UNIVERSIDAD
COMPLUTENSE
MADRID

FACULTY OF ECONOMICS
AND BUSINESS

DEGREE IN ECONOMICS

BACHELOR'S THESIS

TITLE: **Development of a Synthetic Leading Composite Indicator for the U.S. Economy**

AUTHOR: Diana Shilovskaya

TUTOR: Marcos Bujosa Brun

ACADEMIC YEAR: 2023-2024

June 2024

INDEX

ABSTRACT

The objective of this paper is to develop a synthetic composite leading indicator for the U.S economy. Firstly, an explorative analysis is performed, which is aimed at identifying those monthly time series that frequently lead the economic cycle of the United States. Secondly, for each time series selected in the explorative analysis, the trend is estimated using the method of Linear Dynamic Harmonic Regression. Lastly, principal component analysis is employed to construct a linear combination of the estimated trends from the selected time series. This linear combination forms the leading synthetic indicator that aims to improve economic forecasting and provide valuable insights for informed decision making.

INTRODUCTION

In today's complex global economic landscape, the ability to predict economic trends and patterns with accuracy is of utmost importance for policymakers, investors and businesses alike. Leading indicators play a vital role in this endeavour, as they offer insights into the future of economic activity before official data on GDP growth rates, unemployment and inflation are released. Understanding and constructing robust leading indicators is essential for informed decision making and risk management in private and public sectors both.

The objective of this thesis is to construct a composite leading indicator for the economy of the United States of America, capable of predicting economic expansions and recessions. The motivation behind this project comes from a desire to delve deep into time series and econometric analysis. It is also the goal of this thesis to apply and understand the methodology of Linear Dynamic Harmonic Regression, elaborated by Young (1999) and Bujosa (2007), that allows for the dissection of time series data into its three major components: the trend, the seasonal component and the irregular component. Specifically, the component of interest for this project is the trend, uncovered for each economic dataset and then used to develop a composite synthetic leading indicator for the U.S economy. The extracted trends will be aggregated into one composite economic trend, leveraging their respective weights. For this approach to be effective, the datasets selected must exhibit shared behaviours.

2.1 Importance and Criticism of Leading Indicators

Leading indicators are vital instruments for economic performance forecasts, offering crucial insights into future trends before they reveal themselves. In comparison with lagging indicators that simply confirm past occurrences, leading indicators are capable of providing early signs of potential changes in economic activity. By capturing miniature changes in economic behaviour, such as shifts in private consumption patterns or changes in business investment, these indicators aspire to avert economic catastrophes.
Furthermore, leading indicators play an important role in policy making and implementation, as Central Banks and government agencies around the world rely on these indicators to execute informed monetary and fiscal policy decisions that affect interest rates, taxation patterns and spending.

However, it is no secret that the reliability of leading economic indicators in forecasting and preventing significant economic crises, such as the 2008 recession, has been criticised throughout the 20th and 21st centuries. Paul Samuelson, a Noble laureate in economics (1970) has once stated, “In economics, the majority is always wrong’’. What he really referred to is that any sort of economic forecasting is inherently uncertain and subject to error due to the complex nature and dynamics of the economy. Samuelson later explored the challenges of accurately predicting economic phenomena by stating, “The stock market has forecast nine of the last five recessions”, explaining that modern forecast strategies often overpredict recessions by focusing on the short-term fluctuations and noise of the time series, leading to false alarms and a more pessimistic view of the economy.

Another well-known economist, John Maynard Keynes, had also criticised the reliability and use of leading economic indicators by stating that “It is better to be roughly right than precisely wrong”. Keynes argued that when modelling and forecasting economic data, perfect results are unachievable and even unpractical. According to him, there is no universal correct forecasting algorithm that would produce perfect results. Instead, there are different statistical procedures that stem different results, and the challenge presented is to know which procedure best answers the questions that are being asked. Keynes highlighted the importance of flexibility in economic analysis, which is especially justified in the face of economic uncertainty.

Even though the examples above do not refute the use of leading indicators, they call attention to the inherent uncertainties and limitations that are connected with economic forecasting. It is also important to take into account behavioural biases and the complexity of market dynamics that significantly challenge the effectiveness of leading indicators. Human decision making is often irrational and influenced by herd mentality which introduces unpredictability and lack of reason into economic forecasting.

For example, the 2008 crisis was complex to predict due to speculative bubbles and the upturn of the housing market, which led the american society to feel a false sense of economic stability. Moreover, external shocks, such as natural disasters, geopolitical tensions and global health crises often interfere with economic trends and diminish the effectiveness of leading indicators.

2.2 Economic Landscape of the United States

The United States, renowned as one of the world's foremost economies, exhibits a dynamic nature characterised by multifaceted interactions among diverse factors. As a global economic linchpin, the performance of the U.S. economy reverberates across financial markets, trade dynamics, and geopolitical stability worldwide.

