch the focus of interest from the past to the future. As represented in Illustration 11 below, forecasts leverage x_n time-series to evaluate the future evolution of a variable y starting from the time t+1. Stochastic processes, such as Monte Carlo simulations, focus on short to mid-term projections. Their uncertainty (the error in estimation) tends to grow with the temporal distance from the current moment t. Classical static forecast methods do not use high frequency data. As a result, they are often based on structural or statistical models that focus on the future prediction without necessarily leveraging any real time data.

Nowcasting is a very specific type of estimation that focuses explicitly between the time between t and t+1 to recognize trends based on objective real-time information. As stated by Ineichen, "nowcasting is fact-based, focuses on the known and knowable, and therefore avoids forecasting,[6]" (2005, p.45).

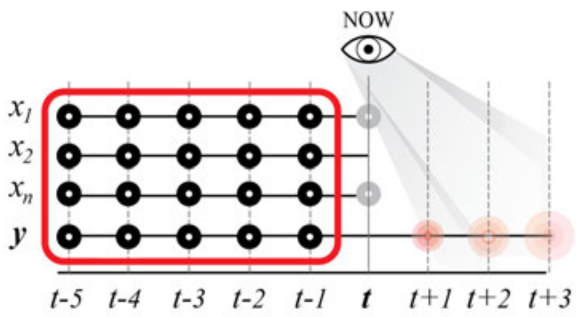


Illustration 11: Forecast diagram

The *Handbook on Rapid Estimates* (UN 2017, p.34) describes now-casting as "the real-time or quasi real-time estimation of the evolution of a given variable."

[6] in this case the term "forecast" is intended as the activity dedicated to the assumption of future events.
`

To put it in a discursive way, imagine a quarterly economic index (y) has been released today (t) referring to observations that have been acquired with certain lag. Not only we miss information between the last observation included in the statistic and today, but starting from now and for the next 90 days we will lack any knowledge about the trend of the index. However, from time t to time t+1, plenty of other economic indicators are potentially available and each of them can hypothetically suggest some insight about the changes of variable y. These more timely data can include low frequency data (i.e. yearly, quarterly) or high frequency data (i.e. monthly, daily) available after the last publication of y. As represented in Illustration 12, nowcasting uses all x_n time-series including the most recent observations to estimate the value of the variable y at the end of the time t and the beginning of the time $t+1$.

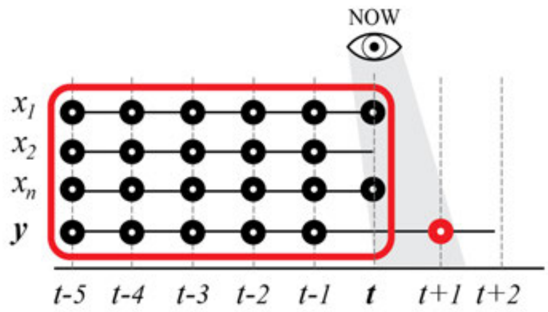


Illustration 12: Nowcast diagram

As soon as new observations become available, the nowcast that embeds this new information can be instantaneously produced. This is a big advantage of the model, especially when timely estimates and forecasts are needed (Guagliano & Mantovani, 2014, p.11)

2.2 Nowcasting for macroeconomics

After a failure to anticipate the global financial crisis, economic forecasting has been in the center of the debate between the orthodox economists who defend their models and methods, and the heterodox economists who criticizes the plausibility of their scope. As the famous quote from the Canadian economist John Kenneth Galbraith states, "There are two kinds of forecasters: those who don't know, and those who don't know they don't know." Our minds and the predicting methods that we develop are both affected by bias and unconscious blindness. The awareness that any attempt of future prediction is a leap into the dark, leads to focus on the here and now. The following analysis takes part in this debate by affirming that state of the art statistical models (in this case, the nowcast model) might not provide exact prediction, but they can help interpret the complex interaction of observable changes in various phenomena and provide insight to support decisions.

The first formalization of the nowcast statistical methodology applied to econometrics has been written in the milestone paper, "Nowcasting: The real-time informational content of macroeconomic data," by Domenico Giannone (ECB), Lucrezia Reichlin (ECB), and David Small (FED) and published in the *Journal of Monetary Economics* in 2008 (pp.665-676). A wide part of our analysis is based on this important work. The model developed in this paper was implemented at the Board of Governors at the Federal Reserve Bank in a project that was started in 2003. It was also used at the European Central Bank in 2008. The model calculates the actual status of U.S. GDP from the statistical analysis of 382 different data series. This panel data included financial data releases, CPI indexes, interest rates, money and credit rates, labor indexes, production rates, surveys and diverse indexes that all provide a live-picture of the U.S. economy. Their model bridges many time-series with different frequencies and creates a dimensionality reduction process followed by a Kalman smoothing process (formally named "*two step*" process).

Other authors have further elaborated this method, creating more sophisticated variations. In the publication "ECB Working Paper Series – Nowcasting" (Banbura, Giannone & Reichlin , 2010, p.12), the authors demonstrate how to trace the sources of forecasts revisions back to individual predictors. In a later paper Doz, Giannone and Reichlin (2012, pp.1014-1020) introduce advanced statistical learning techniques (Expectation-Maximization algorithm) for increasing the robustness of the results. The introduction of the Expectation Maximization (EM) algorithm created the conditions for estimating models also with a large portion of missing observations.

Castle, Hendry and Kitov (2013, pp.8-15) performed a methodological overview of the contemporary statistical nowcasting methods. They compare different models to provide a detailed recommendation checklist to produce a "good" GDP nowcasting model (pp.38-50). They also introduce an automatic model selection method that forecasts all the variables before the end of each period and recognize eventual location shifts (changes in the trend of each variable) to detect single variable breaks and anticipate general shocks (p.28-31).

Many others studies have developed the nowcasting method introduced by Giannone's work. Here below a short list of papers worthy of mention:

- Guagliano and Mantovani (2014, pp.7-26) provide a clear discursive explanation of the method and compare the results of the calculation of the GDP of different European countries
- D'agostino, Lenza, Giannone and Modugno. (2015, pp.2-24) focus on nowcasting as tool for outline business cycle
- Tiffin (2015, pp.3-17) writes a comparison of different modern advanced machine learning (Lasso Regression, Decision Tree, Random Forest) technics to assess the GDP nowcast of Lebanon
- Ferrara and Simoni (2019, pp.1-28) evaluates how Google search data can be integrated in the nowcasting of GDP models adding a gain in accuracy and enlarging the real time data flow

The application of the nowcasting model in macroeconomic analysis has seen much growth during the last 10 years. Today it is common practice for most of the major central banks to adopt this methodology in the real time assessment of their country's GDP.

2.3 Nowcasting for real estate economics

In the following analysis, nowcasting is used to study the development of macro-economical and social variables leveraging them to make a short-term prediction of real estate prices.

Nowcasting is a relatively a young methodology. Currently, its application in real estate is exceedingly rare. Assuming that macroeconomic aspects have an influence on the real estate economy, the nowcasting methodology can be an effective analysis tool to study them together holistically. The connection between macroeconomic trends and real estate prices is complex, bidirectional and difficult to describe with a traditional quantitative method. Many studies have considered macroeconomic and demographic variables as endogenous factors in order to explain the development of real estate markets. DiPasquale and Wheaton (1992, p.188) proposed an elegant representation of the complexity of the real estate economic equilibrium with their four-quadrant model shown in Illustration 13.

This model divides the real estate market into two main sub-markets. The two right-hand quadrants represent the property sub-market describing the demand and supply of space for use. The two left-hand quadrants represent the asset sub-market describing the ownership of real estate.

Illustration 13: DiPasquale&Wheaton four quadrants diagram

Rental prices are determined in the short run in the NE quadrant, and are affected by the supply and demand of rented spaces. When in equilibrium, the demand of rental spaces is equal to the supply, setting the market rental price. The NW quadrant represents the first part of the asset market and it is explained by the rent/price ratio usually called cap-rate. The NW quadrant takes the rent level R and combines it with the cap rate to determine the price P. The SW quadrant represents the construction activities and the supply of new apartment space. Construction costs and building volume directly impact the price P of the built space. The SE quadrant represents the net change in the stock of apartment spaces, which is the difference between the new stock that enters the market and the old stock that exits the market due to obsolescence, natural causes or change of preferences (DiPasquale & Wheaton,1992, pp.187-190).

Each quadrant is potentially exposed to macroeconomic and demographic shocks that can lead to a loss of equilibrium. For example, a change in the immigration rate leads to a shift of the demand curve for apartment space in the NE quadrant. A monetary policy fluctuation might alter the interest rate, affecting cap-rates and finally rotating the curve in the NW quadrant. Alternatively, a change in the inflation rate can affect construction prices and shift the supply curve for new apartment construction in the SW quadrant. It should be noted that the cap-rate is a “catch-all” exogenous factor that takes in consideration inflation, growth, risk free investment options, risk of illiquidity, specific risks of the single property and location related risks. Many of these aspects are also strongly related to the macro economical evolution.

Karl Case has contributed extensively to academic literature by studying the relationship between macroeconomic fundamentals and real estate market trends. In his paper "Real Estate and Macro-economy" (Case, 2000, pp. 119-145) he illustrates a relationship between American housing pricing and employment, household's wealth, mortgages, banking activities and stocks. With a quantitative approach, he shows a close relation between U.S. real estate value, household's wealth, consumption and stock market. Case and others also demonstrate that changes in housing markets trigger more significant variations on aggregate consumption than stock market (Case, Quigley & Shiller, 2005, p.26). Girouard, Kennedy, van den Noord and Andreé (2006, pp. 9-32) analyze macro economical fundamental such as GDP, inflation, the price to income ratio, household accounts, and house price fluctuations. They reveal linkages between the studied variables and demonstrate a similar pattern in a different international property market. Chaney and Hosely (2014, pp. 50-76) study the long-run equilibrium relationship between multifamily residential prices and the economy. They use a structural model assessing the relationship between capitalization rates and macro-economic indicators such as inflation, interest rates, GDP, monetary supply and construction expenditure as endogenous variables. Their study suggests a bidirectional relation between the cap rates and the macroeconomic fundamentals.

The above-mentioned empirical analyses are a small selection of a wide literature regarding the linkage between economy and real estate. A description of the academic research on this topic will be extended in section 5.1 below. These analyses offer a qualitative or quantitative evaluation of the relation between economic indicators and real estate prices. Depending on the type of model applied, these studies focus on a limited selection of indicators, concentrating on specific dynamics of the economy and the real estate market without embracing this link in its entirety.

The nowcast offers a statistical technique to consider a multitude of different variables in a holistic and comprehensive way. It describes the economy in a broad sense and the real estate market as a part of it. This statistical method evaluates a large set of time-series data and infers the collected information onto a dependent variable of interest (property prices). Referring again to the DiPasquale and Wheaton diagram, the nowcasting model allows capturing the co-movements of a myriad of indicators that impact the equilibrium of each of the four quadrants and finally allows estimating short-term predictions.

3. Statistical Methodology

In this section, the definition of statistical learning is clarified, as are the formal aspects of the data. Finally an overview of the algorithm used in the nowcasting model is explained.

3.1 Statistical learning definition

Statistical learning is a recently developed area of statistics that blends computer science and machine learning (Tibshirani, Gareth & Witten, 2013, p. vii). In this work we use the Tibshirani's terminology and we considerer for our application the term *statistical learning* more appropriate than *machine learning*. Statistical learning refers to a vast set of tools for understanding data and being able to predict or estimate an output based on one or more input.

Statistical learning processes applied to large-scale public or private datasets can improve the way economic activities are measured, tracked, described and it can also enable to trace the consequences of different events or policies. Statistical learning leverages two main components: statistical data and statistical algorithms.

3.2 Statistical data

The recent expansion of the data landscape and its accessibility underpins the growth of statistical learning efforts and their empirical applications. Only in recent years digital networks provided efficient access to large databases allowing the utilization of historical data and generating greater opportunities to extract relevant information.

Private businesses and public institutions are traditional generators of data while users, social networks, computers and sensors have became sources of new information streams (Buono, Kapetanios, Marcellino, Mazzi & Papailias, 2018, pp. 4-5). Modern connectivity and new technologies now make real-time data collection possible, and have even generated new types of variables, such as geo-located data and mobile statistics (Einav & Levin 2013, p. 4). The data available is expanding rapidly and the ability to leverage it to gain additional knowledge is becoming a key competitive factor among businesses. Data can now be sourced in different forms (i.e. images, numeric series). For the following analysis, we leverage data as a set of time-series variables.

3.2.1 Time-series and panel data

In descriptive statistics, a time-series is defined as a vector of t observations, ordered in time and frequency and expressing the dynamics of a certain phenomenon. These historical series are studied both to interpret phenomenon, identifying the trend, cyclicality, or seasonality components in order to predict their future movements. A combination of n time-series (or variables[7]) into one indexed matrix, n x t, creates a *panel data*. Panel data contains observation of multiple phenomena during a defined *time window*. Panel data are defined as *balanced* when each variable is observed in each period, and *unbalanced* when one or more of these observations are missing.

3.2.2 Unbalanced panel data, ragged edges and vintages

Nowcasting models can leverage a multivariate panel data composed of time-series that describe different economic, demographic and real estate related phenomena. The time-series used in the following analysis, however, contain variables that have different frequencies that are updated at different moments in time. This means that as time progresses, some variables are updated with new observations and others are not. The resulting panel data is therefore *unbalanced*, because observations are unsynchronized or occasionally missing, as represented in the Illustration 14 below.

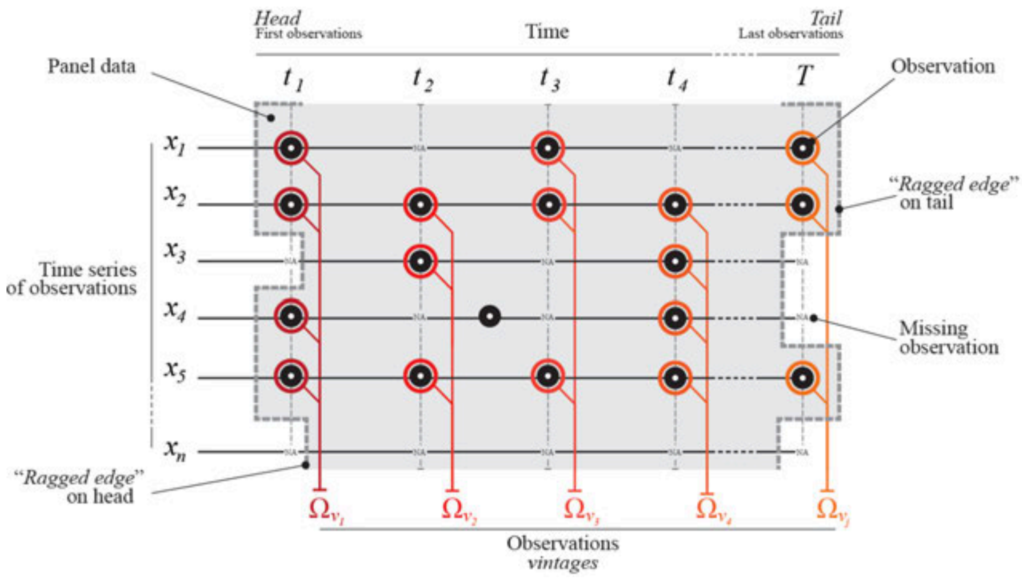


Illustration 14: Unbalanced panel data diagram

The panel of data has "*ragged edges*" because the earliest observations (the "*heads*") and the latest observations (the "*tails*") are not matching in time. The global frequency used in our model is defined in months (frequency = 12). Those "cells" in the panel of data without an observation recorded is labeled as "NA" (Not Available). Paraphrasing

[7] In this analysis, the terms time-series and variables are interchangeable.

Guagliano (2014 , p.13), this process will “bridge monthly and quarterly data by treating the latter as monthly series with missing observations.”

Each monthly interval represented in our model is described by a *vintage* v_j that includes a unique set of observation Ω_{v_i} defined as:

$$\Omega_{v_i} = \left\{ \mathrm{x}_{i_t|v_j} \; ; \; t = 1, \dots , \mathrm{T}_{iv_j} \;\; ; i = 1 \, , \dots \, , n \right\} \tag{1}$$

where:

$\mathrm{x}_{i_t|v_i}$: are the i variables $\mathrm{x}_{i_t} = \mathrm{x}_{1_t}, \mathrm{x}_{2_t}, \mathrm{x}_{3_t}, \dots , \mathrm{x}_{n_t}$

t: is the time from the first until the last observation T_{iv_j} (time of the last observation of the variable i for the vintage v).

For each vintage the nowcasting algorithm extrapolates valuable information to make a forecast.

3.3 Statistical learning algorithms

If data are the “new oil,” then algorithms are the “new engine.” Algorithms are the mechanical component of statistical learning and are formulated on the basis of statistical, probability and linear algebra principles. As represented in Illustration 15, the algorithms family includes regression, Lasso, and Bayesian techniques, dimensionality reduction techniques, neural networks, supporting vector machine, and a growing number of new sub-branches and hybrid compositions (Brownlee, 2016, pp. 25-148).

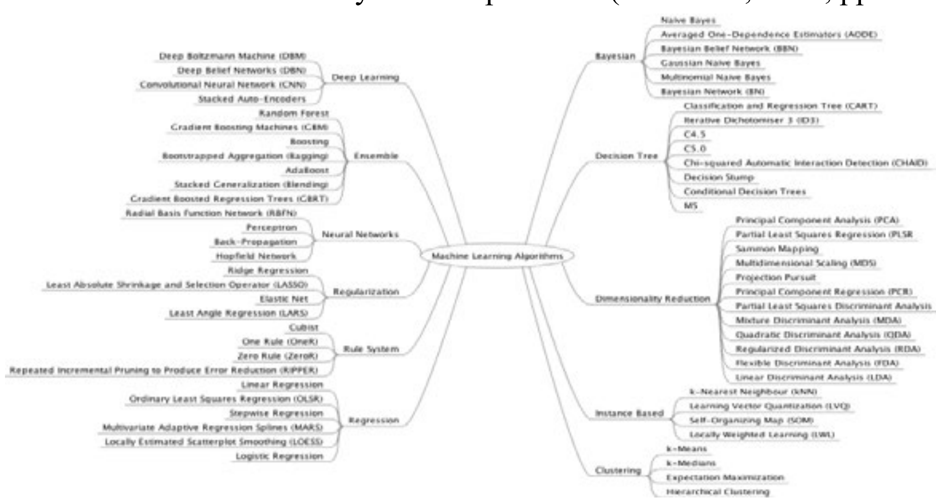


Illustration 15: Machine learning algorithm classification [8]

[8]Source: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/ (without date).

`

The nowcasting model[9] leverages a combination of statistical learning algorithms as described in the followings sections.

3.3.1 Regression

The aim of the nowcasting model is making a prediction of a continuous quantity based on the analysis of different time-series. Because the ultimate goal is the calculation of a quantitative measure, regressions are the suitable family of algorithms for this purpose. Therefore, ordinary linear regression will be used.

For a single time-series, the regression is described as follows:

$$y_{it} = \alpha + \beta x_{it} + \varepsilon_{it} \tag{2}$$

where for the i^{th} variable and the time t :

y_t	is the dependent variable
x_{it}	is the independent variable (where x changes over time)
α	is the constant intercept
β	is the constant regression parameters
ε_{it}	is an equation error term (that varies with unit and time) with variance σ_ε^2 .

A simple example of an application of a simple linear regression based on a single time-series can be the prediction of real estate transaction volumes based on their observed seasonality.

For a panel data with more than one time-series it will take the form of a multivariate regression as follows:

$$y_t = \alpha_t + \beta_t x_{it} + \gamma_t z_{it} + \varepsilon_{it} \tag{3}$$

Where y_t are the dependent variables, x_{it}, z_{it} are the independent variables as above and α_t, β_t, γ_t and ε_{it} are the regression parameters and the uncertainty changing over time.

The Nowcasting model analyzes a large set of variables and incurs the so-called "*curse of dimensionality*" due to the collinearity issue. This makes calculating the parameters of a multi-regression with an ordinary least squares (OLS) methodology problematic. For this reason, it is necessary to use techniques that reduce the dimensionality of the regression equation (Donoho, 2000, p. 18).

[9] For simplicity, the term *algorithm* and *model* are used interchangeably, although a model can technically be a combination of algorithms.

`

3.3.2 Dimensionality reduction algorithm

Dimensionality reduction is the process of reducing the number of observed variables by obtaining a set of principal variables that "summarizes" them. In a large panel data, variables have a certain degree of correlation and redundancy. This interdependency can be used "to reduce the number of variables to a smaller number of components or factors and capture the most of the variance in the observed variables," (Wuensch, 2019, p. 10).

One of the common methods is called Principal Component Analysis (PCA) invented by Karl Pearson and presented by Hotelling (1933, p. 417). PCA finds a set of q principal components (λ_q) ranked in a linear combinations $\lambda_1 x, \lambda_2 x, \ldots, \lambda_q x$ that have maximum variance for the data set and minimum correlation between successive combinations $\lambda_q x$ (Jolliffe, 2011, p. 1094). A variation of the PCA is the Factor Analysis, which will be used in our model it is extensively described in the chapter 4.1.

Much literature explains the suitability of this method in econometrics (Forni, Giannone, Lippi &Reichlin, 2005, p. 8-21; Giannone, Reichlin & Sala, 2004, pp. 161-200 ; Stock&Watson, 2002, p. 161-200).

3.3.3 Kalman Filter

The Kalman Filter (KF) is an algorithm that was developed in the middle of the last century (Kalman, 1960, pp. 35-45). It estimates the state of an unknown dynamic variable based on the recursive analysis of observations and their statistical noise. Kalman filters are ideal for analyzing phenomenon that change continuously over time and was first applied to spacecraft navigation for trajectory optimization. The KF process has the advantage of being able to calculate a stable estimation in only a few iterations, making it computationally inexpensive. (suffice it to say that that KF had an essential role in the navigation system of the Apollo space ship equipped in 1969 with a computer with only 2k RAM of memory and chip speed of only 100 kHz). Thanks to its efficiency, KF has subsequently been applied to many other fields, including econometrics and especially time-series analyses.

