Time Series Smoothing by Penalized Least Squares with Applications

Time Series Smoothing by Penalized Least Squares with Applications

Víctor Manuel Guerrero Guzmán
Alejandro Islas Camargo
Willy W. Cortez Yactayo
Eliud Silva Urrutia

Center for the Study of Development Problems (CEPRODE)
Department of Quantitative Methods
School of Economics and Administrative Sciences
University of Guadalajara
September, 2023

Primera edición 2024

© D.R. 2024 UNIVERSIDAD DE GUADALAJARA
Centro Universitario de Ciencias Económico Administrativas
Periférico Norte N° 799 Núcleo Universitario
C. Prol. Belenes, 45100 Zapopan, Jalisco, México

ISBN: 978-607-581-194-9

Hecho en México
Made in Mexico

Contents

Acknowledgments 7

About the authors 9

Prologue 11

Chapter 1. Introduction. .13

Chapter 2. Some Basic Concepts. 23
2.1. Matrix Algebra 23
2.2. Generalized Least Squares 27
2.3. Time Series. 29
2.3.1. The backshift operator 29
2.3.2. ARMA model 30
2.3.2.1. Stationarity and Invertibility 31
2.3.2.2. Model mean .31
2.3.2.3. Model variance 31
2.3.2.4. Model autocorrelation function 32
2.3.2.5. The AR(1) process 32
2.4. Filtering Time series. 34

Chapter 3. Trend Estimation of Univariate Time Series with Controlled Smoothness. 35
3.1. Time series smoothing through Penalized Least Squares . . .35
3.1.1. A statistical solution 37
3.1.2. A measure of smoothness 41

3.2. Penalized Least Squares to second-order integrated processes with percentage of smoothness chosen by the user 42
  3.2.1. TREND REPRESENTATION AND ESTIMATION 42
  3.2.2. CHOOSING THE SMOOTHING PARAMETER TO ACHIEVE SOME DESIRED PERCENTAGE OF SMOOTHNESS 44
  3.2.3. A SIMULATION EXERCISE 52
3.2.3.1. Same time series with different sample sizes 65
3.3. Penalized Least Squares to second-order integrated processes with the percentage of smoothness chosen by the user in the presence of noise autocorrelation 71
  3.3.1. SMOOTHING PARAMETER SELECTION IN PRESENCE OF AUTOCORRELATION 72
  3.3.2. THE IMPACT OF SMOOTHNESS ON THE TREND WHEN THE ERRORS FOLLOW AN $AR(1)$ MODEL 73
  3.3.3. A GRAPHICAL ANALYSIS 77

Chapter 4. Penalized Least Squares to first-order integrated processes with percentage of smoothness chosen by the user 93
4.1. Statistical description of the exponential filter 94
4.2. A measure of smoothness 96
4.3. Daily data and extension to other frequencies 103
4.4. A simulation study. 106
  4.4.1. SAME TIME SERIES WITH DIFFERENT SAMPLE SIZES 117

Chapter 5. Empirical Applications 119
5.1 Case 1: Economic growth and crime 119
5.2. Case 2: Regional Effects of Monetary Policy in Mexico . . 128
5.3. Case 3. The long-run evolution of oil prices 139

Appendix. Technical details. 147

References. 157

Tables Content. 163

Figures Content 165

Acknowledgments

We would like to thank the people who help us in the preparation of this manuscript. We are especially indebted to the University of Guadalajara for the financial support it gave us through the PRO-SNI 2023 program and to Dr. Guillermo Sierra, chair of the Department of Quantitative Methods, who believed this project would be of great help for students learning new techniques within time series analysis.

We would also like to thank The Center for the Study of Development Problems (CEPRODE) for all the facilities it gave us throughout the entire project; in particular, for sponsoring our work and providing the site where the Application for the Univariate Controlled Smoothing (UCS) will be located. https://ceprode.cucea.udg.mx/

The authors gratefully acknowledge the comments and suggestions of two anonymous referees, Víctor Guerrero and Alejandro Islas thank the financial support provided by Asociación Mexicana de Cultura, A. C. to carry out this work.

The following link leads to Shiny app, which automates the controlled smoothing methodology presented in the book.

bit.ly/TSSPLSsoftware24

After clicking the link, you will be redirected to the posit web page, where you will need to complete the registration process to access the shiny.You can also use the following QR Code to get to the app.

About the authors

Víctor Manuel Guerrero Guzmán

He is an Emeritus Professor at the Statistics Department, Autonomous Technological Institute of Mexico (ITAM). He is a member of the National System of Researchers, Level III. He holds a PhD in Statistics from the University of Wisconsin at Madison, USA. His research interests focus on model-based statistical analysis, particularly of time series, and developing methods to generate and analyze official statistical data. ORCID: https://orcid.org/0000-0003-2184-5216

Alejandro Islas Camargo

He is a Professor and Researcher at the Statistics Department, Autonomous Technological Institute of Mexico (ITAM). He is a member of the National System of Researchers, Level II. He holds a PhD in Economics from the University of New Mexico, USA. His academic work focuses on interdisciplinary research questions related to finance, labor market, environment and health economics. ORCID: https://orcid.org/0000-0003-0910-313X

Willy W. Cortez Yactayo

He is a Professor and Researcher at the Department of Quantitative Methods, School of Economics and Administrative Sciences, University of Guadalajara. He is a member of the National System of Researchers, Level III. He holds a PhD in Economics from the University of Notre Dame, Notre Dame, IN, USA. His main research lines are economics of crime, labor markets and macroeconomic performance, income inequality and economic development. ORCID: https://orcid.org/0000-0001-8839-4064.

Eliud Silva Urrutia

He is a Professor at the Faculty of Actuarial Sciences, University Anahuac. He is a member of the National System of Researchers, level I. He holds a PhD in Mathematics Engineering from University Carlos III of Madrid, Spain. His lines of research focus on topics where statistics and demography are linked. In particular, he is interested in making inferences through statistical models applied to social phenomena in Mexico. ORCID: https://orcid.org/0000-0003-0499-0446

Prologue

As the research journey progresses, tools and methodologies emerge that revolutionize how we approach and understand real-world issues. This book, through its range of non-parametric proposals, testifies to a progressive research development guided by Dr. Victor Guerrero's strong leadership. It embodies ideas that are increasingly adapted to the needs of analysts and users of statistics.

Since 2007, the research presented here has evolved, adopting and evolving innovative and complementary methodological approaches. Nevertheless, what distinguishes this compendium is its accessibility. Although deeply rooted in rigorous research, its approach is accessible to doctoral students and those seeking a master's degree, researchers, and analysts. Thus, it is presented as an anteroom, an open invitation for those who wish to dive deeper into time series smoothing issues.

One of the fundamental pillars of the work is the controlled univariate smoothing technique, which is exploratory and allows estimations with specificities for the estimation of trends with different orders of differentiation, such as "d=1" and "d=2", and variants "with constant" and "without constant". The reader will discover how these techniques arise, among other things, from the evaluation of business cycles. The authors present detailed and valuable methods for determining crucial parameters such as the 'lambda', and the impact this has on research.

A distinctive element is the approach to trend estimation. Although there are problems at the extremes or "tails" in previous applications of the proposals, this book proposes solutions to correct these deviations. Moreover, unlike other traditional techniques, such as X12-ARIMA, this book provides tools to generate estimation intervals, giving the analyst a broader and more nuanced perspective of their respective data and estimates.

The book is also characterized by its intuitive part, by proposing a computational tool programmed in free statistical software called R. Some of the benefits of this tool include, for example, allowing analysts to manipulate various parameters such as the order of differentiation, the smoothing parameter, the width of the estimation intervals in real-time. By accessing this online platform, the user can load time series and experiment with different degrees of smoothing, visualizing the changes immediately and gaining a deeper understanding of how these parameters modify the results. In addition, a value added to what is proposed in the book is that the techniques do not impose rigid distribution assumptions to estimate trends, thus offering a remarkable flexibility, especially in data series of less than 500 observations. Through this work, the ideas embodied in the time series smoothing problem proposed by Guerrero (2007) are generalized. In this way, the book offers innovative and simple ways to calculate trends, allowing the analyst to objectively choose the desired smoothness to make possible comparisons between trends with the same smoothness.

In conclusion, more than a simple collection of proposals, this book is a window to the future of research. It invites the reader to enter, question, and continue exploring, always with a reliable tool in hand and the contributions of Dr. Victor Guerrero as a guide, accompanied by his notable collaborators.

María Rosa Nieto Delfín
Research Professor,
School of Business and Economics
University Anáhuac, Mexico

Chapter 1

Introduction

There is a wide agreement among specialists that a time series is made up of different components and therefore that it can be decomposed. A conventional decomposition considers that the observed series contains two elements: trend (T) and cycles (C). Although there are proposals where the series is composed of additional elements such as seasonality and irregular components. There are several possible reasons why a researcher might want to decompose a series. One of them may be that the researcher is interested in analyzing the cyclical behavior of the series. Or, perhaps, she wants to study the series' long-term pattern so that she can do some forecasting. Another possible motive is when the time series contains too much noise that does not allow a clear analysis of the series. In this case, an estimation of the series' trend or cyclical behavior would help the analyst.

In general, any set of ordered and equally spaced observations (in time, age, or any other dimension) could be filtered (or smoothed) to see their underlying trend more clearly. Thus, data smoothing and trend estimation are usually related when analyzing ordered and equally spaced data. Throughout this book we use the term "Time Series" even when the ordering does not correspond to time.

There are different theories and methodologies to decompose a time series. The large majority are linear models that convert one time series into another. The literature uses different names to describe this process of transformation: to decompose, to filter, to smooth. All these terms, however, describe the process by which the original time series is decomposed into its trend and cyclical components. Throughout this book, a filter is defined by any operation on the observed series $\{y_t\}$ that yields another series, which in the present case will be the estimated trend $\{\hat{g}_t\}$. In other words, a filtering procedure consists of applying a filter L on the observed data y_t, such that $\hat{g}_t = Ly_t$. Since g_t is a random variable, it would be preferable to call $\hat{g}_t$ predictor rather than estimator. Nevertheless, we will employ the usual terminology that uses estimation instead of prediction.

The predominant view in the literature considers the trend as the component of the evolution of a persistent series that cannot be attributed to observable factors. The univariate approaches reviewed in the literature assume that the trend is either a deterministic or a random function of time. When analyzing trends, it is common to use an unobserved component model representation. This assumes that the observed time series can be expressed as a signal-plus-noise model, that is,

$$y_t = g_t + v_t \quad (1.1)$$

where for each value of $t \in [1,N]$, $\{g_t\}$ represents the unobserved trend or signal, which may be a random or deterministic function of time, while $\{v_t\}$ denotes the unobserved stationary noise (where a stationary series is such that its mean, variance and covariance between two observations, are constant along time) of the observed value of the series, $\{y_t\}$.

This decomposition has different meanings in different areas of knowledge. In Experimental sciences, for example, v_t is usually interpreted as a pure measurement error, so that the observed series $\{y_t\}$, is assumed to be the sum of the error term plus a signal $\{g_t\}$. In Actuarial sciences, on the other hand, a researcher may use graduation to make the observed data smoother, but she must ensure that the graduated values maintain some degree of closeness to the original data. One of the most critical tasks in Actuarial sciences is to describe the population's actual but unknown mortality pattern.

In Economics, if $\{y_t\}$ represents Gross National Product (GNP), then $\{v_t\}$ is often interpreted as the stationary stochastic cycle or the cyclical component of GNP, while $\{g_t\}$ is the output's trend. The Trend-Cycle decomposition stems from the idea that total output is the sum of a long-term growth and some stationary, temporary fluctuations around its long-term trend. In Finance, the Efficient Market Hypothesis (EMH) suggests that financial asset prices reflect all available information, making future returns unpredictable. However, since the 1990s, many studies have challenged this theory. One argument is that risk premiums vary over time and depend on the business cycle (Cochrane, 2001). This means that the returns of financial assets are related to slow-moving economic variables that exhibit cyclical patterns during the business cycle (Bruder et al., 2011). Another argument is that some agents are not entirely rational, which can result in asset prices underreacting in the short term but overreacting in the long term (Hong & Stein, 1977). Hence, contrary to the EMH, asset prices may exhibit trends and cycles. Assets´ rate of return may therefore be represented by the following expression, $r_t = g_t + c_t$. Where r_t is the asset´s return rate, g_t is the long-term return and c_t is the cyclical return rate.

Trend estimation has a long history and its methods have been improving over time due to advancements in both computational techniques and in statistical theory and methodology. There is a wide array of methodologies, including nonparametric techniques -e.g., kernel models-, local polynomial regression, high-, low- and band-pass filters, and wavelet multiresolution analysis. Other approaches include semiparametric methods, -like splines and Gaussian random fields-, and parametric methods, -like ARIMA (Auto-Regressive Integrated Moving Average) models, Structural Time Series models (proposed by Harvey in 1989), the X-11 seasonal adjusted procedure (proposed by Cleveland and Tiao, 1976), and the Hodrick and Prescott filter (Hodrick & Prescott, 1997).

Among these different methods to detrend time series, the Hodrick-Prescott (HP) filter is one of the most popular ones. Its popularity is due to the fact that it does not require applying a formal statistical model-building process to estimate the trend, as it happens with ARIMA and Structural Time Series models. The Penalized Least Squares (PLS) approach that gives rise to the HP filter postulates that the trend must minimize the function.

$$\min_{g_t}\{\sum_{t=1}^{N}(y_t - g_t)^2 + \lambda \sum_{t=3}^{N}(\nabla^2 g_t)^2\} \quad (1.2)$$

Where the symbol ∇(Nabla) is defined as the difference $\nabla y_t = y_t - y_{t-1}$. When Nabla is raised to the second power it produces the second degree difference equation $\nabla^2 y_t = y_t - 2y_{t-1} + y_{t-2}$. That is why the second sum in (1.2) goes from t=3 to N.

Detrending (graduation) has two basic opposing characteristics: smoothness and goodness of fit to the observed data. That is, the trend is the result of the interaction between these two characteristics: to achieve one, we must sacrifice the other. The first term of the equation penalizes the cyclic component, while the second penalizes the trend component's growth rate. In this context, λ is a constant that penalizes the lack of smoothness in the trend. As $\lambda \to 0$ the fit of the trend is emphasized over smoothness, so that $g_t \to y_t$. The opposite occurs when $\lambda \to \infty$, in which case the trend follows essentially the straight-line model $\nabla^2 \tau_t = 0$. Hence λ plays a vital role in deciding the smoothness of the trend.

We argue that despite its popularity, the HP filter presents some drawbacks that not all its users are aware of. In particular, the selection of λ becomes critical for the model's smoothness and goodness of fit. To the extent that it defines the trend and the cyclical behavior over time, the entire analysis and conclusions depend on the proper selection of λ.

The main objective of this book is to present a more general approach to the time series smoothing problem suggested by Guerrero (2007). Our goal is to present the ideas in a not so technical language and present some examples using the proposed methodology. To achieve this goal, we developed a friendly routine in "R" that anybody can use for free. This routine will allow users to calibrate the smoothness of their time series according to their interests.

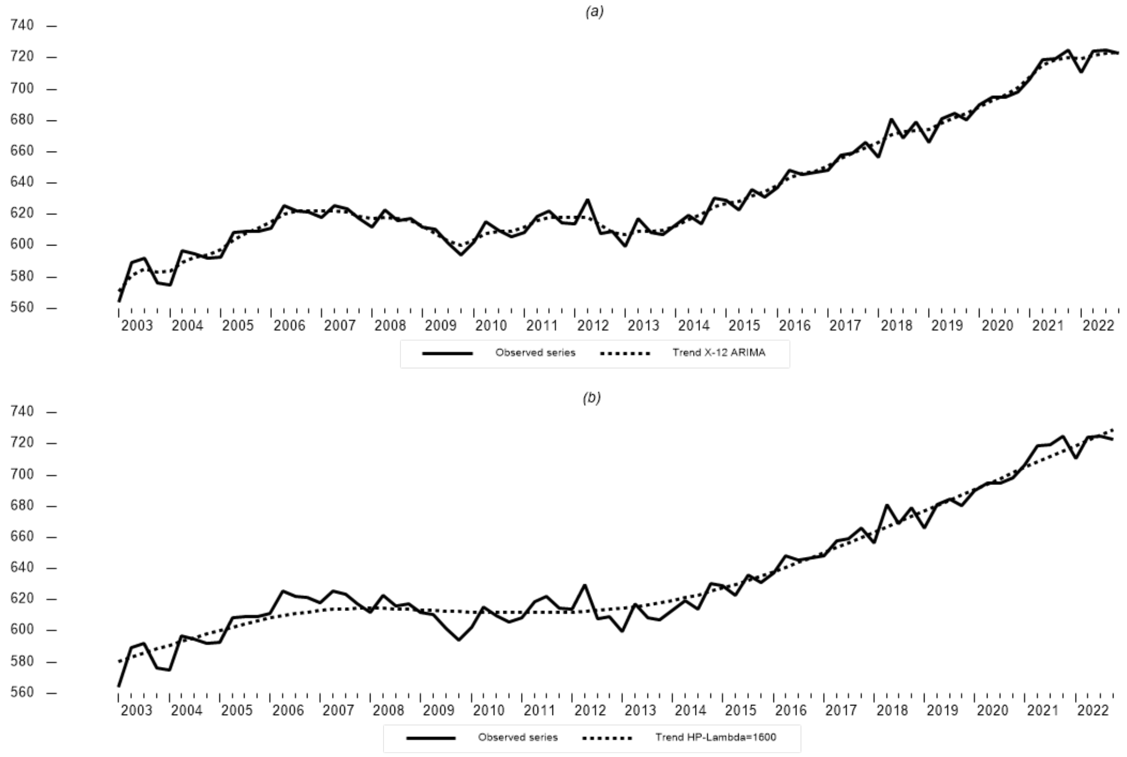
Our approach introduces an index of smoothness as a tool for selecting the smoothing constant. The controlled smoothness approach, is presented as an alternative method for choosing the smoothing parameter objectively. This technique involves estimating the trend of a time series by fixing the desired percentage of smoothness and then determining the λ value that satisfies this criterion. This value must be consistently employed with all time series to ensure valid comparisons. By fixing the same amount of smoothness, the same estimated trend can be obtained when the procedure is applied to another set of observations of the same variable with fewer or more data points than the previous one.

Guerrero's (2007) suggestion is like the interval estimation of a fixed parameter Θ, with an expression like $\hat{\Theta} \pm \varkappa se(\hat{\Theta})$, where $se(\hat{\Theta})$ is the standard error of $\hat{\Theta}$ and $\varkappa$ a percentile of the appropriate distribution. In such a case, it is customary to fix the desired confidence level instead of fixing the value of the constant $\varkappa$ on a priory grounds. This approach provides a better statistical interpretation of the confidence interval and greater comparability with other intervals. By fixing the desired percentage of smoothness of the trend estimator, the same benefits can be achieved for estimating the trend.

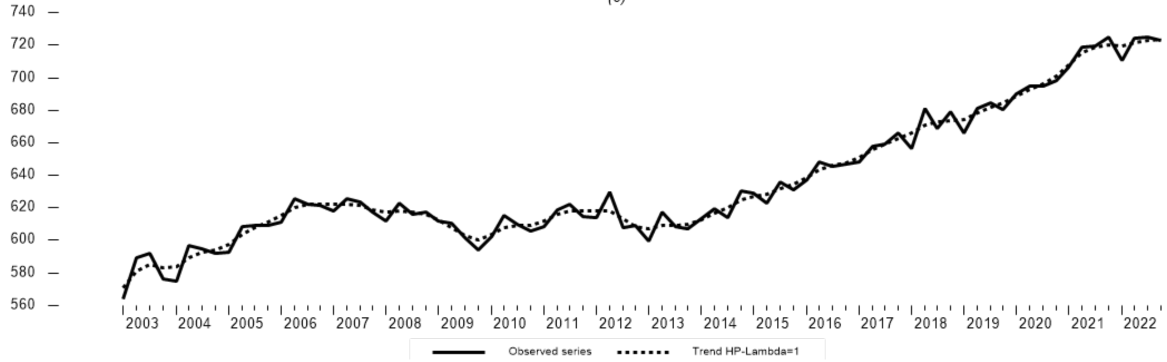
Furthermore, the length of the time series does not significantly impact the estimated trend if the smoothness is controlled. This is a valuable property when estimating trends routinely since trend revisions are minimized when more data are required. This property makes the method ideal for practical applications. Additionally, it is possible to compare estimated trends of time series for different variables, even with different sample sizes and periodicity of data, just by using the same level of smoothness. This is like fixing the confidence level of multiple confidence intervals in order to make valid comparisons. What is important is to keep the same setting of the percentage of smoothness.

Let us present some examples of how our proposed methodology differs from conventional detrending techniques. Figure 1.1 presents the logarithm of the quarterly remittance flows received by the Mexican State named Jalisco during the period 2003-I - 2022-IV. We compare the trend component obtained from applying the seasonal adjustment program X-12 ARIMA-SEATS (Panel a) to the one obtained from using the HP filter, using the standard smoothing parameter $\lambda = 1600$ for quarterly data (Panel b). Panels (c) and (d), on the other hand, show the trend estimated using $\lambda = 1$ and $\lambda = 199$, respectively. We can see that the smoothness of the trend is quite similar in cases (a) and (c), but noticeably different in the other two cases. Therefore, to adequately estimate the trend, we must select the appropriate value for the constant λ in an objective manner.

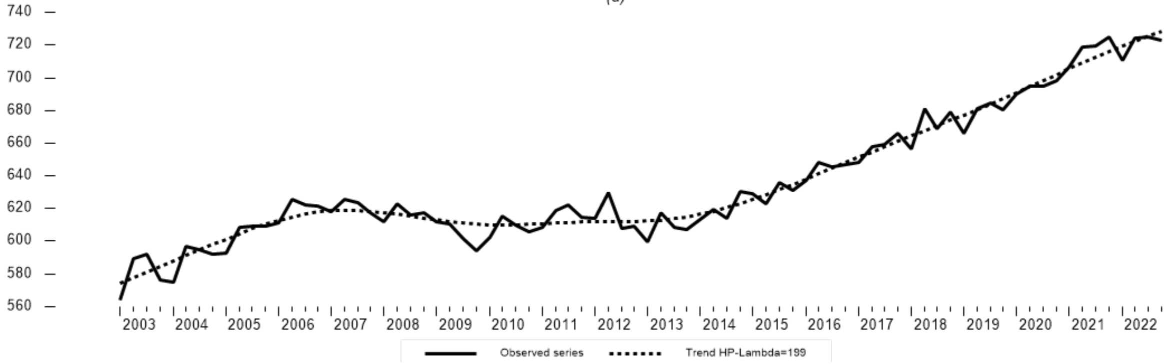
Figure 1.1
The logarithm of Jalisco ´s remittances



(c)



(d)



Source: own estimates

When analyzing quarterly data using the HP filter, it is generally accepted to use the value $\lambda = 1600$. This value was initially proposed by Hodrick and Prescott based on the assumption that $\nabla^2 g_t$ and v_t were independent random variables with a normal distribution $N(0, \sigma_\varepsilon^2)$ and $N(0, \sigma_v^2)$, respectively. They determined that the appropriate values for σ_v and σ_s for the US macroeconomic series they were studying were 5 and 1/8, respectively, resulting in a value of $\sigma_v^2/\sigma_\varepsilon^2 = 1600$. They also tested the results with other values of λ, including 400, 6400, and ∞, and found that only with $\lambda = \infty$ did the estimated trend change significantly. Therefore, the value of $\lambda = 1600$ became the consensus for the smoothing constant when using the HP filter for quarterly data.