To comprehend the economic terrain of the United States necessitates scrutiny of myriad factors, encompassing GDP growth, employment trends, inflation rates, consumer spending, and business investment. These variables collectively shape the trajectory of the economy, influencing consumer confidence, investor sentiment, and broader market dynamics.

The resilience and adaptability of the U.S. economy are manifest in its response to various economic crises and challenges. From enduring recessions and financial crises to periods of robust expansion and innovation, the U.S. economy has weathered diverse storms. Amidst evolving technological advancements, demographic shifts, and geopolitical developments, navigating the intricacies of the U.S. economy mandates a nuanced understanding of its structural dynamics and principal drivers of growth.

THEORETICAL FRAMEWORK FOR INDICATOR CONSTRUCTION

4.1 Introduction to Linear Dynamic Harmonic Regression

The methodology of Dynamic Harmonic Regression is employed in this study because it has proven to be useful for working with non stationary time series data. The Linear Dynamic Harmonic Regression algorithm consists in the identification and the estimation of the unobserved components that form the time series, assigning a sensible model for each of them. This estimation algorithm is based on a spectral approach that decomposes the series into various DHR components who have different variances that are concentrated around certain frequencies.

Considering a univariate time series and applying the LDHR algorithm, the series can be decomposed into the trend, the seasonal, or a periodic component, and the irregular component. The variance of each of those components can be decomposed at different frequencies and this decomposition is known as the spectrum of the series. The trend of the series is the component of most interest for this study, and its variance, which is infinite in the theoretical model, is mainly located at the zero frequency. The trend is also known as a low frequency component; the objective of this paper is to estimate it for each time series chosen and then utilise a weighted aggregation of trends in order to construct a synthetic composite indicator.

Before exploring how LDHR is applied in this specific study, it is essential to first discuss how this algorithm would work on a theoretical model. There are two types of models of interest: mean non-stationary and mean stationary models. A mean non-stationary model is a time series model where the average value of the series is non-constant over time, which would imply that the average value of the series may increase or decrease substantially over time. A mean stationary time series is the opposite: the average value of the series is unchanged over time, and even though data points fluctuate around this mean, the average value remains constant along different time periods. It is crucial to mention that stationary time series are much easier to analyse and work with because they exhibit a well defined variance, autocovariance and mean, which are the statistical properties necessary to build a model over which the time series can be interpolated and extrapolated.

However, it must be noted that the time series subject to analysis and identification in this particular study are non stationary: this type of data is most commonly encountered in economics due to the dynamic and non-constant growth of most business and financial variables over time. GDP components, inflation rates, unemployment rates and stock prices manifest systematic patterns with non - constant mean and variance over time, affected by complex consumer behaviour choices and external shocks. This is why non-stationary time series are challenging to model and forecast.

4.2 Theoretical framework of non stationary and stationary time series data

In a theoretical framework with a mean stationary time series, LDHR is set in motion with the concept of the autocovariance generating function.
A time series inherently displays a degree of correlation between data points across time. This correlation can be quantified by calculating covariances between a data point at time 't' and previous data points at 't-1', 't-2', 't-3', and so forth. The computed autocovariances can be recorded into a sequence of numbers which forms the autocovariance generating function, which is actually a vector with it´s first element being the variance of the time series.

A visual representation of the autocovariance generating function can be presented in a correlogram, which is a depiction of autocovariances that are uniformly graphed in bar format, and each bar is commonly referred to as a lag. This is a widely used tool in ARMA model specification where the speed with which these lags decrease with time or the statistical significance of each lag is crucial in identifying whether the model has significant MA (moving average) or/and AR (autoregressive) components.
However, in order to specify a reasonable model for a specific time series, it is also necessary to look at the Partial Autocorrelation Function, which computes the correlation between two observations while holding constant the effect of the remaining lags of the time series. Therefore, PACF is an exceptionally useful tool for identifying an appropriate lag order of a time series model.

A stationary time series that may be identified with an ARMA (Autoregressive Moving Average) model may be expressed in polynomial form, where time series data multiplied by the autoregressive (AR) polynomial are equal to a white noise process multiplied by a moving average (MA) polynomial. For a series to be mean stationary, the autoregressive polynomial has to be invertible (it has to possess roots outside of the unit circle), which means it would be possible to multiply this polynomial by its own **summable** inverse and the result would yield '1'. This result is known as a unique stationary solution for an infinite stationary sequence of time series data. **Reference.** It is also important to mention that there is an infinite number of inverses for this polynomial, but only one of those inverses is summable.