As was mentioned before, the time-series variables of the nowcasting model are organized in panel data. Along with the panel data, the Kalman filter needs a *state-space representation*, which is a mathematical model that estimates the *state* of the variables at the time of the observation *t*. The *state* is defined as a vector of parameters (observed and calculated) that provide a representation of the model at a specific instance in time. The state space representation of a system is given by two equations called *state equation* and *output equation* that will be described later on in chapter 4.2. For the algorithm, a state space model represents both the memory of the past from which it can learn, and a tool for dynamically storing descriptive elements about the evolution of the *state* over time.

In summary, the Kalman filter (KF) is an iterative mathematical process that populates a state space representation by calculating an estimation of the *state* (the vector of parameters describing a single moment) for each interval. During each iteration, the KF makes a prediction, a measurement and a correction to find the optimum averaging factor for each subsequent state. The model below uses the KF to estimate the *state* parameters after the dimensionality reduction as described in the chapter 4.2.2.

3.3.4 Expectation-Maximization (EM)

The nowcasting model must deal with time-series of different frequencies and starting times. This requires solving the problem of missing data points along the observed time window. EM method is used to back-up the regression calculation finding with an iterative function the maximum-likelihood estimates for the regression parameters also in the moment some of the observations are missing.

This iteration is based on 2 steps:

1. The "E-Step" (Expectation) first runs the Kalman filter to estimate the state of the space representation and derive its probability distribution.
2. The "M-Step" (Maximization) uses the estimated state and its distribution to process a maximum-likelihood estimation (MLE) to obtain updated parameter estimates.

This iteration repeats (100 times in this model) and for each loop the model increases the likelihood of the estimate, which stabilizes the prediction of the parameters. The explanation of the integrations and conditional probabilities involved during the iterative processes of the EM and the MLE is behind the scope of this effort. The EM calculation allows incomplete time-series to be used in the estimation, but it has the drawback of being extremely expensive in terms of computing resources.

4. Model mechanics

The nowcasting model has been programmed in R, based on the CRAN open source nowcasting package ‘nowcasting’ version 1.1.3 (DeMatto, Gomes, Ferreira & de Valk, 2019), that was developed as initiative of the Center for Statistical and Computational Methods (NMEC) for the Brazilian Institute of Economics (IBRE) and the Getulio Vargas Foundation (FGV). The open source code has been deeply studied to understand the mechanics of the method (see Annex D and E) and has occasionally been modified to efficiently run with data sets used in this model. Improvements have been done introducing in the code calculation of a sparse matrix allowing a faster calculation using the Estimated-Maximization method with a large data set.

Since each data set provides relevant information for a projection, the core challenges of a nowcasting algorithm are the ability to:

- handle the fact that in each *vintage* Ω_{v_i}, the predictors set (the available independent variables) is different
- interpret how much each predictor *x* have influence on the dependent variable *y*
- provide a new updated projection

$$\text{Proj}\left[y_t | \,\Omega_{v_i}\right] \tag{4}$$

as soon as new information is made available and flows into the model.

Since each new data can help to improve the prediction’s accuracy, using this model “there is no reason to throw away any information” (Giannone, 2008, p. 666).

4.1 Dimensionality Reduction with Factor Analysis

High-dimensional data often contain correlated or redundant variables. Dimensionality reduction methods transfer the significant information into a low-dimensional data set, removing the noise while retaining the signal of interest. In this model, the Factor Analysis (FA) process extracts *p* linear components (the common factors) from the *t* observations of *n* variables, where *p* is much smaller than *n*. The unique feature of the method is that FA explicitly specifies a model relating the observed variables to a smaller set of underlying, unobservable (latent) common factors.

To further explain the meaning of common factors, consider the following example: a direct measurement of the variable "*real estate market shocks*" is not possible because it is unobservable. However, this unobservable variable is quantifiable through the observations of other variables that affect it. While these market shocks may not be measured directly, the model can deduce them by assessing other, observable variables that are related to it (i.e. the change of transactions number, the increase of new developments, the variations in real estate expenditure, the spread fluctuation between the asking price and the transaction price, etc.). Assuming that market shocks impact all the observed variables with a different weight, these unobserved event can be considered as common factors. Summarizing, common factors can be explained as indirect, unobserved but still measurable variables.

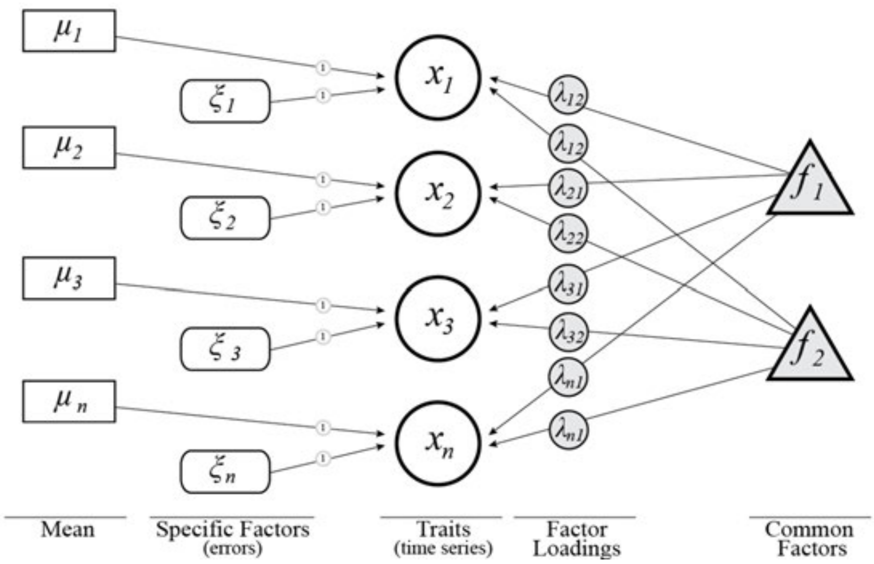


Illustration 16: Path diagram representation

Though it may seem counter-intuitive, **the model switches the point of view and assumes that the changes of the factors are the cause of the shifting of the variables**. Following this approach, the central assumption of the model is that it is possible to predict the variables $\boldsymbol{x}_i$ with a multiple linear regression equation, where: (a) the common factors f_i are weighted by specific loadings ($\boldsymbol{\lambda}_{ij}$), (b) the intercept μ_i is the mean of the observed stationarized time-series, and (c) the idiosyncratic component is $\boldsymbol{\xi}_i$ (with i = 1...n and j = 1...r):

$$
\begin{aligned}
\boldsymbol{x}_1 &= \mu_1 + \boldsymbol{\lambda}_{11} f_1 + \boldsymbol{\lambda}_{12} f_2 + \ldots + \boldsymbol{\lambda}_{1r} f_r + \xi_1 \\
\boldsymbol{x}_2 &= \mu_2 + \boldsymbol{\lambda}_{21} f_1 + \boldsymbol{\lambda}_{22} f_2 + \ldots + \boldsymbol{\lambda}_{2r} f_r + \xi_2 \\
&\ldots \\
\boldsymbol{x}_n &= \mu_n + \boldsymbol{\lambda}_{n1} f_1 + \boldsymbol{\lambda}_{n2} f_2 + \ldots + \boldsymbol{\lambda}_{nr} f_r + \xi_n
\end{aligned}
\qquad (5)
$$

The path diagram in the Illustration 16 above shows the relation and the causal direction between the terms in the equations (5).

With the assumptions drawn from the above equations, the number of studied parameters is drastically reduced from n observed variables ($\boldsymbol{x}_n$) to r unobserved parameters f_r and their weights λ_{nr}. Most of the mechanics of the nowcasting models are designed to find the λ_{nr} , f_r and ξ_n components to solve the equations (5). These components are formally described below.

4.1.1 Vector of traits

The vector of traits $\boldsymbol{x}$ is a vector of n time-series x_i (with i = 1...n) present in the panel data:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \text{ vector of traits}$$

Assuming stationarity of the time-series, the expected value of each trait can be defined as $E(x_i) = \mu_i$ and it is presumed to be constant. The population means of the variables x_i are represented in the vector $\boldsymbol{\mu}$:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{pmatrix} = \text{ population mean vector}$$

4.1.2 Common factors

The r common factors f_j (with j = 1...r) can also be represented as a vector:

$$\boldsymbol{f} = \begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_r \end{pmatrix} = \text{ vector of common factors}$$

with the following assumptions:

- r is substantially less than the number of time-series in the panel in order to achieve a dimensionality reduction ($r << n$)
- The mean of f_j is centered ($\mu_f = 0$)
- The standard deviation of f_j is unitary ($\sigma_f = 1$)
- The covariance is an identity matrix (Cov(f) =1)

- The single factors are independent from each other.

As in the equations (5) all the variables x_i (at the time t) share the same r common factors.

4.1.3 Factor loadings (weights)

The factor loadings are the regression coefficients λ_{ij} (the partial slopes) of the multiple regressions in (5). Formally, λ_{ij} is the *loading* of the i^{th} variable on the j^{th} factor. This means each single variable has a specific loading for each factor. The factor loadings populate a matrix, as shown below:

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{11}\lambda_{12} \cdots \lambda_{1r} \\ \lambda_{21}\lambda_{22} \cdots \lambda_{2r} \\ \vdots \quad \vdots \qquad \vdots \\ \lambda_{n1}\lambda_{n2} \cdots \lambda_{nr} \end{pmatrix} = \text{matrix of factor loadings}$$

The role of factor loadings is to explain the weight common factors have on each variable at a specific time.

4.1.4 Specific factors (errors)

The errors ξ_i are called the specific factors for each variable x_i and t. They are represented in a vector $\boldsymbol{\xi}$:

$$\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \cdots \\ \xi_n \end{pmatrix} = \text{ vector of specific factors}$$

with the following characteristics:

- The mean of ξ_i is centered $\mu_{\xi_i} = 0$
- The covariance is a diagonal matrix (Cov(ξ) =ψ)
- the error ξ term is independent from x and f.

This last assumption of independency leads to state that factor model framework is represented as the sum of two mutually orthogonal unobservable stochastic processes: the *weighted common factors* and the *idiosyncratic component (Giannone, 2008; Banbura, 2013).*

In summary, the Factor Analysis model is a regression model where each variable x_i is predicted with a linear function described with unobserved common factors f_j weighted by factor loadings λ_{ij} and a specific factors ξ_i that represents the idiosyncratic error term "driven by variable-specific shocks" (Giannone, 2008, p.666).

4.2 Nowcasting model

The dimensionality reduction elements described in the previous section are the foundation of the nowcasting model that is expressed as follows:

$$\boldsymbol{x_t} = \boldsymbol{\mu} + \boldsymbol{\Lambda\, f_t} + \boldsymbol{\xi_t} \qquad (6)$$

where the equations (6) is the matrix notation of the formulas (5), μ is costant while Λf_t and ξ_t are orthogonal stochastic processes. The model assumes that Λf_t captures "almost all" co-movements in the economy, while the idiosyncratic linear processes ξ_t are driven by variable-specific shocks. The common r factors "aim at capturing the lead and lag relations among variables along the business cycle" (Giannone et al. , 2008, p.668).

It is important to notice that the factors are not estimated in to explain which variable are more suitable to forecast real estate prices development, "the purpose is rather to gather information from the series and build a monthly series that can be thought of as a (latent) indicator of the business cycle" (Guagliano, 2014, p.11).

4.2.1 Two-step estimation process - Step 1: Dimensionality reduction

The methodology proposed by Giannone et al. (2008) is called the two-step (2S) estimation process. In the first version proposed by Giannone, the 2S process defines an important precondition: the model processes a panel data with variables observed at the same frequency, with the exception of the latest time periods (the "*tail*")[10]. Namely the data set **X** can be divided in two portions:

- a long balanced portion **Z** where observations are available at each point in time from the beginning until recent periods
- an unbalanced portion **W** of most recently observations where some of the them are missing since not yet published.

[10] Giannone's model is based on a "bridged" monthly dataset. The panel data is balanced (each variable is observed over time) from the year 1982-Q1 until the 2007-Q2, and it is unbalanced (some variable has missing observations creating a "ragged edge in the tail") in the 2007-Q3.

`

During the first step of the 2S method, the ragged edges are trimmed from the data set **X** taking in consideration just the balanced subset **Z**.

After standardizing **Z**, a first estimation of the common factors, factor loading, and specific factors is done with the following calculations as specified in the appendix of the Giannone's publication:

$\mathbf{S} = \text{cov}(\mathbf{Z})$	(7)	n x n covariance matrix of the subset **Z**
$\mathbf{D} = \text{Eigenvalue}(\mathbf{S})$	(8)	r x r diagonal Matrix with the major r eigenvalues
$\mathbf{V} = \text{Eigenvector}(\mathbf{S})$	(9)	n x r diagonal matrix with the r largest eigenvectors
$\widehat{\mathbf{F}} = \mathbf{V} \cdot \mathbf{Z}$	(10)	n x r matrix of the $\widehat{f}$ common factor estimation
$\widehat{\mathbf{\Lambda}} = \mathbf{V}$	(11)	r x t matrix of the $\widehat{\lambda}$ factor loadings estimation
$\widehat{\boldsymbol{\psi}} = \text{diag}(\mathbf{S\text{-}VDV})$	(12)	n x n covariance diagonal matrix of the idiosyncratic component (with the assumption mentioned in section 4.1.4)

After this first approximation, the model captures the dynamics (the changing over time) of the factors with a vector auto regression (VAR) with length *p,* as follows:

$$\boldsymbol{f_t} = \mathbf{A}\boldsymbol{f_{t-1}} + u_t \tag{13}$$

where **A** are the calculated parameters of the VAR(1) and u_t is the idiosyncratic component of the shocks that affect the commons factors.

The process continues with the definition of the idiosyncratic components ξ_t:

$$\text{E}(\xi_t \xi_t') = \psi_t = diag(\widehat{\psi}_{1,t}, \widehat{\psi}_{2,t}, \dots, \widehat{\psi}_{n,t}) \tag{14}$$

where the variance is equal to ψ_i when the observation is not present, and zero when the observation is available. Both error terms $\boldsymbol{\xi_t}$ and u_t are normally distributed white noise with independent variance-covariance matrices Σ_ξ and $\Sigma_{\mathbf{u}}$.

At the end of the first step, the estimated parameters of each point in time of the balanced panel (**Z**) are stored in the state space array, as represented in the diagram in Illustration 17 below.

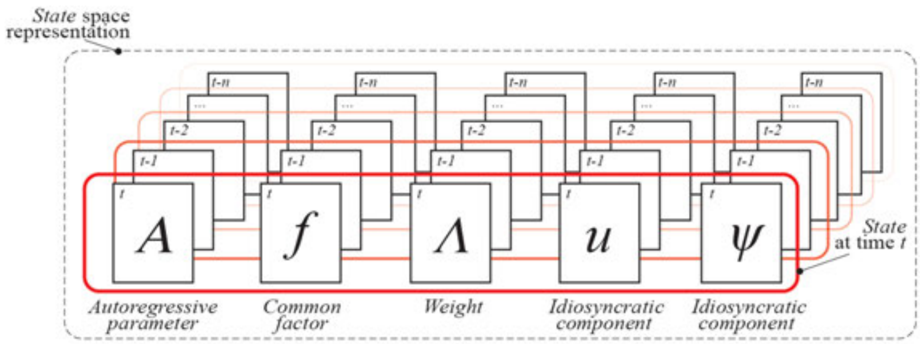


Illustration 17: State space array diagram

4.2.2 Two-step estimation process - Step 2: Kalman filter

In the second step, all the parameters passed in the state space array are re-estimated by applying the Kalman filter. With the KF the re-estimation is done considering the dynamics of the factors (their changing in time) from the previous estimation. The KF takes in consideration both the common shocks u_t and the idiosyncratic component the error terms ξ_t to finally compute the estimation of the factors weighting with the Kalman Gain the *innovation content* of each variable at each time interval. The KF operates in two stages summarized as follows:

Prediction stage

At time t the KF calculates a first estimation of each variable of the state space model leveraging the information available at t-1 and evaluate the predicted error covariance[11]:

Predicted state estimate $\hat{x}_t^- = F\,\hat{x}_{t-1}^+ + B\,u_{t-1}$ *(15)*

Predicted error covariance $P_t^- = F\,P_{t-1}^+\,F^T + Q$ *(16)*

Update stage

During this stage the KF evaluate the observed variable z_t (the variables passed in the state space model), calculates the residuals from the first estimate and calculates the factor K_t called *Kalman Gain:*

Residual measurement $\tilde{y}_t = z_t - H\,\hat{x}_t^-$ *(17)*

Kalman gain $K_t = P_t^-\,H^T(R + H\,P_t^-\,H^T)^{-1}$ *(18)*

[11] The superscript – and + indicate when the estimation is predicted (prior) or updated (posterior).

The Kalman Gain evaluates how accurate is the estimation versus the observation and sets a relative importance in the error in the estimate versus the error in the observation to update the state estimate and the error covariance:

Updated state estimate $\hat{x}_t^+ = \hat{x}_t^- + K_t \tilde{y}_t$ *(19)*

Updated error covariance $P_t^+ = (I - K_t H)P_t^-$ *(20)*

This last parameter is delivered to the next iteration and the cycle is repeated again. Thanks to the fact that the common shocks are cross-sectional and they affect the variation of all the variables in a single point in time with different weights, it is possible to use the Kalman filter to estimate the rest of the common factors when the vintage has unobserved variables (the portion **W** where the tail has ragged edges).

4.2.3 Two-step estimation process - Nowcast calculation

With the assumption that the observations x (the independent variables) are jointly normal with the investigated element y (the dependent variable) we obtain that the nowcast at time t of the investigated element is a linear function of the expected common factors:

$$\widehat{y_t} = \alpha + \beta\widehat{f_t} \qquad (21)$$

Where $\widehat{f}_t$ are the estimated parameters for each vintage with the KF. Given the estimate of the common factors, the nowcast of the real estate prices can be computed calculating the coefficients α and β of equation (21) by OLS regression

In our model, with the assumption that real estate prices are related to the dynamics of a wide array of time-series, projecting on the common factors (instead of projecting on all the time-series) is not only parsimonious and feasible but it also “provides a good approximation for the full, but unfeasible and over-parameterized, projection on all the variables”. (Giannone, 2008, p. 668).

As introduced in the chapter 3.3.3 the state space representation model is finally given by the *state equation* (6) and the *output equation* (21).

4.2.4 Advantage and limitation of the Two-step estimation process

Although the 2S process, as it has been conceived originally, is a sophisticated method with a relatively low computational cost, one of the drawbacks is the requisite of a data set that is mainly composed by a balanced panel (the portion **Z**) allowing missing observations just in the last periods (the final portion **W**). In other words, this method requires time-series with same starting point and no missing observation in the middle of the panel. Yet, the publicly available data series are highly heterogenic, have different frequencies, present missing observations and start mostly from different years. This precondition of the original 2S model is too restrictive and the algorithm needs an adaptation to solve the problem.

One way to overcome this obstacle is to implement the 2S method with the EM process introduced in the chapter 3.3.4. Briefly described, this variation of the 2S method begins running a spline interpolation of the time-series before proceeding with the Step 1 as described in 4.2.1 for the calculation a first estimation of the *state* parameters. At this point the EM algorithm evaluates the estimated factors likelihood in case of variables are observed or unobserved iterating the following two steps:

1) it computes the expectation of the log-likelihood (sufficient statistics) conditional on the state parameter estimated in the previous iteration
2) it re-estimates the state parameters trough the maximization of the expected log-likelihood.

During the iteration the likelihood increases until it converges to the maximum likelihood solution. During each iteration the process correct the uncertainty associate with the estimation of the common factors reducing the results of state parameters to the most likely (Banbura, Giannone, Modugno & Reichlin. , 2013, p. 204). The detailed explanation of the mechanics of the EM method is outside the scope of this work.[12]

[12] For more details, see Dempster, Laird & Rubin (1977, pp. 1-21), Engle & Watson (1981, pp. 774–781) , Quah & Sargent (2004, pp. 161-200) and Doz,Giannone&Reichling. (2011, pp. 192-194)
`

5. Data specifications

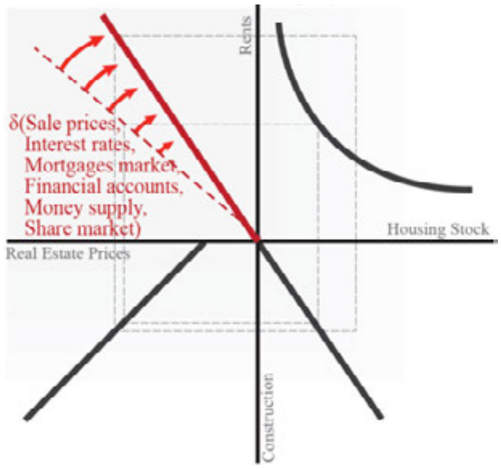
This chapter explains how the time-series are selected tested for stationarity and transformed before proceeding to the nowcast calculation.

5.1 Data selection and taxonomy

The literature offers many studies focused on revealing specific determinants of real estate price fluctuations. The identification of such determinants has strong relation with the underlying models as Girouard et al. (2006, p. 10-15) show in their paper drawing a panorama of the existing studied econometrics models. The nowcasting approach introduces an additional method that captures the co-movements of a large number of potentially significant variables into common unobserved factors. Considering the infinite possibilities of suitable datasets, the DiPasquale and Wheaton four-quadrants model is used as a “compass” to select and organize the potentially meaningful variables for this analysis.