However, consensus disappears when other frequencies of observations are used. For example, for annual data, authors like Baxter & King (1999) recommended the value $\lambda = 10$, while Backus & Kehoe (1992), Giorno et al. (1995) and the European Central Bank (2000) used the value $\lambda = 100$. Regarding monthly data, Dolado et al. (1993) used $\lambda = 4800$, while the econometric software E-views uses the default value $\lambda = 14400$.

Many researchers have suggested statistical methods to estimate the smoothing constant, λ, in various situations, as discussed in literature. Kohn et al. (1992) looked at a regression function with ARMA errors, while Lee (2003) compared multiple methods for selecting λ via Monte Carlo simulation. However, it is crucial to note that these methods are computationally complex and lack interpretation for the numerical value of λ. As such, they may not be suitable for estimating time series trends routinely and on a large scale.

To take a rigorous statistical approach, one needs to formulate a model, estimate its parameters (including λ), and ensure that the assumptions underlying it are valid. Harvey & Jager (1993) proposed a structural time series model that employs maximum likelihood estimation. However, this approach may not be suitable for large-scale applications, even with modern computers and fast algorithms, as it requires an explicit statistical model.

On the other hand, Young (1994), Pedersen (2001), and Kaiser & Maravall (2001) took a different approach for choosing the smoothing constant. They analyze the results produced by different values of λ, and used this information to select the most appropriate value. They looked at the effects of λ in the frequency domain and provided criteria for choosing it correctly. This ensures that the HP filter can eliminate cycles with a periodicity that is less than the minimum value required for accurate business cycle analysis.

Regarding the so-called automatic methods to choose the smoothing constant, λ, they are: Cross Validation, Generalized Cross Validation (Craven & Wahba, 1979), Akaike´s Information Criteria (Hurvich et al., 1998), and the Bayesian Information Criteria (Schwarz, 1978) all of them are in the context of cubic smoothing splines. These methods determine the value of λ by optimizing a function that depends on the smoothing parameter. Unfortunately, in the presence of high positive correlation in the noise component of model (1.1), standard smoothing parameter selectors fail to work and overfit the data (Krivobokova & Kauermann, 2007).

This book is structured as follows. In Chapter 2, the basic time series concepts are reviewed to ensure the book is self-contained. Chapter 3 discusses the time series smoothing problem more generally than usual. Section 3.1 presents the problem of smoothing time series of any order of integration, and the estimation method used is Penalized Least Squares. An index of smoothness is deduced and suggested as a tool for selecting the smoothing constant. Section 3.2 covers a general version of the filter for a time series with an integration order of two and uncorrelated noise in the model of the unobserved component. Section 3.3 studies the effect of autocorrelation on the smoothness of the trend when the noise follows an autoregressive process of order one. Chapter 4 presents the case of a time series whose integration order is one, in which case the estimated trend employs the exponential smoothing filter.

Chapter 5 presents three examples of how the proposed filter can be used in the analysis of trend and cyclical components of economic and financial time series.

While there is a fair amount of notation included, the book strives to maintain a narrative and informal writing style. Only brief and essential proofs are presented formally within the text, while longer and more technical proofs are located in the Appendix for optional study.

Chapter 2

Some Basic Concepts

In this chapter we provide an overview of essential concepts and results that facilitate the understanding of the suggested approach to smoothing time series. Those familiar with this material can skip the respective section or read sub-headings for specific details.

2.1. Matrix Algebra

Definitions

An $(n\text{x}m)$ matrix, denoted by an upper bold letter, is an array of numbers ordered into n rows and m columns.

$$\boldsymbol{A}_{(n\text{x}m)} = \begin{bmatrix} a_{11} & a_{12} & \dots. & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ & & \dots. & \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

If there is only one column ($m=1$), then $\boldsymbol{A}$ is described a column vector, whereas if there is only one row ($n=1$), then $\boldsymbol{A}$ is called a row vector. If the number of rows equals the number of columns ($n=m$), the matrix is said to be square. The diagonal running through $a_{11}, a_{22}, \dots, a_{nn}$ in a square matrix is called the principal diagonal. If all elements off the principal diagonal are zero, the matrix is said to be diagonal. A matrix is sometimes specified by describing the element in row i and column j as $\boldsymbol{A} = [a_{i,j}]$.

Summation and multiplication

Two (*n*x*m*) matrices are added element by element as follows,

$$\begin{bmatrix} a_{11} & a_{12} & \ldots. & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ & & \ldots. & \\ a_{n1} & a_{n2} & \ldots & a_{nm} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \ldots. & b_{1m} \\ b_{21} & b_{22} & \ldots & b_{2m} \\ & & \ldots. & \\ b_{n1} & b_{n2} & \ldots & b_{nm} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \ldots. & a_{1m}+b_{1m} \\ a_{21}+b_{21} & a_{22}+b_{22} & \ldots & a_{2m}+b_{2m} \\ & & \ldots. & \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & \ldots & a_{nm}+b_{nm} \end{bmatrix}$$

or, more compactly

$$\boldsymbol{A}_{(n\mathrm{x}m)} + \boldsymbol{B}_{(n\mathrm{x}m)} = \left[a_{ij} + b_{ij}\right]$$

The product of an (*n*x*m*) matrix and an (*m*x*q*) matrix is an (*n*x*q*) matrix.

$$\boldsymbol{A}_{(n\mathrm{x}m)}\mathrm{x}\boldsymbol{B}_{(mxq)} = \boldsymbol{C}_{(n\mathrm{x}q)}$$

where the row *i*, column *j* element of $\boldsymbol{C}$ is given by $\Sigma_{l=1}^{n} a_{il}b_{lj}$. Notice that multiplication requires that the number of columns of $\boldsymbol{A}$ be the same as the number of rows of $\boldsymbol{B}$.

To multiply **A** by a scalar α, each element of **A** is multiplied by α

$$\alpha\boldsymbol{A}_{(n\mathrm{x}m)} = \boldsymbol{C}_{(n\mathrm{x}m)}$$

with

$$\boldsymbol{C} = \left[\alpha a_{ij}\right]$$

Identity matrix

The identity matrix of order *n*, denoted I_n, is an (*n*x*n*) matrix with *1*s along the principal diagonal and *0*s elsewhere.

$$\boldsymbol{I}_{(n)} = \begin{bmatrix} 1 & 0 & \ldots. & 0 \\ 0 & 1 & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

For an (nxn) matrix $\boldsymbol{A}$: $\boldsymbol{A}$ x $\boldsymbol{I}_n = \boldsymbol{A}$ and also, $\boldsymbol{I}_n$ x $\boldsymbol{A} = \boldsymbol{A}$.

Power of matrices

For an (nxn) matrix $\boldsymbol{A}$, the expression $\boldsymbol{A}^2$ denotes $\boldsymbol{A}$x$\boldsymbol{A}$**.** The expression $\boldsymbol{A}^k$ indicates the matrix $\boldsymbol{A}$ multiplied by itself k times. $\boldsymbol{A}^0$ is interpreted as the (nxn) identity matrix. $\boldsymbol{A}$ is said to be idempotent if $\boldsymbol{A}$x$\boldsymbol{A} = \boldsymbol{A}$.

Transposition

The transpose of matrix $\boldsymbol{A}_{(nxm)}$, written $\boldsymbol{A}'$ is the (mxn) matrix whose rows contain the elements of the columns of $\boldsymbol{A}$. Let a_{ij} denote the row i, column j element of matrix $\boldsymbol{A}$, that is, $\boldsymbol{A} = [a_{ij}]$, the transpose of $\boldsymbol{A}$ is given by $\boldsymbol{A}' = [a_{ji}]$. A square matrix $\boldsymbol{A}$ satisfying $\boldsymbol{A} = \boldsymbol{A}'$ is said to be symmetric. Transposition of sum and product is governed by the rules: $(\boldsymbol{A}+\boldsymbol{B})' = \boldsymbol{A}'+\boldsymbol{B}'$; $(\boldsymbol{A}\text{x}\boldsymbol{B})' = \boldsymbol{B}'\text{x}\boldsymbol{A}'$.

Trace of a matrix

The sum of the diagonal elements of a square matrix $\boldsymbol{A}$ is called the trace of $\boldsymbol{A}$, written $tr(\boldsymbol{A})$. That is, $tr(\boldsymbol{A}) = a_{11}+a_{22}+\ldots+a_{nn}$. For square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$**,** $tr(\boldsymbol{A}+\boldsymbol{B}) = tr(\boldsymbol{A}) + tr(\boldsymbol{B})$, whereas if $\boldsymbol{A}$ is (nxm) and $\boldsymbol{B}$ is (mxn), then $\boldsymbol{A}$x$\boldsymbol{B}$ is an (nxn) matrix whose trace is

$$tr(\boldsymbol{A}\text{x}\boldsymbol{B}) = \sum_{j=1}^{m} a_{1j}b_{j1} + \sum_{j=1}^{m} a_{2j}b_{j2} + \cdots + \sum_{j=1}^{m} a_{nj}b_{jn} = \sum_{k=1}^{n}\sum_{j=1}^{m} a_{kj}b_{jk}$$

The product $\boldsymbol{B}$x$\boldsymbol{A}$ is an (mxm) matrix whose trace is

$$tr(\boldsymbol{B}\text{x}\boldsymbol{A}) = \sum_{k=1}^{n} b_{1k}a_{k1} + \sum_{k=1}^{n} b_{2k}a_{k2} + \cdots + \sum_{k=1}^{n} b_{mk}a_{km} = \sum_{j=1}^{m}\sum_{k=1}^{n} b_{jk}a_{kj}$$

Thus,

$$tr(\boldsymbol{A}\text{x}\boldsymbol{B}) = tr(\boldsymbol{B}\text{x}\boldsymbol{A})$$

Matrix inversion

The inverse of a square matrix $\boldsymbol{A}_{(nxn)}$, when it exists, is the (nxn) matrix denoted $\boldsymbol{A}^{-1}$ such that $\boldsymbol{A}\text{x}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\text{x}\boldsymbol{A} = \boldsymbol{I}_n$. A matrix whose inverse exists is called nonsingular. The rules for inverting products and transposes are: $(\boldsymbol{A}\text{x}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\text{x}\boldsymbol{A}^{-1}$ and $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$.

To give a rule for computing the inverse of a matrix requires the definition of the determinant of a square matrix $\boldsymbol{A}$, usually denoted $|\boldsymbol{A}|$. This is a scalar quantity that is calculated from the elements of $\boldsymbol{A}$ as follows. Let $\boldsymbol{A}$ denote an (nxn) matrix and $\boldsymbol{A}_{ji}$ be the (n-1)x(n-1) matrix that results from deleting row j and column i of $\boldsymbol{A}$. The adjoint of $\boldsymbol{A}$ is the (nxn) matrix whose row i, column j element is given by $(-1)^{i+j}|\boldsymbol{A}_{ji}|$. Therefore, if the determinant of $\boldsymbol{A}$ is not equal to zero, its inverse exists and is found by dividing the adjoint by the determinant:

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}(-1)^{i+j}\left|\boldsymbol{A}_{ji}\right|$$

Positive definite matrices

An (nxn) real symmetric matrix $\boldsymbol{A}$ is said to be positive semidefinite if for any real nonzero (nx1) vector $\boldsymbol{x}$, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$. We make the stronger statement that $\boldsymbol{A}$ is positive definite if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$.

- *Lemma 2.1.1* If $\boldsymbol{A}$ (nxn) is symmetric and positive definite, $\boldsymbol{A}^{-1}$ is symmetric and positive definite.
- *Lemma 2.1.2.* If $\boldsymbol{A}$ (nxn) is symmetric and positive definite, there exists a nonsingular matrix $\boldsymbol{P}$ (nxn) such that $\boldsymbol{PAP}' = \boldsymbol{I}_n$ and $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{A}^{-1}$.

2.2. Generalized Least Squares

A general multiple linear regression considers the following model,

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \qquad E(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{0}, \qquad Var(\boldsymbol{\varepsilon}|\boldsymbol{X}) = E(\boldsymbol{\varepsilon\varepsilon}') = \sigma_\varepsilon^2\boldsymbol{\Omega} \quad (2.2.1)$$

In these equations, $\boldsymbol{Y}$ represent an (nx1) column vector, $\boldsymbol{X}$ is an (nxk) matrix, $\boldsymbol{\beta}$ is a (kx1) vector of parameters, $\boldsymbol{\varepsilon}$ is a vector of order (nx1) called the disturbance or error term, and $\boldsymbol{\Omega}$ is a symmetric and positive definite (nxn) matrix. This model allows for the errors to be heteroskedastic, autocorrelated or both.

- If $\boldsymbol{\Omega}$ is diagonal with non-constant diagonal elements, the errors terms are uncorrelated, but they are heteroskedastic.
- If $\boldsymbol{\Omega}$ is not diagonal ($Cov(\varepsilon_i, \varepsilon_j) = \Omega_{i,j} \neq 0$ *for some* $i \neq j$) and the diagonal elements are constant, then the errors are autocorrelated and homoskedastic.
- If $\boldsymbol{\Omega}$ is not diagonal and the diagonal elements are no constant, then the errors are autocorrelated and heteroskedastic.

It is well known that under the following assumptions:

Assumption 2.2.1. $E(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{0}$

Assumption 2.2.2. $Var(\boldsymbol{\varepsilon}|\boldsymbol{X}) = E(\boldsymbol{\varepsilon\varepsilon}'|\boldsymbol{X}) = \sigma_\varepsilon^2\boldsymbol{I}_n$

Assumption 2.2.3. $rank(\boldsymbol{X}) = k$

the Gauss-Markov theorem holds and states that the Ordinary Least Squares (OLS) estimator $\widehat{\boldsymbol{\beta}}_{OLS}$ is the Best Linear Unbiased Estimators (BLUE) of $\boldsymbol{\beta}$.

If OLS is used when $Var(\boldsymbol{\varepsilon}|\boldsymbol{X}) = E(\boldsymbol{\varepsilon\varepsilon}') = \sigma_\varepsilon^2\boldsymbol{\Omega} \neq \sigma_\varepsilon^2 I_n$, $\widehat{\boldsymbol{\beta}}$ is still unbiased, that is, since the OLS estimator can be expressed as

$$\widehat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{XY}$$

then,

$$E(\widehat{\boldsymbol{\beta}}_{OLS}|\boldsymbol{X}) = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}E(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{\beta}$$

but

$$Var(\widehat{\boldsymbol{\beta}}_{OLS}|\boldsymbol{X}) = E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon\varepsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}] = \sigma_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \neq \sigma_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

Since $Var(\widehat{\boldsymbol{\beta}}_{OLS}|\boldsymbol{X}) \neq \sigma_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$, statistical inferences based on the following assumptions,

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}_{OLS}) = \hat{\sigma}^2_{\varepsilon OLS}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$\hat{\sigma}^2_{\varepsilon OLS} = \frac{\widehat{\boldsymbol{\varepsilon}}'_{OLS}\widehat{\boldsymbol{\varepsilon}}_{OLS}}{n-k} = \frac{(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{OLS})'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{OLS})}{n-k}$$

are invalid since, in general $\sigma^2_{\varepsilon\ OLS}$ is a biased and inconsistent estimator of σ^2_ε, consequently $Var(\widehat{\boldsymbol{\beta}}_{OLS}|\boldsymbol{X})$ is also biased and inconsistent.

In other words, the OLS estimators are unbiased and consistent, but their variance estimators are biased and inconsistent, leading to incorrect statistical inference results.

To demonstrate how to handle these cases, assume $\boldsymbol{\Omega}$ is known. By applying Lemmas 2.1 and 2.2, we can factorize $\boldsymbol{\Omega}^{-1} = \boldsymbol{P}'\boldsymbol{P}$. Multiplying both sides of (2.2.1) by $\boldsymbol{P}$ yields a linear relation between transformed data, with the same parameter vector but a transformed error vector:

$$\boldsymbol{PY} = \boldsymbol{PX\beta} + \boldsymbol{P\varepsilon} \quad (2.2.2)$$

Note the properties,

$$E(\boldsymbol{P\varepsilon}|\boldsymbol{X}) = \boldsymbol{P}E(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{0}$$

now, by using lemma 2.2.1 and the properties of inverting and transposing matrix products, we can deduce the following:

$$Var(\boldsymbol{P\varepsilon}|\boldsymbol{X}) = E(\boldsymbol{P\varepsilon\varepsilon}'\boldsymbol{P}|\boldsymbol{X}) = \boldsymbol{P}E(\boldsymbol{\varepsilon\varepsilon}'|\boldsymbol{X})\boldsymbol{P}' = \sigma_\varepsilon^2\boldsymbol{P\Omega P}' = \sigma_\varepsilon^2\boldsymbol{P}(\boldsymbol{\Omega}^{-1})^{-1}\boldsymbol{P}$$

$$= \sigma_\varepsilon^2\boldsymbol{P}(\mathrm{P}'\mathrm{P})^{-1}\boldsymbol{P}' = \sigma_\varepsilon^2\boldsymbol{PP}^{-1}\boldsymbol{P}'^{-1}\boldsymbol{P}' = \sigma_\varepsilon^2\boldsymbol{I}_n$$

Hence, the transformed regression satisfies Assumptions 2.2.1 and 2.2.2. The OLS estimators from the regression of $\boldsymbol{PY}$ on $\boldsymbol{PX}$ are:

$$\widehat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{PX})^{-1}\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{PY} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y} \quad (2.2.3)$$

This method is called Generalized Least Squares (GLS). By construction, since the conditions of the Gauss-Markov theorem have been induced to hold by the transformation, it produces estimators that are BLUE for this model. Nevertheless, GLS is not a feasible estimator as it requires knowledge of Ω, which is usually not available. However, the formula provides the basis for various feasible procedures involving estimates of $\boldsymbol{\Omega}$.

The $\widehat{\boldsymbol{\beta}}_{GLS}$ is unbiased.

$$E\left(\widehat{\boldsymbol{\beta}}_{GLS}|\boldsymbol{X}\right) = E[(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y}|\boldsymbol{X}] = E[(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})|\boldsymbol{X}]$$

$$= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{E}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{\beta}$$

The GLS variance-covariance matrix is:

$$Var\left(\widehat{\boldsymbol{\beta}}_{GLS}|\boldsymbol{X}\right) = \sigma_\varepsilon^2(\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X})^{-1} = \sigma_\varepsilon^2(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}$$

with an estimator of σ_ε^2, given by

$$\hat{\sigma}_\varepsilon^2 = \frac{\left(\boldsymbol{PY} - \boldsymbol{PX}\widehat{\boldsymbol{\beta}}_{GLS}\right)'\left(\boldsymbol{PY} - \boldsymbol{PX}\widehat{\boldsymbol{\beta}}_{GLS}\right)}{n-k} = \frac{\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{GLS}\right)'\boldsymbol{P}'\boldsymbol{P}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{GLS}\right)}{n-k}$$

$$= \frac{\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{GLS}\right)'\boldsymbol{\Omega}^{-1}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{GLS}\right)}{n-k}$$

2.3. Time Series

Time series data refer to observations on a variable that occurs in a time sequence; usually, the observations are equally spaced in time. A time series model provides a convenient, simple, probabilistic description of a process of interest. In this section, we will describe the class of models known as Auto-Regressive Moving Average (ARMA). Further details of these models can be found in the textbook by Guerrero (2009).

2.3.1. The backshift operator

The backshift operator, B, plays a useful role in carrying out algebraic manipulations in time series analysis. It is defined by the transformation.

$$By_t = y_{t-1} \quad (2.3.1)$$

Applying B to y_{t-1} yields $By_{t-1} = y_{t-2}$. Substituting in (2.3.1) gives $B(By_t) = B^2y_t = y_{t-2}$ and so, in general,

$$B^j y_t = y_{t-j},\ j = 1,2, \dots.. \text{ (2.3.2)}$$

To complete the definition, let B^0 have the property $B^0y_t = y_t$ so that (2.3.2) holds for all non-negative integers.

Let $\nabla = 1 - B$ denotes the difference operator, such that, $\nabla^0 y_t = y_t$, $\nabla y_t = y_t - y_{t-1}$, $\nabla^2 y_t = \nabla(\nabla y_t) = \nabla(y_t - y_{t-1}) = y_t - 2y_{t-1} + y_{t-2}$ and so on. Therefore, $\nabla^d y_t$ denotes the *dth* order difference of $\{y_t\}$, with d a non-negative integer.

If y_t is a deterministic linear trend, as in $y_t = a + bt$, then

$$\nabla y_t = y_t - y_{t-1} = (a + bt) - \big(a + b(t-1)\big) = b$$

$$\nabla^2 y_t = \nabla(\nabla y_t) = \nabla(b) = b - b = 0$$

In general, it can be seen that ∇^d will reduce a polynomial of degree d to a constant.

2.3.2. ARMA model

The non-seasonal ARMA(p,q) model can be written as

$$\phi_p(B)y_t = \theta_0 + \theta_q(B)a_t \text{ (2.3.3)}$$

where Θ_0 is a constant term

$$\phi_p(B) = \big(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\big)$$

is the *pth*-order autoregressive (AR) operator, and

$$\theta_q(B) = \big(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\big)$$

is the *qth*-order moving average (MA) operator.

The variable a_t in equation (2.3.3) is the random shock term, also sometimes called the innovation term, which is assumed to be independent over time and normally distributed with mean 0 and variance σ_a^2.

The ARMA model can also be written as

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} - \theta_1 a_{t-1} - \phi_2 a_{t-2} + \cdots + \theta_q a_{t-q}$$

2.3.2.1. *Stationarity and Invertibility*

If $p = 0$, the model is a pure MA model, and y_t is always stationary. An ARMA model with $p > 0$ AR terms is stationary if all roots of the polynomial in B $\Phi_p(B)$ lie outside the unit circle. If $q = 0$, the ARMA model is pure AR, and y_t is always invertible. An ARMA model with $q > 0$ MA terms is invertible if all roots of $\Phi_q(B)$ lie outside the unit circle.

2.3.2.2. *Model mean*

The mean of a stationary AR(p) model is derived as

$$E(y_t) = E\left(\theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + a_t\right)$$

$$= E(\theta_0) + \phi_1 E(y_{t-1}) + \phi_2 E(y_{t-2}) + \cdots + \phi_p E(y_{t-p}) + E(a_t)$$

$$= \theta_0 + (\phi_1 + \phi_2 + \cdots + \phi_p) E(y_t)$$

$$= \frac{\theta_0}{\phi_1 - \phi_2 - \cdots - \phi_p}$$

Because $E(\Theta_k a_{t-k}) = 0$ for any k, it is easy to show that this same expression also gives the mean of a stationary ARMA(p,q) model.