To expand, every inverse of this AR polynomial is an infinite sequence of numbers, but only one of those inverses has an infinite sequence of numbers such that these numbers are summable to a **finite** quantity. Therefore, in order to adequately express the time series in polynomial form, the autoregressive polynomial must be multiplied by a summabe inverse and the product of the multiplication would yield '1'.

Assuming that the above requirements are satisfied, a theoretical ARMA model is the one for which the original data is equal to an MA polynomial multiplied by the inverse of an AR polynomial, multiplied by white noise process.

To visualise the frequency content of a theoretical ARMA model, the autocovariance generating function must be passed through the Fourier Transform in order to yield a spectrum of a stationary time series. The spectrum of a univariate time series can be defined as a representation of the frequency content of the variance of this series. The analysis and the characteristics of the spectrum will reveal the distribution of the variance frequencies in the data.

It is possible to depict the spectrum between 0 and pi, and the area underneath, or the integral of the spectrum function is the variance of the time series. The regions where the spectrum function exhibits large values indicate the frequencies that significantly contribute to the variance of the series. The frequency of interest for this study is the zero frequency, at which the spectrum function possesses a maximum.

Although it is possible to transform the autocovariance generating function into the spectrum of a stationary time series, non - stationary stochastic processes present a dilemma inside this theoretical framework. This is because the autocovariance generating function is not defined for a non-stationary stochastic process. Since non-stationary series possess no explicitly defined mean and have infinite variance, it is not possible to compute the autocovariance generating function because it does not actually exist. Therefore, there is a level of abstraction regarding the statistical methods applied to analyse and model non stationary data. There is no unique stationary solution, nor is there a **summable** inverse of an AR polynomial which would multiply this polynomial and yield '1'. So, whenever the AR polynomial has roots on the unit circle, the spectrum is not defined.

However, for a non-stationary time series, it is possible to define a certain 'pseudo covariance function', which can be passed through the extended Fourier Transform in order to yield the pseudospectrum of the time series.
The pseudospectrum is a function that describes the distribution of variance over frequency for a non stationary stochastic process. The area underneath, or the integral of this function, is the variance of the time series, just like in the mean stationary case, but with non-stationary data this variance is infinite for some frequencies. **Reference.** Specifically, the representation of the pseudospectrum is similar to the one of the stationary time series spectrum, except that each root of the unit circle of non - stationary data results in a peak with infinite variance for some frequencies represented in the pseudospectrum.

4.3 Application of LDHR to time series framework

The goal of Linear Dynamic Harmonic Regression within time series modelling is to adjust the 'real' pseudospectrum function of the data to an estimated spectrum or the periodogram of the series. LDHR achieves this by minimising the distance between the pseudospectrum of the data and its estimated spectrum.

In order to estimate the spectrum of time series data, LDHR first decomposes the series into its three main components, which are the trend, the seasonal component and the irregular component. Then, for each of those components, LDHR identifies an adequate model with the use of sine and cosine functions, which are applicable for describing recurring and periodic patterns in the data. Later, LDHR identifies a sensible model for the sine and cosine components estimated above, where each harmonic component corresponds to a certain frequency and may be represented on a periodogram with the spectrum function.

The challenge of this algorithm, however, lies in minimising the distance between the 'real' pseudospectrum of the time series and the estimated spectrum of the data. This minimization is impossible to achieve when the pseudospectrum has spectral peaks where the variance is infinite. The area underneath those poles is infinite which means that the distance to minimise is not defined.

Linear Dynamic Harmonic Regression addresses this challenge by using a certain function that is able to transform the poles of infinite variance of the pseudospectrum into points with zero variance. In other words, the minimization problem is multiplied by a certain function that allows for a finite minimal distance between the pseudospectrum and the estimated spectrum. LDHR uses an Ordinary Least Square algorithm in order to minimise this distance.

PRACTICAL DEVELOPMENT OF SYNTHETIC INDICATOR

5.1 The application of Moving Average filter for trend identification

Before using the Linear Dynamic Harmonic Regression method to extract the smoothed trend for each of the chosen datasets, a more common technique known as the Moving Average smoothing has been applied to this study. Moving Average smoothing is a well known algorithm used in time series analysis to reduce noise and reveal the underlying patterns within the data. This method is applied by taking an average of a sample of close - by data points and replacing each data point in this sample with the average. Within the Octave framework, the 'movmean' function had been used to smoothen each one the datasets by replacing the original data with a set of moving average values, which revealed the trend of each series.

However, the moving average filter has not proven to be useful in this study because of the way that the filter has manipulated the data. Once the Moving Average filter was applied, it showed a trend that exhibited significant noise, and it was desired to eliminate such irregularities during the smoothing process. These irregularities were enhanced when the first difference of the trend was taken. Consequently, in order to achieve a smoothed and different trend, another moving average filter was applied to the data, further distorting the true pattern of the series.