For clarity we divide the variables into thematic groups. With the support of the literature the analysis assumes that each of the selected variables can affect with different weight the equilibrium of the real estate market as represented in the Illustration 18 and explained in the next paragraphs. The indicators included in the nowcasting model are schematically reported in Tables 1-71 while a detailed list of all the variables including sources, frequencies and applied transformations is listed in Annex A.

(a) Shift of rents



(b) Shift of prices

(c) Shift of construction activities

Illustration 18: Potential impact of studied variables on the D&W diagram

In the following sections we argument the selection of the variables for each thematic group.

5.1.1 Economic activity

Economic activities include producing, selling, distributing, buying, and consuming goods or services. They exist at all levels of the society and they leverage resources and producing factors (i.e. capital, land, materials, workforce) to create added value in the economy and redistribute it between stakeholders. Economic activities are allocated across Primary (raw-materials), Secondary (manufacturing), Tertiary (services) and Quaternary sectors (knowledge). Gross Domestic Product (GDP) is the broadest quantitative measure of a country's total economic activity and is calculated as the sum of private consumption, expenditure, government investment and the difference in value between exports and imports.

	Indicators
Expenditure	*Government expenditure* *Housesholds expenditure* *Gross fix.cap.form. Construction* *Gross fix.cap.form. Equipment*
Import/ Export	*Export of goods* *Export of services* *Import of goods* *Import of services*
Income	*Compensation of employees* *Consumption of fixed capital* *Net operating surplus* *Property Income* *Subsides* *Taxes on production and import*
Mixed	*Gross national Income (GNI)* *Gross domestic product* *GDP pro Capita*
Labour	*Jobless Rate* *Job Vacancies* *FTE Secondary / Tertiary*

Table 1: Economic activity indicators

Empirical studies show that economic growth is positively associated with job creation (Khan, 2007, pp. 14-18), demand for spaces (Vries & Boelhouwer, 2016, pp. 7-12) and appreciation of real estate prices (Wheaton, 1999, pp. 214-215). As explained by Adams and Füss (2010, p.41) and represented in Illustration 18(a), an increase of economic activity tends to increases the demand for spaces and shift the demand curve depicted in the NE quadrant to the right. Following the DiPasquale and Wheaton model, with an inelastic supply, rents will than increase together with sale prices in the asset market. The economic activity related time-series included in the nowcasting model are summarized in Table 1.

5.1.2 Interest rates, financial accounts and credits

Interest rates have a major impact on investment decisions. When interest rates are low, mortgages are more affordable and investing in properties becomes more attractive. This leads to an increase in demand of owned properties, new construction and renovation of existing properties. For this reason short, middle and long-term interest rate time-series are included in this selection, as showed in Table 2.

The account balances of financial, non-financial corporate, government and household entities also offer an outlook on the investment purchase capability of these stakeholders. Therefore, these are important factors for explaining the evolution in the demand for investments in real estate assets.

	Indicators
Interest rates	*Swiss Bond 10Y* *LIBOR 12M* *LIBOR 3M* *SARON*
Money Supply	*Monetary aggregate M2*
Accounts	*Average Bank Deposit* *Fin. Corp. Assets* *Fin. Corp. Liabilities* *Non-Fin. Corp. Assets* *Non-Fin. Corp. Liabilities* *Gov. Corp. Assets* *Gov. Corp. Liabilities* *Household Assets Real Estate* *Household Disp. Income* *Household Savings*
Credit	*Mortgage - Construction* *Mortgage - Households* *Mortgage - Real Estate activities* *Domestic Mortgage All banks*

Table 2: Interest rates, financial accounts and credit indicators

Money supply measured with the M2 indicator, describes from an aggregated point of view similar dynamics. Zhou, Chen and Cheng (2014, p.416) used a co-integration analysis to study the influences of the money supply to find it had a positive influence on the real estate prices. Liu (2014, p. 365) analyzed the effects of money supply, bank credit and real estate interest rates on land prices in the both before and during the bubble period in Japan and China. Their research found that the money supply and total lending shocks can have positive effects on land prices.

Credit data series are also included in the model not only because mortgages development permeates real estate economy, but also because credit dynamics in general affect the real activity via different conduits as studied by Constantinescu and Nguyen (2018, p. 26).

Therefore the model assumes that shifts in interest rates, the money supply and credit volumes all could have an impact in the NW quadrant of the DiPasquale and Wheaton schema, as represented in the Illustration 18(b).

`

5.1.3 Shares market

Whether a correlation exists between security markets, real estate fund shares and housing price development has been debated in the literature. These studies include Gyourko and Keim (1992, pp. 457-485), Ling and (1999, pp. 483–515) and Qaun and Titman (1999, pp. 183–207). Algieri states, "houses can be regarded as both investment goods and consumption goods," (2013, p. 319), in contrast with pure investment assets, such as stocks. She points out that stocks and investment properties can be interpreted as substitutes. When the shares market creates low rates of returns, investors increase their interest in real estate, leading to greater demand and higher market prices. From another perspective, when households, corporations or investment funds receive unexpected gains from stocks, this increases their total capital and enhances their capacity to invest in other assets, including new properties (Koivu, 2012, pp. 307-309 in Yuan, Hamori& Chen, 2014, p. 2).

	Indicators
Stock Market	*SPI*
Real estate funds	*WUPIX-A* *WUPIX-E* *Rüd-Blass Immobilienfonds Index* *OTC-X Immobilien* *SXI-RE*

Table 3: Share market indicators

Different studies about portfolio hedging, such as Hoesli and Hamelink (1997, p. 369), examine the relationship between stocks, real estate indirect funds and real estate investment. There is also much literature analyzing the relationship between REITS, or real estate specific funds, and property prices. For example, Hufnage and Kloess (2017, p.36) illustrate how the variations of the rental prices are related to the fluctuations of Swiss real estate funds share prices. As indicated in the table 3, time-series describing the Swiss stock market and the fluctuation of Swiss real estate related funds are both included in the model.

5.1.4 Demographics

Demographic trends have indeed a significant impact on consumption in general and also in private and commercial housing demand. Absolute population variations due to births, deaths and net-immigrations have a direct relation with the demand for housing, as well as with house rental prices real estate stock production and renovations (assuming an inelastic supply), Variation in the size of each demographic class often leads home users to seek new home typologies, challenging commercial real estate developers and stimulating new construction (i.e. multifamily housing, mixed generations developments, senior apartments, etc.). Recent studies have analyzed the impact of changing population composition in a more sophisticated way. Examples include Geanakoplos, Magill and Quinzii (2004, p. 272) and Liu and Spiegel (2011, p. 2), who introduced the MY ration and the MO ratio, respectively.

	Indicators
Demographics values	*Tot. Population at 31. Dec* *Population Class (0-19 years)* *Population Class (20-29 years)* *Population Class (30-39 years)* *Population Class (40-49 years)* *Population Class (50-59 years)* *Population Class (60-69 years)* *Population Class (70-99 years)* *Immigration Saldo*
Demographic ratios	*MO Ratio* *MY Ratio*

Table 4: Demographic indicators

The MY ratio (middle-aged population 40-49 years versus young adults 20-29 years) has a strong correlation with private equity market fluctuation over the years. Geanakoplos et al. have found evidence that middle-aged populations have higher purchasing power that supports the demand of investments and drive up market prices. From the real estate perspective, it is reasonable to assume that when the MY ratio is small, young generations support the demand for new apartments (see Illustration 6). Further, as the MY ratio is larger, their greater purchasing power might be correlated with a larger demand of ownership and direct investment. Conversely, the MO ratio (middle-aged 40-49 years versus old 60-69 years) has a strong correlation with the price/earning trends. The explanation may be that middle-aged adults save primarily for safe investments rather than for risky investments such as stock shares.

A specific explanation of the relation between demography and real estate is out of the scope of this analysis. However, the literature mentioned above supports including the indicators listed in the table 4 into the model. These variables may have an influence on variations in the demand for spaces, and consequently may cause a shift of the demand curve in the NW quadrant of the diagram of DiPasquale and Wheaton.

5.1.5 Sentiment and web search data

	Indicators
Sentiment	*Outlook for prices* *Outlook for the general economic situation* *New consumer confidence index* *Job security sentiment* *Likelihood of major purchases*
Web Data	*Google Trends analytics*

Table 5: Sentiment and web-based indicators

Sentiment indexes are "soft" indicators that measure the degree of optimism among survey participants about a specific thematic. Examples include the "state of the economy," or their "personal finance outlook." The importance of those indexes is based on the assumption that consumer sentiment can impact their behavior and ultimately influence their future demand and consumption of goods and services.

Classical economic theory states that in a relatively frictionless market, the mispricing of assets is quickly neutralized by well informed and rational "*homo economicus*," who compete to capture such irregularities. The fact that the classical theory could not adequately explain the persistence of market bubbles led to the rise of a "behavioral" finance approach which recognizes that investors can be irrational when they are subjected to bias that is driven by sentiment. A theory of behavioral economic named "limit to arbitrage" examines how in case of market frictions investor sentiment might hold prices for a long time in a non-equilibrium state. In this way, behavioral economic theory studies apply particularly well to real estate markets, where transactions are often highly inefficient and non-transparent. Examples of those studies include Shilling and Sing (2007, p. 11) and Clayton, Ling and Naranjo (2007, pp.16-21). The first work examines the rationality of investor expectations and the total return in private commercial real estate markets. The second work examines the influence of macroeconomic fundamentals and investor sentiment in explaining the time-series variation in property cap rates by using a set of sentiment proxies[13]. Table 5 lists the selection of sentiment related indexes issued by the SECO and the KOF that are included in the data set used in this analysis.

Another approach to understand changes in the consumer interest in the real estate market is by analyzing real time web generated data streams. New methods of "mining"

[13] "CLN Sentiment is the first principal component of a group of sentiment proxies: commercial mortgage flows as a % of GDP, transaction activity from the NCREIF Property Index (% of properties sold from the NPI), the ratio of the MIT transaction based (TBI) and constant liquidity versions of the NPI value index, the NPI total return over the past 4 quarters, and the most recent quarterly TBI total return. RERC Sentiment is the first principal component extracted from RERC investment condition question survey responses for nine property types." (Clayton, 2007, p. 33).

`

this real-time data include social network web-scraping, machine learning and natural language algorithms or search engine analytics. For example, Google Trends generates an index describing the popularity of a specific search term over time. This index ranges in value from 0 to 100, has a monthly frequency and it is available from the year 2004. Dietzel, Braun and Schäfers (2014, pp.547-552) show how to leverage these data to catch additional real-time insight about the evolution of the real estate market. They analyze the real estate investment process and list what are the common terms researched through Google during each phase (i.e. "Buy Apartment," "Mortgage", "Sale contract", etc.). They define a set of categories of keywords, analyze their frequency in time and finally generate an index that captures the sentiment of real estate investment. In the following analysis, the popularity index generated with Google Trends (ie. term "Hauskaufen", "Makler", "Hypothek") are included, with the aim of providing to the algorithm more information concerning the evolution of the demand and supply of private apartments.

5.1.6 Construction data

To spot the movement into the construction market, various time-series listed in Table 6 are include into the model. These variables describe the evolution of construction prices, the expenditures from the private and public sectors, and the number of building permit approvals. We include in this group also the inflation after observing in the Illustration 7 a linkage with the construction prices.

	Indicators
Building Permits	*Num. approved apartments building permit*
Construction Expenditure	*Private Sector - Housing* *Public Sector - Civil Engineering*
Construction Price Index	*Construction Price Index CH*
Labour	*FTE- Secondary Construction*
Inflation	*Inflation index*

Table 6: Construction activities indicators

5.1.7 Property market data

Finally, the model includes a set of index and price observations directly related with the property market in a broad sense, as illustrated in the table 7.

	Indicators
Rent	*Ask-price for commercial space rental* *Ask-price for office space ren* *Ask-price for retail space ren*
Sale	*Number of transactions*

Table 7: Real estate market direct indicators

5.1.8 Time-series frequency and number

The collected time-series have yearly, quarterly, monthly or daily frequencies, as well as different starting points generating the "*ragged edge"* both at the beginning and at the end of the time window being analyzed. Some of the selected variables have observation recorded starting from the year 1948, but the data-set begins to be consistent from the year 1970. Illustration 19 below shows that as time progresses, the available number of time-series increases, reaching a peak of 113 selected variables over 302 analyzed.

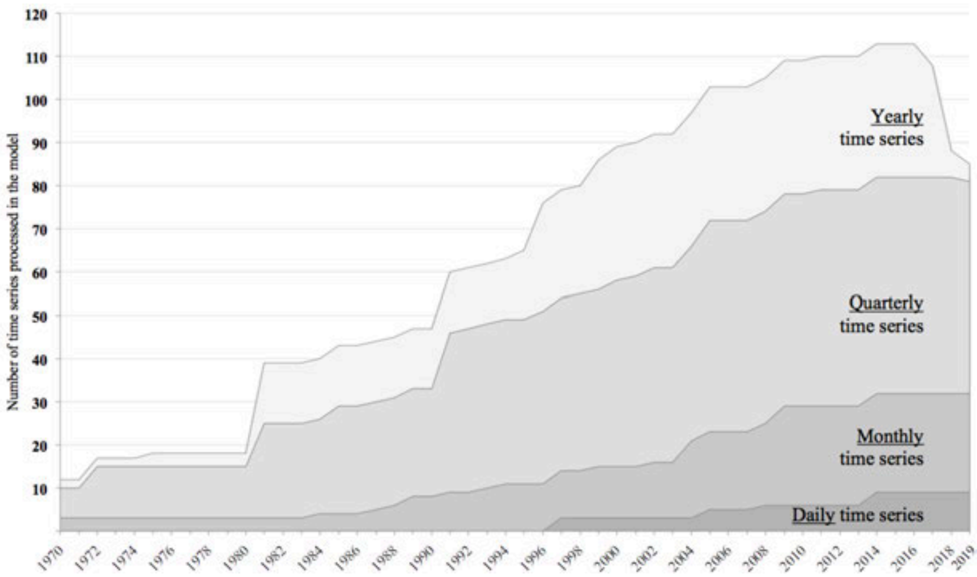


Illustration 19: Number of available time-series over time

It is essential for the calculation of the nowcast to have high frequency variables available. To produce a nowcast on a monthly basis the algorithm needs to process a data set that is composed mainly of (but not limited to) high frequency variables. For instance, economic measures observed daily, weekly or monthly can all be used to provide insights for estimating an updated monthly nowcast. As shown in Illustration 19, the proportion of high frequency variables increases over time. In the year 2019, the algorithm can process 23 monthly updated and 9 daily updated variables. This stable data stream of frequently up to date variables allows the estimation of the nowcast.

5.2 Data stationarization

Time-series variables are a sequence of observations ordered in time over a fix sampling interval. Time-series can be described as being a sum of three components: trend, seasonality and random influences. This basic decomposition (additive, multiplicative or logarithmic) is useful to visualize each component's underlying constant pattern, as exemplified in Illustration 20. Stationa-rization can be explained as a process that aims to isolate the random component with the assumption that the other components remain constant over time. This random component is important because it is defined with statistical terms as mean and variance, rather than using mathematical relations. Therefore, stationarity is a fundamental precondition of most forecasting methods, including nowcasting, because they all follow probabilistic rather deterministic principles. There are three main features that show if a time-series is stationary:

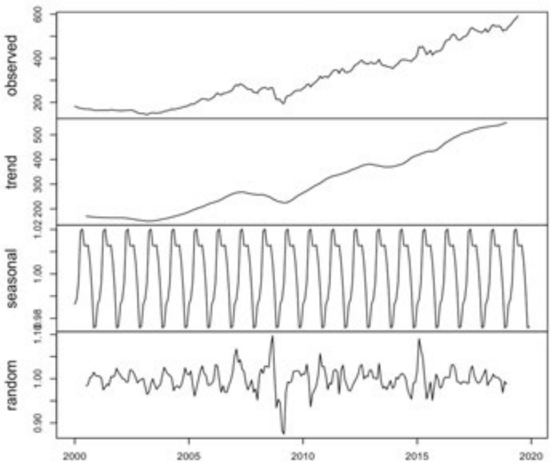


Illustration 20: Multiplicative decomposition of the WupixA index.

1) The mean μ is constant
2) The variance σ is constant, and it doesn't shows homoscedasticity
3) The time-series doesn't show sign of auto-correlation.

If these conditions are satisfied, it means the above-mentioned statistical properties are not a function of time, and the stochastic process is defined as "strictly stationary."

Each time-series selected in this model has been visualized and the hypothesis of stationarity has been systematically tested using the augmented Dickey Fuller test (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS). These are statistical hypothesis tests that are conceived to determine whether a transformation of the time-series is efficient[14]. The explanation of the stationarity tests is outside the scope of this work.

[14] ADF test p-value < 0.05 support stationarity hypothesis. The more negative ADF test statistic, the stronger the evidence for rejecting the null hypothesis of a unit root. The closer to zero are KPSS test statistic, the stronger the evidence of stationarity.

5.2.1 Time-series transformation

All of the time-series variables have been subject to examination to determine whether a transformation is needed for reaching a greater stationarity. The applied differencing formulas are reported here below:

- Rate of change $$\frac{x_{i,t} - x_{i,t-1}}{x_{i,t-1}} \qquad \text{(Trans.1)}$$
- First difference $$x_{i,t} - x_{i,t-1} \qquad \text{(Trans.2)}$$
- Difference in year-over-year rate of change $$\frac{x_{i,t} - x_{i,t-12}}{x_{i,t-12}} - \frac{x_{i,t-1} - x_{i,t-13}}{x_{i,t-13}} \qquad \text{(Trans.3)}$$
- First difference in year difference $$(x_{i,t} - x_{i,t-12}) - (x_{i,t-1} - x_{i,t-13}) \qquad \text{(Trans.4)}$$
- Year difference $$x_{i,t} - x_{i,t-12} \qquad \text{(Trans.5)}$$
- Yearly rate of change $$\frac{x_{i,t} - x_{i,t-12}}{x_{i,t-12}} \qquad \text{(Trans.6)}$$
- Second difference $$x_{i,t} - 2 \cdot x_{i,t-1} - x_{i,t-2} \qquad \text{(Trans.7)}$$
- The original time-series is preserved (Trans.0)

For each transformation test, the ADF and KPSS test has been done together with the visualization of the ACF and PACF correlograms. The appropriate transformation method has been finally decided after a comparison of the results obtained with these tests. An example of this proceed has been reported in the Annex B.

6. Application of the Nowcasting model to the Swiss real estate variables

The nowcasting algorithm has been used to produce monthly frequency estimation of three differently variables:

1) The quarterly residential property prices index of Wuest&Partner for privately owned apartments (time window: from 2000-Q1 until 2019-Q2)
2) The quarterly residential property asking-prices index of Wuest&Partner for privately owned apartments (time window: from 1980-Q1 until 2019-Q2)
3) A calculated half-year square meter median price of commercial housing based on REIDA data (time window: from 2011-Q1 until 2019-Q2)

6.1.1 Data set update and pre-processing

An R script was developed to automate the collection, updating and processing of the data (Annex C). During the first step, the script loads an index spreadsheet with the complete list of all the variables being considered. This index file indicates the variables that are selected to flow into the nowcasting model and provides for each of them the following specifications:

- On-line / Off-line address[15]
- Frequency
- Effective release month for quarterly and yearly observations
- Stationarity transformation[16]
- File format (CSV, XLSX)
- Data structure information (i.e. table header, separator sign, date format, etc.).

As soon as the set of variables and their specifications are loaded, a second script corrects the outliers[17] and finally transforms the series into stationarity. A third script automatically controls the presence of errors[18] and correct them. After the data update and preparation the nowcasting algorithm can then be triggered.

[15] Most variables are automatically updated by downloading data online in real time, however due to difficulties in reading and interpreting some files automatically, a small set of variables is updated manually.
[16] See chapter 5.2.1 Time-series transformation, on page 36.
[17] Outliers are defined as observations of a time series greater than 4 times the interquartile range.
[18] In some rare cases, missing observations or zero values caused the transformation to produce an infinite value that cannot be processed by the nowcasting algorithm.

6.1.2 Nowcast of the Wuest&Partner apartments sale price index

The first variable object of a nowcast is a large-scale and quality adjusted index that is published by Wuest&Partner. This variable represents the evolution of effective transactions of owner-occupied apartments in the whole Switzerland. This variable focuses mainly on single apartment transactions and therefore reflects the private market. This variable is updated on a quarterly base with provisional data and is founded on more than 20’000 market-based transactions. This time-series is available starting from the first quarter of the year 2000 (index = 100).

After processing the variables listed in the Annex A, the nowcasting algorithm estimates the rate of change of the index over time starting from the beginning of the series. Illustration 21 below represents the rate of change in first differences of both the Wuest&Partner index and of the estimated nowcast.