2.3.2.3. *Model variance*

There is no simple general expression for the variance of a stationary ARMA(p,q) model. There are, however, simple formulas for the variance of the AR(p) and MA(q) models.

The variance of the AR(p) model is:

$$\gamma_0 = \sigma_y^2 = Var(y_t) = E(\tilde{y}^2) = \frac{\sigma_a^2}{\phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p}$$

where $\tilde{y} = y_t - E(y_t)$ and $\rho_1,\ldots,\rho_p$ are autocorrelations (see below)

The variance of the MA(q) model is:

$$\gamma_0 = \sigma_y^2 = Var(y_t) = E(\tilde{y}^2) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_a^2$$

2.3.2.4. Model autocorrelation function

The model autocorrelations

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov(y_t, y_{t-k})}{Var(y_t)} = \frac{E(\tilde{y}_t\tilde{y}_{t-k})}{Var(y_t)}, \quad k = 1,2,\ldots$$

are the theoretical correlations between observations separated by k time periods. This autocorrelation function, where the correlation between observations separated by k time periods is a constant, is defined only for stationary time series. The true autocorrelation function depends on the underlying model and the parameters of the model. Standard expectation operations can derive the covariances and variances needed to compute the autocorrelation function. The formulas are simple for pure AR and pure MA models. The formulas become more complicated for mixed ARMA models.

2.3.2.5. The AR(1) process

The variable y_t follows an AR(1) process if

$$y_t = \phi y_{t-1} + a_t \quad for\ t = 1,2,\ldots,N\ and\ y_1 = a_1, where\ |\phi| \leq 1$$

Although the time series is observed at time t = 1, the process is regarded as having started at some time in the remote past. Substituting repeatedly for lagged values of y_t gives

$$y_t = \sum_{j=0}^{J-1} \phi^j a_{t-j} + \phi^J y_{t-J} \text{ (2.3.4)}$$

Now, since $|\Phi| \leq 1$ the component Φ^j is negligible if j is large As $j \to \infty$, it effectively disappears and so if the process is regarded as having started at one point in the remote past, it is possible to write (2.3.4) in the form

$$y_t = \sum_{j=0}^{\infty} \phi^j a_{t-j} , \qquad t = 1,2,\dots,N,.. \text{ (2.3.5)}$$

and, since summing the squared coefficients as a geometric progression yields

$$\sum_{j=0}^{\infty} \phi^{2j} = 1/(1-\phi^2)$$

Therefore,

$$E(y_t) = \sum_{j=0}^{\infty} \phi^j E(y_{t-j}) = 0,$$

$$\gamma(0) = E(y_t^2) = E\left(\sum_{j=0}^{\infty} \phi^j a_{t-j}\right)^2 = \sigma_a^2 \sum_{j=0}^{\infty} \phi^{2j} = \sigma_a^2/(1-\phi^2)$$

$$\gamma(k) = E(y_t y_{t+k}) = \sigma_a^2 \left(\sum_{j=1}^{\infty} \phi^j \phi^{k+j}\right) = \sigma_a^2 \phi^k \sum_{j=0}^{\infty} \phi^{2j} = \sigma_a^2 \phi^k/(1-\phi^2), k = 1,2,\dots$$

and

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k, k = 0,1,\dots$$

$$= \phi^{|k|}, k = 0, \pm 1, \pm 2, \dots$$

2.4. Filtering Time series

In the model

$$x_t = \sum_{j=-\infty}^{\infty} \gamma_j y_{t-j}$$

the collection of $\{\gamma_j\}$ is called a linear filter. Clearly, x_t is a linear function of y_t and a filtered version of y_t. Linear filtering, where γ_j is a known collection of numbers, is often used to identify patterns and signals in a noisy time series (in this case, y_t). The weights γ_j could be found in such a way as to capture the relevant variation associated with the particular component of interest. Thus, a filter for the trend would capture the variation related to the long-run term movement of the series, and a filter for the seasonal component would capture the variation of a seasonal nature. A filter designed in this way, with a prior choice of the weights, is an "ad hoc" fixed filter in the sense that it is independent of the particular series to which it is going to be applied.

Over time, the use of ad hoc filtering has shown severe limitations. One major drawback is its fixed nature, which can lead to spurious results, and for some series, the component may be overestimated, while for others, it may be underestimated. To overcome this limitation, an alternative approach is suggested in this book.

Chapter 3

Trend Estimation of Univariate Time Series with Controlled Smoothness

3.1. Time series smoothing through Penalized Least Squares

This chapter heavily draws on the work of Guerrero (2007, 2008) and Guerrero et al. (2010, 2018).

When conducting statistical analysis of time series, it is natural for the trend concept to arise. This is because the trend of a time series serves as a descriptive measure equivalent to the centrality measure of a data set. However, it is essential to note that the center of a time series behaves dynamically, and analysts often need to distinguish between short-term and long-term movements. The common notion of trend is that it reflects the long-term behavior of an underlying component of the observed time series and evolves smoothly. To extract information from observed data about concepts such as permanent income or potential output, smoothing procedures are regularly applied. Economic theories frequently use empirically unobservable concepts like expectations and equilibrium variables. The results provide stylized facts about business cycles (Kydland & Prescott 1990, Björnland 2000), or they are used as artificial data in econometric analysis; for example, permanent income explains private consumption, and the output gap serves as an explanatory variable in a Phillips type equation for the dynamics of inflation.

As stated in the introduction, it is common to use an unobserved component model representation when analyzing trends. This model assumes that the observed time series can be expressed as a signal-plus-noise model, not because the data was actually generated this way, but to account for the patterns found in the data. That is,

$$y_t = g_t + v_t \quad (3.1.1)$$

where for each value of t ranging from 1 to N, $\{g_t\}$ represents the unobserved trend or signal, which may be a random or deterministic function of time, and $\{v_t\}$ refers to the unobserved stationary noise present in the observed value of the series being studied, denoted as $\{y_t\}$. Researchers in fields like economics, demography, and finance have being employing this type of representation for some time. In an early use of (3.1.1), Whittaker (1923) and Henderson (1924) suggested to smooth (graduate) actuarial data by solving the following Penalized Least Squares (PLS) problem for $\mu = 0$ and $\lambda > 0$. PLS produces a method that minimizes a sum of squares function that considers fidelity to the original data, plus a penalty for the lack of smoothness on the trend. The penalty is given by the sum of squared differences of order d of the trend, weighted by a smoothing parameter (λ) that trades off smoothness against goodness of fit, as follows:

$$\min_{g_t}\left\{\sum_{t=1}^{N}(y_t - g_t)^2 + \lambda \sum_{t=d+1}^{N}(\nabla^d g_t - \mu)^2\right\} \quad (3.1.2)$$

Here, μ denotes the mean, if exists, or just a reference level for $\{\nabla^d g_t\}$, where $\nabla = 1 - B$ denotes the difference operator and B is the backshift operator such that $Bg_t = g_{t-1}$ for every subindex t. Therefore, $\nabla^d g_t$ denotes the *dth* order difference of $\{g_t\}$, with d a nonnegative integer. That is, $\nabla^0 g_t = g_t$, $\nabla g_t = g_t - g_{t-1}$, $\nabla^2 g_t = \nabla(\nabla g_t) = \nabla(g_t - g_{t-1}) = g_t - 2g_{t-1} + g_{t-2}$ and so on, in such a way that the second term of (3.1.2) is related to the smoothness of g_t.

The parameter $\lambda > 0$ is a constant that penalizes the lack of smoothness in the trend. That is, as $\lambda \to 0$, the trend resembles more closely the original data, so that $g_t \to y_t$ for all t, and no smoothness is achieved. The opposite occurs when $\lambda \to \infty$, in which case the trend follows essentially the (smooth) polynomial implied by $\nabla^d g_t = \mu$. Thus, in the latter case, when $d = 0$, the trend will be constant, that is, $g_t = \mu$ and when $d \geq 1$ the trend will be given by the polynomial $g_t = \beta_0 + \beta_1 t + \ldots + \beta_{d-1} t^{d-1} + (\mu/d!)t^d$, where the constants β_i, for $i = 0,1,\ldots, d - 1$ depend on the d values of $\{g_t\}$. Therefore, assuming $\mu = 0$ has implications on the degree of the polynomial.

Intuitively, the minimization problem has two opposing forces. One force attempting to minimize the sum of squared cyclical noise. The other force is attempting to minimize the sum of squared $\nabla^d g_t - \mu$. The smoothing parameter λ, gives relative weight to these two forces.

The value of d in (3.1.2) is usually chosen by the analyst on a priori grounds as $d = 2$, and seldom is $d \geq 3$ used in practice. When $\mu = 0$ and $d = 1$ or 2, the corresponding solution to the minimization problem is well-known in the financial and economic literature. They are called exponential smoothing and Hodrick-Prescott (HP) filtering, respectively.

For the $\mu = 0$ and $d = 2$ case, several approaches lead to stochastic models generally used to represent trends of economic time series. They are based on: a) Auto-Regressive Integrated Moving Average models (ARIMA); b) Structural Time Series Models, as proposed by Harvey (1989); c) The X-11 Seasonal Adjustment procedure (Cleveland & Tiao, 1976); and d) the HP filter (Hodrick & Prescott, 1997). The general model form employed is $\nabla^2 \tau_t = (1 - \Theta_1 B - \Theta_2 B^2)\alpha_t$, where the parameters Θ_1 and Θ_2 are constant and $\{\alpha_t\}$ is a white noise Gaussian process, that is, a sequence of independent and identically distributed random errors with normal distribution.

3.1.1. A statistical solution

The minimization problem discussed in (3.1.2) is more general than the typical problem studied in the literature, as it does not assume μ to be zero beforehand. However, a solution for $\mu = 0$ is available in Kitagawa & Gersch (1996), where they approached it through a least squares computational perspective. King & Rebelo (1993) also obtained the same solution using optimal linear filtering tools. In this book, we propose estimating a random vector to develop a statistical solution and derive an index of smoothness. Therefore, we consider the following tentative statistical model for $\{g_t\}$, which is similar to the one used by Hodrick & Prescott (1997) or Kitagawa & Gersch (1996).

$$\nabla^d g_t = \mu + \varepsilon_t \quad \text{for } t = d + 1, \dots, N \, (3.1.3)$$

With $\{\varepsilon_t\}$ a sequence of serially uncorrelated and identically distributed random errors with mean zero and $Var(\varepsilon_t) = \sigma_\varepsilon^2$.

Now, define the following arrays: $\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_N)$, $\boldsymbol{g} = (g_1, g_2, \dots, g_N)'$ and $v = (v_1, v_2, \dots, v_N)'$ are Nx1 vectors; $\boldsymbol{\varepsilon} = (\varepsilon_{d+1}, \varepsilon_2, \dots, \varepsilon_N)'$ and $\mathbf{1}_{N-d} = (1,1,\dots,1)'$ are N-dx1 vectors (from now on, a prime ´is used to denote the transpose of a vector or matrix). And K_d is the matrix representation of the difference operator ∇^d, so that,

$$K_d = \begin{pmatrix} \boldsymbol{k_d} & & \mathbf{0}_{N-d-1} \\ 0 & \boldsymbol{k}_d & \mathbf{0}_{\boldsymbol{N-d-2}} \\ & & \cdots \\ \mathbf{0}_{\boldsymbol{N-d-1}} & & \boldsymbol{k}_d \end{pmatrix} \quad (3.1.4)$$

is an $(N\text{-}d)\text{x}N$ matrix, with $\boldsymbol{k_d}$ the $1\text{x}(d+1)$ vector given by

$$\boldsymbol{k}_d = \left((-1)^d \binom{d}{d}, (-1)^{d-1} \binom{d}{d-1}, \ldots, (-1) \binom{d}{1}, \binom{d}{0} \right)$$

Where $\binom{d}{l} = \frac{d!}{l!(d-l)!}$, is the binomial coefficient, and $\mathbf{0}m$ is the $1\text{x}m$ zero vector. Thus, the vector of observed data $\boldsymbol{Y}$ can be expressed as

$$\boldsymbol{Y} = \boldsymbol{g} + \boldsymbol{v}, \quad (3.1.5)$$

where the trend component is given, for μ known, by

$$K_d \boldsymbol{g} = \mu \mathbf{1}_{N-d} + \boldsymbol{\varepsilon}. \quad (3.1.6)$$

Therefore, combining (3.1.5) and (3.1.6) we have the system of equations:

$$\begin{pmatrix} \boldsymbol{Y} \\ \mu \mathbf{1}_{N-d} \end{pmatrix} = \begin{pmatrix} I_N \\ K_d \end{pmatrix} \boldsymbol{g} + \begin{pmatrix} \boldsymbol{v} \\ -\boldsymbol{\varepsilon} \end{pmatrix} \quad (3.1.7)$$

The following assumptions will be made throughout the book: (i) the matrix $\binom{I_N}{K_d}$ has column rank N, (ii) the random vectors $\boldsymbol{v}$ and $\boldsymbol{\varepsilon}$ have zero mean vector with positive definite Variance-Covariance matrices and are uncorrelated, that is, $E(\boldsymbol{v}) = \mathbf{0}_N$, $Var(\boldsymbol{v}) = \sigma_v^2 V$, $E(\boldsymbol{\varepsilon}) = \mathbf{0}_{N-d}$, $Var(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 I_{N-d}$ and $E(\boldsymbol{v}'\boldsymbol{\varepsilon}) = \mathbf{0},$ with V a known positive definite matrix. It follows from assumptions (i) and (ii) that the vector mean and the Variance-Covariance matrix of the combined error vector in (3.1.7) are

$$\boldsymbol{E}\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix} = \mathbf{0}_{2N-d} \quad \text{and } Var\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix} = \boldsymbol{E}\left[\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}(\boldsymbol{\nu}' \quad -\boldsymbol{\varepsilon}')\right] = \begin{pmatrix}\sigma_\nu^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)}\\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d}\end{pmatrix} = \boldsymbol{\Omega}$$

Where the matrix $\boldsymbol{\Omega}$ is positive definite. Therefore, there exists a matrix P such that,

$$P'P = \Omega^{-1} \text{ and } P\Omega P' = I_{N-2}$$

Now, by using the matrix P, (3.1.7) can be expressed as follows:

$$P\begin{pmatrix}\boldsymbol{Y}\\ \mu\mathbf{1}_{N-d}\end{pmatrix} = P\begin{pmatrix}I_N\\ K_d\end{pmatrix}\boldsymbol{g} + P\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}$$

$$\boldsymbol{E}\left[P\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}\right] = P\boldsymbol{E}\left[\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}\right] = \mathbf{0}_{2N-d}$$

and

$$\boldsymbol{E}\left[P\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}(\boldsymbol{\nu}' \quad -\boldsymbol{\varepsilon}')P'\right] = P\Omega P' = \boldsymbol{I}_{N-2}$$

Then, by applying Least-Squares we obtain the following estimation equation.

$$(I_N{}' \quad K_d{}')\begin{pmatrix}\sigma_\nu^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)}\\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{Y}\\ \mu\mathbf{1}_{N-d}\end{pmatrix}$$

$$= (I_N{}' \quad K_d{}')\begin{pmatrix}\sigma_\nu^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)}\\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d}\end{pmatrix}^{-1}\begin{pmatrix}I_N\\ K_d\end{pmatrix}\widehat{\boldsymbol{g}},$$

where $\hat{\boldsymbol{g}}$ is the resulting estimator of the trend $\boldsymbol{g}$. The unbiased estimator for the trend and its respective MSE are:

$$\hat{\boldsymbol{g}} = (V_{N\text{x}N}^{-1} + \lambda K_d{}'K_d)^{-1}(V_{N\text{x}N}^{-1}\boldsymbol{Y} + \lambda\mu K_d{}'\mathbf{1}_{N-d}) \quad (3.1.8)$$

$$\Gamma^{-1} = Var(\hat{\boldsymbol{g}}) = \sigma_v^2(V_{N\text{x}N}^{-1} + \lambda K_d{}'K_d)^{-1} \quad (3.1.9)$$

A proof of these results is provided in the Appendix.

The proposed model can incorporate several possible statistical and/ or economic modeling assumptions. For instance, taking $\mu = 0$, Reeves et al. (2000) consider a diagonal matrix V with non constant variance $\sigma_{v\,t}^2$, for $t = 1, 2, ..., N$. Then, the resulting estimated trend corresponds to a time-varying smoothing parameter λ_t. The matrix V could also be used to account for the presence of cycles in the observed component by considering that $\{v_t\}$ is a stationary and invertible $ARMA(p,q)$ process. Harvey (1985) and Harvey & Jaeger (1993) suggest specifying $p=2$, which allows the cycle process to be periodic in the sense of having a peak in its spectral density function. These are examples of how to incorporate prior information about the economy's structure. A case that will be considered in Section 3.3 is an ARMA(1,0), that is an AR(1) stationary process for the $\{v_t\}$ component.

A feasible trend estimator must consider that μ is commonly unknown and must be estimated from the data. Thus, an unbiased estimator of μ is given by the sample mean of $\{\nabla^d y_t\}$, that is,

$$\hat{\mu} = (N-d)^{-1}\mathbf{1}'_{N-d}K_d\boldsymbol{Y} = \frac{\sum_{t=d+1}^{N}\nabla^d y_t}{N-d}, \quad (3.1.10)$$

therefore, (3.1.8) becomes,

$$\hat{\boldsymbol{g}} = (V_{N\text{x}N}^{-1} + \lambda K_d{}'K_d)^{-1}(V_{N\text{x}N}^{-1}\boldsymbol{Y} + \lambda K_d{}'\mathbf{1}_{N-d}\hat{\mu}) \quad (3.1.11)$$

$$= (V_{N\text{x}N}^{-1} + \lambda K_d{}'K_d)^{-1}(V_{N\text{x}N}^{-1}\boldsymbol{Y} + \lambda K_d{}'\mathbf{1}_{N-d}(N-d)^{-1}\mathbf{1}'_{N-d}K_d\boldsymbol{Y})$$

$$= (V_{N\text{x}N}^{-1} + \lambda K_d{}'K_d)^{-1}(V_{N\text{x}N}^{-1} + \lambda(N-d)^{-1}K_d{}'\mathbf{1}_{N-d}\mathbf{1}'_{N-d}K_d)\boldsymbol{Y}$$

Further, to measure variability around the estimated series $\{\hat{\boldsymbol{g}}\}$ we need to estimate σ_v^2 in (3.1.9). Then, under assumptions (i) and (ii), an unbiased estimator of σ_v^2 is given by

$$\tilde{\sigma}_{\nu}^{2} = [\boldsymbol{Y}'V_{N\mathrm{x}N}^{-1}\boldsymbol{Y} - \hat{\boldsymbol{g}}'(V_{N\mathrm{x}N}^{-1} + \lambda K_d{}'K_d)\hat{\boldsymbol{g}}]/(N-d) + \lambda\mu^{2} \quad (3.1.12)$$

The proof is provided in the Appendix.

To appreciate the effect of the constant μ, it is essential to note the following remarks. (i) By wrongly assuming $\mu = 0$, the variance σ_{η}^{2} will be underestimated by an amount that grows as $\lambda\mu^{2}$. (ii) We should notice that the array $K_d{}'\,\mathbf{1}_{N-d}$ is an N-dimensional vector of zeros, except for the first d and last d elements, that is, when $d \geq 1$ and given that $\sum_{i=0}^{d}(-1)^{i}\binom{d}{i} = 0$ we have,

$$K_d{}'\mathbf{1}_{N-d} = \left[\sum_{i=d}^{d}(-1)^{i}\binom{d}{i}, \ldots, \sum_{i=1}^{d}(-1)^{i}\binom{d}{i}, 0, \ldots, 0, \sum_{i=0}^{d-1}(-1)^{i}\binom{d}{i}, \ldots, \sum_{i=0}^{0}(-1)^{i}\binom{d}{i}\right]'$$

Therefore, the observed values of the original series $\{Y_t\}$ enter the formula of the estimator $\hat{\boldsymbol{g}}$ modified in both of its extremes by the value of μ, weighted by λ. (iii) It is worth reminding that when extrapolating the trend, $\mu \neq 0$ implies the trend follows a polynomial of degree d. In contrast, $\mu = 0$ implies a polynomial of degree d-1. Besides, the extrapolated values will depend critically on the least d estimated trend values. That is, let $\hat{g}_N(h)$ be the h-period ahead forecast of g_{N+h}, with origin at N, then for $h \geq 1$ we get $\hat{g}_N(h)=\mu$ if $d=0$, $\hat{g}_N(h)=h\mu+g_N$ if $d=1$, and $\hat{g}_N(h)=[h(h+1)/2]\mu+(h+1)g_N-hg_{N-1}$ if $d=2$.

3.1.2. A measure of smoothness

To apply the proposed method in practice, deciding the value of the smoothing parameter λ is the only thing that is required. By looking at the precision matrix of $\hat{\boldsymbol{g}}$, Γ in (3.1.9), we see that it is composed by two precision matrices, $\sigma_{\nu}^{-2}V^{-1}_{NxN}$ associated with expression (3.1.5) for the observations, and $\sigma_{\varepsilon}^{-2}K_d{}'K_d$ associated with expression (3.1.6) for the smooth component of the series. Measuring the precision contributed by the smooth component to the total precision is now interesting. This amounts to deriving a scalar measure to quantify the share of $\sigma_{\varepsilon}^{-2}K_d{}'K_d$ in (3.9). Such a measure is given by the following expression.

$$\Lambda(\sigma_{\varepsilon}^{-2}K_d{}'K_d, \Gamma) = 1 - tr[V_{N\mathrm{x}N}^{-1}(V_{N\mathrm{x}N}^{-1} + \lambda K_d{}'K_d)^{-1}]/N \quad (3.1.13)$$

In the Appendix, it is demonstrated that equation (3.1.13) possesses the following characteristics: (1) symmetry; (2) values ranging from zero to one; (3) invariance under linear nonsingular transformations of the variable concerned; and (4) linear behavior.

This index evaluates the precision of the smoothness component in relation to the total precision. It can be expressed as a percentage to represent the degree of smoothness in the estimated trend, when multiplied by 100%. The higher the index, the smoother the trend.

3.2. Penalized Least Squares of second-order integrated processes with percentage of smoothness chosen by the user

As previously mentioned, analysts typically choose the value of d in equation (3.1.2) to be $d \leq 2$, and rarely $d \geq 3$ in practice. When $\mu = 0$ and $d = 2$, the solution to the minimization problem is known as the Hodrick-Prescott (HP) filter, which is commonly used to estimate trends and detrend economic time series. However, the HP filter has been criticized for its endpoint sensitivity, generation of spurious cycles, and arbitrariness in the choice of the smoothing parameter λ. This section presents a slightly more general solution by considering $\mu \neq 0$.