The reason why the Moving Average filter had revealed such undesirable results is because Moving Average filters manifest a poor frequency response in time series data smoothing. Time series data exhibit various frequency components, including the low-frequency trend and the high-frequency noise. Seasonal components display a range of frequencies between the highest and the lowest of them. The Moving Average filter aims to diminish the high frequency noise in the data and preserve the low frequency components, such as the trend, or any other gradual variations in time series data. However, the issue is that this filter uniformly weakens all frequencies, resulting in insufficient reduction of high frequencies during the smoothing process. Consequently, the trend pattern does not appear as smooth as desired.

This issue is aggravated when applying the first difference to the moving average trend. Taking the first difference is a widely used technique to put emphasis on short-term fluctuations to better capture the rate of change of the data. That said, the first difference filter amplifies high-frequencies once again, and since the Moving Average filter does not reduce the high frequencies sufficiently, once they are amplified, the differenced trend exhibits noisy and unclear behaviour.

In this particular case study, the Moving Average filter had also exhibited poor performance when it came to handling anomalies in the data, and this was particularly harmful because these anomalies provided valuable insights into the underlying patterns of the data. The filter smoothed out important spikes and dips, discarding vital information about the behaviour of the time series. Furthermore, the ‘movmean’ function in Octave operated by sliding a window of values across the data in order to compute the average for each window. This meant that the initial and the final observations of the sample were gravely affected by the smoothing process. At the beginning of the sample, only succeeding values were averaged, while at the end of the sample, only preceding values were averaged. This asymmetric averaging distorted the edges of the sample, resulting in the higher effectiveness of the filter in the middle of the sample but providing inaccuracies on the edges.

In contrast, the Linear Dynamic Harmonic Regression algorithm had proven to effectively reduce high frequencies of the original data to zero in the smoothed trend. Consequently, when the first difference was applied to the trend, higher frequencies were amplified, but since these high frequencies had been eliminated successfully in the previous stage, the outcome was a smoothed differenced trend. This is because the annulled high noise frequencies remained unaffected during the amplification process of applying the first difference.

5.2 Practical Application of Linear Dynamic Harmonic Regression

After the selection of time series datasets from the Federal Reserve of Economic Data (FRED), the primary tool employed in this study was the Octave platform. The objective was to apply the LDHR algorithm (through custom made Octave functions) to each, monthly, not seasonally adjusted dataset, and estimate each one of the DHR components for every dataset. Consequently, the estimated trend of each time series was taken and used for the construction of the leading synthetic indicator.

Initially, monthly time series data that has not been seasonally adjusted, has been selected. It was vital for this study to employ raw data that has not been modified or filtered using other statistical techniques since the objective was to use the LDHR algorithm only to transform and dissect the data.

Data used in this analysis was taken from mean non - stationary, infinite stochastic processes. However, the samples extracted from these time series are composed of a finite set of observations. Therefore, these samples have a defined mean and finite variance, and so it is possible to define the autocovariance generating function and the corresponding spectrum that can be represented on the periodogram. The spectral peaks in this study would take very large values (however, they are not infinite), and these peaks are located at frequencies with a significant contribution to the variance of the time series.

For each dataset, the practical application of the LDHR algorithm has been set in motion with the 'autodhr' custom made function. This is a key function that identifies and estimates a model for each DHR component of the series. For the correct execution of this function, one of the inputs used was a vector that contains the harmonics corresponding to seasonality of the series. This vector instructed the 'autodhr' program to seek models with a trend and a seasonal component with the corresponding harmonics.

Specifically, for monthly observations, the seasonal component follows a harmonic pattern with oscillations at periods of 12, 6, 4, 3, 2.4, and 2. The spectral peaks for the seasonal component of monthly data are accordingly found at frequencies such as '2pi/12', '2pi/6', '2pi/4', '2pi/3', '2pi/2.4' and '2pi/2'.

Another input for the ‘autodhr’ function was a matrix that indicated the modulus of the roots of the autoregressive process of the dataset, which would help model the amplitude of the oscillations for each model component. This matrix was key to capturing the dynamic behaviour of the time series components.

The ‘autodhr’ program processed the data and had identified and estimated the most sensible model for each DHR component of each dataset.

The next step was to compute the ratios of the variance of innovations for each component relative to the trend variance for each series. In the theoretical LDHR framework, the variance of innovations of the DHR components are the coefficients from the OLS estimation that was used to minimise the distance between the pseudospectrum of the data and the estimated spectrum.

The final step to extract DHR components for each time series involved applying the ‘dhrfilt’ function. This program decomposed the data into its trend, its seasonal and its irregular component. This custom instruction had used the Kalman Filter that was able to accurately extract the necessary components from the series. Although not part of the Linear Dynamic Harmonic regression algorithm, the ‘dhrfilt’ function used the parameters obtained by the ‘autodhr’ function.