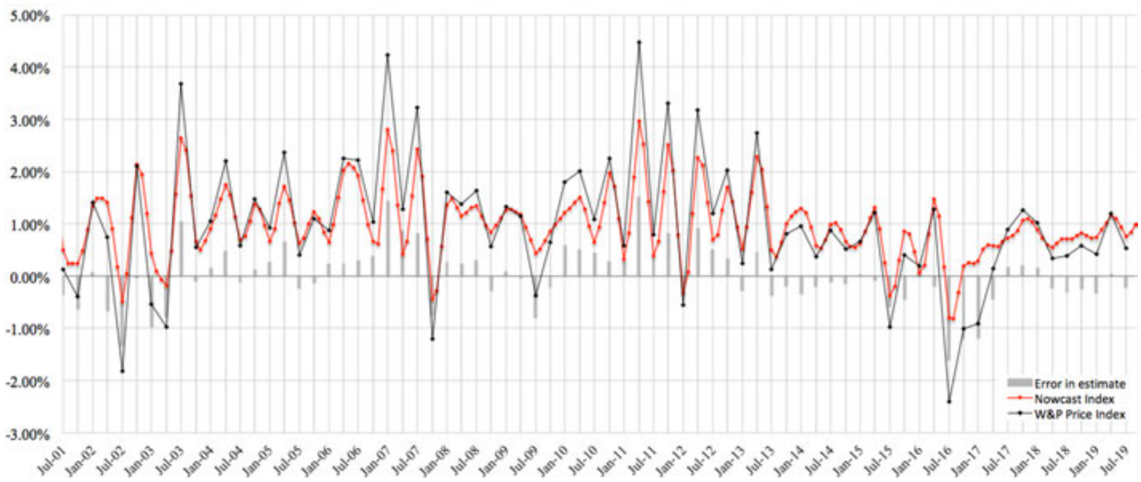


Illustration 21: W&P apartment price index rate of change and nowcast index (Jul01~Aug19)

The black line represents the quarterly price index rate of change. The red line represents the nowcast, as calculated on a monthly frequency. For each month t the nowcasting uses the available data calculating the common factors, the loadings, the idiosyncratic components and finally estimating the new rate of change at the time $t+1$. Referring to Illustration 22 (a close-up view of a shorter time period taken from the previous graphic), a simplified description of the behavior of the algorithm and the way it estimates the Wuest&Partner Price monthly nowcast index is presented.

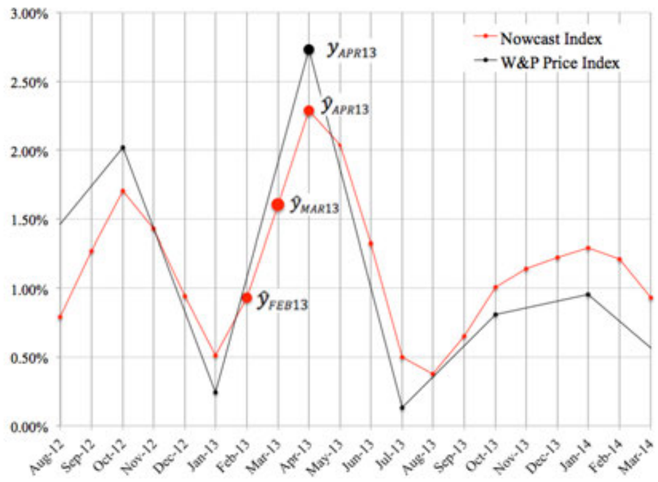


Illustration 22: W&P Apartment price index (Aug12~Mar14)

In the month of February, 2013, the model evaluates all of the available information at that time processing the *vintage* Ω_{FEB13}[19]; it calculates the common factors f_{FEB13}, and estimates their value for the next month $\hat{f}_{MAR13}$. Using the estimated common factors and the relative loadings, the model creates a nowcast of the rate of change of the following month, $\hat{y}_{MAR13}$. During March 2013 the model has at its disposal a new *vintage* of variables Ω_{MAR13} but no new direct observation of the price index. With the new available data the algorithm calculates the common factors f_{MAR13} and compares them with the previously estimated $\hat{f}_{MAR13}$ obtaining new information to update their error terms u_{MAR13}. With this new set of parameters, the model can then calculate the new set of common factors $\hat{f}_{APR13}$ and estimate the rate of change for April ($\hat{y}_{APR13}$). During April, the process repeats as before, but the algorithm now can also use the published value of the price index. The algorithm can now use the observed rate of change (y_{APR13}) and the previous estimate ($\hat{y}_{APR13}$) to calculate the error in the estimate ξ_{APR13}. With this updated values the model proceed to estimate the index $\hat{y}_{MAY13}$.

Following this process, leveraging the dimensionality reduction and the Kalman filter, the model "learns" from the present and estimates the near future. In this way, the model creates a good estimation of the trends, either positive or negative. For example, even after robust economic growth in May of 2014, the model correctly estimates a trend inversion. With this specific set of variables the algorithm also has a fair estimation of the slopes. However, the model struggles to correctly estimate the levels (the actual index rate of change), especially in moment of peak, as shown by the error in estimate bars charts in Illustration 21.

[19] For clarity of explanation, in this section we will use the term Feb13 ,Mar13 and Apr13 in the subscript of the formulas instead that the value *t, t+1, t+2*.
`

Finally, at the present time (15th August 2019), leveraging the available data at the current time *vintage* Ω_{AUG19} the algorithm has calculated the nowcast of August $\hat{y}_{AUG19}$ and September ($\hat{y}_{SEP19}$) and estimate a forecast of October ($\tilde{y}_{OCT19}$).

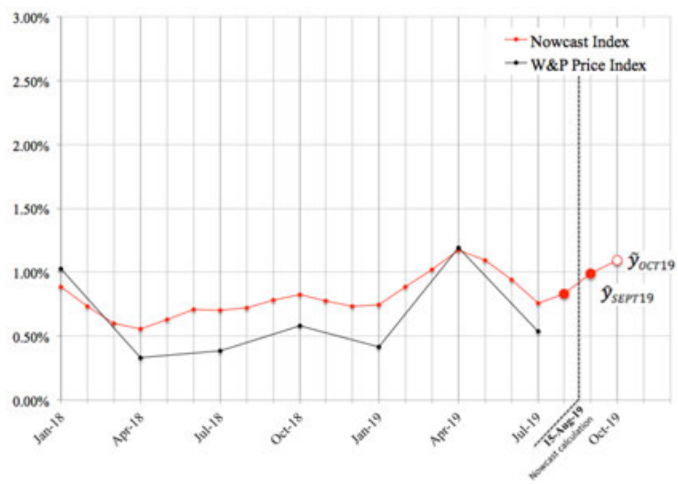


Illustration 23: Apartment price index (Jan18~Oct19)

An actualized version of the nowcast can be visualized at following link below:

tabsoft.co/2jZVSUK

or scanning the QR code in the Illustration 24 on the side.

Illustration 24: QR link to the actualized apartment price index nowcast

6.1.3 Nowcast of the Wuest&Partner apartments sale asking price index

The second variable of concern is a price index of owner occupied apartments purchased and sold in Switzerland. The index is published quarterly by Wuest&Partner and is based on the evaluation of online and printed ads for properties. This index is available since 1970 and is quality adjusted and calculated using the Lowe procedure.

As the Asking Price index provides data since 1970, we have tried to use the nowcasting algorithm to observe how the estimates evolve over a longer period. We decided to let the algorithm process the data starting from the year 1980 because from this point in time the number and the diversity of available time-series is more consistent as represented in the Illustration 19. The monthly nowcast of the rate of change compared with the observed rate of change of the quarterly published index are represented in the Illustration 25.

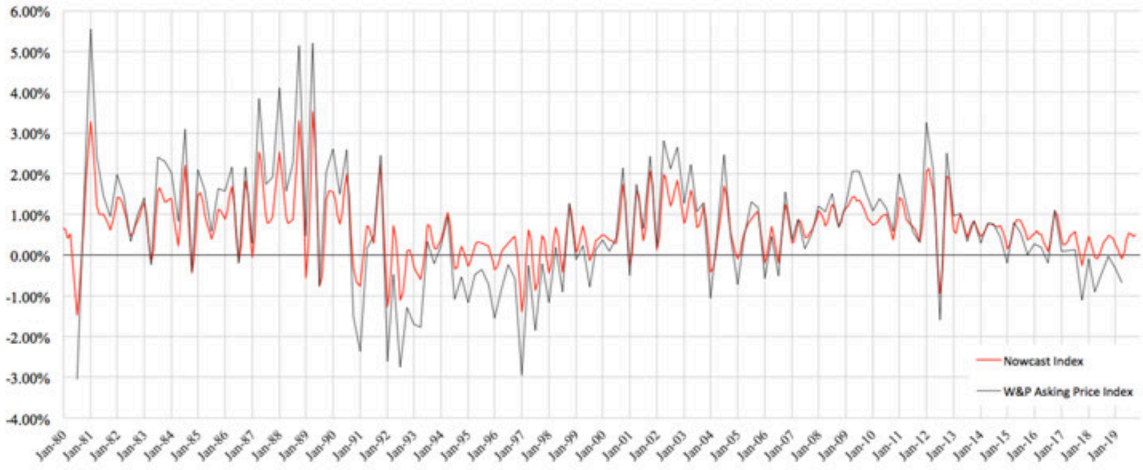


Illustration 25: W&P apartment asking price index rate of change and nowcast index (Jan80~Aug19)

The nowcast model frequently predicts trends in the correct direction, but exhibits difficulty accurately estimating the levels. Generally, the rate of change is under-estimated especially during periods with higher variance. A shift of level prediction is also obvious during the period 1992-1997 and from the year 2017 until the present. Also in these periods the nowcast algorithm tracks the trends quite well but fails to predict the rate of change with an average 0.4% deviation. The latest available data was published in April 2019 and showed a downward trend. The current nowcast estimate (August 2019) depicts a trend inversion.

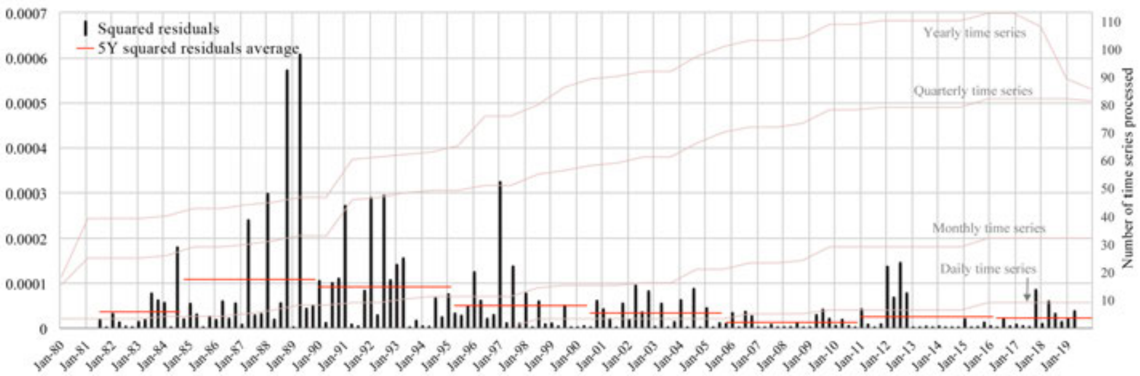


Illustration 26: Squared residual errors versus number of available time-series

The squared residuals between the effective observed value and the nowcast are presented separately in Illustration 26 above. Since the nowcast calculates monthly values and the observations are quarterly, we calculate the residuals subtracting the quarterly observation at time *t* with the average of the monthly estimations at the time *t*-1, *t*, *t*+1.

The squared residuals, represented by black bars, show a tendency to decrease. This is confirmed by the red horizontal line, representing a 5 year residuals average. In the same chart we overlay a graphic with the sum of available data-series year after year (as

Illustration 19). We can observe that in period of similar volatility, as soon as the number of data available increases the squared errors tend to decrease.

6.1.4 Nowcast of commercial housing price based on REIDA data

The third variable of concern focuses on housing properties (occasionally reflecting mixed-use) that are sold and purchased in Switzerland by institutional investors, as provided by REIDA. This data is composed of 792 observations from the time Q1-2011 to the Q2-2019.

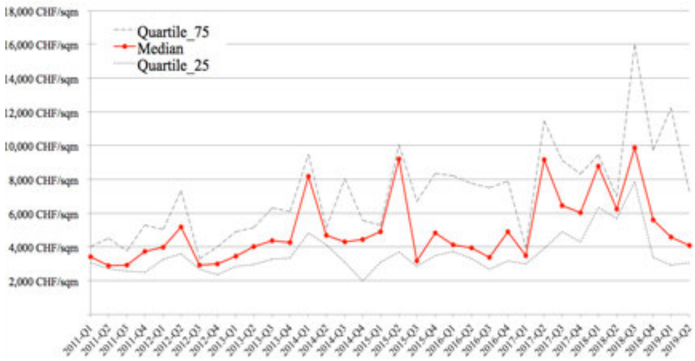


Illustration 27: REIDA commercial apartments median square meter sale price quarterly aggregated

As represented in the Illustration 27, we aggregated the REIDA median square meter prices on a quarterly level observing high interquartile values and irregular skewness in the data. Due to the short data range and small number of observations during certain time periods, the quarterly median price data series might appear to be unsuitable.

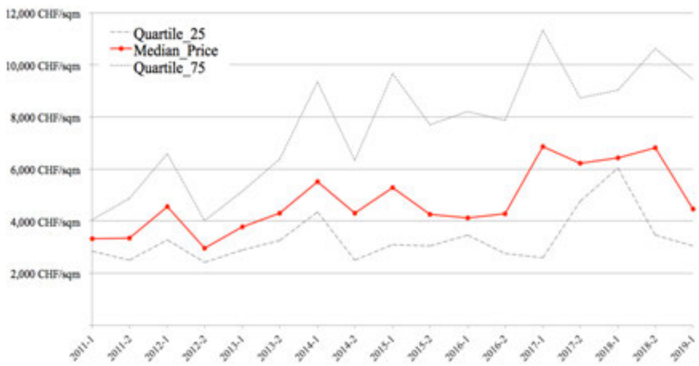


Illustration 28: REIDA commercial apartments median square meter sale price half-yearly aggregated

As a workaround, the data are aggregated on a half-year base, calculating the median of a larger base of price observations (six months observations bin instead than three). This time-series appears to be more regular with clearer trends as showed in Illustration 28. Differently from previous calculations in this case the analysis proceeds to test the nowcasting model using a reference time-series with a 6-months frequency.

The nowcast processes the same data set as in the section 6.1.2. The results are represented in Illustration 29 below. From a year-to-year perspective, the nowcasted prices appear to follow a similar trend as the REIDA median prices. However, the output is noisy due to the fact that the algorithm was conceived to calculate results on a monthly basis.

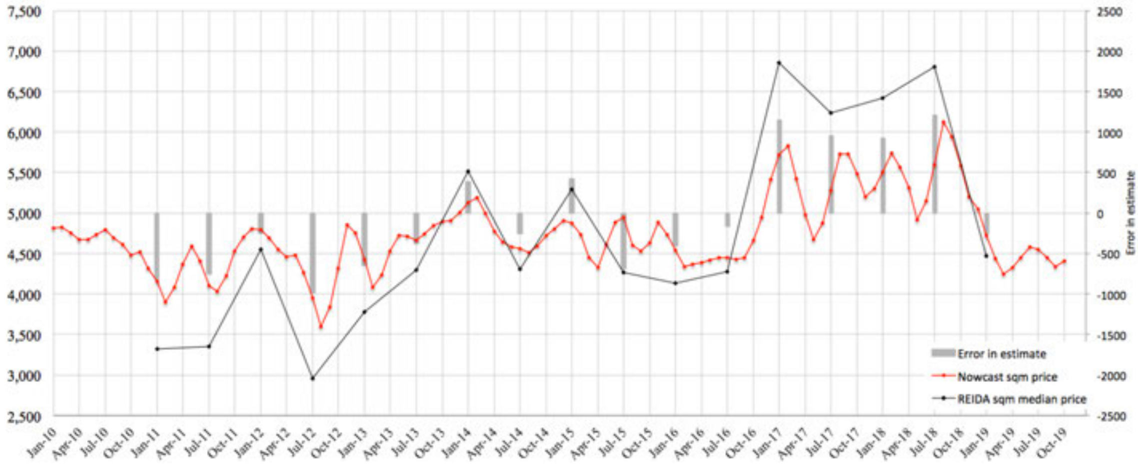


Illustration 29: REIDA commercial apartments square meter median sale price versus nowcast (Jan18~Oct19)

In this specific case, a quarterly nowcast that considers the aggregated trend of the independent variables might provide a smoother result. The utilization of the tool in this application looks to be sub-optimal for this specific aggregated data set.

7. Can nowcasting be a new tool for portfolio managers?

For a better understanding of the potential of the model and its application in practice, the author conducted short interviews with portfolio managers active in the Swiss real estate market. Interviews were conducted either during a meeting or over the phone. During the conversation the author focused on:

- Issues related to market and operating risk assessment
- Macroeconomic indicators and their impact on the decision processes
- Quality and frequency of data used during portfolio management processes.

The interviewees requested anonymity.

7.1 Investment foundation

The interview has been held with a portfolio manager active for of a group of investment foundations directly controlling a portfolio mainly consisting of apartments. Below is a summary of the discussion conversation:

> *Looking in real time at property market "there are trends that you can recognize trough your own experience although they are represented into public statistic with a three or six months lag". The understanding of these trends has an important role especially at a regional level during the process of properties acquisition and appraisal. Those dynamics are often slow, the transactions number is limited, information are not always transparent, as result the data available is scattered and fragmented.*
>
> *We assess a specific market, its trends and risks through consultation within a team of managers and experts evaluating different qualitative and quantitative indicators. Above all interest and vacancy rates are central factors while surveys and real estate portals can help providing more current information supporting recognizing risks and opportunities. It might be useful to have a more frequent and better-structured view of the dynamics of asking prices of properties and the evolution of their vacancy rate.*

7.2 Cantonal pension fund

The interview has been done with a portfolio manager of a cantonal pension found. The managed portfolio consists mainly of rented apartments with a lower percentage of leased space used for administration.

Each year together with the valuation of the portfolio we assess in a qualitative way the risks we are exposed to. We take in consideration the evolution of demand and offer, the risk of recessions, the risk of fluctuation of properties values. The source of information is heterogenic: we refer in some case to institutional national sources as BSF that provides us valuable insight. Some time we need to be critic with those data because in some specific case (i.e. vacancy rate) they have a wide divergence with the reality we observe. On the other side, we leverage our local network and the communication with other cantonal department accessing to a stream of information rooted in the local territory. Although this information exchange activity is not institutionalized into specific processes, it has a high value to perceive local market trends. It would be of great value if we could evaluate more frequently the level of demand and supply in the real estate market. It would also be desirable to have greater transparency in real estate transaction as it happens in the Anglo-Saxon markets to be updated in real time on the evolution of prices.

7.3 Private thematic fund

The portfolio manager interviewed operates for an investment group that manages different thematic funds.

Our fund is characterized by a large portion of assets located in AAA locations that provide stable performances. The properties in this portion of the fund have a mixed functionality with a good percentage of apartments. In the context of diversification we also have a specific fund related to the development of projects for health care and elderly homes. In this case the risk of investments is more significant and we try to mitigate it by examining economic indicators and, in particular, demographic trends. Although demographic data allows a certain degree of predictability, the most recent observations are published with a great time delay. More than a higher data frequency, however, in our case it would be very useful to have more localized information to describe the development of the *economy and population at the neighborhood level and to provide more granular forecasts of local future demand.*

Communication with portfolio managers has revealed a widespread need for more frequent data in order to better recognize apartments demand trends. The quality of the desired information is not only linked to frequency, but also to its geographical localization.

7.4 Private insurance real estate fond

The managed portfolio comprehend assets from different fonds and institutional investors counting totally 190 properties with single value from 10 to 100 Million CHF.

One of the main challenges in portfolio management is the ability to take right decisions at the right time. A condition to reach this goal is to have access to a very up-to-date information stream. This is not trivial nor for the data supplied by external sources, nor for the information sourced within the company.

In most cases, the data are reported with a delay, sometimes of several months, which leaves us unaware of the trends that are taking place today. Nevertheless, we observe systematically the data available, looking at the macro-economy fundamentals as well as demographic and sentiment indicators to support market risk assessments and internal decision processes.

An efficient and timely collection of the most current data can provide the basis for spotting trends inversions and underpin more accurate projections.

8. Conclusion

Real estate economy constantly fluctuates striving to reach equilibrium. It reacts to shifts in both the demand and supply of housing spaces. It is influenced by the competition between financial and real estate investment opportunities. It is conditioned by the new construction activities or the retrofit of existing stock. Meanwhile, economic activities, finance, demography, politics, taxes, sentiments and many other macro-economical forces daily prevent the real estate market from settling into balance.

Previous studies examining the linkage between exogenous macro-economic fundamentals and real estate values have focused on a short-list of main influencing indicators. In contrast this analysis applies a robust statistical model that enable to include a broad set of economical indicators with different frequencies in quasi-real time. This method also overcomes common data limitations due to time inconsistencies and missing observations. The model processes a large data set of time-series describing the co-movements of real estate market and macro-economy as a whole. As a result, the evolution of the real estate market is represented in a holistic way, considering it as both a cause and a consequence of economy dynamics.

This work leverages a nowcasting model, a statistical method originally used in meteorology and recently applied to real-time GDP estimates by many national banks. With this method the analysis evaluates how apartment prices fluctuate along with the fundamentals of the national economy. The real estate market has been interpreted according to the DiPasquale&Wheaton diagram and variables have been considered significant when they could affect at least one of their four quadrants. Using this approach and referring to relevant literature, 113 time-series has been selected and integrated in the nowcasting model.

All the selected variables have been analyzed, the outliners excluded and the optimal transformation for a greater stationary been assigned. The algorithm developed gathers automatically the newest data available, update the time-series with the most recent values, proceed with the stationarization and pass them to the nowcasting calculation. This model elaborates the time-series and parameterizes their co-movements over time using a process of dimensionality reduction synthetizing their dynamics into a small number of unobservable common factors. Common factors and their uncertainty are projected in the time $t+1$ through the creation of a state space model and the application of the Kalman Filter. With the assumption that each variable shift applies a force to the market equilibrium and has a weighted impact on the price dynamic, every time a variable is updated and new real-time data are made available, the model can extract useful information and leverage the dynamic common factors to update real estate price estimations.

As many time-series variables have different frequencies and their new observations are published in different moments throughout the year, this model is exposed to a continuous and variegated stream of new information. The nowcasting model leverages an Expectation-Maximization algorithm to cope with unbalanced panel data due to uneven reporting frequencies or a different publication calendar. In this way, as soon new information flows into the model, the nowcast "captures" the co-movement of the variables into the common factors, describing in a broad sense the fluctuation of the real estate economy and finally infer the housing prices trends.