3.2.1. Trend representation and estimation

The penalized approach that gives rise to the HP filter postulates that the trend must minimize the function.

$$\min_{\tau_t}\left\{\sum_{t=1}^{N}(y_t - g_t)^2 + \lambda\sum_{t=3}^{N}(\nabla^2 g_t - \mu)^2\right\} \quad (3.2.1)$$

As in (3.1.2), λ is a constant that penalizes the lack of smoothness in the trend. That is, as $\lambda \to 0$ the trend resembles more closely the original data, so that $g_t \to y_t$ for all t, and no smoothness is achieved. Conversely, when $\lambda \to \infty$, the trend follows the smooth polynomial model $\tau_t - 2\tau_{t-1} + \tau_{t-2} = \mu$ which represents the trend growth expressed as a second difference. Therefore, λ plays a crucial role in determining the smoothness, while μ is a reference level for the trend growth. It is important to note that the trend follows a second-degree polynomial given by this equation

$$g_t = \beta_0 + \beta_1 t + (\mu/2)t^2$$

which becomes a straight line when $\mu = 0$. Thus, using $\mu = 0$ as is usual in practice has important consequences on the trend behavior, particularly at the endpoint of the time series, as will be seen below.

Therefore, for $d = 2$, $\mu \neq 0$ and non-autocorrelated noise component, that is $Var(\eta) = \sigma_v^2 I$, (3.1.8), (3.1.9), (3.1.11) and (3.1.12) becomes.

$$\hat{\boldsymbol{g}} = (I_N + \lambda K_2' K_2)^{-1}(\boldsymbol{Y} + \lambda\mu K_2' \mathbf{1}_{N-2}) \quad (3.2.2)$$

$$\Gamma^{-1} = Var(\hat{\boldsymbol{\tau}}) = \sigma_v^2 (I_N + \lambda K_2' K_2)^{-1} \quad (3.2.3)$$

$$\hat{\boldsymbol{g}} = (I_N + \lambda K_2' K_2)^{-1}(I_N + \lambda(N-2)^{-1} K_2' \mathbf{1}_{N-2} \mathbf{1}'_{N-2} K_2)\boldsymbol{Y} \quad (3.2.4)$$

$$\tilde{\sigma}_v^2 = [\boldsymbol{Y}'\boldsymbol{Y} - \hat{\boldsymbol{g}}'(I_N + \lambda K_2' K_2)\hat{\boldsymbol{g}}]/(N-2) + \lambda\mu^2 \quad (3.2.5)$$

where K_2 is the $(N\text{-}2)$xN matrix representation of the second difference operator appearing on the above formulas

$$K_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ & & & & & \ddots & & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 & 1 \end{pmatrix}$$

It is important to note that the estimator $\tilde{\sigma}_v^2$ was obtained assuming that μ was a known parameter, even though it actually needs to be estimated. Therefore, when replacing $\hat{\mu}$ with μ, it is necessary to adjust (3.2.5) accordingly. In such a scenario, the suggested estimator would be: $\tilde{\sigma}_v^2 = (N - 2)\tilde{\sigma}_v^2/(N - 2 - 1)$, that is.

$$\hat{\sigma}_v^2 = \left[\sum_{t=1}^{N}(Y_t - \hat{g}_t)^2 + \lambda \sum_{t=3}^{N}(\nabla^2 \hat{g}_t - \hat{\mu})^2\right] \Big/ (N-3) \quad (3.2.6)$$

In order to understand the impact of the constant μ, it is important to note that the array $K_2'\mathbf{1}_{N-2}$ in equation (3.2.2) is a vector of zeros with dimension N, except for the first two and last two elements. This means that $K_2'\mathbf{1}_{N-2} = (1,1,0,...,0,1,1)'$. As a result, the values observed in the original series $\{y_t\}$ are included in the formula for the estimator $\hat{\mathbf{g}}$, which is adjusted at both extremes by the value of μ and weighted by λ. Equation (3.2.2) shows that the smoother matrix $(I_N+\lambda K_2'K_2)^{-1}$ is applied to $\boldsymbol{Y}+\lambda\mu K_2'\mathbf{1}_{N-2} = (y_1 + \lambda\mu, y_2 + \lambda\mu, y_3, \dots, y_{N-2}, y_{N-1} + \lambda\mu, y_N + \lambda\mu\ )$. By doing this, the filter adjusts the first and last two values of the series with the aim of reducing end-point bias.

It is widely acknowledged that when data is revised or new observations are made, all previously estimated trend values will change. The HP filter has been criticized for being too sensitive at the actual end of the sample, which makes it difficult to interpret the estimated trend. This is particularly problematic for policy applications that focus on current development (see for example, Mohr 2001). A more comprehensive solution was proposed by Guerrero (2007) (3.2), where $\mu \neq 0$ helps to correct the end-point sensitivity.

3.2.2. Choosing the smoothing parameter to achieve some desired percentage of smoothness

The smoothness index (3.1.13) for the non-autocorrelated case becomes.

$$S(\lambda, N) = 1 - tr\,[(I_N + \lambda K'_2 K_2)^{-1}]/N \quad (3.2.7)$$

In order to measure the level of precision attributed to the trend smoothness induced by model (3.1.6) for $d=2$, it is recommended to use (3.2.7) as a smoothness index. It is important to note that this index only relies on the values of λ and N, as K_2 remains constant. It is worth mentioning that since K_2 is a matrix of rank N-2, the matrix K'_2K_2 has two eigenvalues of zero, while the remaining N-2 eigenvalues can be arranged in descending order as $e_1 \geq e_2 \geq \dots \geq e_{N-2}$. Thus, the trace in (3.2.7) can be expressed as:

$$tr(I_N + \lambda K'_2 K_2)^{-1} = (1 + \lambda e_1)^{-1} + \ \dots + \ (1 + \lambda e_{n-2})^{-1} + 2$$

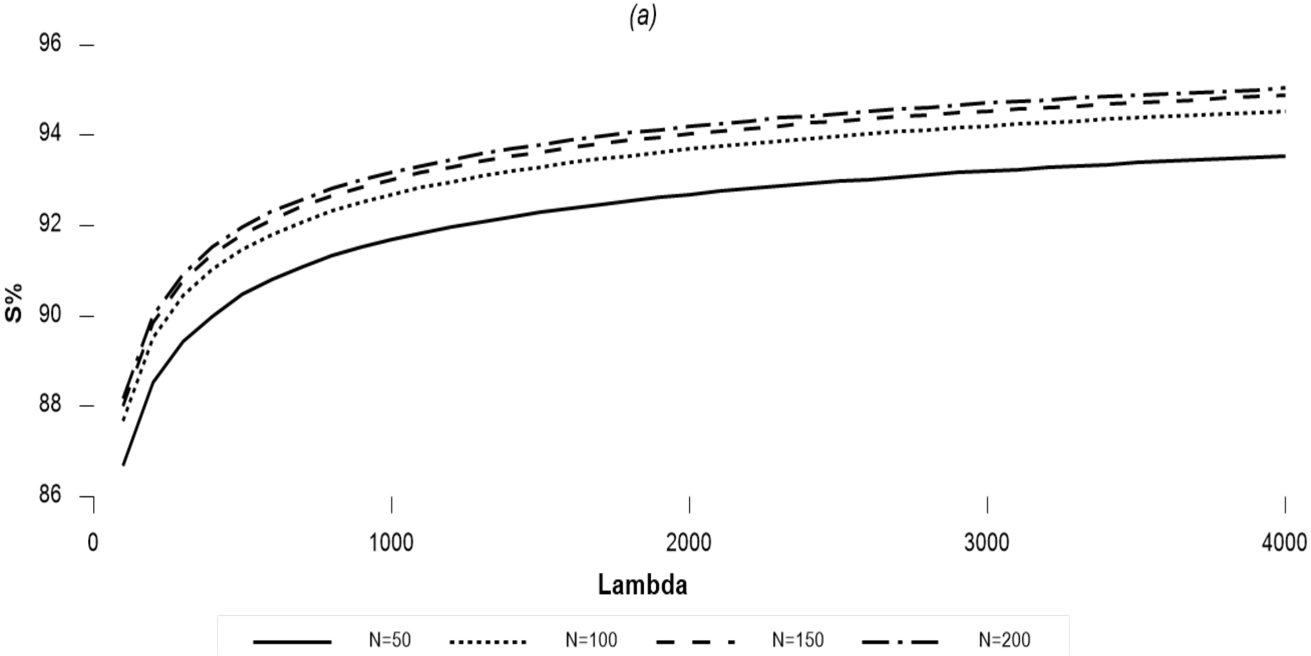
and it can be observed that $S(\lambda,N) \to 0$ as $\lambda \to 0$ and $S(\lambda,N) \to 1-2/N$ as $\lambda \to \infty$.

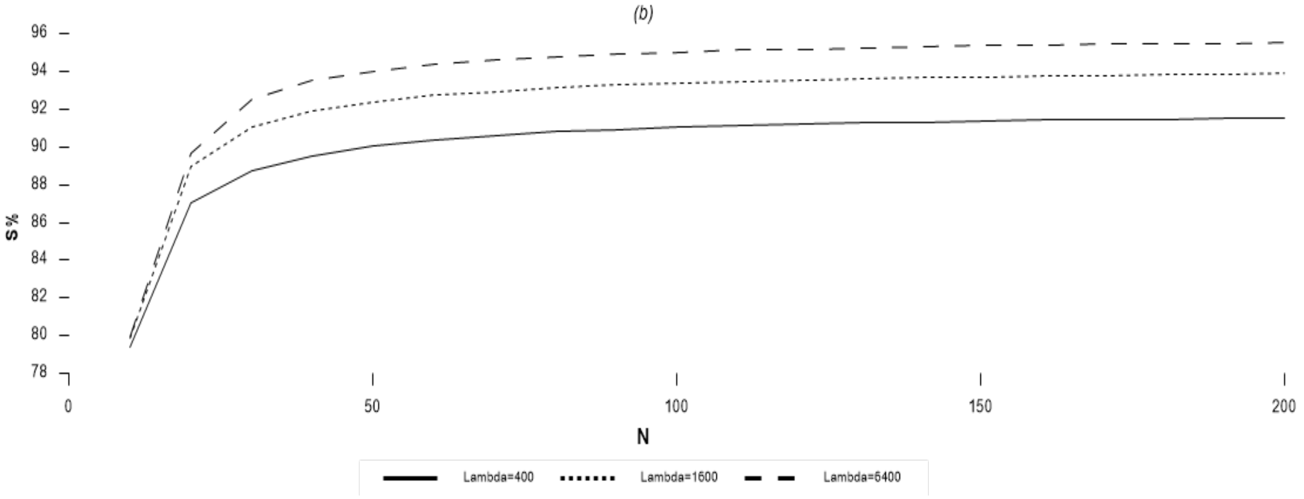
This means that no matter how large the value of λ is, the trend will never achieve 100% smoothness. However, the larger the sample size (N), the smoother the trend can be. To express the index in terms of percentage, we can write $S(\lambda,N)\%$ to indicate the percentage of smoothness achieved by the filter. We must specify the smoothness $S(\lambda,N)\%$ and determine the corresponding λ for fixed values of N. It is important to note that there is no analytical solution for λ from expression (3.2.7), so it cannot be calculated directly. Instead, the calculations are done numerically from (3.2.7) while keeping N and $S(\lambda,N)\%$ fixed.

The chart in Figure 3.2.1 displays how $S(\lambda, N)\%$ behaves for various values of N and λ. Figure 3.2.1 (a) depicts that $S(\lambda, N)\%$ increases rapidly as λ grows, but then slows down considerably around λ =1000, regardless of the sample size. On the other hand, Figure 3.2.1 (b) demonstrates how the sample size affects the results for fixed λ values, which were the same ones used by Hodrick and Prescott (1997). In all three cases depicted in each graph, the percentage of smoothness is greater than 90%, even when the sample size is as small as $N=50$ or the smoothing constant is as small as λ =400.

Figure 3.2.1

Behavior of $S(\lambda, 0)\%$ for: (a) N=50, 100, 150, 200, and (b) λ = 400, 1600, 6400





Source: own estimates

Regrettably, it is impossible to derive an analytical expression for λ as a function of N and $S\%$ from equation (3.2.7). Consequently, Table 3.2.1 presents values of λ for various sample sizes and percentages of smoothness. These values were obtained numerically by solving equation (3.2.7) for λ based on the given N and $S\%$.

Table 3.2.1

Values of λ as a function of sample size N and percentage of smoothness S% (quarterly series)

N	60%	70%	72.5%	75%	77.5%	80%	82.5%	85%	87.5%	90%	92.50%	95%
20	1.35	4.66	6.91	10.79	17.96	32.56	66.31	159.66	511.97	-	-	-
21	1.32	4.51	6.66	10.33	17.08	30.67	61.63	145.63	444.15	4314.41	-	-
22	1.30	4.38	6.44	9.94	16.33	29.08	57.75	134.15	396.02	2406.37	-	-
23	1.27	4.26	6.24	9.60	15.68	27.72	54.48	124.63	359.08	1768.13	-	-
24	1.25	4.16	6.07	9.30	15.12	26.54	51.70	116.65	329.30	1445.12	-	-
25	1.24	4.07	5.92	9.04	14.62	25.52	49.30	109.89	304.60	1246.60	-	-
26	1.22	3.98	5.79	8.80	14.18	24.62	47.21	104.10	283.75	1109.28	-	-
27	1.20	3.91	5.67	8.59	13.79	23.82	45.37	99.09	265.98	1006.34	49779.6	-
28	1.20	3.85	5.56	8.41	13.45	23.12	43.76	94.72	250.71	924.69	13619.2	-
29	1.18	3.79	5.46	8.24	13.13	22.48	42.32	90.88	237.49	857.34	8455.76	-
30	1.17	3.73	5.37	8.08	12.85	21.91	41.03	87.47	225.96	800.30	6387.05	-
32	1.15	3.63	5.21	7.81	12.35	20.92	38.81	81.69	206.87	708.33	4558.67	-
34	1.13	3.55	5.08	7.58	11.93	20.09	36.99	76.99	191.74	637.59	3696.99	-
36	1.12	3.48	4.96	7.38	11.57	19.39	35.46	73.10	179.46	582.00	3171.85	-
38	1.10	3.42	4.86	7.21	11.27	18.79	34.16	69.84	169.30	537.46	2802.36	-
40	1.09	3.36	4.77	7.07	11.00	18.27	33.04	67.05	160.79	501.12	2520.04	-
42	1.08	3.31	4.70	6.93	10.77	17.82	32.07	64.66	153.54	470.93	2294.54	68885
44	1.07	3.27	4.63	6.82	10.56	17.42	31.22	62.58	147.32	445.48	2110.21	38407
46	1.06	3.23	4.57	6.71	10.38	17.07	30.47	60.75	141.91	423.74	1957.48	28209
48	1.06	3.19	4.51	6.62	10.21	16.75	29.80	59.13	137.17	404.98	1829.54	23046
50	1.05	3.16	4.46	6.54	10.06	16.47	29.20	57.70	132.99	388.64	1721.22	19871
52	1.04	3.13	4.41	6.46	9.93	16.21	28.66	56.41	129.28	374.29	1628.54	17673

N	60%	70%	72.5%	75%	77.5%	80%	82.5%	85%	87.5%	90%	92.50%	95%
54	1.04	3.11	4.37	6.39	9.80	15.98	28.18	55.25	125.95	361.60	1548.41	16025
56	1.03	3.08	4.33	6.33	9.69	15.77	27.73	54.20	122.96	350.29	1478.48	14717
58	1.03	3.06	4.30	6.27	9.59	15.57	27.33	53.25	120.26	340.17	1416.95	13637
60	1.02	3.04	4.26	6.21	9.49	15.39	26.96	52.38	117.81	331.05	1362.41	12723
62	1.02	3.02	4.23	6.16	9.41	15.23	26.62	51.58	115.57	322.80	1313.76	11935
64	1.01	3.00	4.21	6.12	9.32	15.08	26.31	50.85	113.52	315.30	1270.12	11249
66	1.01	2.99	4.18	6.07	9.25	14.93	26.02	50.17	111.64	308.46	1230.77	10647
68	1.01	2.97	4.15	6.03	9.18	14.80	25.75	49.54	109.90	302.18	1195.13	10116
70	1.00	2.96	4.13	6.00	9.11	14.68	25.50	48.96	108.30	296.42	1162.71	9645
72	1.00	2.94	4.11	5.96	9.05	14.57	25.26	48.42	106.81	291.10	1133.09	9225
74	1.00	2.93	4.09	5.93	9.00	14.46	25.04	47.92	105.42	286.18	1105.95	8850
76	0.99	2.92	4.07	5.90	8.94	14.36	24.84	47.44	104.13	281.62	1080.99	8512.9
78	0.99	2.91	4.05	5.87	8.89	14.26	24.65	47.00	102.93	277.37	1057.96	8208.1
80	0.99	2.89	4.04	5.84	8.84	14.18	24.46	46.59	101.80	273.41	1036.65	7931.4
82	0.99	2.88	4.02	5.81	8.80	14.09	24.29	46.20	100.74	269.71	1016.88	7679.2
84	0.98	2.87	4.01	5.79	8.75	14.01	24.13	45.83	99.75	266.25	998.49	7448.4
86	0.98	2.87	3.99	5.77	8.71	13.94	23.98	45.48	98.81	263.00	981.35	7236.6

λ values calculated numerically by solving (3.1.7) given S% and N

- Denotes an unreliable value

Table 3.2.1

Values of λ as a function of sample size N and percentage of smoothness S% (quarterly series, continuation)

N	60%	70%	72.5%	75%	77.5%	80%	82.5%	85%	87.5%	90%	92.50%	95%
88	0.98	2.86	3.98	5.74	8.68	13.87	23.84	45.15	97.92	259.94	965.33	7041.33
90	0.98	2.85	3.96	5.72	8.64	13.80	23.70	44.84	97.09	257.07	950.32	6860.93
92	0.98	2.84	3.95	5.70	8.60	13.73	23.57	44.55	96.30	254.35	936.25	6693.77
94	0.97	2.83	3.94	5.68	8.57	13.67	23.45	44.27	95.55	251.78	923.01	6538.47
96	0.97	2.83	3.93	5.66	8.54	13.61	23.33	44.00	94.84	249.36	910.55	6393.86
98	0.97	2.82	3.92	5.65	8.51	13.56	23.22	43.75	94.16	247.06	898.80	6258.90
100	0.97	2.81	3.91	5.63	8.48	13.51	23.11	43.50	93.52	244.87	887.69	6132.68
102	0.97	2.81	3.90	5.61	8.45	13.46	23.01	43.27	92.91	242.80	877.19	6014.40
104	0.97	2.80	3.89	5.60	8.43	13.41	22.91	43.05	92.32	240.82	867.23	5903.35
106	0.97	2.79	3.88	5.58	8.40	13.36	22.82	42.84	91.76	238.94	857.79	5798.92
108	0.96	2.79	3.87	5.57	8.38	13.32	22.73	42.64	91.23	237.15	848.81	5700.53
110	0.96	2.78	3.86	5.56	8.35	13.27	22.64	42.45	90.72	235.44	840.28	5607.70
112	0.96	2.78	3.85	5.54	8.33	13.23	22.56	42.27	90.23	233.80	832.15	5519.97
114	0.96	2.77	3.85	5.53	8.31	13.19	22.48	42.09	89.77	232.24	824.40	5436.94
116	0.96	2.77	3.84	5.52	8.29	13.15	22.41	41.92	89.32	230.74	817.00	5358.26
118	0.96	2.76	3.83	5.51	8.27	13.12	22.33	41.75	88.89	229.30	809.94	5283.59
120	0.96	2.76	3.83	5.50	8.25	13.08	22.26	41.60	88.47	227.92	803.18	5212.65

λ values calculated numerically by solving (3.1.7) given S% and N

- Denotes an unreliable value

Table 3.2.2 offers a useful tool for selecting λ in practical applications with sample sizes greater than 120 quarters. The table shows a parsimonious function of N that provides a reliable estimate of λ, based on the results of estimating a function that approximates the λ values produced by (3.2.7). The coefficients of determination (R^2) for the estimated simple linear regression model $\widehat{log}(\lambda) = \hat{\beta}_0 + \hat{\beta}_1(1/N)$ are also provided in the table, all of which are very close to unity. This function gives a close approximation to the true value of λ, making it a handy resource for simplifying the selection process.

Table 3.2.2

Estimation of an approximating function that relates
λ to S% and N, N > 120 (quarterly series)

Approximating function: $\lambda = Exp(\hat{\beta}0 + \hat{\beta}_1 1/N)$			
S%	**$\hat{\beta}_0$**	**$\hat{\beta}_1$**	**$100R^2$%**
60%	-0.113914	8.210552	99.951
70%	0.908608	12.324897	99.924
72.5%	1.221603	13.888427	99.91
75%	1.567039	15.791137	99.89
77.5%	1.951985	18.170203	99.88
80%	2.385697	21.217659	99.85
82.5%	2.880539	25.265800	99.80
85%	3.453495	30.918669	99.72
87.5%	4.127088	39.477104	99.55
90%	4.205282	102.804613	95.23
92.5%	5.847323	87.131279	89.72
95%	6.113325	67.069052	99.83

3.2.3. A simulation exercise

To test the effectiveness of the suggested approach, a simulation exercise was conducted. Three dynamic behaviors were analyzed, two nonlinear ones and one linear. The first nonlinear series simulated the trend function as a specific instance of a piecewise function with the following expression.

$$g_t = \begin{cases} \alpha_0 + \beta_0 t + \gamma_0 t^2, & t = 1, \dots, n_1 \\ \alpha_1 + \beta_1 t + \gamma_0 t^2, & t = n_1 + 1, \dots, n_2 \\ \alpha_2 + \beta_2 t + \gamma_0 t^2, & t = n_2 + 1, \dots, N \end{cases}$$

The selection of parameters a_i's and β_i's is such that we get smooth joints. That is, $\alpha_{i-1} + \beta_{i-1}n_i + \gamma_0 n_i^2 = \alpha_i + \beta_i n_i + \gamma_0 n_i^2$, for $i = 1,2$. The function specification was decided in order for the trend model implied $\nabla^2\tau_t = \mu$ to resemble the behavior of an observed time series. Therefore,

$$\begin{aligned}\nabla^2 g_t &= 2\gamma_0, & t &= 1,\dots,n_1,n_1+3,\dots,n_2,n_2+3,\dots,N\\ &= 2\gamma_0 + (\alpha_1-\alpha_0) + (n_1+1)(\beta_1-\beta_0), & t &= n_1+1\\ &= \gamma_0 + (\alpha_0-\alpha_1) + n_1(\beta_0-\beta_1), & t &= n_1+2\\ &= 2\gamma_0 + (\alpha_2-\alpha_2) + (n_2+1)(\beta_2-\beta_1), & t &= n_2+1\\ &= 2\gamma_0 + (\alpha_1-\alpha_2) + n_2(\beta_1-\beta_2), & t &= n_2+2\end{aligned}$$

Thus, $\nabla^2 g_t = 2\gamma_0$ except for the discontinuity points located at the observations $t = n_1 + 1, n_1 + 2, n_2 + 1, n_2 + 2$.