The nowcast model is used to estimate three endogenous variables describing the evolution of housing prices: the W&P apartment sale price index, the W&P apartment asking price index, the REIDA commercial apartment median price. The first two are published quarterly, while the last one is aggregated half-yearly. The algorithm was designed to create a nowcast with a monthly frequency, adding granularity to the original time-series and providing a quasi-real time estimation of their variations.

As result, the model recognizes trends very well, but it has difficulty predicting levels, especially in times of high economic volatility. While the trends are often estimated correctly, the magnitude of their movement is generally underestimated. This lack in precision can be attributed to both the quality and the quantity of the input observations. Higher frequency in the reported data or a greater number of indicators related to the real estate economy could both improve the quality of the estimate.

This work offers a base for a wide set of further researches and developments. A statistical method could be developed to increase the robustness of this methodology and trace the sources of forecast revisions back to individual predictors. The choice of variables could be reconsidered, expanded or narrowed, depending on other indicators to improve estimates. The method could also be applied to the estimation of other variables other than prices indexes.

9. References

9.1 Citations

Adams, Z., & Füss, R. (2010). *Macroeconomic determinants of international housing markets*. Journal of Housing Economics, 19(1), pp. 38–50

Algieri, B. (2013). *House prices determinants: Fundamentals and underlying factors*. Comparative Economic Studies, 55(2), pp. 315–341

Bańbura, M., Giannone, D., & Reichlin, L. (n.d.). *Nowcasting*. Working Papers European Central Bank, No. 1275(December 2010), pp. 5-28

Bańbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). *Now-casting and the real-time data flow*. Handbook of Economic Forecasting (Vol. 2), p. 196-227

Brownlee, J. (2016). *Master Machine Learning Algorithms Discover How They Work and Implement Them From Scratch*. Machine Learning Mastery With Python, (1), pp. 34-148.

Buono, D., Kapetanios, G., Marcellino, M., Mazzi, G. L., & Papailias, F. (2018). *Evaluation of Nowcasting / Flash Estimation based on a Big Set of Indicators*. 16th Conference of the International Association of Official Statisticians, pp. 1-52

Case, K. E. (2000). *Real Estate and the Macroeconomy*. Brookings Papers on Economic Activity, 2, 119–145

Case, K. E., Quigley, J. M., & Shiller, R. J. (2005). *Comparing Wealth Effects: The Stock Market versus the Housing Market*. Advances in Macroeconomics, pp. 1–29

Castle, J. L., Hendry, D. F., & Kitov, O. I. (2013). *A methodology overview Forecasting and Nowcasting Macroeconomic Variables.* Oxford university discussion papers, (674). pp. 3-65

Chaney, A., & Hoesli, M. (2015). *Multifamily residential asset and space markets and linkages with the economy*. Journal of Property Research, 32(1), pp. 50–76

Clayton, J., Ling, D. C., & Naranjo, A. (2007). *Investor Sentiment and Commercial Real Estate Valuation Commercial Real Estate Valuation: Fundamentals versus Investor Sentiment*, (December 2010), pp. 1-5 /16-21

Constantinescu, M., & Nguyen, A. D. M. (2018). *A century of gaps,* Working Paper Series No 51 / 2018. Lietuvos Bankas Working Paper Series, (51), pp. 4–26

D'Agostino A. , Giannone D. , Lenza M., Modugno M. , *Nowcasting Business Cycles : a Bayesian Approach to Dynamic Heterogeneous Factor Models*, Finance and Economics Discussion Series (2015–066). pp. 2-24

David C. Ling, A. N. (1999). *The integration of commercial Real Estate Markets and Stock Markets,* Real Estate Economics, 27(3), pp. 483–515

Dempster, A., Laird, N., and Rubin, D. (1977), *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, no. 1, pp. 1-21

Dietzel, M. A., Braun, N., & Schäfers, W. (2014). *Sentiment-based commercial real estate forecasting with Google search volume data*. Journal of Property Investment and Finance (Vol. 32), pp. 540-552

DiPasquale, D., & Wheaton, W. (1992). *The markets for real estate assets and space: A conceptual framework*. Journal ofthe American Real Estate & Urban Economics Association, 20, pp. 181–197

Dipasquale, D., & Wheaton, W. C. (1992). The Markets for Real Estate Assets and Space: A Conceptual Framework Denise DiPasquale* and William C. Wheaton**, pp. 181–197

Donoho, D. L. (2000). *High-dimensional data analysis: The curses and blessings of dimensionality*. Aide-Memoire of a Lecture at. AMS Conference on Math Challenges of the 21st Century, pp. 17–23

Doz, C., Giannone, D., & Reichlin, L. (2011). *A two-step estimator for large approximate dynamic factor models*. Journal of Econometrics, 164(1), pp. 192–194

Doz, C., Giannone, D., & Reichlin, L. (2012). *A quasi-maximum likelihood approach for large, approximate dynamic factor models*. Review of Economics and Statistics, 94(4), pp. 1014–1024

Einav, L., & Levin, J. D. (2013). *The data revolution and economic analyses*. Climate Change 2013 - The Physical Science Basis, 53(9), pp. 1–30

Elke, H., & Kloess, S. (2017). *Zusammenhang und Zeitverschiebung von Mietmarktzyklen und Ergebnisentwicklung für Schweizer Immobilienfonds und Immobilienanlagestiftungen*. pp. 5-16

Engle, Robert F., Watson, Mark, 1981. *A one-factor multivariate time series model of metropolitan wage rates*. Journal of the American Statistical Association 76 (376), pp. 774–781

Eurostat, & U.N. (2017). *Handbook on Rapid Estimates*, pp. 17-280, 286-314, 348-516

Faragher R., *Understanding the Basis of the Kalman Filter via a Simple and Intuitive Derivation* [Lecture Notes]. Sept. 2012 Signal Processing Magazine, IEEE , vol.29, no.5, pp. 128-132

Ferrara, L., Simoni, A. (2019). *When Are Google Data Useful to Nowcast Gdp? An Approach Via Pre-Selection and Shrinkage*. SSRN Electronic Journal, pp. 1–28

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2005. *The generalized dynamic factor model: one-sided estimation and forecasting*. Journal of the American , pp. 830-839

Gareth James, Daniela Witten, Trevor Hastie, R. Tibshirani (2013). *Elements in statistical learning with R application*, pp. 15-28, 59-120, 228-248

Geanakoplos, J., Magill, M., & Quinzii, M. (2004). *Demography and the Long-Run Predictability of the Stock Market.* Brookings Papers on Economic Activity, 2, pp. 241–247

Giannone, D., Reichlin, L., Sala, L., 2004. *Monetary policy in real time*. In: Gertler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual. MIT Press, Cambridge, MA, pp. 161–200

Girouard N, M. Kennedy, P. van den Noord, C. Andreé (2006). *Recent house price developments: The role of fundamentals.* OECD Economic Outlook, OECD Economics Department Working Papers No. 475, pp. 9–32

Guagliano, C., & Mantovani, C. (2014). *Real-Time Evaluation of GDP in Some Eurozone Countries.* Ssrn, (June), pp. 7-26

World Meteorological Organization, *Guidelines for Nowcasting Techniques*, No. 1198, 2017 p. xi, pp.10-11

Gyourko J., Keim D.B. (1992),*What Does the Stock Market Tell Us About Real Estate Returns?,* Real Estate Economics, 20(3), pp. 457-485

Hoesli, M., & Hamelink, F. (2007). *An examination of the role of Geneva and Zurich housing in Swiss institutional portfolios*. Journal of Property Valuation and Investment, 15(4), pp. 354–371

Hotelling, H. (1933) *Analysis of a complex of statistical variables into principal components.* J Educ Psychol 24: 417-441, pp. 498-520

Ineichen, A. (n.d.). *Nowcasting : A Risk Management Tool*. Alternative Investment Analyst Review, (Summer 2015), pp. 45–50

Jolliffe, I. (2011). *Principal component analysis*. In International encyclopedia of statistical science, pp. 1094–1096

Kalman, R. E. 1960. "*A New Approach to Linear Filtering and Prediction Problems*," Transaction of the ASME—Journal of Basic Engineering, pp. 35-45 (March 1960)

Khan, A. R. (2007). *Growth, employment and poverty: An analysis of the vital nexus based on some recent UNDP and ILO/SIDA studies*, 49(October). pp. 14-18

King, D. E. (2009). *Dlibml: A Machine Learning Toolkit*. Journal of Machine Learning Research, p. 10

Koivu, T. (2012). *Monetary policy, asset prices and consumption in China*. Economic Systems, 36(2), 307–325

Liu FY. (2014). *The Dynamic Effects of Money Supply on Real Estate Prices in the Japanese Pre-bubble and Bubble Period : Compared with Recent China*, pp. 350-359

Liu, Z., & Spiegel, M. (2011). *Boomer Retirement: Headwinds for U.S. Equity Markets?* FRBSF Economic Letters, pp. 1–4

Quah, Danny, Sargent, Thomas J., 2004. *A dynamic index model for large cross-section.* In: Stock, James, Watson, Mark (Eds.), Business Cycle. Univeristy of Chicago Press, 161–200

Quan, D. C., & Titman, S. (1999). *Do real estate prices and stock prices move together? An international analysis*. Real Estate Economics, 27(2), pp. 183–207

Shilling, J. D., & Sing, T. F. (2007). *Do Institutional Real Estate Investors have Rational Expectations?* , (January 2007), pp. 1–25

Stock, J.H., Watson, M.W., 2002b. *Macroeconomic forecasting using diffusion indexes.* Journal of Business and Economics Statistics 20 (2), pp. 147–162

Tiffin, A. (2016). *Seeing in the Dark : A Machine-Learning Approach to Nowcasting in Lebanon*. International Monetary Fund Working Paper, (No. 16/56), pp. 3-17

Vries, P. De, Boelhouwer, P., & Classification, J. (2014). *House Price, House Quality and Economic Growth*, pp. 1–16

Wheaton, W. C. (1999). *Real Estate "Cycles": Some Fundamentals*. Real Estate Economics, 27, pp. 214-215

Wuensch, K. (2019). *An Introduction to Multivariate Statistics*, pp. 10-13

Yuan, N., Hamori, S., & Chen, W. (2014). *House Prices and Stock Prices: Evidence from a Dynamic Heterogeneous Panel in China*. Kobe University Discussion Paper 1428 , pp. 1-10

Zhou, L., Cheng, X., & Chen, S. (2014). *Analysis of the Effect of Money Supply on Real Estate Price*. Journal of Applied Science and Engineering Innovation ISSN, 1(7), pp. 412-416

9.2 Additional references

Bańbura, M., & Rünstler, G. (2007). *A look into the factor model black box - Publication lags and the role of hard and soft data in forecasting GDP*. Working Papers European Central Bank, No. 751(May 2007), pp. 4-21

Bou, J. C., & Satorra, A. (2018). *Univariate Versus Multivariate Modeling of Panel Data : Model Specification and Goodness-of-Fit Testing* (Vol. 21), pp. 150-171

Browning, K.A., 1981: *A total system approach to a weather radar network, proceedings of the International Symposium "Nowcasting: Mesoscale Observations and Short-Range Prediction",* Hamburg, Germany, pp. 115-122

Forni M., Giannone, D., Lippi, M., & Reichlin, L. (2007). *Structural Factors models with large cross -sections*. ECB Working Papers, pp. 8–21

Gupta, M. & Chen, Y. (2010) *Theory and Use of the EM Algorithm. Foundations and Trends in Signal Processing,* Vol. 4, No. 3, pp. 223–296

Piazzesi, M., Schneider, M., & Tuzel, S. (2007). *Housing, consumption and asset pricing. Journal of Financial Economics*, 83(3), pp. 531–569

Rhoads, N., *Predicting Market Data Using The Kalman Filter*, January 2010 and Part 2 , February 2010, pp. 1–10

Rossi, Richard J. (2018). *Mathematical Statistics: An Introduction to Likelihood Based Inference.* New York: John Wiley & Sons. pp. 227

Tepper M., *Weather now-casting,* Bulletin of the American Meteorological Society, Vol. 52, Amer Meteorological Soc 45 Beacon St, Boston, MA 02108-3693, 1971, pp. 1173

Xu, T. (2017). *The Relationship between Interest Rates, Income, GDP Growth and House Prices.* Research in Economics and Management, 2(1), p. 30

9.3 On-line resources

Athey, Susan, and Guido Imbens (2015) NBER Lectures on Machine Learning
http://www.nber.org/econometrics_minicourse_2015/

Davis E. King (2009) Journal of Machine Learning Research
http://dlib.net/ml.html

DeMatto, Gomes, Ferreira & de Valk (2019) CRAN open source nowcasting package 'nowcasting' version 1.1.3,
https://cran.r-project.org/web/packages/nowcasting/index.html

Eigenvectors and Eigenvalues Explained Visually, no date
http://setosa.io/ev/eigenvectors-and-eigenvalues/

STAT 505: Applied Multivariate Statistical Analysis, no date
https://newonlinecourses.science.psu.edu/stat505/

Annex A: Time-series list

Group	Variable Name	Description	Frequ ency	Transfor mation	Source
Accounts	JOBLESS_RATE	Jobless Rate	12	Transf. 1	SNB
Accounts	BANK_PRVT_DEPST_AVG_CH	Average Bank Deposit Nominal Value CH	4	Transf. 4	SNB
Accounts	FIN_CORP_ASST_T1	Financial assets - All financial instruments	1	Transf. 1	SNB
Accounts	FIN_CORP_LBL_T1	Liabilities - All financial instruments	1	Transf. 1	SNB
Accounts	FIN_HSHLD_DSPSBINCOM_OECD	Household Disposable Income OECD	1	Transf. 1	OECD
Accounts	FIN_HSHLD_FINASST_OECD	Household financial assets OECD	1	Transf. 1	OECD
Accounts	FIN_HSHLD_ASSET_IMMO_SNB	Household Assets - Real Estate SNB	1	Transf. 1	SNB
Accounts	FIN_HSHLD_SAVINGS_OECD	Household savings	1	Transf. 1	OECD
Construction	CONSTR_APPROVED_PROJECT_NUM_APT_CH	Number of apartment in approved building permit	12	Transf. 1	W&P
Construction	CONSTR_PRICE_INDEX_CH	Construction Price Index CH	4	Transf. 1	SNB
Consumption	RETAIL_TURNOVER_REAL-TOTAL	Retail totoal turnover	12	Transf. 1	SNB
Debt	FIN_HSHLD_DEBT_OECD	Household debt (mostly mortgages) OECD	1	Transf. 1	OECD
Debt	FIN_HSHLD_HYPO_SNB	Household Morgages SNB	1	Transf. 1	SNB
Debt	HYPO_ALLBANKS_CH	Domestic Mortgage Loan, All currency included, All banks	12	Transf. 4	SNB
Debt	HYPO_CONSTR	Mortgage loans-Total maturity-Construction	12	Transf. 4	SNB
Debt	HYPO_PRIV_HSHOLD	Mortgage loans-Total maturity-Private households	12	Transf. 2	SNB
Debt	HYPO_RE_ACTIVITIES	Mortgage loans-Total maturity- Real estate activities	12	Transf. 2	SNB
Demographic	DEMOGR_IMMIGR_SALDO_CH	Immigration Saldo	1	Transf. 2	BFS
Demographic	DEMOGR_MO_RATIO_CH	Middle-aged population (40-49 years) relative to old adults (60-69 years)	1	Transf.	BFS
Demographic	DEMOGR_MY_RATIO_CH	Middle-aged population (40-49 years) relative to young adults (20-29 years)	1	Transf.	BFS
Demographic	DEMOGR_POPULATION_CH	Total population at the 31. December	1	Transf. 1	BFS
Demographic	DEMOGR_POPULATION_CLASS_0-19Y_CH	Population Class (0-19 years)	1	Transf.	BFS
Demographic	DEMOGR_POPULATION_CLASS_20-29Y_CH	Population Class (20-29 years)	1	Transf.	BFS
Demographic	DEMOGR_POPULATION_CLASS_30-39Y_CH	Population Class (30-39 years)	1	Transf.	BFS
Demographic	DEMOGR_POPULATION_CLASS_40-49Y_CH	Population Class (40-49 years)	1	Transf.	BFS
Demographic	DEMOGR_POPULATION_CLASS_50-59Y_CH	Population Class (50-59 years)	1	Transf.	BFS
Demographic	DEMOGR_POPULATION_CLASS_60-69Y_CH	Population Class (60-69 years)	1	Transf.	BFS
Demographic	DEMOGR_POPULATION_CLASS_70-99Y_CH	Population Class (70-99 years)	1	Transf.	BFS
Economic Activity	CONSTR_EXPEND_PRIV_HSNG	Construction expenditure - Private Sector - Housing	1	Transf. 1	SNB
Economic Activity	CONSTR_EXPEND_PUB_CIVIL	Construction expenditure - Public Sector - Civil Engineering	1	Transf. 1	SNB
Economic Activity	GDP_NOM_QRT_CHANGE_EXPENDITURE_GOVERNMENT	Final consumption expenditure - Government	4	Transf.	SNB
Economic Activity	GDP_NOM_QRT_CHANGE_EXPENDITURE_HOUSEHOLDS_NPIESH	Final consumption expenditure - Households and NPISHs	4	Transf.	SNB
Economic Activity	GDP_NOM_QRT_CHANGE_EXPORT_GOODS	Exports - Exports of goods excluding non-monetary gold	4	Transf.	SNB
Economic	GDP_NOM_QRT_CHANGE_EXPORT_S	Exports - Exports of services	4	Transf.	SNB

Group	Variable Name	Description	Frequency	Transformation	Source
Activity	ERVICES				
Economic Activity	GDP_NOM_QRT_CHANGE_FIXED_CAPIT_FORMATION_CONSTRUCTION	Gross fixed capital formation - Construction	4	Transf.	SNB
Economic Activity	GDP_NOM_QRT_CHANGE_FIXED_CAPIT_FORMATION_EQUIPMENT	Gross fixed capital formation - Equipment	4	Transf.	SNB
Economic Activity	GDP_NOM_QRT_CHANGE_GROSS_DOMESTIC_PRODUCT	Gross domestic product Quarterly Series	4	Transf.	SNB
Economic Activity	GDP_NOM_QRT_CHANGE_IMPORT_GOODS	Imports - Imports of goods excluding non-monetary gold	4	Transf.	SNB
Economic Activity	GDP_NOM_QRT_CHANGE_IMPORT_SERVICES	Imports - Imports of services	4	Transf.	SNB
Economic Activity	GDP_NOM_YR_CHANGE_COMPENS_EMPLOYEE	Compensation of employees	1	Transf.	SNB
Economic Activity	GDP_NOM_YR_CHANGE_COMPENS_EMPLOYEE_RECEIVED_RESTWORLD	Compensation of employees received from the rest of the world	1	Transf.	SNB
Economic Activity	GDP_NOM_YR_CHANGE_CONSUMPT_FIX_CAP	Consumption of fixed capital	1	Transf.	SNB
Economic Activity	GDP_NOM_YR_CHANGE_NET_OPER_SURPLUS	Net operating surplus	1	Transf.	SNB
Economic Activity	GDP_NOM_YR_CHANGE_PROPERTY_INCOME_RECEIVED_RESTWORLD	Property income received from the rest of the world	1	Transf.	SNB
Economic Activity	GDP_NOM_YR_CHANGE_SUBSIDIES	Subsidies	1	Transf.	SNB
Economic Activity	GDP_NOM_YR_CHANGE_TAXES_PRODUCTION_AND_IMPORT_	Taxes on production and imports	1	Transf.	SNB
Economic Activity	GNI_NOM_YR_CHANGE	Gross national income (GNI)	1	Transf.	SNB
Economic Activity	GDP_PRO_CAPITA	GDP pro Capita	4	Transf. 1	SNB
Economic Activity	INFLATION	Montly frequency with 2015 = 100	12	Transf. 4	OECD
Interest Rates	CH_TBOND_10Y	CHF Swiss Confederation bond issues 10Y	12	Transf. 2	SNB
Interest Rates	LIBOR_12M	London Interbank Offered Rate.	12	Transf. 3	SNB
Interest Rates	LIBOR_3M	London Interbank Offered Rate.	12	Transf. 3	SNB
Interest Rates	REAL_M2	Monetary aggregate M2	12	Transf. 4	SNB
Interest Rates	SARON	Swiss Average Rate Overnight, 12.00 noon fixing	12	Transf. 2	SNB
Labour	LABOUR_ACT_BB	FTE-Secondary Construction	4	Transf. 4	SNB
Labour	JOB_VACANCIES	Notified Job Vacancies	12	Transf. 1	SNB
Labour	LABOUR_ACT_ED	FTE-Other service activities	4	Transf. 4	SNB
Labour	LABOUR_ACT_EF	FTE-Provision of financial services	4	Transf. 4	SNB
Labour	LABOUR_ACT_EU	FTE-Education	4	Transf. 4	SNB
Labour	LABOUR_ACT_GBG	FTE-Hospitality/Accommodation / F&B service activities	4	Transf. 4	SNB
Labour	LABOUR_ACT_GS	FTE-Health and social work activities	4	Transf. 4	SNB
Labour	LABOUR_ACT_GW	FTE-Real estate activities	4	Transf. 4	SNB
Labour	LABOUR_ACT_HIRK	FTE-Trading, maintenance and repair of motor vehicles	4	Transf. 4	SNB
Labour	LABOUR_ACT_OV	FTE-Public administration	4	Transf. 4	SNB
Labour	LABOUR_ACT_V	FTE-Insurance	4	Transf. 4	SNB
Labour	LABOUR_ACT_VGHW	FTE-Secondary Manufacturing	4	Transf. 4	SNB
Labour	LABOUR_ACT_VL	FTE-Transport, storage and warehousing	4	Transf. 4	SNB
Consumption	RE_APARTAM__VACANCY	Apartments Vacancy Rate	1	Transf. 1	SNB
Rent Prices	RE_INDEX_ASKPRICE_APT_RENT_CH	Ask-price for apartments rental	4	Transf. 1	W&P
Rent Prices	RE_INDEX_ASKPRICE_APT_SELL_CH	Ask-price for apartments sale	4	Transf. 1	W&P
Rent Prices	RE_INDEX_ASKPRICE_COMM_RENT_CH_SNB_WP	Ask-price for commercial space rental	4	Transf. 4	SNB

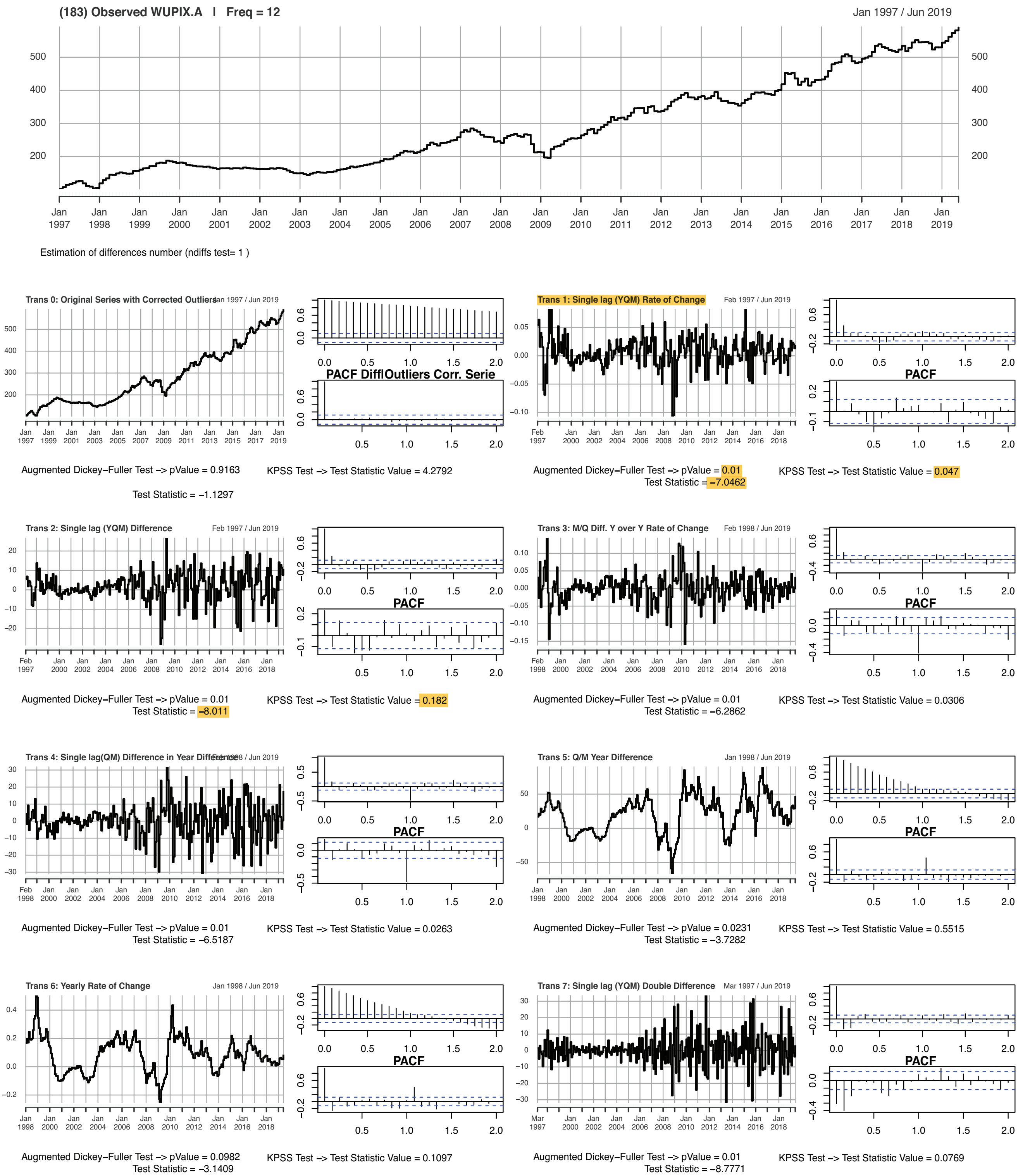
Group	Variable Name	Description	Frequency	Transformation	Source
Rent Prices	RE_INDEX_ASKPRICE_OFF_RENT_CH_SNB_WP	Ask-price for office space rental	4	Transf. 4	SNB
Rent Prices	RE_INDEX_ASKPRICE_RETAIL_RENT_CH_WP	Ask-price for retail space rental	4	Transf. 4	SNB
Rent Prices	RE_INDEX_RENTED_APT_GS	Price index for rental housing units - Total Switzerland	1	Transf. 4	SNB
Sale Prices	RE_INDEX_ASKPRICE_SNGLHOUSE_CH	Ask-price for single house sale	4	Transf. 4	W&P
Sale Prices	RE_INDEX_PRIV-APT_GS	Price index for privately owned apartments - Total Switzerland	1	Transf. 4	SNB
Sale Prices	RE_INDEX_TRANS_APT_ASKPRICE_CH_SNB_WP	Transaction price index for residential property prices	4	Transf. 4	SNB
Sale Prices	RE_INDEX_TRANS_APT_CH_SNB_WP	Residential property prices index Privately owned apartments - Wüest Partner	4	Transf. 1	SNB
Sale Prices	RE_INDEX_TRANS_APT_TRANS_CH_SNB_FP	Transaction price index for privately owned apartments - Fahrländer Partner	4	Transf. 4	SNB
Sale Prices	RE_INDEX_TRANS_APT_TRANS_CH_SNB_IAZI	Transaction price index for privately owned apartments - IAZI Privately owned apartments IAZI	4	Transf. 4	SNB
Sale Prices	RE_INDEX_TRANS_EFH_ASKPRICE_CH_SNB_WP	Asking price index for privately owned apartments - Single-family houses Wüest Partner	4	Transf. 4	SNB
Sale Prices	RE_INDEX_TRANS_RENDOBJ_TRANS_CH_SNB_WP	Transaction price index for rented apartments - Wüest Partner	4	Transf. 4	SNB
Sale Prices	RE_MEDIAN_REIDA_PRICE	REIDA Median Price	4	Transf.	REIDA
Sentiment	KOF_ECON_BAROMETER	KOF Composite Indicator for the Swiss Business Cycle	12	Transf. 1	KOF
Sentiment	KOF_BUSINESS_SITUATION_TOT_INDEX	KOF Survey about Business Situationa and sentiment	12	Transf. 3	KOF
Sentiment	KOF_BUSINESS_SITUATION_CONSTRUCTION	KOF Survey about Construction related companies Situationa and sentiment	12	Transf. 3	KOF
Sentiment	KOF_EMPLOYMENT_INDICATOR	Based on the quarterly KOF Business Tendency Surveys	4	Transf. 1	KOF
Sentiment	KOF_BAUBLATT_OUTLOOK	Outlook of Construction Activities	4	Transf. 2	KOF
Sentiment	CONS_CONFIDENCE_EXPECTED_PRICES	Outlook for prices	4	Transf. 2	SECO
Sentiment	CONS_CONFIDENCE_GENERAL_ECON_OUTLOOK	Outlook for the general economic situation	4	Transf. 2	SECO
Sentiment	CONS_CONFIDENCE_CONS_CONF_INDEX_NEW	New consumer confidence index	4	Transf. 2	SECO
Sentiment	CONS_CONFIDENCE_JOB_SECURITY	Job security	4	Transf. 2	SECO
Sentiment	CONS_CONFIDENCE_MAJOR_PURCHASE_LIKELIHOOD	Likelihood of major purchases	4	Transf. 2	SECO
Sentiment	GOOGLE_TREND_Immobilien_Bewertung	Google Trends of the term "Immobilien Bewertung"	12	Transf. 4	GOOGLE
Sentiment	GOOGLE_TREND_Wohnung_Kaufen	Google Trends of the term "Wohnung Kaufen"	12	Transf. 4	GOOGLE
Sentiment	GOOGLE_TREND_Hypothek	Google Trends of the term "Hypothek"	12	Transf. 4	GOOGLE
Sentiment	GOOGLE_TREND_Hauskauf	Google Trends of the term "Hauskauf"	12	Transf. 4	GOOGLE
Sentiment	GOOGLE_TREND_Makler	Google Trends of the term "Makler"	12	Transf. 4	GOOGLE
Sentiment	GOOGLE_TREND_Wohnung_Verkaufen	Google Trends of the term "Wohnung verkaufen"	12	Transf. 4	GOOGLE
Shares Market	SPI	Swiss Performance Index	360	Transf. 3	W&P
Shares Market	WUPIX-A	Index of dividend of listed real estate companies	360	Transf. 3	W&P
Shares Market	WUPIX-F	Index of dividend of listed real estate funds	360	Transf. 1	W&P
Shares Market	BEKB_OTIX_IMMO	OTC-X Immobilien	360	Transf. 2	BEKB

`

Group	Variable Name	Description	Frequency	Transformation	Source
Shares Market	SXI_RE_ALL_SHARES_TOT-RET	SXI Real Estate - All Shares Index Total Return	360	Transf. 2	SIX
Shares Market	SXI_RE_ALL_SHARES_PRICE	SXI Real Estate - All Shares Index Price	360	Transf. 2	SIX
Shares Market	RUD-BLASS_IMMOFOND_NAV	Rüd Blass Immobilienfonds Index NAV	360	Transf. 1	FUW
Shares Market	RUD-BLASS_IMMOFOND_GG	Rüd Blass Immobilienfonds Index Gleichgewichted	360	Transf. 1	FUW
Shares Market	SWX_SP_CONST-MAT	Construction and Material related SXI Listed companies	360	Transf. 3	FUW

Annex B: Stationarization test

The page is exemplificative of a single time-series (WUPIX_A). The same graphic has been automatically exported for all the time-series analyzed.



`

Annex C: Data selection and update algorithm R code

```
##################################
### NOWCAST EM WITH LARGE DATASET
##################################

Nowcast.EM <- function(Index.Path = "DataIndex_EM.xlsx" , Y_VarName, TimeWindow =
"1980/2021"){
    #load and unload libraries
    library(matlab)
    library(xts)
    library(stats)
    library(Matrix)
    library(dplyr)
    detach("package:dplyr",TRUE) # attach and detach to ensure that DPLYR is detached
    # needed loaded r files:
    # my.Nowcast, my.method_EM, method_2s, month2qtr, qtr2month
    my.data <- DataIndex_Reader (Index.Path)
    ## interpolate the Yearly value into Quarterly
    ObsMQ_InterpY2Q <- year2qtr(DataObserved = my.data$Observed[TimeWindow], DataInterpolated
= my.data$Interpolated[TimeWindow] ,  Frequency = my.data$Frequency )
    my.data$ObsMQ_InterpY2Q <- ObsMQ_InterpY2Q$ObsMQ_InterpY2Q
    my.data$Frequency_InterpY2Q <- ObsMQ_InterpY2Q$New.Frequency
    data.base.EM <- stationarized.bpanel(base = my.data$ObsMQ_InterpY2Q[TimeWindow],trans =
my.data$Transformation ,Blocks = my.data$Blocks, Frequency = my.data$Frequency_InterpY2Q,
NA.replace = F, na.prop = 1)
    #AUTOMATIC DEBUG PANEL from Infinite or #NUM values

data.base.EM$TimeSerieStationarized.ts[which(is.infinite(data.base.EM$TimeSerieStationarized.t
s))] <- NA
    # as for example in:
    # data.base.EM$TimeSerieStationarized.ts [177,"CONSTR_APPROVED_PROJECT_NUM_APT_CH"] <- NA
    # data.base.EM$TimeSerieStationarized.ts [339,"FIN_CORP_LBL_T2"] <- NA
    #DEBUG SUPPORT
    #xlsfilename = "data.base.EM"
    #openxlsx::write.xlsx(as.data.frame(data.base.EM[1]),  file = paste0(xlsfilename,".xlsx")
, sheetName = "data.base.EM", row.names= TRUE, Append = TRUE, overwrite=T)
    #### NOWCAST FUNCTION
    #### Careful! It work with STATS not with DPLYR (Dplyr must switch OFF!!)
    my.nowcast.EM <- nowcast(formula =  as.formula(paste(Y_VarName , paste("."), sep="~")),
                             data =  data.base.EM$TimeSerieStationarized.ts,
                             r=1, p=1 ,
                             method = 'EM', blocks = data.base.EM$Blocks,
                             frequency = data.base.EM$Frequency)
    #prepare the XY base xts based on the nowcast function results
    nowcast_yfcst <- my.nowcast.EM$yfcst
    nowcast_yfcst <- as.xts(nowcast_yfcst)
    #transform the Y Observed into Quarterly TS to syncronize with the fitted Y()
    # Make it TS to prepare it to qtr2month function
    Y_Observed_qtr.ts <-ts(data = my.data$Observed[,Y_VarName],start =
start(my.data$Observed[,Y_VarName]),frequency = frequency(my.data$Observed[,Y_VarName]))
    Y_Observed_qtr <- month2qtr(x = Y_Observed_qtr.ts)
    #Merge the Nowcasted Y (Stationarized values)  and the Original Observed Y
    XY <- merge.xts(nowcast_yfcst, Y_Observed_qtr, join = "left")
    XY <- merge.xts(XY,NA,NA,NA)
    names(XY)[5:7] <- c("Y_NOWCAST_DESTATIONARIZED" ,"Y_RESIDUALS", "Y_FORECAST")
    TT <- nrow(XY)
    #### DESTATIONARIZE
    y_pos <- which(colnames(my.data$Observed)==Y_VarName)
    #DeStationarize_Trans 1
    if (my.data$Transformation[y_pos] == 1) {
          for (i in 2:TT) {
            if(is.na(XY[i,2])){
              #If Observation are availabele, calculate the nowcast
              if(!is.na(XY[(i-1),4])){
                XY[i,5] <- as.numeric(XY[i,3])*as.numeric(XY[(i-1),4])+as.numeric(XY[(i-1),4])
              }
              #If Observation are not available, calculate the Forecast (Should be just in the
tail)
              if(is.na(XY[(i-1),4])){
                if(!is.na(XY[(i-1),5])){XY[(i-1),7] <- XY[(i-1),5]}
                XY[i,7] <- as.numeric(XY[i-1,5])*as.numeric(XY[(i-1),4])+as.numeric(XY[(i-
1),4])
              }
            }
            if(is.na(XY[i,5])){
              XY[i,5] <- as.numeric(XY[i,2])*as.numeric(XY[(i-1),4])+as.numeric(XY[(i-1),4])
            }
          }
    }
    #DeStationarize_Trans 4
    if (my.data$Transformation[y_pos] == 4) {
          for (i in 6:TT) {
            if(is.na(XY[i,2])){
              #If Observation are availabele, calculate the nowcast
              if(!is.na(XY[(i-1),4])){
                XY[i,5] <- as.numeric(XY[i,3])+as.numeric(XY[(i-4),4])+as.numeric(XY[(i-
1),4])-as.numeric(XY[(i-5),4])
              }
              #If Observation are not available, calculate the Forecast (Should be just in the
tail)
              if(is.na(XY[(i-1),4])){
                if(!is.na(XY[(i-1),5])){XY[(i-1),7] <- XY[(i-1),5]}
                XY[i,7] <- as.numeric(XY[i,3])+as.numeric(XY[(i-4),4])+as.numeric(XY[(i-
1),7])-as.numeric(XY[(i-5),4])
              }
            }
            if(is.na(XY[i,5])){
              XY[i,5] <- as.numeric(XY[i,2])+as.numeric(XY[(i-4),4])+as.numeric(XY[(i-1),4])-
as.numeric(XY[(i-5),4])
            }
          }
    }
    XY[,6] <- as.numeric(XY[,4])- as.numeric(XY[,5])
    #### EVALUATE THE MODEL
    #Sum of the Residuals between the Y and  Yhat
    #XY_SUM_RESID <- sum(as.vector(na.omit(XY[TimeWindow,6]^2)))
    XY_SUM_RES_STD <- sum(as.vector( scale(XY[TimeWindow,6])^2), na.rm = TRUE)
    return(list(Observations.Panel =my.data,
                Stationarized.Panel = data.base.EM,
                Nowcast =my.nowcast.EM,
                Y.Destationarized = XY,
                Std.Residuals = XY_SUM_RES_STD))
}
##################################
### PLOT THE RESUTLTS
##################################
function() {
my.nowcast <- my.nowcast.EM_ch_2e
nowcasting::nowcast.plot(my.nowcast$Nowcast, type = "fcst")
plot(my.nowcast$Y.Destationarized[,4], ylim = c(95,220))
lines(my.nowcast$Y.Destationarized[,5], col = "blue")
lines(my.nowcast$Y.Destationarized[,7], col = "red")
plot(my.nowcast$Y.Destationarized[,6])
my.nowcast$Std.Residuals
tail(my.nowcast$Y.Destationarized[,4:5],20)
}
##########################################################################
########################     DATAINDEX.READER      ########################
##########################################################################
#Index.Path <- "DataIndex.xlsx"
#my.data <- DataIndex_Reader (Index.Path)
DataIndex_Reader <- function(Index.Path, nForecastMonths = 14, SaveCSV = FALSE) {
  #' @param Index.Path The path of the Index Table
  library(dplyr)
  #import the index table
  my.Index.Table <- readxl::read_xlsx(Index.Path, skip = 1)
  #select just the elements marked with "YES" on the column "INCLUDE"
  my.Index.Table <- subset(my.Index.Table, my.Index.Table$INCLUDE == "YES")
  my.data.transformation <- my.Index.Table$TRANSFORMATION
  my.data.frequency      <- my.Index.Table$FREQUENCY
  #inizialize the my.data lists
  my.data.observed         <- list()
  my.data.interpolated     <- list()
  my.data.observed.STD     <- list()
  my.data.interpolated.STD <- list()
  my.data.Blocks           <- matrix()
  my.data.frequency        <- list()
  for(i in 1:nrow(my.Index.Table) ) {
    # Call the CSV.Data.Loader for the data marked as CSV_ONLINE or CSV OFFLINE
    # call the data loader for each row of the Index Table and store it in Data.Temp
    if (my.Index.Table$EXTRACTION_METHOD[i] == "CSV_ONLINE" |
my.Index.Table$EXTRACTION_METHOD[i] == "CSV_OFFLINE" ) {
                        if (SaveCSV == TRUE) {
                          #If the CSV is online Download the CSV File into the "DATA_DOWNLOAD/"
folder
if(substr(my.Index.Table$CSV_DOWNLOAD[i],1,4)=="http") {
download.file(url = my.Index.Table$CSV_DOWNLOAD[i], destfile =
paste0("DATA_DOWNLOAD/",my.Index.Table$GROUP_NAME[i],".csv"),quiet = TRUE) }
}
Data.ts.Temp <-  CSV.Data.Loader( File.Address = my.Index.Table$CSV_DOWNLOAD[i],
ExcludedColumns =  eval(parse(text = my.Index.Table$COL_EXCLUDE[i])),
Date.Format = my.Index.Table$DATE_FORMAT[i]          ,
Release.Month = my.Index.Table$RELEASE_MONTH[i]      ,
my.frequency = my.Index.Table$FREQUENCY[i]           ,
column.varnames =  my.Index.Table$COL_VAR_NAME[i] ,
column.values = my.Index.Table$COL_VALUE[i],
column.date = my.Index.Table$COL_DATE[i],
group.name = my.Index.Table$GROUP_NAME[i],
Rows.Skip= my.Index.Table$SKIP[i] , Col.Separator= my.Index.Table$COL_SEP[i],
Fill.Method = my.Index.Table$FILL_METHOD[i],
SNB_D0 = my.Index.Table$SNB_D0[i] , SNB_D1 = my.Index.Table$SNB_D1[i],
Transformation = my.Index.Table$TRANSFORMATION[i],
my.Blocks = eval(parse(text = my.Index.Table$BLOCKS[i]))
)
}
if (my.Index.Table$EXTRACTION_METHOD[i] == "XLSX_ONLINE" | my.Index.Table$EXTRACTION_METHOD[i]
== "XLSX_OFFLINE" ) { Data.ts.Temp <-XLSX.Data.Loader(File.Address =
my.Index.Table$CSV_DOWNLOAD[i],
my.frequency = my.Index.Table$FREQUENCY[i]  ,
Date.Format = my.Index.Table$DATE_FORMAT[i],
Release.Month = my.Index.Table$RELEASE_MONTH[i],
column.varnames =  my.Index.Table$COL_VAR_NAME[i] ,
column.values = my.Index.Table$COL_VALUE[i],
column.date = my.Index.Table$COL_DATE[i],
group.name = my.Index.Table$GROUP_NAME[i],
Rows.Skip= my.Index.Table$SKIP[i] ,
Fill.Method = my.Index.Table$FILL_METHOD[i],
Transformation = my.Index.Table$TRANSFORMATION[i],
my.Blocks = eval(parse(text = my.Index.Table$BLOCKS[i])),
Table.Format = my.Index.Table$TABLE_FORMAT[i],
Aggregation.List = my.Index.Table$AGGREG_LIST[i],
Aggregation.Function = my.Index.Table$AGGREG_FUN[i],
Subset.Variable = my.Index.Table$SUBSET_VAR[i],
Subset.Selection = my.Index.Table$SUBSET_SEL[i]
)
    }
   #Populate the lists with the monthly observed TS (without interpolation / with NA)
   #and the monthly Filled TS (with interpolation / without NA) and theier standardized time
series
   my.data.observed[[ my.Index.Table$GROUP_NAME[i] ]] <- Data.ts.Temp$Observations.TS
   my.data.observed.STD[[ my.Index.Table$GROUP_NAME[i] ]] <- Data.ts.Temp$STD.Observations.TS
   my.data.interpolated[[ my.Index.Table$GROUP_NAME[i] ]] <- Data.ts.Temp$Filled.TS
   my.data.interpolated.STD[[ my.Index.Table$GROUP_NAME[i] ]] <- Data.ts.Temp$STD.Filled.TS
   if (i==1) { my.data.Blocks <- eval(parse(text = my.Index.Table$BLOCKS[i])) }
   else      { my.data.Blocks <- rbind(my.data.Blocks,eval(parse(text =
my.Index.Table$BLOCKS[i]))) }
   my.data.frequency[[ my.Index.Table$GROUP_NAME[i] ]] <- Data.ts.Temp$frequency
   print(paste("Added ", i, " Dataset (", my.Index.Table$GROUP_NAME[i] , ")"))
   }
  #Transform the lists into a merged Zoo Object
  #With this process all the TS will have the same Index (same startingpoint, same end,
missing obs on head and tail filled with NA)
  my.data.observed         <- do.call(xts::merge.xts, my.data.observed)
  my.data.interpolated     <- do.call(xts::merge.xts, my.data.interpolated)
  my.data.observed.STD     <- do.call(xts::merge.xts, my.data.observed.STD)
  my.data.interpolated.STD <- do.call(xts::merge.xts, my.data.interpolated.STD)
  ## EXTEND ALL THE PANELS WITH SO MANY "NA" AT THE TAIL AS THE NUMBER OF REQUIRED FORECAST
MONTHS
  forecast.start <- end(my.data.observed)+1/12
  #crete a XTS object populated with NA with the same number of columns as the panels and
number of rows as the fcst months
  forecast.period.matrix <- matrix(data = NA,nrow = nForecastMonths, ncol =
ncol(my.data.observed) )
  forecast.period.ts     <- ts(forecast.period.matrix)
  forecast.period.zoo    <- zoo::zooreg(data =  forecast.period.matrix , start =
forecast.start, frequency = 12)
  forecast.period.xts    <- xts::xts(forecast.period.zoo)
  # append the xts to each panel
  my.data.observed         <- rbind(my.data.observed,forecast.period.xts)
  my.data.interpolated     <- rbind(my.data.interpolated,forecast.period.xts)
  my.data.observed.STD     <- rbind(my.data.observed.STD,forecast.period.xts)
  my.data.interpolated.STD <- rbind(my.data.interpolated.STD,forecast.period.xts)
  my.data.frequency        <- as.vector(unlist(my.data.frequency))
  detach("package:dplyr", unload = TRUE)
  return(list(Observed         = my.data.observed ,
              Observed.STD     = my.data.observed.STD ,
              Interpolated     = my.data.interpolated,
              Interpolated.STD = my.data.interpolated.STD,
              Transformation   = my.data.transformation,
              Frequency        = my.data.frequency,
              Blocks           = my.data.Blocks  ))
}
##########################################################################
########################     CSV.Data.Loader      ########################
##########################################################################
CSV.Data.Loader <- function(   File.Address, ExcludedColumns= c(1,2) , Transformation,
                               my.frequency , Date.Format, Release.Month = 3,
                               column.values = "Value" , column.varnames = "D1" , column.date
= "Date",

                               group.name = NA,
                               Rows.Skip = 3 , Col.Separator = ";",
                               SNB_D0 , SNB_D1, Fill.Method, my.Blocks ) {

  #Read the CSV File
  my.read.table <- read.table( File.Address ,sep = Col.Separator , header = TRUE , skip =
Rows.Skip )
  # Filter for SNB Data: The single Time Serie are Extracted from the downloaded general Panel
  if (!is.na(SNB_D0) &&  is.na(SNB_D1)) { my.read.table <- ( my.read.table %>% subset(D0 ==
SNB_D0) )  }
  if (!is.na(SNB_D0) && !is.na(SNB_D1)) { my.read.table <- ( my.read.table %>% subset(D0 ==
SNB_D0) %>% subset(D1 == SNB_D1)   )  }
  if ( is.na(SNB_D0) && !is.na(SNB_D1)) { my.read.table <- ( my.read.table %>% subset(D1 ==
SNB_D1))  }
  #Transform the long panel into wide panel (key is the column where the var names are listed)
  my.data.frame <- tidyr::spread(my.read.table,  key= column.varnames ,value = column.values)
  # Date.Format <- "%Y-Q%q"
  # Create the Time Index for the XTS Time Serie
  # The Time Index is always in Month
  if (my.frequency==1) {
    #set the column Date from class "Character" to "Date" in format "Yearmon" (monthly)
    #the operation ( + (Release.Month-1)/12 ) allow to shift the time serie and assign it to
the real publication month
    my.data.month <- zoo::as.yearmon(my.data.frame[,column.date] , format = Date.Format) +
(Release.Month-1)/12
  }
  if (my.frequency==4) {
    # extract the Date Index by month. The order must be:
    #first read the quarter (as.yearqtr - format specification)
    #than  get the Date in Month(yearmon))
    # the operation ( + Release.Month%%3 ) allow to shift the time serie and assign it to the
real publication month
    my.data.month <- zoo::as.yearmon ( zoo::as.yearqtr(my.data.frame[,column.date] , format =
Date.Format) ) + (Release.Month-1)%%3/12
  }
  if (my.frequency==12) {
    #set the column Date from class "Character" to "Date" in format "Yearmon" (monthly)
    my.data.month <- zoo::as.yearmon(my.data.frame[,column.date] , format = Date.Format)
  }
  #Create a XTS object
  #my.data.xts <- xts::xts(x = my.data.frame[,-ExcludedColumns], order.by = my.data.month)
  my.data.xts <- xts::xts(x = my.data.frame[,-ExcludedColumns], order.by = my.data.month)
  # CREATE A TS WITH THE FULL 12 months series
  # create a vector with all the missing months (yearmon format)
  my.empty.12m<- seq( start(my.data.xts), end(my.data.xts), 1/12 )
  # create an additional my.data.xts.12m merging my.data.xts and my.empty.12m and fill it with
the selected Fill.Method (na.locf / na.approx / na.spline)
  if (Fill.Method == "na.approx"){
    my.data.xts     <- merge(my.data.xts, my.empty.12m)
    my.data.xts.12m <- merge(my.data.xts, my.empty.12m, fill = zoo::na.approx)
  } else if ( Fill.Method == "na.locf"){
    my.data.xts     <- merge(my.data.xts, my.empty.12m)
    my.data.xts.12m <- merge(my.data.xts, my.empty.12m, fill = zoo::na.na.locf)
  } else if ( Fill.Method == "na.spline"){
    my.data.xts     <- merge(my.data.xts, my.empty.12m)
    my.data.xts.12m <- merge(my.data.xts, my.empty.12m, fill = zoo::na.spline)
  } else {
    my.data.xts.12m <- merge(my.data.xts, my.empty.12m)
  }
  #Check the number of the columns (my.data.xts[2])
  #IF there is just one column assign the name to the only colum
  if (dim(my.data.xts)[2]==1) {
    names(my.data.xts)     <- group.name
    names(my.data.xts.12m) <- group.name
  }
  #ELSE IF there are many column (already labeled) assign them a prefix
  else {
    colnames(my.data.xts)     <- paste(group.name , colnames(my.data.xts), sep = "_")
    colnames(my.data.xts.12m) <- paste(group.name , colnames(my.data.xts), sep = "_")
  }
  #Standardize the Time Series
  std.my.data.xts      <- scale(my.data.xts)
  std.my.data.xts.12m  <- scale(my.data.xts.12m)
  #create a vector long as the number of variables, populated with the transformation value
  ts.Transformation <- rep(Transformation,  ncol(my.data.xts))
  return(list(Observations.TS = my.data.xts ,
```