The second nonlinear time series was also defined as piecewise linear having the expression.

$$g_t = \begin{cases}\alpha_0 + \beta_0 t, & t = 1,\dots,n_1\\ \alpha_1 + \beta_1 t, & t = n_1+1,\dots,n_2\\ \alpha_2 + \beta_2 t, & t = n_2+1,\dots,n_3\\ \alpha_3 + \beta_3 t, & t = n_3+1,\dots,n_4\\ \alpha_4 + \beta_4 t, & t = n_4+1,\dots,N\end{cases}$$

The selection of intercepts α's is such that we get smooth joints. That is $\alpha_i + \beta_i n_i = \alpha_{i+1} + \beta_{i+1}n_{i+1}$, for $i = 1, \dots, 3$. The function specification was decided for the trend model implied by $\nabla^2 g_t = 0$ to resemble the behavior of an observed time series. Therefore

$$
\begin{aligned}
\nabla^2 g_t &= 0, && t = 1, \dots, n_1, n_1 + 3, \dots, n_2, n_2 + 3, \dots, n_3, n_3 + 3, \dots, n_4, n_4 + 3, \dots, N \\
&= (\alpha_1 - \alpha_0) + (n_1 + 1)(\beta_1 - \beta_0), && t = n_1 + 1 \\
&= -(\alpha_1 - \alpha_0) - n_1(\beta_1 - \beta_0), && t = n_1 + 2 \\
&= (\alpha_2 - \alpha_1) + (n_2 + 1)(\beta_2 - \beta_1), && t = n_2 + 1 \\
&= -(\alpha_2 - \alpha_1) - n_2(\beta_2 - \beta_1), && t = n_2 + 2 \\
&= (\alpha_3 - \alpha_2) + (n_3 + 1)(\beta_3 - \beta_2), && t = n_3 + 1 \\
&= -(\alpha_3 - \alpha_2) - n_3(\beta_3 - \beta_2), && t = n_3 + 2 \\
&= (\alpha_4 - \alpha_3) + (n_4 + 1)(\beta_4 - \beta_3), && t = n_4 + 1 \\
&= -(\alpha_4 - \alpha_3) - n_4(\beta_4 - \beta_3), && t = n_3 + 2
\end{aligned}
$$

Thus, $\nabla^2 g_t = 0$ except for the discontinuity points located at the observations $t = n_1 + 1, n_1 + 2, n_2 + 1, n_2 + 2, n_3 + 1, n_3 + 2, n_4 + 1, n_4 + 2$. The parameters of the nonlinear models are specified in Table 3.2.3.

Table 3.2.3

Parameter values for the nonlinear trends

Nonlinear$_1$	$N = 120$ $n_1 = 40$, $n_2 = 77$	$\gamma_0 = {}^{2.4}/_N$	$\alpha_0 = 5$, $\beta_0 = 0.4$ $\alpha_1 = 81.8$, $\beta_1 = 2.32$ $\alpha_2 = 180.36$, $\beta_2 = 3.6$
Nonlinear$_2$	N = 120 $n_1 = 25$, $n_2 = 77$ $n_3 = 40$, $n_4 = 100$	$\beta_0 = 8/N$ $\beta_1 = -5/N$ $\beta_2 = 5/N$ $\beta_3 = -2/N$ $\beta_4 = 8/N$	$\alpha_0 = 0.030$ $\alpha_1 = 2.738$ $\alpha_2 = -0.178$ $\alpha_3 = 4.487$ $\alpha_4 = -3.845$

The *linear* function for the trend is $g_t = 1 + 6t/N$, for $t = 1, 2, \dots, N$.

In Figure 3.2.2, there are three simulated series that are based on the parameters described earlier. The theoretical and trend estimates for these series are shown with different smoothing levels. The smoothing indices used are obtained from Table 3.2.1. The dotted lines represent the true trends. Panels (a) and (b) show the nonlinear time series estimated trends, while panel (c) displays the linear time series estimated trends.

Figure 3.2.2

Examples of simulated time series with estimated trends

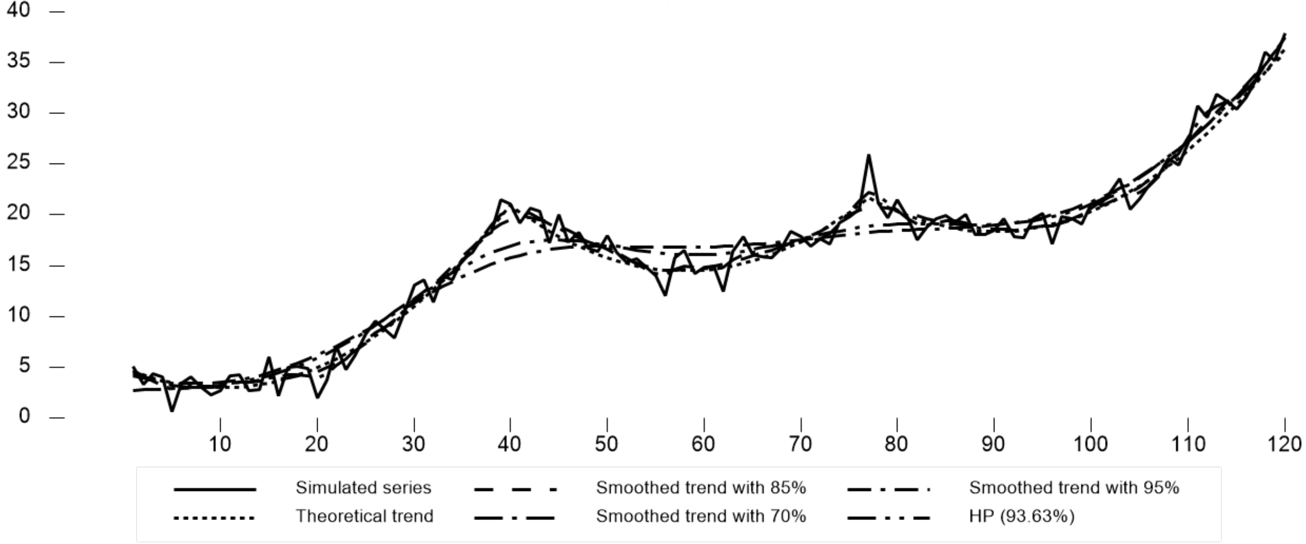
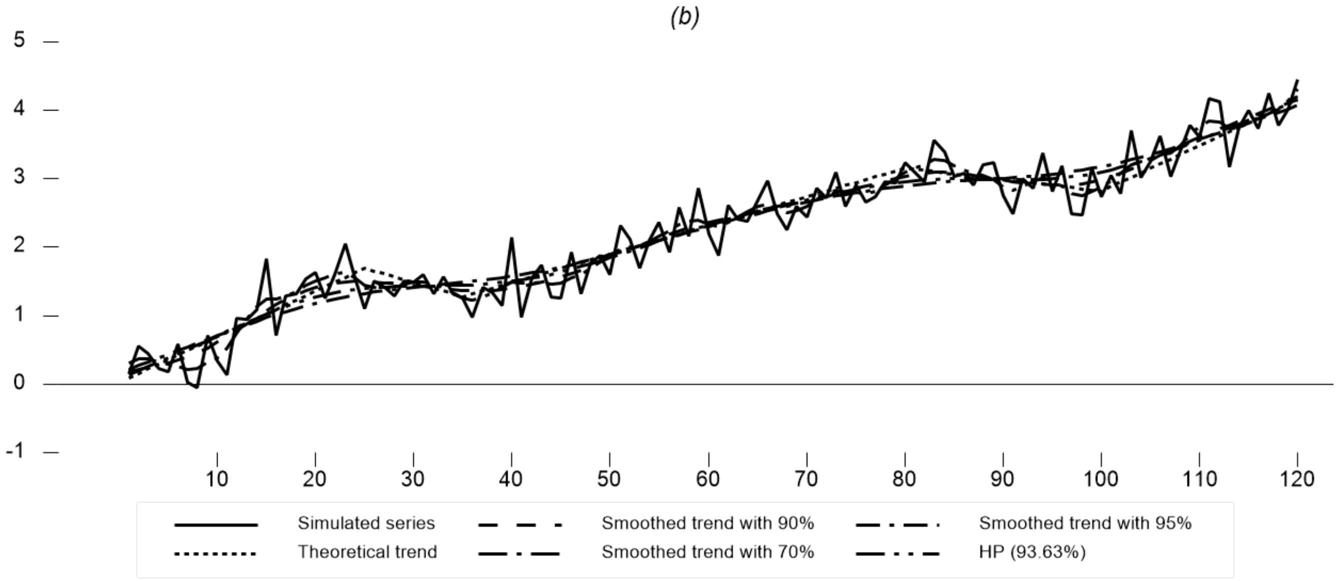
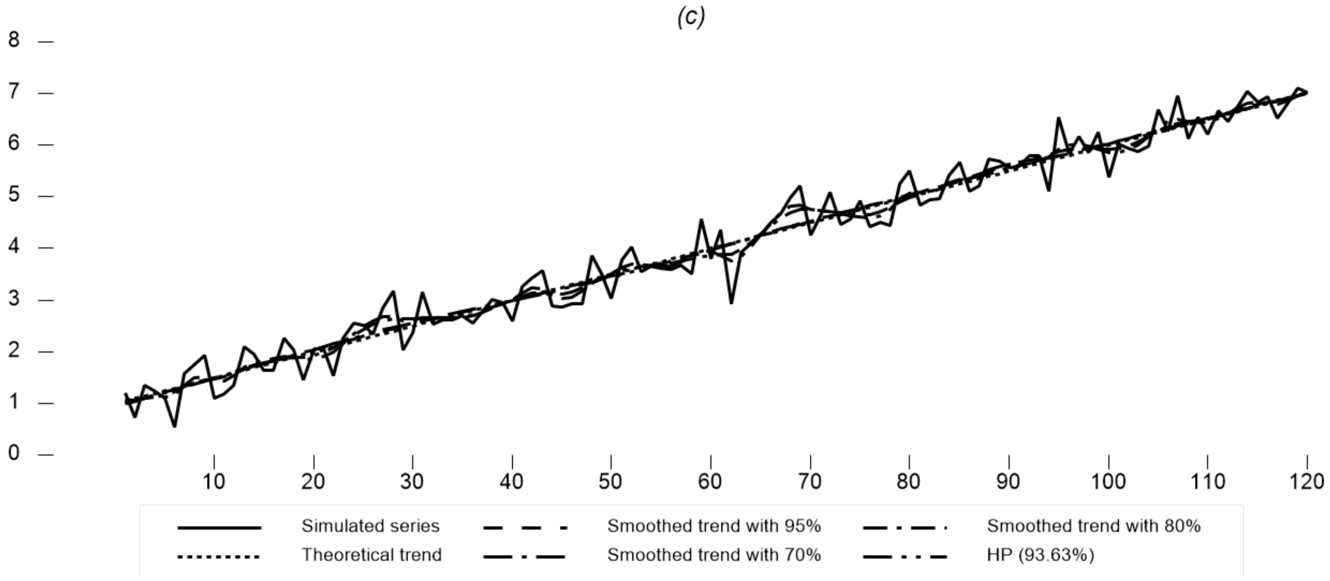

(a)

Figure 3.2.2 (*continuation*)

Examples of simulated time series with estimated trends





Source: own estimates

After analyzing the results achieved with different levels of smoothness, we can provide some guidance on selecting a suitable degree of smoothness in situations where the trend behavior can be identified visually. Guerrero et al. (2017) derived some guidelines based on a simulation study, which can help determine an appropriate percentage of smoothness. Therefore, we recommend the following:

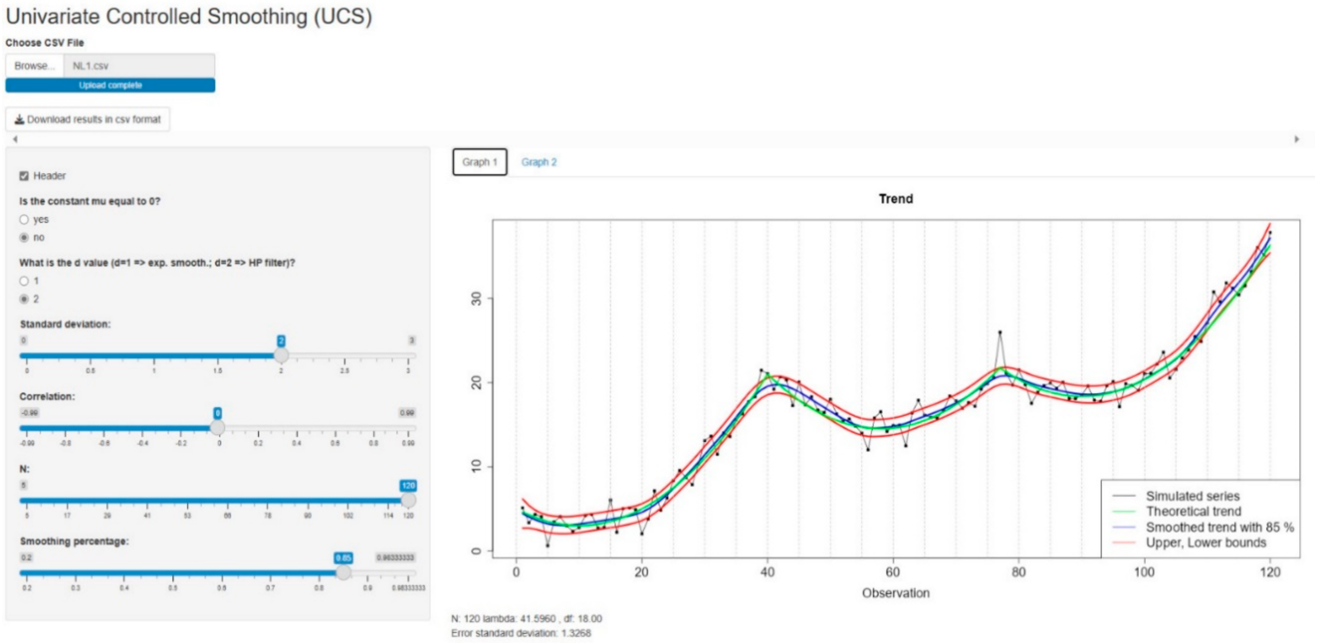
i. if the original series behaves as a straight line, choose a large value of $100S(\lambda,N)\%$, starting from 92.5 to 95 % for $N \geq 48$, and increase it from 95 to 97.5 % for values of $N \geq 100$;
ii. when the series shows a non-straight-line pattern, the percentage of smoothness should start at 80 to 85% for $N \geq 48$, and increase it from 85 to 90 % for values of $N \geq 100$.

Let us use an example to illustrate the guidelines presented by Guerrero et al. (2017). In Figure 3.2.2, we see a simulated series with a sample size of N=120. For $Nonlinear_1$ (Panel (a)), a suggested percentage of smoothness of 85% corresponds to a smoothing constant of λ =41.60. For $Nonlinear_2$ (Panel (b)), a suggested percentage of smoothness of 90% corresponds to a smoothing constant of λ =227.92, while for the linear series (Panel (c)), a suggested percentage of smoothness of 95% corresponds to a smoothing constant of λ =5212.65. All these constants can be found in Table 3.2.1. Figure 3.2.3 shows the same simulated series as Figure 3.2.2, but with the suggested smoothness percentages and theoretical trends with ∓ 2 standard error bands centered at $\hat{\boldsymbol{g}}$. The standard errors are calculated by taking the square root of the diagonal elements of $Var(\hat{g})$ (see equation (3.2.3)).

The Univariate Controlled Smoothing (UCS) web-tool, developed by us, displays trend estimates, $\hat{\boldsymbol{g}}$, (Graph 1) and Cycle estimates (Graph 2) once data in .csv format for univariate time series (with or without header) have been uploaded. To the left of the UCS tool, the analyst selects parameters related to the filter she wants to use, such as: i) whether the constant (μ) is null or not, (ii) difference order (d) (1 or 2); (iii) number of standard deviations (SD) for the interval estimates (0, ±1, ±2 or ±3); (iv) Correlation (ρ) (-1,1) (see, section 3.3); (v) the selected length of data (by default N, although it could be modified); and (vi) smoothing percentage, $S\%$, (0.2, 0.9833).

The steps to estimate the trend, cycle, and error bands were the following. First, we chose and uploaded the data to be analyzed and explore its trend behavior visually. In this case, the $Nonlinear_1$ simulated series was selected. Two significant observations can be made: (i) the *Nonlinear*$_1$ series does not follow a global trend, and the concept of stochastic local trend seems better to describe the underlying dynamics of the trend, so we chose $\mu \neq 0$; (ii) the series shows a non-straight-line pattern. Second, following Guerrero's (2017) guideline, a suggested smoothness percentage of 85% was set to select the corresponding smoothing parameter, which, for a sample size of $N=120$, corresponds to $\lambda=41.60$; $d=2$, $\rho=0$, and two standard deviations for the bands, were also selected. After configuring the parameters, Graph 1 (Figure 3.2.2(a)) displays the estimated trend ($\hat{\boldsymbol{g}}$) with ∓2 standard error bands in the UCS output, while Graph 2 illustrates the estimated cycle component ($y_t - \hat{g}_t$). Finally, once the analyst has finished, the results can be downloaded in .csv format for additional analyses, by just pushing the top-right button ("Download results in .csv format").

We estimated the trend of the series *Nonlinear*$_2$ and *linear* in similar manner.



Univariate Controlled Smoothing (UCS)

Choose CSV File

Browse... NL1.csv

Upload complete

Download results in csv format

☑ Header

Is the constant mu equal to 0?

○ yes

● no

What is the d value (d=1 => exp. smooth.; d=2 => HP filter)?

○ 1

● 2

Standard deviation: 2

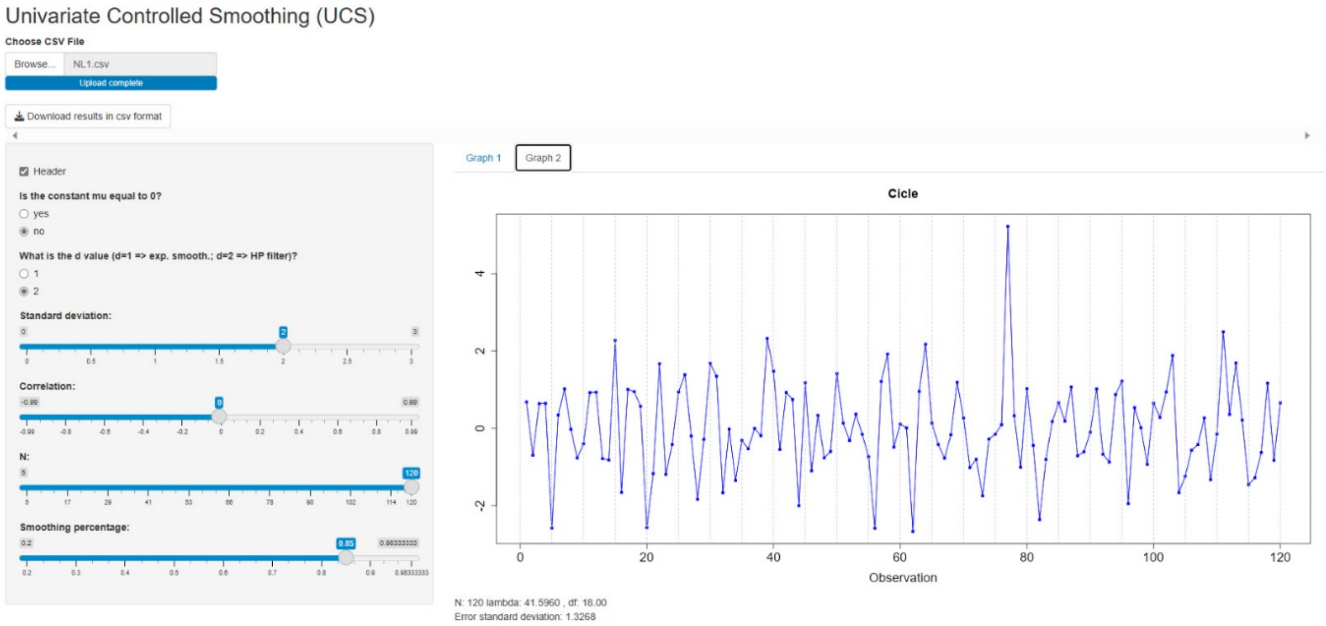
Correlation: 0

N: 120

Smoothing percentage: 0.85

Graph 1
Graph 2
0
3
0
0.5
1
1.5
2
2.5
3
-0.99
0.99
-0.99
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
0.99
5
120
5
17
29
41
53
66
78
90
102
114
120
0.2
0.98333333
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.98333333
Trend
0
10
20
30
0
20
40
60
80
100
120
Observation
Simulated series
Theoretical trend
Smoothed trend with 85 %
Upper, Lower bounds

N: 120 lambda: 41.5960 , df: 18.00
Error standard deviation: 1.3268



Univariate Controlled Smoothing (UCS)

Choose CSV File

Browse... NL1.csv

Upload complete

Download results in csv format

☑ Header

Is the constant mu equal to 0?

○ yes

● no

What is the d value (d=1 => exp. smooth.; d=2 => HP filter)?

○ 1

● 2

Standard deviation: 2

Correlation: 0

N: 120

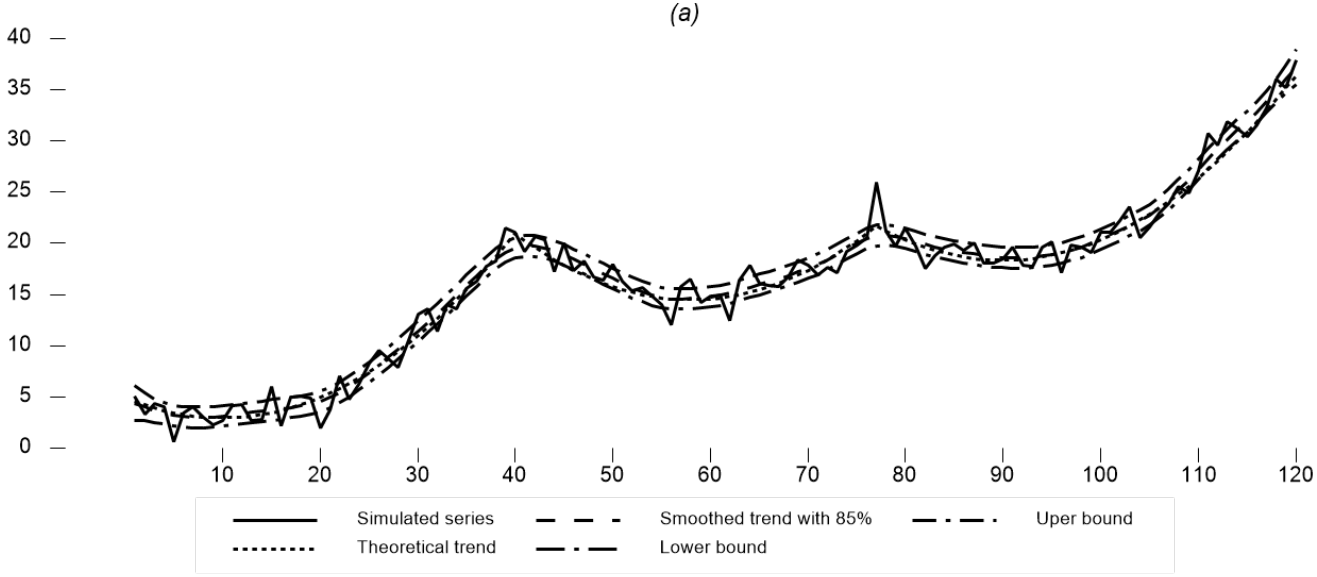
Smoothing percentage: 0.85

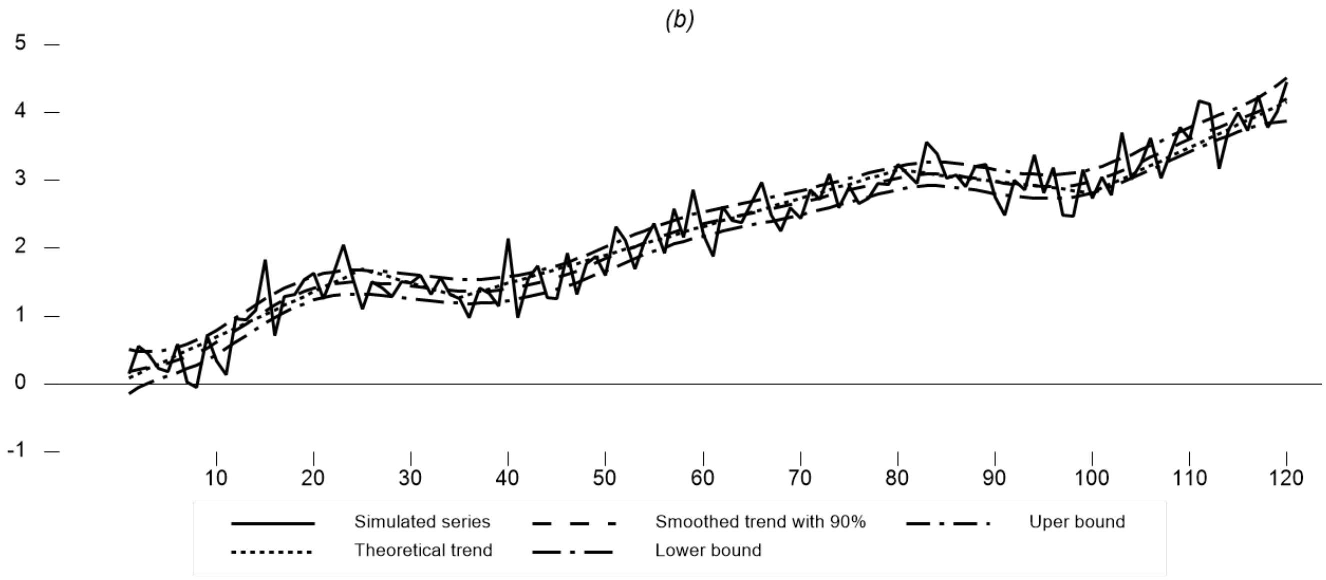
N: 120 lambda: 41.5960 , df: 18.00
Error standard deviation: 1.3268

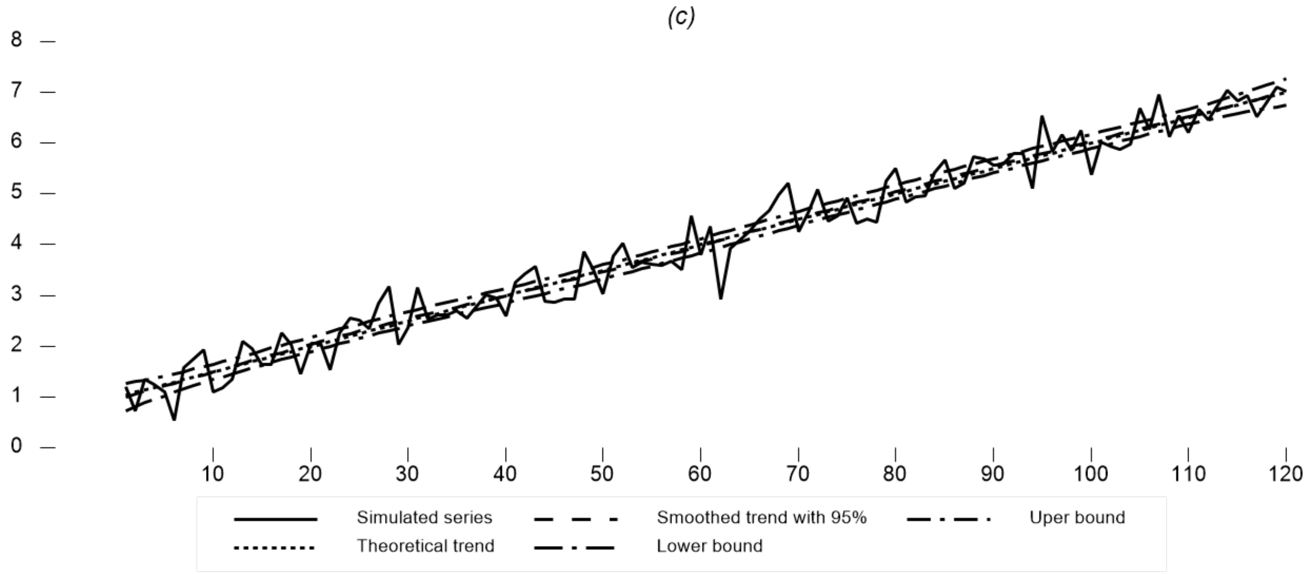
Graph 1
Graph 2
0
3
0
0.5
1
1.5
2
2.5
3
-0.99
0.99
-0.99
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
0.99
5
120
5
17
29
41
53
66
78
90
102
114
120
0.2
0.98333333
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.98333333
Cicle
4
2
0
-2
0
20
40
60
80
100
120
Observation

Figure 3.2.3

Examples of simulated time series with estimated trends based on Guerrero's (2017) guideline for fixing a smoothness percentage. Panels (a) and (b) show the nonlinear time series, while panel (c) shows the linear time series.




(b)
5
4
3
2
1
0
-1
10
20
30
40
50
60
70
80
90
100
110
120
Simulated series
Smoothed trend with 90%
Uper bound
Theoretical trend
Lower bound



Source: own estimates.

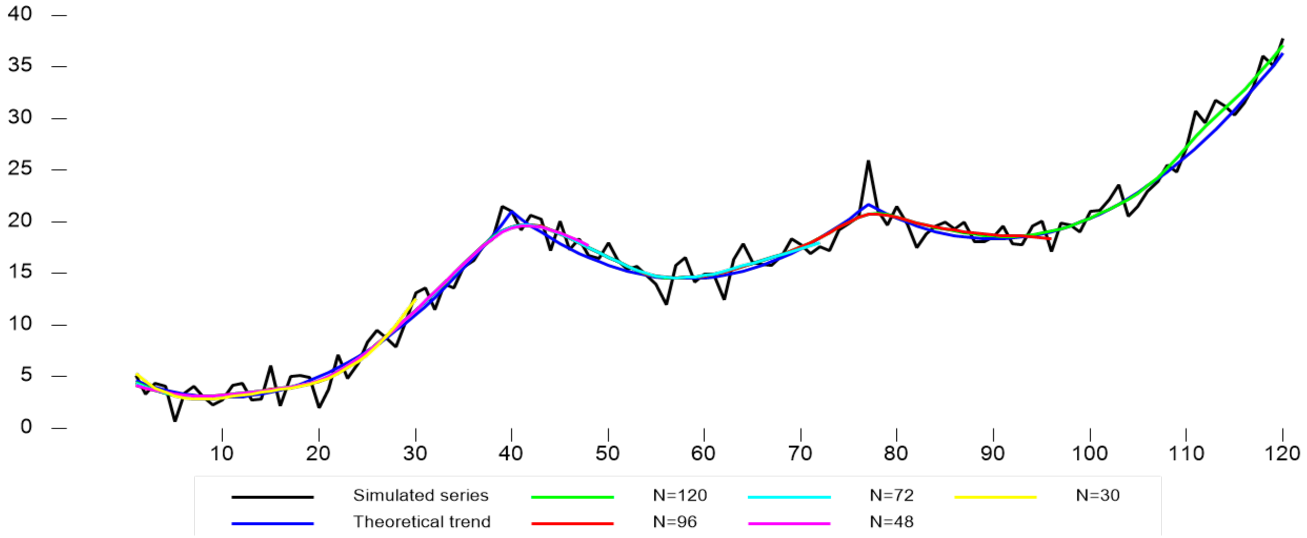
It is clear from Figure 3.2.2 that the estimated trend based on Guerrero's (2017) guidelines provides the most accurate approximation of the theoretical trend compared to all other estimated trends. The smoothness recommended by Guerrero's guideline is particularly effective.

3.2.3.1. Same time series with different sample sizes

It is important to note that trends with the same level of smoothness can be compared even if the sample sizes are different. To illustrate this, let us look at the examples in Figure 3.2.4 ($Nonlinear_1$), where trends were estimated for the same series with varying sample sizes. The λ values used for 85% smoothness were: 41.60 for $N=120$, 44.00 for $N=96$, 48.48 for $N=72$, 59.13 for $N=48$, and 87.47 for $N=30$. From Figure 3.2.4, we can see that the trend estimates are reasonably similar, especially when the sample sizes are comparable.

Figure 3.2.4

Selected trend estimates with 85% smoothness and different sample sizes for the simulated *Nonlinear*$_1$ series

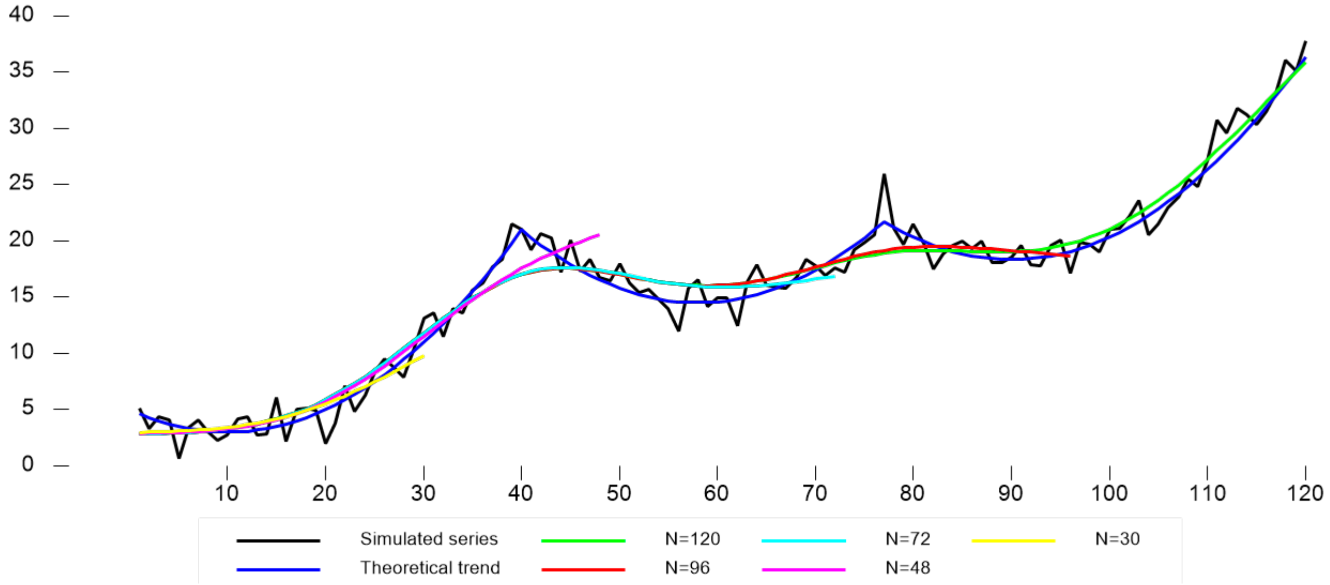


Source: own estimates

We conducted a similar analysis with the HP filter. However, we kept the smoothing parameter $\lambda = 1600$ constant across all sample sizes. Despite criticisms of the HP filter for generating spurious cycles and the ad hoc selection of the smoothing parameter without considering sample size, we found that the trends estimates were not comparable. By fixing $\lambda = 1600$, we observed 93.56% smoothness for $N=120$, 93.35% for $N=96$, 93% for $N=72$, 92.31% for $N=48$, and 91.06 for $N=30$, as shown in Figure 3.2.5.

Figure 3.2.5.

Trend estimates for λ = 1600 and different sample sizes

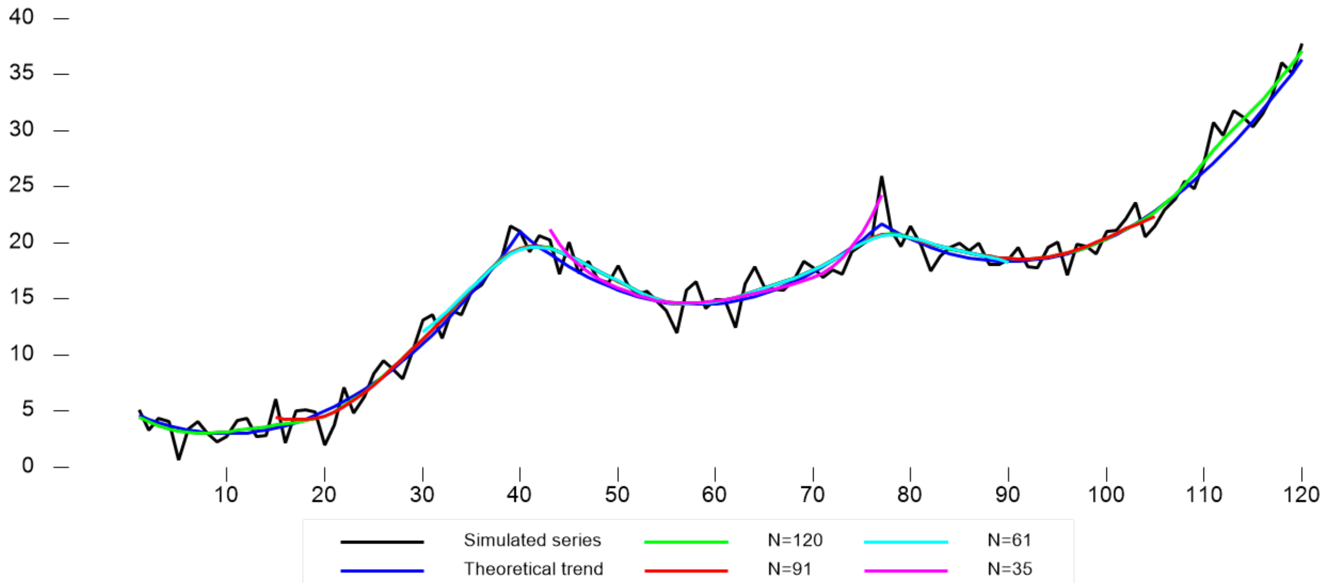


Source: own estimates

Figure 3.2.6 confirms that the trend remains relatively stable despite changes in sample size and a fixed percentage of smoothness. The trend values show only a slight shift when we zoom in on the central observations of the 120 original observations. However, the central 35 observations show a larger difference due to structural changes that occurred in observations 40 and 77. The λ values used for 85% smoothness were: $\lambda = 74.96$ for $N=35$, $=51.97$ for $N=61$, and $\lambda = 44.69$ for $N=91$.

Figure 3.2.6

Trend with 85% smoothness



Source: own estimates

Based on the examples above, we can see that when we set the same specific percentage for the smoothness of a trend for each series being studied, we will obtain almost the same estimated trend when we apply the same procedure to another time series of the same variable but with more or fewer data points compared to the previous one. Moreover, we can accurately compare the estimated trend of time series for different variables, even with different sample sizes, by using the same amount of smoothness. This is similar to fixing the confidence level of various confidence intervals to ensure valid comparisons.

Therefore, the estimated trend will not be significantly impacted by the length of the time series if we control smoothness. This is beneficial when estimating trends on a massive scale, as trend revisions can be minimized when more data are acquired. The key is to maintain a constant percentage of smoothness.

3.3. Penalized Least Squares of second-order integrated processes with the percentage of smoothness chosen by the user in the presence of noise autocorrelation

It is widely known that the accuracy and smoothness of the trend component in smoothing techniques depend largely on the chosen smoothing parameter value, which varies for each technique. To avoid the need for trial-and-error parameter selection, several data-driven methods have been developed to assist researchers. However, if the presence of autocorrelation in the noise component is ignored, commonly used automatic tuning parameter selection methods such as Cross Validation (CV), Generalized Cross Validation (GCV) (Craven & Wahba, 1979), Akaike's Information Criterion (AIC) (Akaike, 1973), corrected Akaike's Information Criterion (Hurvich et al. 1988), and Bayesian Information Criterion (BIC) (Schwarz, 1978) among others, fail to work and could cause overfitting of the data.

According to Krivobokova & Kauermann's (2007), smoothing with correlated errors is a significant issue in time series settings, such as macroeconomic time series. Their study, which focused on penalized splines for autocorrelated data, assumes a normal error distribution and uses Restricted Maximum Likelihood (REML) to estimate the smoothing parameter. Through simulations, they demonstrated that REML's smoothing parameter choice is more reliable than automatic methods and that moderate autocorrelation structures do not result in over or under-adjustment. The study also considers problem (3.1.1), which deals with both trend adjustment to the data and trend smoothness. Over (under)-adjustment occurs when an inadequate low (high) value of λ results in less (more) weight than necessary on smoothness. The authors assume an autoregressive structure of order 1, or AR(1), for the errors with an autocorrelation coefficient between 0.0 and 0.4.

Krivobokova & Kauermann's (2007) research on smoothness suggests that accurately estimating trends is possible when adjustments are not excessive or insufficient, even if there is moderate misspecification of autocorrelation. However, their simulation study only considered autocorrelations ranging from 0 to 0.4.

To avoid overfitting, it's best to consider the correlation structure when selecting a smoothing parameter. This means taking into account the variance-covariance matrix of the noise component ($Var(\boldsymbol{v}) = \sigma_v^2 V$), which can help to account for correlated noise. In this section, we use the smoothness index to study and measure the effect of first-order autocorrelated noise and choose a smoothing parameter accordingly.

3.3.1. Smoothing parameter selection in presence of autocorrelation

The smoothing index for AR(1) noise can be derived by utilizing model (3.1.5)-(3.1.6) to represent the univariate time series. It is important to note that the vector $\boldsymbol{v}$ is assumed to be a function of the N-dimensional vector of serially uncorrelated errors $\boldsymbol{v} = (v_1,\ldots, v_N)'$ in this particular case, that is

$$v_t = \rho v_{t-1} + \zeta_t \quad \text{for } t = 1, 2, \ldots, N \text{ and } v_1 = \zeta_1, \text{where } |\rho| \leq 1 \tag{3.3.1}$$

with

$$E(\boldsymbol{v}) = \boldsymbol{0}_N, \qquad Var(\boldsymbol{v}) = \sigma_v^2 \boldsymbol{I}_N, \qquad E(\boldsymbol{v}\zeta') = 0 \;\text{ and }\; E(\zeta\boldsymbol{\varepsilon}') = 0.$$

Besides, the variance-covariance matrix is of the form (see Appendix)

$$Var(\boldsymbol{v}) = \sigma_v^2 V = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \dots & \rho^{N-2} \\ \rho & 1 & \dots & \rho^{N-2} \\ & \dots & & \\ \rho^{N-1} & \rho^{N-2} & \dots & 1 \end{pmatrix} \quad (3.3.2)$$

Then, the index of smoothness (3.1.13) that takes into account the presence of the AR(1) autocorrelation becomes

$$S(\lambda,\rho,N) = 1 - tr[V^{-1}(V^{-1} + \lambda K_2{}'K_2)^{-1}]/N \quad (3.3.3)$$

3.3.2. The impact of smoothness on the trend when the errors follow an AR(1) model

To assess the impact of model misspecification or correlated noise (v) on the sensitivity of the smoothing level, we obtained various values of $S(\lambda,\rho,N)$ as explained in (3.3.3) under different scenarios. Tables 3.3.1, 3.3.2, 3.3.3, 3.3.4 and 3.3.5 demonstrate the changes in the smoothness index by fixing the value of λ to achieve 70%, 80%, 82.5%, 85%, 87.5%, 90%, 92.5%, and 95% of smoothness for sample sizes of 20, 50, 100, 200, 300, and 500 and ρ values of -0.9, -0.8, -0.6, -0.2, 0.0, 0.2, 0.4, 0.6, 0.8, and 0.9. Since (3.3.3) relies on N, λ, and ρ, we first fixed the number of observations and then set the smoothness percentage (S%) assuming no autocorrelation ($\rho = 0$). After that, we calculated the value of λ corresponding to those indices. However, as mentioned before, there is no analytical solution for λ from (3.3.3), and therefore, it cannot be calculated directly. Therefore, the calculations were performed numerically from (3.3.3) by keeping the values of $\rho = 0$, N, and S% fixed. Finally, with the λ value fixed, we calculated the corresponding smoothness index directly from (3.3.3) for the different values of ρ.