`

```
                Filled.TS = my.data.xts.12m,
                STD.Observations.TS = std.my.data.xts,
                STD.Filled.TS =  std.my.data.xts.12m,
                Transformation = ts.Transformation,
                frequency = my.frequency,
                Blocks = my.Blocks))
}
##########################################################################
########################    XLSX.Data.Loader     #########################
##########################################################################
XLSX.Data.Loader <- function(  File.Address, Transformation,
                               my.frequency , Date.Format, Release.Month = 3,
                               column.values = "Value" , column.varnames, column.date =
"Date",
                               group.name = NA,
                               Rows.Skip = 5 ,
                               Fill.Method, my.Blocks,
                               Table.Format = "WIDE",
                               Aggregation.List = NA,
                               Aggregation.Function = NA,
                               Subset.Variable = NA,
                               Subset.Selection = NA) {
if (Table.Format == "WIDE") {
    #Read the XLSX File
    my.read.table <- openxlsx::read.xlsx( xlsxFile = File.Address, sheet = 1,
startRow=Rows.Skip , detectDates = TRUE)
    my.read.table.subset <- subset(x = my.read.table, select = column.varnames )
    colnames(my.read.table.subset) <- group.name
  }
  if (Table.Format == "LONG") {
    my.read.table <- openxlsx::read.xlsx( xlsxFile = File.Address, sheet = 1,
startRow=Rows.Skip , detectDates = TRUE)
    # AGGREGATE THE DATA
    my.read.table   <-      aggregate( as.formula(paste(column.values ,
paste(Aggregation.List), sep="~"))  ,
                                      data=my.read.table   ,
                                      FUN = eval(parse(text = Aggregation.Function))  ,
                                      na.rm=TRUE )
    #The above line does the same as described here below:
    #           column.values  Aggregation.List                    Aggregation.Function
    #  aggregate(TRANS_PRICE   ~  Date+Kanton,   data=my.read.table,     FUN = mean )
    # SUBSET THE AGGREGATED DATA
    if (!is.na(Subset.Variable)){
      #Transform the strings Subset.Variable and Subset.Selection into vectors
      Subset.Variable <- unlist(strsplit(Subset.Variable, ","))
      Subset.Selection <- unlist(strsplit(Subset.Selection, ","))
      #Verify if all the number of subset variable are equal to the subset selection
      Num.subset.Variable <- length(Subset.Variable)
      Num.subset.Selection <- length(Subset.Selection)
      if(Num.subset.Variable!=Num.subset.Selection) {stop("The number of elements in the
SUBSET_VAR and SUBSET_SEL column in the index Excele file must be equal!") }
      #Loop the subset for each Subset.Variable
      for (k in 1:Num.subset.Variable) {
        my.read.table <- subset(my.read.table , subset = eval(parse(text =
Subset.Variable[k])) == Subset.Selection[k] )
      }
    }
    #if (!is.na(Subset.Variable)){
    #  my.read.table <- subset(my.read.table , subset = eval(parse(text = Subset.Variable)) ==
Subset.Selection  )
    #}
    my.read.table.subset <- subset(x = my.read.table, select = column.values ) # select just
the column of interest after the aggregation
    colnames(my.read.table.subset) <- group.name
  }
  # Create the Time Index for the XTS Time Serie
  # The Time Index is always in Month
  if (my.frequency==1) {
    #set the column Date from class "Character" to "Date" in format "Yearmon" (monthly)
    #the operation ( + (Release.Month-1)/12 ) allow to shift the time serie and assign it to
the real publication month
    my.data.month <- zoo::as.yearmon(my.read.table[,column.date] , format = Date.Format) +
(Release.Month-1)/12
  }
  if (my.frequency==4) {
    # extract the Date Index by month. The order must be:
    #first read the quarter (as.yearqtr - format specification)
    #than  get the Date in Month(yearmon))
    # the operation ( + Release.Month%%3 ) allow to shift the time serie and assign it to the
real publication month
    my.data.month <- zoo::as.yearmon ( zoo::as.yearqtr(my.read.table[,column.date] , format =
Date.Format) ) + (Release.Month-1)%%3/12
  }
  if (my.frequency==12) {
    #set the column Date from class "Character" to "Date" in format "Yearmon" (monthly)
    my.data.month <- zoo::as.yearmon(my.read.table[,column.date] , format = Date.Format)
  }
  if (my.frequency== 360 ) {
    #aggregate the data monthly as average of each month
    #set the column Date from class "Character" to "Date" in format "Yearmon" (monthly)
    my.data.month <- zoo::as.yearmon(my.read.table[,column.date] , format = Date.Format)
  }
  #Create a XTS object
  #my.data.xts <- xts::xts(x = my.data.frame[,-ExcludedColumns], order.by = my.data.month)
  my.data.xts <- xts::xts(x = my.read.table.subset, order.by = my.data.month)
  #Aggregate at monthly level the daily observations ( monthly mean ) and reassign the
frequency value to 12
  if (my.frequency== 360 ) {
    my.data.xts <- xts::apply.monthly(my.data.xts , FUN = mean)
    my.frequency <- 12
  }
  # CREATE A TS WITH THE FULL 12 months series
  # create a vector with all the missing months (yearmon format)
  my.empty.12m<- seq( start(my.data.xts), end(my.data.xts), 1/12 )
  # create an additional my.data.xts.12m merging my.data.xts and my.empty.12m and fill it with
the selected Fill.Method (na.locf / na.approx / na.spline)
  if (Fill.Method == "na.approx"){
    my.data.xts     <- merge(my.data.xts, my.empty.12m)
    my.data.xts.12m <- merge(my.data.xts, my.empty.12m, fill = zoo::na.approx)
  } else if ( Fill.Method == "na.locf"){
    my.data.xts     <- merge(my.data.xts, my.empty.12m)
    my.data.xts.12m <- merge(my.data.xts, my.empty.12m, fill = zoo::na.na.locf)
  } else if ( Fill.Method == "na.spline"){
    my.data.xts     <- merge(my.data.xts, my.empty.12m)
    my.data.xts.12m <- merge(my.data.xts, my.empty.12m, fill = zoo::na.spline)
  } else {
    my.data.xts.12m <- merge(my.data.xts, my.empty.12m)
  }
#Standardize the Time Series
  std.my.data.xts       <- scale(my.data.xts)
  std.my.data.xts.12m  <- scale(my.data.xts.12m)
  #create a vector long as the number of variables, populated with the trasnformation value
  ts.Transformation <- rep(Transformation,  ncol(my.data.xts))
  return(list(Observations.TS = my.data.xts ,
              Filled.TS = my.data.xts.12m,
              STD.Observations.TS = std.my.data.xts,
              STD.Filled.TS =  std.my.data.xts.12m,
              Transformation = ts.Transformation,
              frequency = my.frequency,
              Blocks = my.Blocks))
}

##########################################################################
#######################       DATAPANEL.PDF       ########################
##########################################################################
datapanel.pdf.export <- function (my.data,
                                  path = "/Users/Fabio/Library/Mobile
Documents/com~apple~CloudDocs/CUREM/00_MasterArbeit/Daten/MY_DATA_PANEL/",
                                  file_name =  "_DataPanelGraphics.pdf" )   {
  file_name <- paste0(  format(Sys.Date(),"%y") , format(Sys.Date(),"%m") ,
format(Sys.Date(),"%d") , file_name)
  pdf (file = paste0(path, file_name ))
  for (i in 1: ncol(my.data$Observed)) {
  TS.frequency <- frequency(na.omit(my.data$Observed.STD[,i]))
  par(mfcol=c(4,1))
  print( plot(na.omit(my.data$Observed[,i]), type = "s", main = paste0("Observed " ,
names(my.data$Observed[,i]), "   |   Freq = ", TS.frequency ) ) )
  print( plot(na.omit(my.data$Observed.STD[,i]), type = "s", main = paste0("Observed STD " ,
names(my.data$Observed[,i]) , "   |   Freq = ", TS.frequency ) ) )
  print( plot(na.omit(my.data$Interpolated[,i]), main = paste0("Interpolated " ,
names(my.data$Interpolated[,i]) , "   |   Freq = 12" )) )
  print( plot(na.omit(my.data$Interpolated.STD[,i]), main = paste0("Interpolated STD " ,
names(my.data$Interpolated[,i]) , "   |   Freq = 12" )) )
  }
  dev.off()
}
year2qtr <- function(DataObserved , DataInterpolated, Frequency ){
  library(zoo)
  x <- DataObserved
  z <- DataInterpolated
  TT<- ncol(x)
  Panel.Y2Q.Interpolated <- x
  New.Frequency <- Frequency
    for (k in 1:TT) {
        if (Frequency[k] == 1) {
              Newest.Value.Position <-
Newest_Oldest_Value.position(x[,k])$Newest.Value.Position
              Oldest.Value.Position <-
Newest_Oldest_Value.position(x[,k])$Oldest.Value.Position
              TimeSerie.Lenght <- length(x[,k])
              Panel.Y2Q.Interpolated[1:(Oldest.Value.Position-1),k] <- NA              # Fill
with NA the space before the Oldest Observation
              for (i in Oldest.Value.Position:Newest.Value.Position) {
                MonthCheck <- lubridate::month(index(x[,k])[i])
                if ( MonthCheck%%3 == 0) {
                  Panel.Y2Q.Interpolated[i,k] <- z[i,k]                                #
Catch the Interpolated Value of the months 3,6,9,12
                } else {
                  Panel.Y2Q.Interpolated[i,k] <- NA                                    # Fill
the other months with NA
                }
              }
              Panel.Y2Q.Interpolated[(Newest.Value.Position+1):TimeSerie.Lenght,k] <- NA #
Fill with NA the space after the Last Observation
              New.Frequency[k] <- 4
              print(paste0("The yearly time serie",k , " has been interpolated into
quarterly"))
        }
    }
  #output
  return(list(ObsMQ_InterpY2Q=Panel.Y2Q.Interpolated, New.Frequency=New.Frequency))
}
##########################################################################
#################     PLOT FUNCTIONS              ##################
##########################################################################
PLOT.Stationarity.Tests <- function(x) {
  #     The function works with a layout with
  #     par(mar=c(0.1,0.5,0.5,0.1)) on a Quadrant with heights proportion 0.2 (Ex: c(1, 0.2,
0.15 , 1, 0.2 , 1), # Heights of the two rows) )
   #     Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski, Phillips, Schmidt, &
Shin, 1992)
   #     https://otexts.com/fpp2/stationarity.html#fn14
   #     p-value < 0.05 indicates the TS is stationary
   #     The more negative the DF test statistic, the stronger the evidence for rejecting the
null hypothesis of a unit root.
   #     The more close to zero are the test statistic of the KPSS, the stronger the evidence
of stationarity
  adf.test  <- tseries::adf.test(na.omit(x))
  kpss.test <- tseries::kpss.test(na.omit(x))
  text(0.0,0.7, adj = 0, paste0("Augmented Dickey-Fuller Test -> pValue = ",
round(adf.test$p.value,4)) ,cex=1)
  text(0.1,0.5, adj = 0, paste0("                          Test Statistic = ",
round(adf.test$statistic,4)) ,cex=1)
  text(0.53,0.7, adj = 0, paste0("KPSS Test -> Test Statistic Value = ",
round(kpss.test$statistic,4)) ,cex=1)
}
PLOT.ACF.PACF <- function(x, title){
  par(mar = c(2, 0.5, 1, 0.5)) # Set the margin on all sides
  acf( na.omit(x)  , main = paste0("ACF ",title)  , cex = 0.25)
  pacf( na.omit(x) , main = "",cex = 0.25)
  title(paste0("PACF ",title) , cex = 0.25)
}
##########################################################################
#################      COUNT VARIABLE PER YEAR       #################
##########################################################################
Count_Var_Year <- function(x){
  #get the frequency from the orginal index data
  my.Index.Table <- readxl::read_xlsx(Index.Path, skip = 1)
  #select just the elements marked with "YES" on the column "INCLUDE"
  my.Index.Table <- subset(my.Index.Table, my.Index.Table$INCLUDE == "YES")
  frequency      <- my.Index.Table$FREQUENCY
  #if (!(frequency(x)==12)) {paste0("Error: The frequency of the time series must be 12!")}
  start.year <- lubridate::year(x[1,1])
  end.year <- lubridate::year(x[nrow(x),1])
  num.year <- end.year-start.year+1
  end.row <- nrow(x)
  Var_count_Y <- xts(x = rep(0,num.year), order.by = as.yearmon(seq(start.year,end.year)))  #
prepare the count variable as a xts object of populated with 0
  Var_count_Q <- Var_count_Y
  Var_count_M <- Var_count_Y
  Var_count_D <- Var_count_Y
  y <- 0                             # support variable
  for (j in seq(97,  to = (end.row-12), by = 12)){  # from 1 to end increasing of 12 months
per loop
    y = y+1
        for (k in 1:ncol(x)){                   # loop for each variable
          # if any observation betwee the row t and t+11 is a value return y and increase
the Var_count of 1
          if(any(!is.na(x[j:(j+11),k]))==TRUE) {
            if (frequency[k] == 1)   { Var_count_Y[y] <- Var_count_Y[y]+1 }
            if (frequency[k] == 4)   { Var_count_Q[y] <- Var_count_Q[y]+1 }
            if (frequency[k] == 12)  { Var_count_M[y] <- Var_count_M[y]+1 }
            if (frequency[k] == 360) { Var_count_D[y] <- Var_count_D[y]+1 }
          }
        }
  }
  Var_count <- merge.xts(Var_count_Y,Var_count_Q,Var_count_M,Var_count_D)
  return(Var_count)
}
##########################################################################
##################      OUTLIERS CORRECTION CALL     ####################
##########################################################################
Call.OutliersCorrection <- function(x, NA.replace=T, k.ma=3){
  # length(x)
  Newest.Value.Position <- Newest_Oldest_Value.position(x)$Newest.Value.Position
  Oldest.Value.Position <- Newest_Oldest_Value.position(x)$Oldest.Value.Position
  x.df <- as.data.frame(x)
  x.Outliers_corrected <- c( rep(NA, Oldest.Value.Position-1) ,

outliers_correction(x.df[(Oldest.Value.Position):Newest.Value.Position,1],k.ma, NA.replace) ,
                             rep(NA, nrow(x) - Newest.Value.Position) )
  # the step abobe recreates the lenght of the differenced td binding one NA, the outliner
correction(calculated just whereValues are Present), a NA serie at the Tail
  # (in this way the position of NA is the same as the original Timeseries and it is possible
to match with the Index(x) ).
  # recreate the XTS and match the dates
  # length(x.Outliers_corrected)
  x.Outliers_corrected <- xts::xts(x = x.Outliers_corrected,order.by = zoo::index(x) )
  #fill the missing values in the Head with the first observation available
  if (NA.replace == TRUE) {
    x.Outliers_corrected[1:Oldest.Value.Position] <- na.locf(x[1:Oldest.Value.Position],
fromLast = TRUE)}
  names(x.Outliers_corrected) <- names(x)
  return(x.Outliers_corrected)
  }
##########################################################################
#################     STATIONARITY CHECK              ##################
##########################################################################
# my.trans.was=1  trans = 0: the original serie is preserved,
https://www.symbolab.com/solver/ (copy the form below in this web page)
# my.trans.was=2  trans = 1: monthly rate of change
\frac{x_{i,t} - x_{i,t-1}}{x_{i,t-1}}
# my.trans.was=3  trans = 2: monthly difference       x_{i,t} - x_{i,t-1}
# my.trans.was=5  trans = 3: monthly difference in year-over-year rate of change
\frac{x_{i,t} - x_{i,t-12}}{x_{i,t-12}} - \frac{x_{i,t-1} - x_{i,t-13}}{x_{i,t-13}}
# my.trans.was=6  trans = 4: monthly difference in year difference (x_{i,t} - x_{i,t-12}) -
(x_{i,t-1} - x_{i,t-13})
# my.trans.was=7  trans = 5: yearly difference      (x_{i,t} - x_{i,t-12})
# my.trans.was=   trans = 6: yearly rate of change  \frac{x_{i,t} - x_{i,t-12}}{x_{i,t-12}}
# my.trans.was=4  trans = 7: Double Difference   (never used!)

Stationarity.Check <- function (          #my.data = my.data,
                                          path = "/Users/Fabio/Library/Mobile
Documents/com~apple~CloudDocs/CUREM/00_MasterArbeit/Daten/MY_DATA_PANEL/Stationararity_Transfo
rmations_Check/",
                                          file_name =  "_DataPanel_Decomposition.pdf",
                                          from , to,
                                          Transformation)   {
  library(xts)
  file_name <- paste0(file_name,".pdf")
  #pdf (file = paste0(path, file_name ), width = 11.69, height = 8.27 )
  pdf (file = paste0(path, file_name ), width = 11.69, height = 13.31 )
  for (i in from:to) {
    print(paste0("Variable n.",i))
    TimeSerie     <- my.data$Observed[,i]
    TS.frequency <- frequency(na.omit(TimeSerie))
    #colnames(TimeSerie.Stationarized) <- colnames(my.data$Observed)
    par(mar=c(0.05,0.5,0.05,0.5))
    layout.matrix <- matrix(c( 1 , 1 , 1 , 1,
                               2 , 2 , 2 , 2,
                               3 , 5 , 7 , 9,
                               3 , 6 , 7 ,10,