Table 3.3.1

S% values as a function of λ for N=20 and different values of ρ

ρ	λ					
	4.663	**32.561**	**66.318**	**159.669**	**511.974**	**4999999999.9**
-0.9	0.652	0.760	0.790	0.821	0.853	0.900
-0.8	0.657	0.764	0.793	0.823	0.855	0.899
-0.6	0.669	0.773	0.801	0.829	0.859	0.899
-0.4	0.679	0.781	0.808	0.836	0.864	0.900
-0.2	0.690	0.790	0.816	0.842	0.869	0.900
0.0	0.700	0.800	0.825	0.850	0.875	0.900
0.2	0.708	0.809	0.833	0.857	0.880	0.899
0.4	0.716	0.819	0.842	0.865	0.885	0.899
0.6	0.721	0.829	0.852	0.873	0.889	0.900
0.8	0.721	0.839	0.861	0.880	0.893	0.900
0.9	0.718	0.842	0.865	0.883	0.894	0.899

Table 3.3.2

S% values as a function of λ for N=50 and different values of ρ

ρ	λ							
	3.163	**16.468**	**29.201**	**57.697**	**132.991**	**388.642**	**1721.221**	**19871.323**
-0.9	0.644	0.749	0.779	0.809	0.840	0.873	0.906	0.939
-0.8	0.651	0.755	0.783	0.813	0.844	0.875	0.907	0.940
-0.6	0.664	0.765	0.793	0.822	0.851	0.881	0.911	0.942
-0.4	0.677	0.777	0.803	0.830	0.858	0.887	0.915	0.945
-0.2	0.689	0.788	0.814	0.840	0.883	0.893	0.920	0.947
0.0	0.700	0.800	0.825	0.850	0.875	0.900	0.925	0.950
0.2	0.709	0.811	0.836	0.860	0.883	0.907	0.930	0.952
0.4	0.716	0.824	0.848	0.871	0.893	0.915	0.935	0.954
0.6	0.720	0.836	0.860	0.883	0.904	0.924	0.942	0.957
0.8	0.719	0.847	0.873	0.896	0.916	0.934	0.950	0.958
0.9	0.714	0.851	0.878	0.902	0.923	0.940	0.953	0.959

Table 3.3.3

S% values as a function of λ for N=100 and different values of ρ

ρ	λ							
	2.812	**13.506**	**23.111**	**43.505**	**93.519**	**244.872**	**887.692**	**6132.677**
-0.9	0.641	0.746	0.775	0.805	0.836	0.868	0.901	0.935
-0.8	0.648	0.751	0.780	0.810	0.840	0.872	0.904	0.936
-0.6	0.662	0.763	0.791	0.819	0.848	0.878	0.908	0.939
-0.4	0.676	0.775	0.801	0.829	0.856	0.885	0.913	0.943
-0.2	0.688	0.787	0.813	0.839	0.865	0.892	0.919	0.946
0.0	0.70	0.80	0.825	0.850	0.875	0.900	0.925	0.950
0.2	0.709	0.812	0.837	0.861	0.885	0.908	0.931	0.953
0.4	0.716	0.825	0.849	0.873	0.895	0.917	0.938	0.958
0.6	0.720	0.837	0.863	0.886	0.908	0.927	0.946	0.963
0.8	0.717	0.849	0.876	0.900	0.921	0.940	0.956	0.970
0.9	0.712	0.852	0.881	0.907	0.929	0.947	0.962	0.974

Table 3.3.4

S% values as a function of λ for N=200 and different values of ρ

ρ	λ							
	2.657	**12.287**	**20.684**	**38.092**	**79.384**	**198.265**	**662.841**	**6132.677**
-0.9	0.640	0.744	0.773	0.803	0.834	0.866	0.899	0.933
-0.8	0.647	0.750	0.778	0.808	0.838	0.870	0.902	0.935
-0.6	0.662	0.762	0.789	0.818	0.847	0.876	0.907	0.938
-0.4	0.675	0.774	0.801	0.828	0.855	0.884	0.912	0.942
-0.2	0.688	0.787	0.801	0.838	0.865	0.891	0.918	0.945
0.0	0.70	0.80	0.825	0.850	0.875	0.900	0.925	0.950
0.2	0.709	0.812	0.837	0.861	0.885	0.908	0.931	0.954
0.4	0.716	0.825	0.850	0.874	0.897	0.918	0.939	0.959
0.6	0.719	0.838	0.864	0.888	0.909	0.929	0.948	0.965
0.8	0.716	0.850	0.877	0.902	0.924	0.943	0.959	0.973
0.9	0.711	0.853	0.883	0.909	0.931	0.950	0.966	0.978

Table 3.3.5

S% values as a function of λ for N=300 and different values of ρ

ρ	λ							
	2.608	**11.914**	**19.950**	**36.483**	**75.288**	**185.273**	**603.786**	**3321.041**
-0.9	0.639	0.743	0.772	0.802	0.833	0.866	0.899	0.932
-0.8	0.647	0.749	0.778	0.807	0.838	0.869	0.901	0.934
-0.6	0.662	0.762	0.789	0.817	0.846	0.876	0.906	0.937
-0.4	0.675	0.774	0.800	0.827	0.855	0.883	0.912	0.941
-0.2	0.688	0.787	0.812	0.838	0.865	0.891	0.918	0.945
0.0	0.70	0.80	0.825	0.850	0.875	0.900	0.925	0.950
0.2	0.709	0.813	0.837	0.861	0.885	0.909	0.932	0.954
0.4	0.716	0.826	0.851	0.874	0.897	0.919	0.939	0.960
0.6	0.719	0.839	0.864	0.888	0.910	0.930	0.949	0.966
0.8	0.716	0.850	0.878	0.903	0.925	0.944	0.960	0.974
0.9	0.710	0.853	0.883	0.909	0.932	0.951	0.967	0.979

Table 3.3.6

S% values as a function of λ for N=500 and different values of ρ

ρ	λ							
	2.570	**11.626**	**19.387**	**35.259**	**72.205**	**175.651**	**561.489**	**2969.306**
-0.9	0.639	0.743	0.772	0.802	0.833	0.865	0.898	0.932
-0.8	0.647	0.749	0.777	0.807	0.837	0.869	0.901	0.933
-0.6	0.661	0.761	0.789	0.817	0.846	0.876	0.906	0.937
-0.4	0.675	0.774	0.800	0.827	0.855	0.883	0.912	0.941
-0.2	0.688	0.787	0.812	0.838	0.864	0.891	0.918	0.945
0.0	0.70	0.80	0.825	0.850	0.875	0.900	0.925	0.950
0.2	0.709	0.813	0.837	0.862	0.885	0.909	0.932	0.954
0.4	0.716	0.826	0.851	0.874	0.897	0.919	0.940	0.960
0.6	0.719	0.839	0.865	0.888	0.910	0.931	0.949	0.966
0.8	0.716	0.850	0.878	0.903	0.925	0.944	0.961	0.975
0.9	0.710	0.853	0.883	0.910	0.933	0.952	0.968	0.980

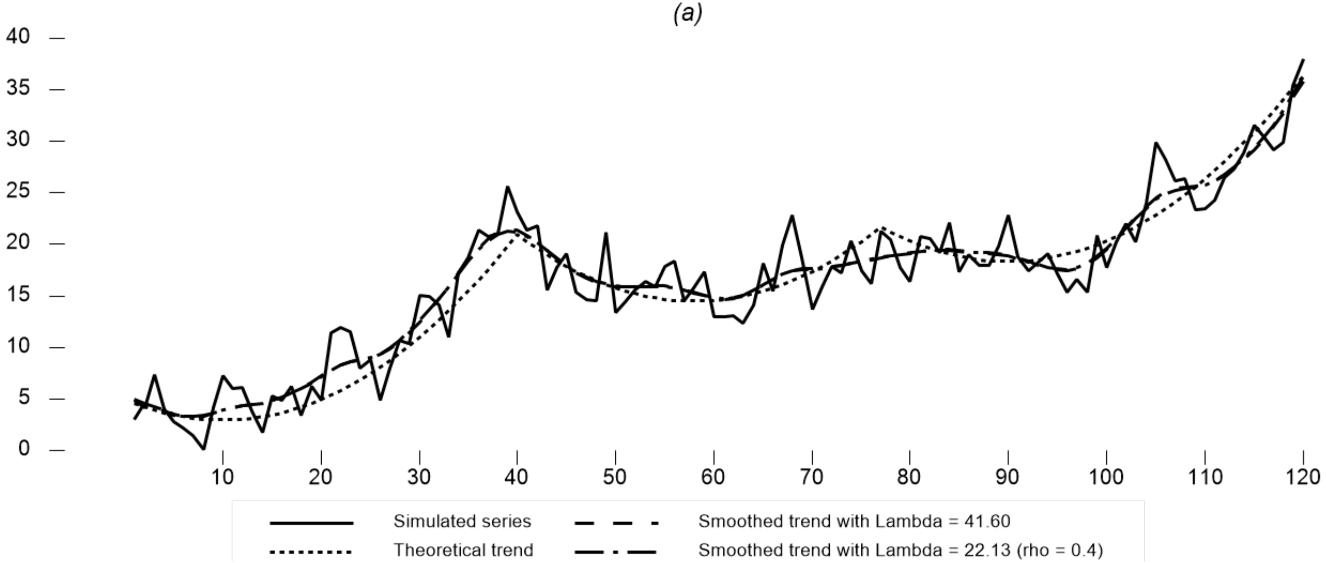
Based on the above tables, it can be generally observed that when there are negative autocorrelations, the smoothness achieved with the λ value corresponding to ρ equal to zero decreases as ρ becomes more negative, but increases as ρ becomes more positive. Additionally, when the values of ρ become more negative, there is a decrease in the percentage of smoothness as N increases; whereas for increasing values of ρ, the percentage of smoothness grows as N increases.

3.3.3. *A graphical analysis*

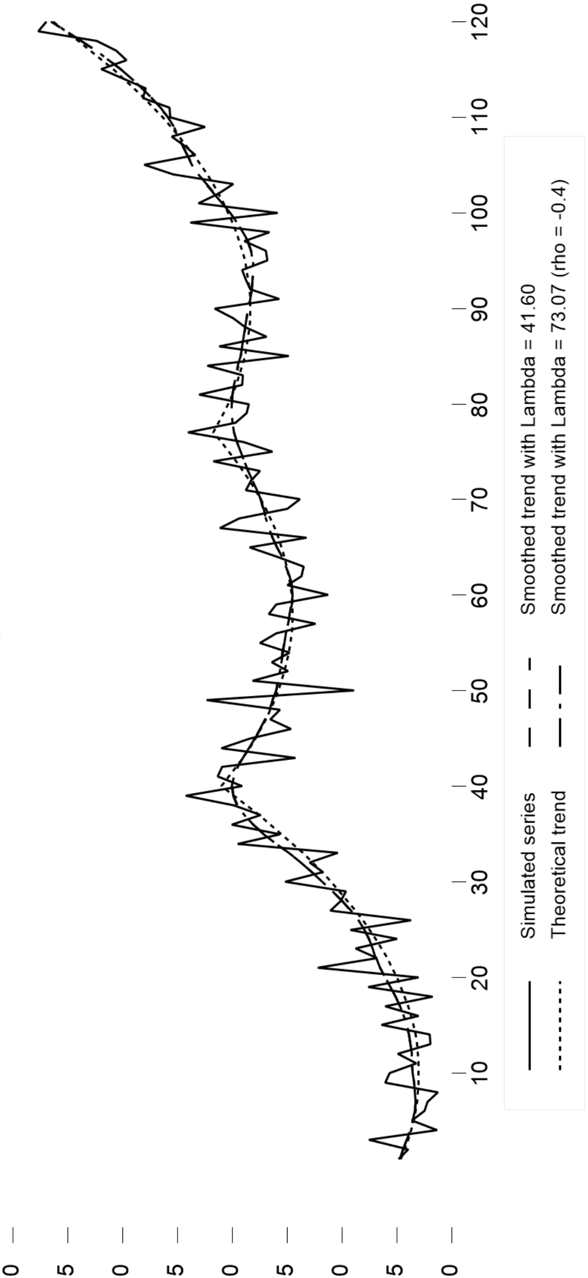
We now use the simulated series *Nonlinear*$_1$ and *Nonlinear*$_2$ from the simulation study in section 3.2, but now with autocorrelated noises. It is important to note that both series contain $N=120$ observations, and the errors are now of type AR(1) with autocorrelation coefficients of ∓0.4 and ∓0.9. To smooth out the two simulated time series (with and without autocorrelation), we used PLS with autocorrelated smoothness. Following Guerrero's (2017) guideline, we set the percentage of smoothness to 85% and 90% to choose the appropriate smoothing parameters based on the assumptions of $\rho = 0$ and $\rho \neq 0$, as determined by equation (3.3.3).

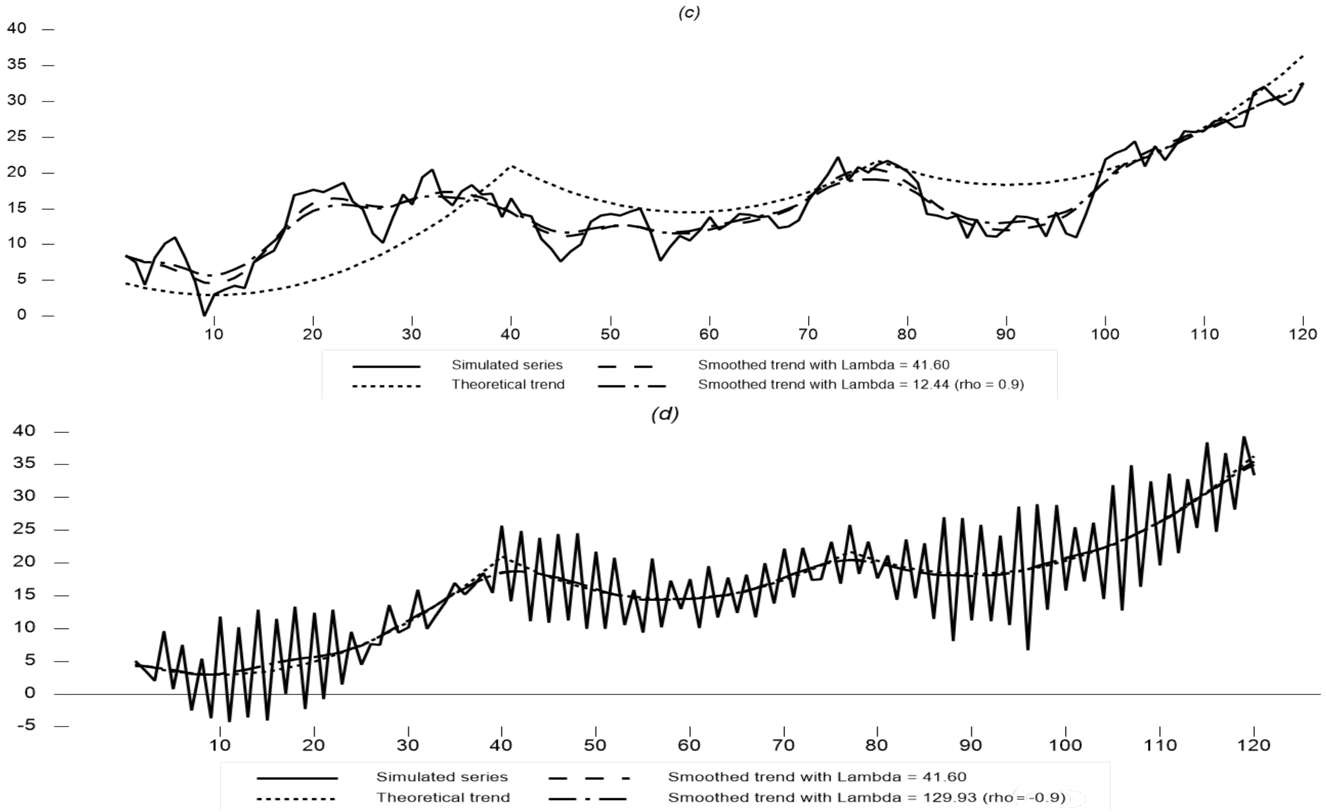
Figure 3.3.1

Time series simulated with theoretical piecewise quadratic trend and smoothed with 85% smoothness and respective autocorrelations: (*a*) ρ=0.4, (*b*) ρ=-0.4, (*c*) ρ=0.9, (*d*) ρ=-0.9



(b)

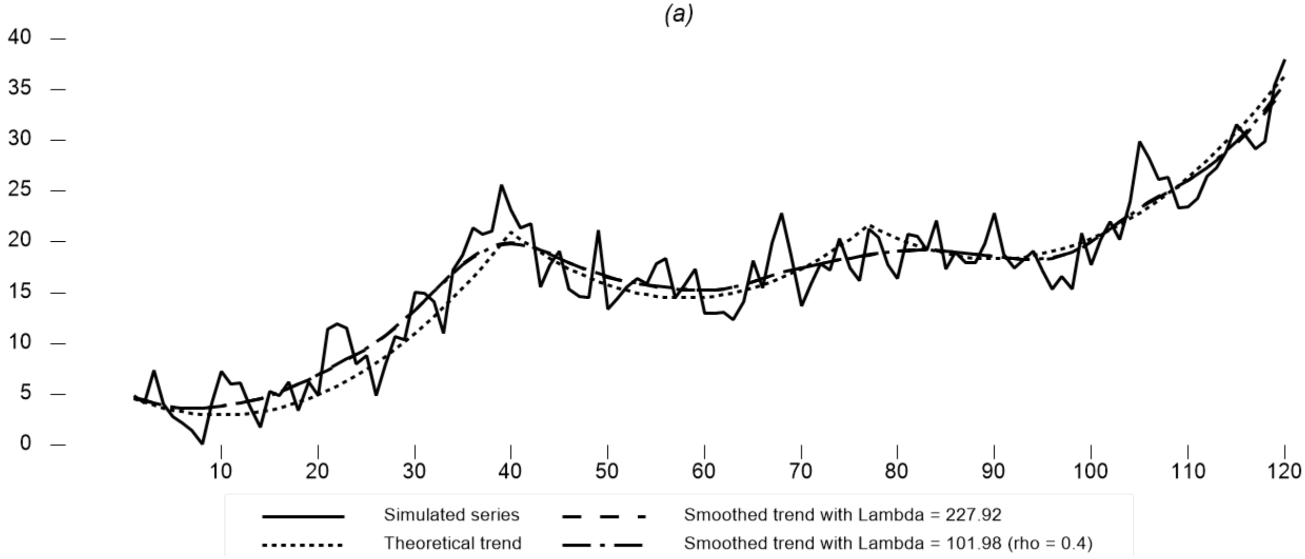

40
35
30
25
20
15
10
5
0
10
20
30
40
50
60
70
80
90
100
110
120
Simulated series
Theoretical trend
Smoothed trend with Lambda = 41.60
Smoothed trend with Lambda = 73.07 (rho = -0.4)

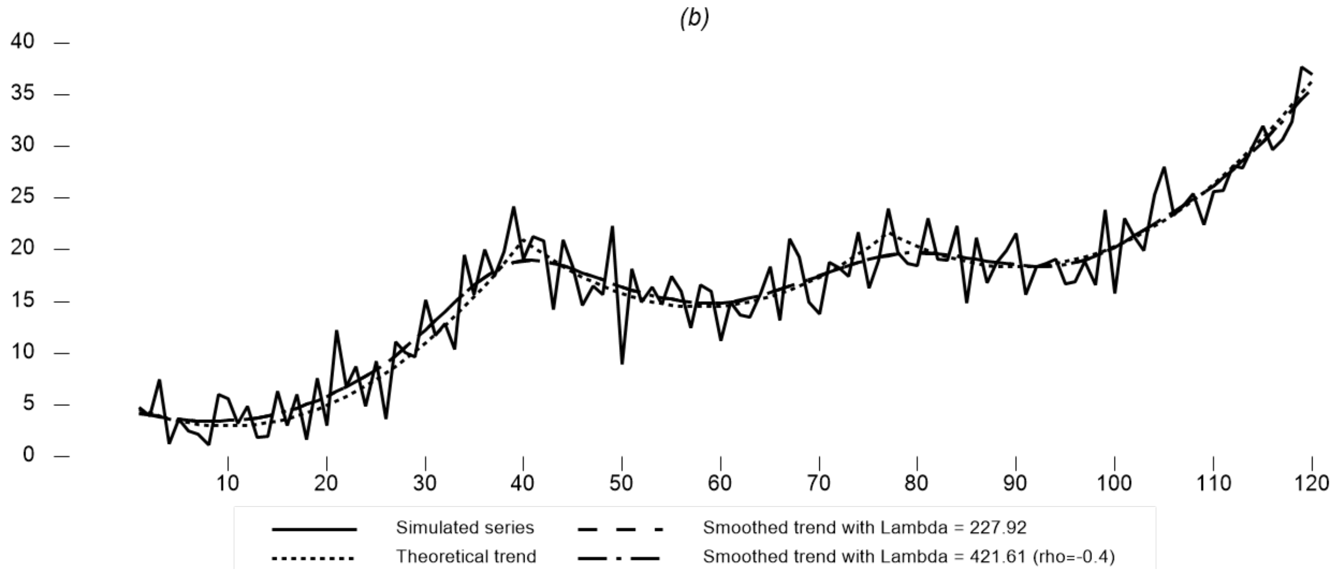


Source: own estimates

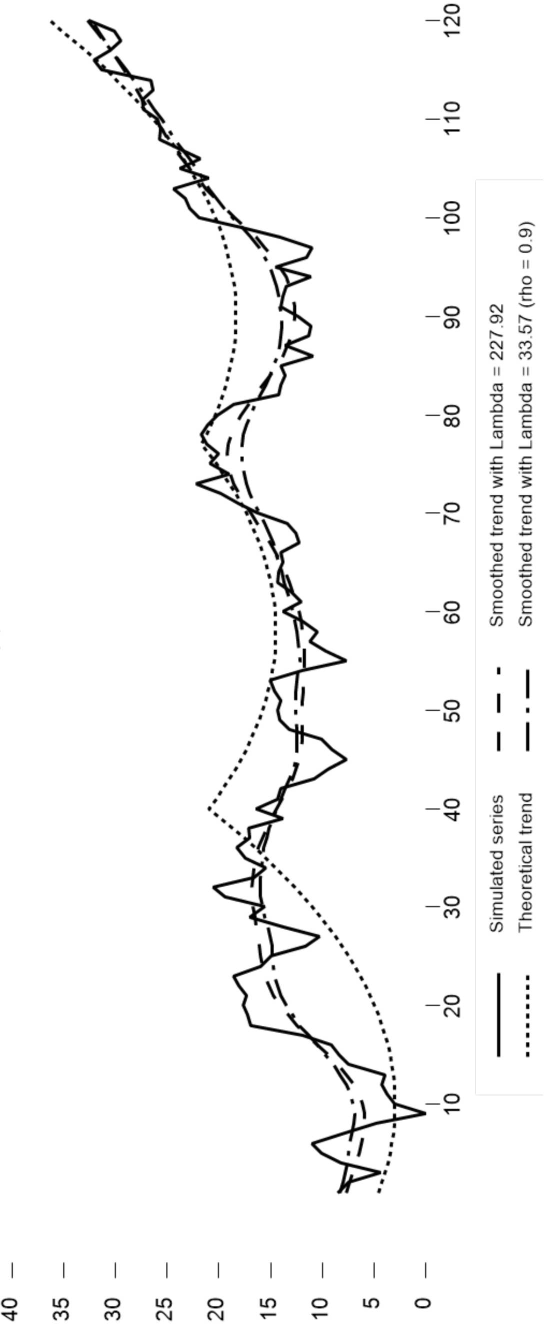
Figure 3.3.2

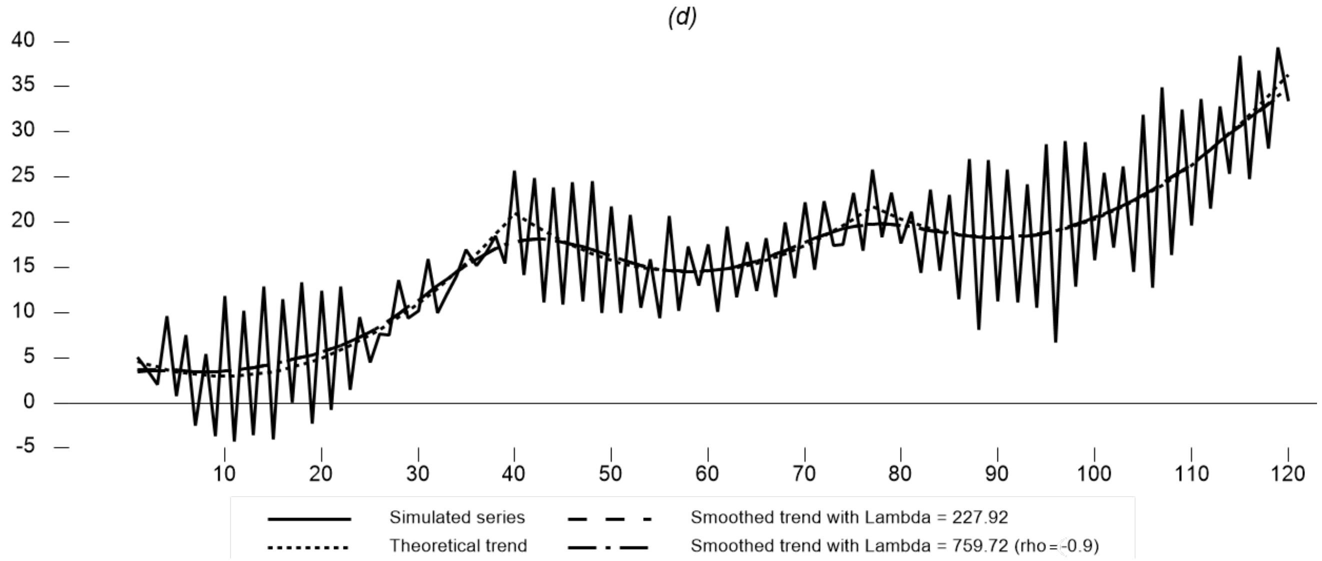
Time series simulated with theoretical piecewise quadratic trend and smoothed with 90% smoothness and respective autocorrelations: (*a*) ρ=0.4, (*b*) ρ=-0.4, (c) ρ=0.9, (d) ρ=-0.9