                               4 , 4 , 8 , 8,
                               11, 13, 15,17,
```

```
                         11, 14, 15,18,
                         12, 12, 16,16,
                         19, 21, 23,25,
                         19, 22, 23,26,
                         20, 20, 24,24,
                         27, 29, 31,33,
                         27, 30, 31,34,
                         28, 28, 32,32
                         ) ,
                       nrow = 14, byrow = TRUE)
    layout( mat = layout.matrix,
            heights = c(1.5, 0.3, 0.5 , 0.5, 0.4 , 0.5, 0.5, 0.4,0.5, 0.5, 0.4,0.5, 0.5, 0.4),
# Heights of the two rows
            widths = c(0.8, 0.5, 0.8, 0.5)) # Widths of the two columns
# QUADRANT 1 - Print Original TS
         print( plot(na.omit(TimeSerie), type = "s", main = paste0("(",i,") Observed " ,
names(my.data$Observed[,i]) , "   |   Freq = ", TS.frequency ) , cex = 0.5) )
    # QUADRANT 2 - Calculate and Print the number of appropriate first differences if Freq > 1
         plot.new()
         numb.first.diff <- forecast::ndiffs(na.omit(TimeSerie))
         if(TS.frequency==1) {
           text(0.0,0.5, adj = 0, paste0("With  annual data, there is no seasonal component,
therefore seasonal decomposition \nof time series would not be appropriate. (ndiffs test=
",numb.first.diff," )" ) ,cex=1 )
         } else {
           #Calculate and Print the number of appropriate first differences
           text(0.0,0.7, adj = 0, paste0("Estimation of differences number (ndiffs test= ",
numb.first.diff ," )" ) ,cex=1)
         }
    if (0 %in% Transformation) {
      ###############################
      ###    TRANSFORMATION  0   ###
      ###  STATIONARITY AT LEVEL ###
      ###############################
      # With  annual data, there is no seasonal component, therefore seasonal decomposition of
time series would not be appropriate
      # However, the stats::stl function calls the loess function to estimate trend, which is
a local polynomial regression
      # and  it's possible to  call loess directly and estimate the trend directly.
      TimeSerie.Outliers_corrected <- Call.OutliersCorrection(na.omit(TimeSerie))
      print( plot(na.omit(TimeSerie.Outliers_corrected), type = "s", main =  "Trans 0:
Original Series with Corrected Outliers" , cex = 0.35, yaxis.right = FALSE) )
      plot.new()
      PLOT.Stationarity.Tests (TimeSerie.Outliers_corrected)
      PLOT.ACF.PACF(x = TimeSerie.Outliers_corrected , title = "Diff|Outliers Corr. Serie")
    }
      if (1 %in% Transformation) {
      ########################################
      ########     TRANSFORMATION 1    ########
      ########     Single LAG (YQM)    ########
      ########       RATE OF CHANGE    ########
      ########################################
      #  \frac{x_{i,t} - x_{i,t-1}}{x_{i,t-1}}
      if (TS.frequency==12){TimeSerie.Stationarized   <- diff(TimeSerie, 1)
/xts::lag.xts(TimeSerie,1)  } #ok
      if (TS.frequency==4 ){TimeSerie.Stationarized   <- diff(TimeSerie, 3)
/xts::lag.xts(TimeSerie,3)  } #ok
      if (TS.frequency==1 ){TimeSerie.Stationarized   <- diff(TimeSerie,
12)/xts::lag.xts(TimeSerie,12) } #ok
      TimeSerie.Stationarized <- na.omit(TimeSerie.Stationarized) # Delete NA on the head for
printing
      TimeSerie.Outliers_corrected <- Call.OutliersCorrection(TimeSerie.Stationarized)
      print( plot(TimeSerie.Outliers_corrected, type = "s", main =  "Trans 1: Single lag (YQM)
Rate of Change" , cex = 0.35, yaxis.right = FALSE) )
      plot.new()
      PLOT.Stationarity.Tests (TimeSerie.Outliers_corrected)
      PLOT.ACF.PACF(x = TimeSerie.Outliers_corrected , title = "")
    }
      if (2 %in% Transformation) {
      ########################################
      ########     TRANSFORMATION 2    ########
      ########     Single lag (YQM)    ########
      ########         DIFFERENCE      ########
      ########################################
      #            x_{i,t} - x_{i,t-1}
      if (TS.frequency==12){TimeSerie.Stationarized   <-    diff(TimeSerie,1)  } #ok
      if (TS.frequency==4) {TimeSerie.Stationarized   <-    diff(TimeSerie,3)  } #ok
      if (TS.frequency==1) {TimeSerie.Stationarized   <-    diff(TimeSerie,12) } #ok
      TimeSerie.Stationarized <- na.omit(TimeSerie.Stationarized) # Delete NA on the head for
printing
      TimeSerie.Outliers_corrected <- Call.OutliersCorrection(TimeSerie.Stationarized)
      print( plot(TimeSerie.Outliers_corrected, type = "s", main =  "Trans 2: Single lag (YQM)
Difference" , cex = 0.35, yaxis.right = FALSE) )
      plot.new()
      PLOT.Stationarity.Tests (TimeSerie.Outliers_corrected)
      PLOT.ACF.PACF(x = TimeSerie.Outliers_corrected , title = "")
    }
      if (3 %in% Transformation) {
        ##########################################
        ########       TRANSFORMATION  3  ########
        ###### SINGLE LAG (M/Q) DIFFERENCE ########
        #######        YEAR OVER YEAR     ########
        ########       RATE OF CHANGE     ########
        ##########################################
        # \frac{x_{i,t} - x_{i,t-4}}{x_{i,t-4}} - \frac{x_{i,t-1} - x_{i,t-5}}{x_{i,t-5}
        # \frac{x_{i,t} - x_{i,t-12}}{x_{i,t-12}} - \frac{x_{i,t-1} - x_{i,t-13}}{x_{i,t-13}
        if (TS.frequency==12){TimeSerie.Stationarized  <-
diff(TimeSerie,12)/xts::lag.xts(TimeSerie,12)  -
diff(xts::lag.xts(TimeSerie,1),12)/xts::lag.xts(TimeSerie,13)} #ok
        if (TS.frequency==4) {TimeSerie.Stationarized  <-
diff(TimeSerie,12)/xts::lag.xts(TimeSerie,12)  -
diff(xts::lag.xts(TimeSerie,3),12)/xts::lag.xts(TimeSerie,15)} #ok
        TimeSerie.Stationarized <- na.omit(TimeSerie.Stationarized) # Delete NA on the head
for printing
        TimeSerie.Outliers_corrected <- Call.OutliersCorrection(TimeSerie.Stationarized)
        print( plot(TimeSerie.Outliers_corrected, type = "s", main =  "Trans 3: M/Q Diff. Y
over Y Rate of Change" , cex = 0.35 , yaxis.right = FALSE) )
        plot.new()
        PLOT.Stationarity.Tests (TimeSerie.Outliers_corrected)
        PLOT.ACF.PACF(x = TimeSerie.Outliers_corrected , title = "")
      }
      if (4 %in% Transformation) {
        ##########################################
        ########       TRANSFORMATION 4   ########
        ###### SINGLE LAG (M/Q) DIFFERENCE ########
        #######      IN YEAR DIFFERENCE   ########
        ##########################################
        #(x_{i,t} - x_{i,t-4}) - (x_{i,t-1} - x_{i,t-5})
        #(x_{i,t} - x_{i,t-12}) - (x_{i,t-1} - x_{i,t-13})
        if (TS.frequency==12){TimeSerie.Stationarized   <-    diff( diff(TimeSerie,12) ) } #ok
        if (TS.frequency==4) {TimeSerie.Stationarized   <-    diff(TimeSerie,12) -
diff((xts::lag.xts(TimeSerie,3)),12)} #ok
        TimeSerie.Stationarized <- na.omit(TimeSerie.Stationarized) # Delete NA on the head
for printing
        TimeSerie.Outliers_corrected <- Call.OutliersCorrection(TimeSerie.Stationarized)
        print( plot(TimeSerie.Outliers_corrected, type = "s", main =  "Trans 4: Single lag(QM)
Difference in Year Difference" , cex = 0.35, yaxis.right = FALSE) )
        plot.new()
        PLOT.Stationarity.Tests (TimeSerie.Outliers_corrected)
        PLOT.ACF.PACF(x = TimeSerie.Outliers_corrected , title = "")
      }
if (5 %in% Transformation) {
          ##########################################
          ########      TRANSFORMATION 5    ########
          #######        YEAR DIFFERENCE    ########
          ##########################################
          #(x_{i,t} - x_{i,t-4})
          #(x_{i,t} - x_{i,t-12})
          TimeSerie.Stationarized   <-    diff(TimeSerie, 12) #ok
          TimeSerie.Stationarized <- na.omit(TimeSerie.Stationarized) # Delete NA on the head
for printing
          TimeSerie.Outliers_corrected <- Call.OutliersCorrection(TimeSerie.Stationarized)
          print( plot(TimeSerie.Outliers_corrected, type = "s", main =  "Trans 5: Q/M Year
Difference" , cex = 0.35, yaxis.right = FALSE) )
          plot.new()
          PLOT.Stationarity.Tests (TimeSerie.Outliers_corrected)
          PLOT.ACF.PACF(x = TimeSerie.Outliers_corrected , title = "")
}
          if (6 %in% Transformation) {
            ##########################################
            ########      TRANSFORMATION  6  ########
            ######              YEARLY        ########
            ########       RATE OF CHANGE     ########
            ##########################################
            # \frac{x_{i,t} - x_{i,t-4}}{x_{i,t-4}}
            # \frac{x_{i,t} - x_{i,t-12}}{x_{i,t-12}}
            TimeSerie.Stationarized  <-     diff(TimeSerie,12)/xts::lag.xts(TimeSerie,12)
            TimeSerie.Stationarized <- na.omit(TimeSerie.Stationarized) # Delete NA on the
head for printing
            TimeSerie.Outliers_corrected <- Call.OutliersCorrection(TimeSerie.Stationarized)
            print( plot(TimeSerie.Outliers_corrected, type = "s", main =  "Trans 6: Yearly
Rate of Change" , cex = 0.35 , yaxis.right = FALSE) )
            plot.new()
            PLOT.Stationarity.Tests (TimeSerie.Outliers_corrected)
            PLOT.ACF.PACF(x = TimeSerie.Outliers_corrected , title = "")
          }
          if (7 %in% Transformation) {
            ##########################################
            ########     TRANSFORMATION 7     ########
            #####   SINGLE LAG (M/Q) DOUBLE    #######
            #######        DIFFERENCES        ########
            ##########################################
            if (TS.frequency==12){TimeSerie.Stationarized <- diff(TimeSerie, differences =
2)}#ok
            if (TS.frequency==4){TimeSerie.Stationarized  <- diff(TimeSerie, lag = 3,
differences = 2)}#ok
            if (TS.frequency==1){TimeSerie.Stationarized  <- diff(TimeSerie, lag = 12,
differences = 2)}#ok
            TimeSerie.Stationarized <- na.omit(TimeSerie.Stationarized) # Delete NA on the
head for printing
            TimeSerie.Outliers_corrected <- Call.OutliersCorrection(TimeSerie.Stationarized)
            print( plot(TimeSerie.Outliers_corrected, type = "s", main =  "Trans 7: Single lag
(YQM) Double Difference" , cex = 0.35, yaxis.right = FALSE) )
            plot.new()
            PLOT.Stationarity.Tests (TimeSerie.Outliers_corrected)
            PLOT.ACF.PACF(x = TimeSerie.Outliers_corrected , title = "")
          }
}
    dev.off()
}
##########################################################################
##################        STATIONARIZATION            ##################
##########################################################################
stationarized.bpanel <- function(base = NULL, trans = NULL, Frequency = NULL, Blocks = "" ,
NA.replace = T, k.ma = 3, na.prop = 1/3, h = 12) {
  library(xts)
  library(matlab)
  if (is.null(trans)) {
    stop("trans can not to be NULL")
  }
  if (sum(is.na(trans)) != 0) {
    stop("trans does not support missings values")
  }
  if (length(trans) != ncol(base)) {
    stop("the number of elements in the vector must be equal to the number of columns of
base")
  }
  if (sum( !names(table(trans)) %in% c(0:6)) != 0) {
    stop("the only available transformations are 0, 1, 2, 3, 4, 5 and 6.")
  }
  if (na.prop < 0 | na.prop > 1) {
    stop("na.prop must be between 0 and 1.")
  }
  TimeSerie.Stationarized  <- base
  TimeSerie.Stationarized[,]  <- NA
  colnames(TimeSerie.Stationarized) <- colnames(base)
    for (i in 1: ncol(base)) {
        TimeSerie <- base[,i]
        TS.frequency <-  Frequency[i]
        #TS.frequency <- frequency(na.omit(base[,i]))
        #length(TimeSerie.Stationarized[,i])
        # trans = 0: the original serie is preserved,
        if (trans[i] == 0) {
          TimeSerie.Stationarized[,i] <- base[,i]
        }
        # trans = 1: single lag rate of change                          \frac{x_{i,t} -
x_{i,t-1}}{x_{i,t-1}}
        if (trans[i] == 1) {
          if (TS.frequency==12){TimeSerie.Stationarized[,i]   <- diff(TimeSerie, 1)
/xts::lag.xts(TimeSerie,1)  } #ok
          if (TS.frequency==4 ){TimeSerie.Stationarized[,i]   <- diff(TimeSerie, 3)
/xts::lag.xts(TimeSerie,3)  } #ok
          if (TS.frequency==1 ){TimeSerie.Stationarized[,i]   <- diff(TimeSerie,
12)/xts::lag.xts(TimeSerie,12) } #ok
        }
        # trans = 2: single lag difference                              x_{i,t} - x_{i,t-1}
        if (trans[i] == 2) {
          if (TS.frequency==12){TimeSerie.Stationarized[,i]   <-    diff(TimeSerie,1)  }
          if (TS.frequency==4) {TimeSerie.Stationarized[,i]   <-    diff(TimeSerie,3)  }
          if (TS.frequency==1) {TimeSerie.Stationarized[,i]   <-    diff(TimeSerie,12) }
        }
        #trans = 3: single lag difference in y-over-y rate of change   \frac{x_{i,t} - x_{i,t-
12}}{x_{i,t-12}} - \frac{x_{i,t-1} - x_{i,t-13}}{x_{i,t-13}
        if (trans[i] == 3) {
                                                                 #diff( diff(TimeSerie,12)
)/xts::lag.xts(TimeSerie,12)
          if (TS.frequency==12){TimeSerie.Stationarized[,i]  <-
diff(TimeSerie,12)/xts::lag.xts(TimeSerie,12)  -
diff(xts::lag.xts(TimeSerie,1),12)/xts::lag.xts(TimeSerie,13)} #ok
          if (TS.frequency==4) {TimeSerie.Stationarized[,i]  <-
diff(TimeSerie,12)/xts::lag.xts(TimeSerie,12)  -
diff(xts::lag.xts(TimeSerie,3),12)/xts::lag.xts(TimeSerie,15)} #ok
        }
        #trans = 4: sinlge lag difference in year difference          (x_{i,t} - x_{i,t-12})
- (x_{i,t-1} - x_{i,t-13})
        if (trans[i] == 4) {
          if (TS.frequency==12){TimeSerie.Stationarized[,i]   <-    diff( diff(TimeSerie,12) )
} #ok
          if (TS.frequency==4) {TimeSerie.Stationarized[,i]   <-    diff(TimeSerie,12) -
diff((xts::lag.xts(TimeSerie,3)),12)} #ok
        }
        #trans = 5: yearly difference                               (x_{i,t} - x_{i,t-12})
        if (trans[i] == 5) {
          TimeSerie.Stationarized[,i]   <-    diff(TimeSerie, 12) #ok
        }
        #trans = 6: yearly rate of change                           \frac{x_{i,t} - x_{i,t-
12}}{x_{i,t-12}}
        if (trans[i] == 6) {
          TimeSerie.Stationarized[,i]  <-    diff(TimeSerie,12)/xts::lag.xts(TimeSerie,12)
        }
        #trans = 7: double differencing
        if (trans[i] == 7) {
          if (TS.frequency==12){TimeSerie.Stationarized[,i] <- diff(TimeSerie, differences =
2)}#ok
          if (TS.frequency==4){TimeSerie.Stationarized[,i] <- diff(TimeSerie, lag = 3,
differences = 2)}#ok
          if (TS.frequency==1){TimeSerie.Stationarized[,i] <- diff(TimeSerie, lag = 12,
differences = 2)}#ok
        }
        print(paste0("The TimeSerie ", i, " has been transformed with the tranformation ",
trans[i]) )
    }
  SerOk <-        colSums(is.na(TimeSerie.Stationarized)) < (nrow(TimeSerie.Stationarized) *
na.prop)
  SerOK.vector <- as.vector(SerOk)
  ### TimeSerie.Stationarized_2 = TimeSerie.Stationarized - the ts that has not enough
observations (too many NA)
  TimeSerie.Stationarized_2 <- TimeSerie.Stationarized[, which(SerOk)]
  Frequency_OK <- Frequency[which(SerOK.vector)]
  Trans_OK <- trans[which(SerOK.vector)]
  if (!Blocks=="") {Blocks_OK <- Blocks[which(SerOK.vector),]} else {Blocks_OK<-""}
  if (sum(SerOk) == 1) {
    stop("the procedure can not be done with only one series available.")
  }
  if (sum( !SerOk) > 0) {
    warning(paste(sum(!SerOk), "series ruled out due to lack in observations (more than",
                  round(na.prop * 100, 2), "% NA)."))
  }
  DeletedSeries <- colnames(TimeSerie.Stationarized[, which(!SerOk)])
  print(DeletedSeries)
  TimeSerie.Stationarized_3 <- TimeSerie.Stationarized_2 * NA
        for (i in 1:ncol(TimeSerie.Stationarized_2)) {
          ##  OUTLIERS CORRECTION PROCESS  it consider the data from [1] to the last value
before the recent consecutive NA
          # k.ma : A numeric representing the degree of the moving average correction.
          # h    : A numeric representing the number of steps ahead to forecasting. Default is
12.
          TimeSerie.Stationarized_3[, i] <- Call.OutliersCorrection(x =
TimeSerie.Stationarized_2[,i] ,NA.replace = NA.replace, k.ma = k.ma)
          print(paste0("The timeserie ", i, "has been transformed and its outliers has been
corrected"))
        }
    # Transformation of the zoo/xts object in ts/multits
    data.zooreg  <- zooreg(data = TimeSerie.Stationarized_3, end = last(zoo::index(base)) ,
frequency = 12)
    TimeSerie.Stationarized_4 <- as.ts(x = data.zooreg)
    return(list(TimeSerieStationarized.xts = TimeSerie.Stationarized_3,
                TimeSerieStationarized.ts  = TimeSerie.Stationarized_4,
                Blocks = Blocks_OK,
                Frequency = Frequency_OK,
                Transformation = Trans_OK ) )
}
```

`