(b)
40
35
30
25
20
15
10
5
0
10
20
30
40
50
60
70
80
90
100
110
120
Simulated series
Theoretical trend
Smoothed trend with Lambda = 227.92
Smoothed trend with Lambda = 421.61 (rho=-0.4)

(c)

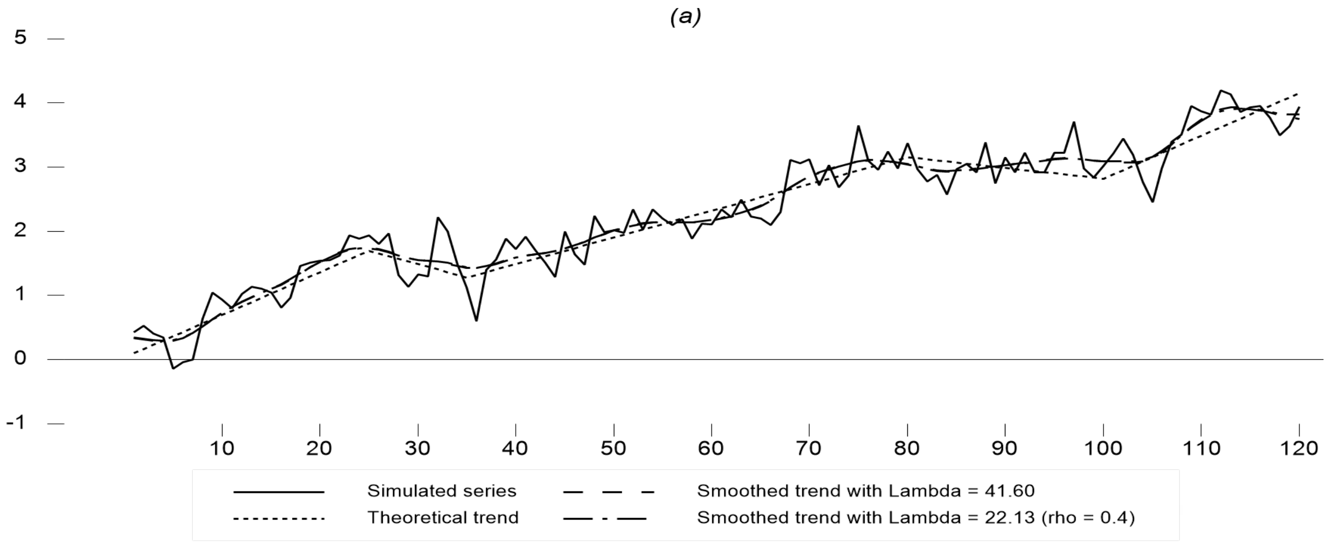

0
5
10
15
20
25
30
35
40
10
20
30
40
50
60
70
80
90
100
110
120
Simulated series
Theoretical trend
Smoothed trend with Lambda = 227.92
Smoothed trend with Lambda = 33.57 (rho = 0.9)

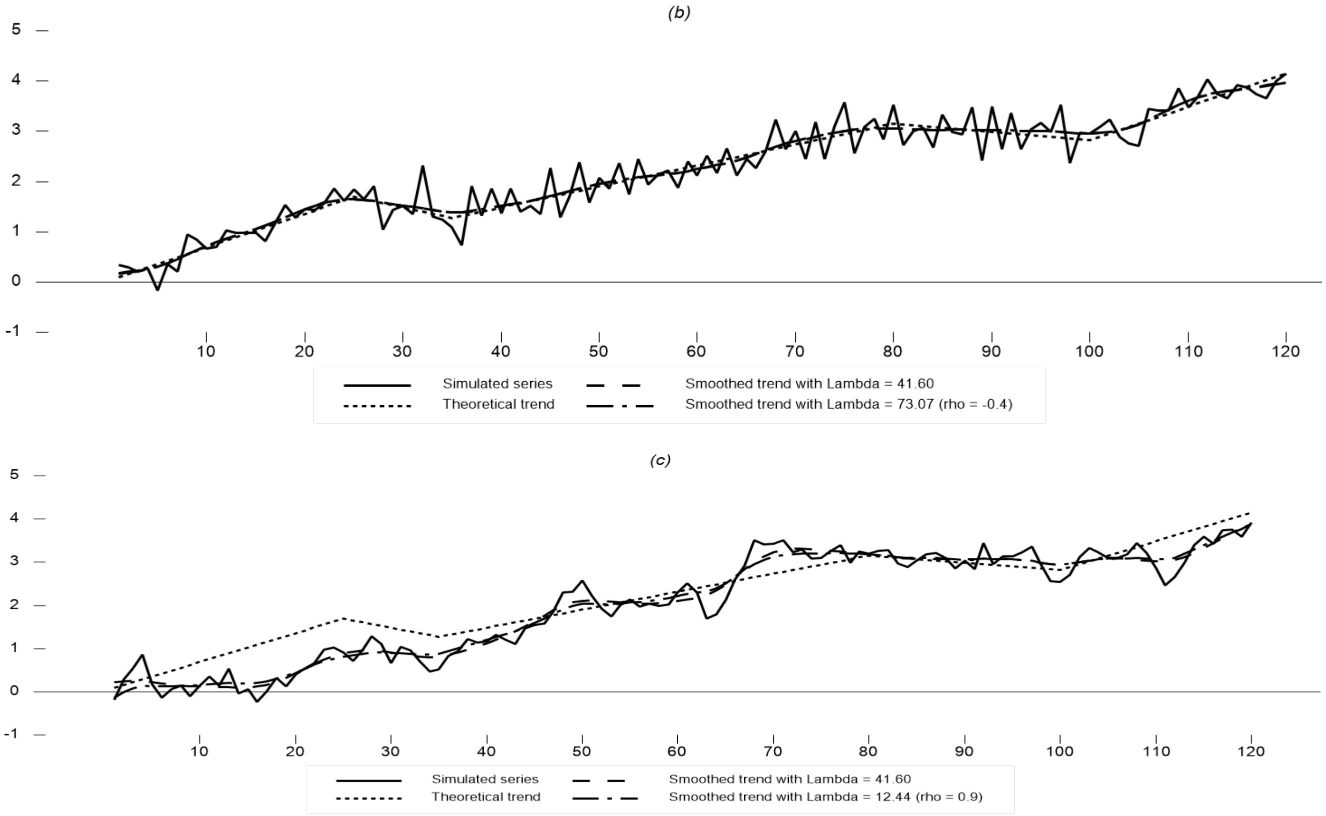


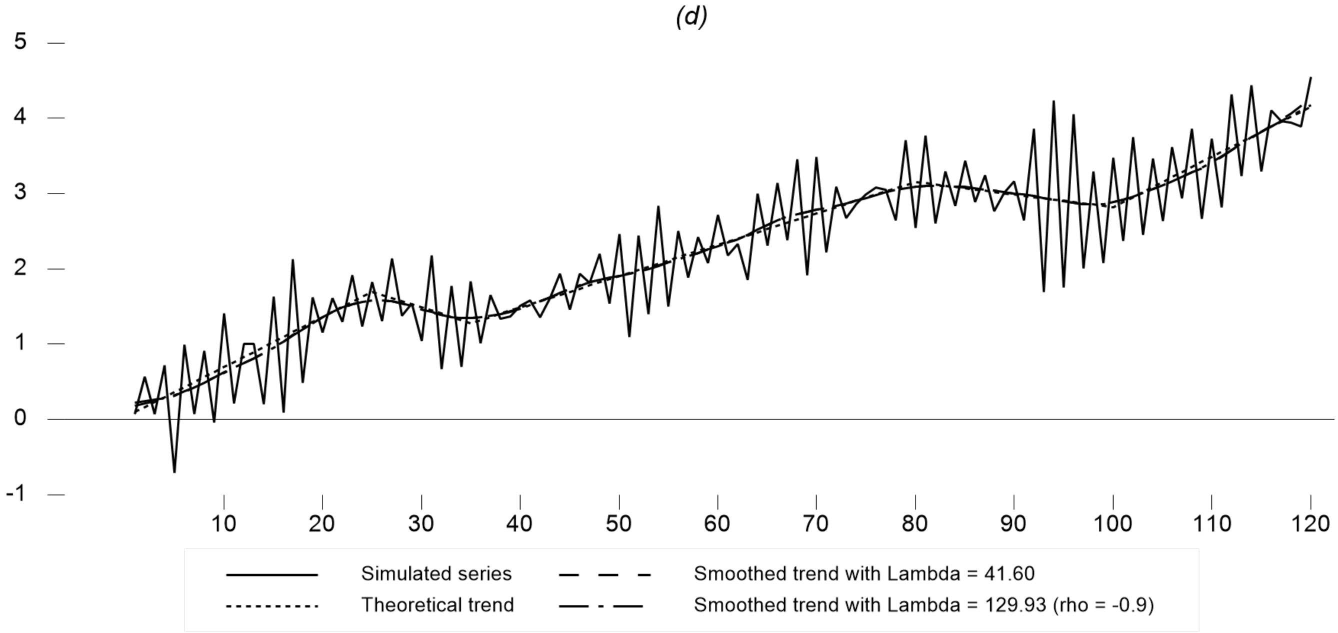
Source: own estimates

Figure 3.3.3

Time series simulated with theoretical piecewise linear trend and smoothed with 85% smoothness and respective autocorrelations: (*a*) ρ=0.4, (*b*) ρ=-0.4, (*c*) ρ=0.9, (*d*) ρ=-0.9



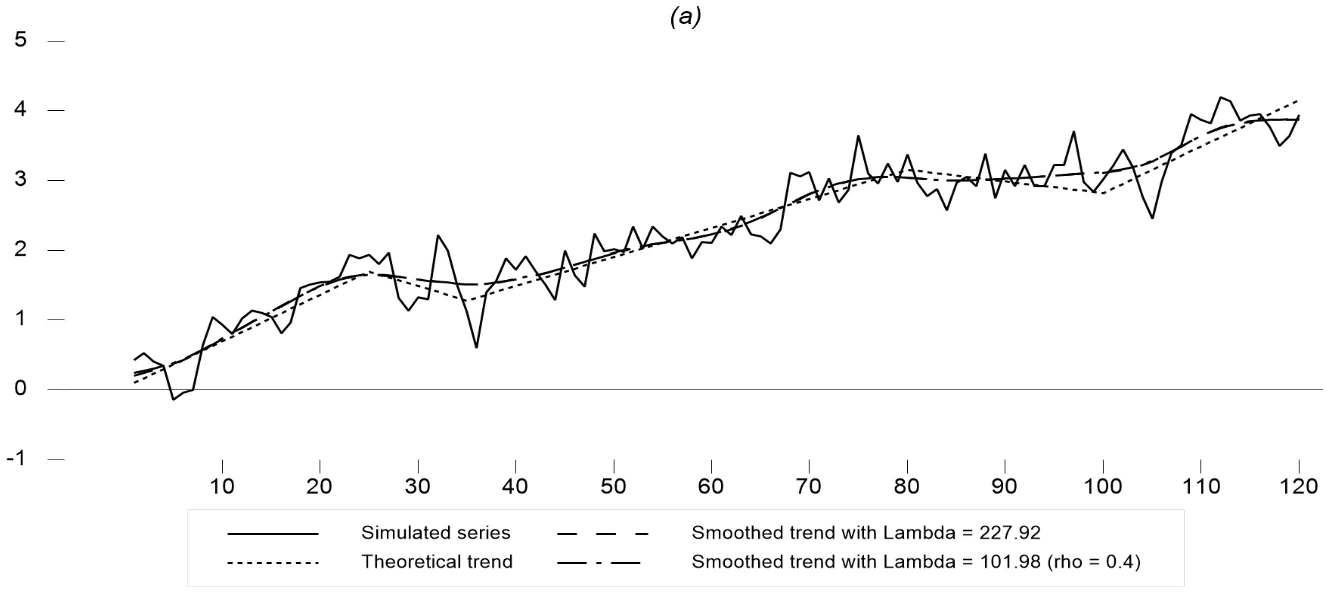

(b)
5
4
3
2
1
0
-1
10
20
30
40
50
60
70
80
90
100
110
120
Simulated series
Theoretical trend
Smoothed trend with Lambda = 41.60
Smoothed trend with Lambda = 73.07 (rho = -0.4)
(c)
5
4
3
2
1
0
-1
10
20
30
40
50
60
70
80
90
100
110
120
Simulated series
Theoretical trend
Smoothed trend with Lambda = 41.60
Smoothed trend with Lambda = 12.44 (rho = 0.9)



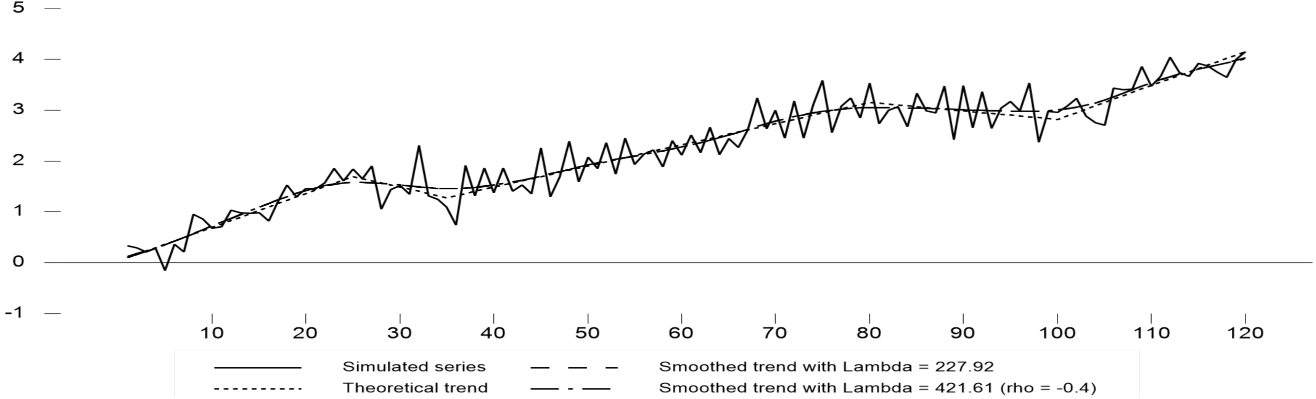
Source: own estimates

Figure 3.3.4

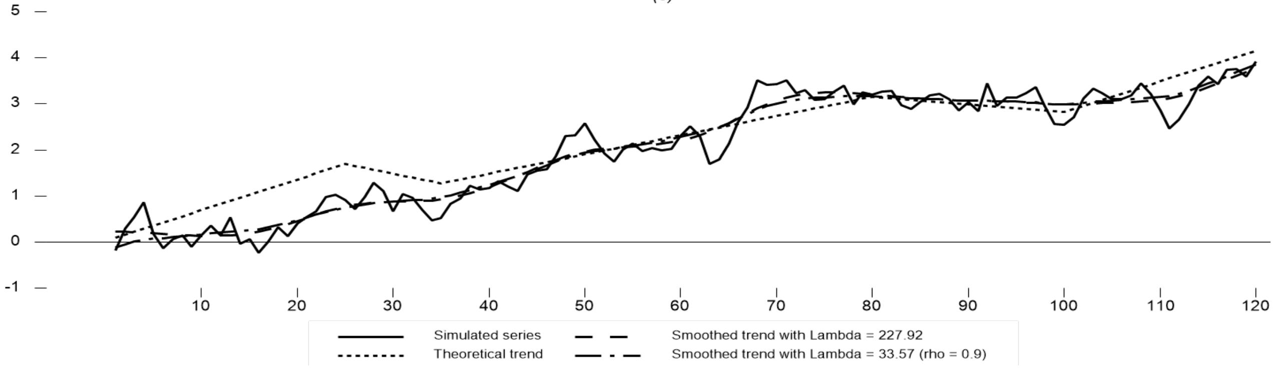
Time series simulated with theoretical piecewise linear trend and smoothed with 90% smoothness and respective autocorrelations: (*a*) ρ=0.4, (*b*) ρ=-0.4, (c) ρ=0.9, (*d*) ρ=-0.9

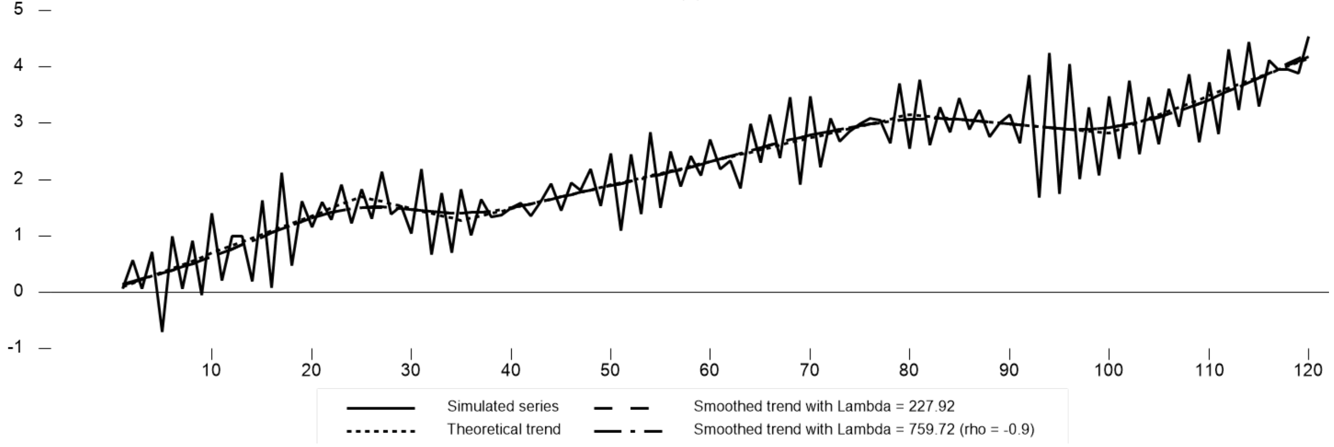


(b)


5
4
3
2
1
0
-1
10
20
30
40
50
60
70
80
90
100
110
120
Simulated series
Theoretical trend
Smoothed trend with Lambda = 227.92
Smoothed trend with Lambda = 421.61 (rho = -0.4)

(c)


5
4
3
2
1
0
-1
10
20
30
40
50
60
70
80
90
100
110
120
Simulated series
Theoretical trend
Smoothed trend with Lambda = 227.92
Smoothed trend with Lambda = 33.57 (rho = 0.9)



(d)

Source: own estimates

In the above-mentioned Figures 3.3.1, 3.3.2, 3.3.3, and 3.3.4, it can be observed that the estimated trends with and without autocorrelation are quite similar when the autocorrelation is ∓0.4 and -0.9. Additionally, when a high percentage of smoothness is considered for the trend estimation, panels (a), (b), and (d) of these figures indicate that the estimated trends with or without autocorrelation, came closer to the theoretical trend.

When the autocorrelation is positive and high (ρ=0.9), it can cause major problems. Panel (c) of Figures 3.3.1, 3.3.2, 3.3.3, and 3.3.4 shows that regardless of whether autocorrelation is considered or not, the estimated trend is far from the theoretical trend. The simulated series poorly represents the theoretical behavior, as it lies far from the theoretical trend in those figures. This has nothing to do with the estimated one. Therefore, it is suggested to avoid using this type of trend when there is a large positive autocorrelation in the noise component.

Chapter 4

Penalized Least Squares to first-order integrated processes with percentage of smoothness chosen by the user

It is commonly accepted that the efficient market hypothesis states that financial asset prices reflect all available information (Fama, 1970). This suggests that future returns cannot be predicted. However, since the 1990s, many researchers have challenged this hypothesis. They argue that risk premiums change over time and depend on the business cycle. Financial asset returns are linked to slow-moving economic variables that follow cyclical patterns according to the business cycle. Another reason to reject the efficient market hypothesis is that some agents are not entirely rational. This means that prices may underreact in the short term but overreact in the long run (Hong & Stein, 1977). Behavioral finance theory (Barberis & Thaler, 2002) can explain this phenomenon.

It is widely accepted that prices can demonstrate trends or cycles, as supported by the above two arguments. Financial time series are often assumed to behave as a random walk, or an integrated of order 1 (I(1)) process. Studies, such as Baillie & Bollerslev (1989), have shown that currency exchange rates against the US dollar behave as random walks. Additionally, Narayan & Smyth's (2005) study found that stock prices of OECD countries should be considered I(1) processes. Tsay's (2002) conventional model for prices also uses a random walk with drift. As such, the random walk model is crucial in financial time series. To decompose a financial time series into trend and noise, we opted for the exponential smoothing (ES) filter instead of the HP filter, aligning with the idea of consistency with the random walk model.

4.1. Statistical description of the exponential filter

The penalized approach that gives raise to the Exponential Filter (ES) postulates that the trend must minimize the function.

$$\min_{g_t}\left\{\sum_{t=1}^{N}(y_t-g_t)^2+\lambda\sum_{t=2}^{N}(g_t-g_{t-1}-\mu)^2\right\} \quad (4.1.1)$$

As in (3.1.2) and (3.2.1), $\lambda > 0$ is a constant that penalizes the lack of smoothness in the trend. That is, as $\lambda \to 0$ the trend resembles more closely the original data, so that $g_t \to y_t$ for all t, and no smoothness is achieved. The opposite occurs when $\lambda \to \infty$, in which case the trend follows essentially the (smooth) polynomial model $g_t - g_{t-1} = \mu$, which represents the trend growth component. As previously mentioned, λ has a significant impact on determining the smoothness, while μ serves as a reference level for the trend growth. It is important to note that the trend follows a first-degree polynomial, which can be expressed as:

$$g_t = \beta_0 + \mu t \qquad \mu \neq 0, \quad (4.1.2)$$

which becomes a constant when $\mu = 0$, so that using this reference level, as is usual in practice has important consequences on the trend behavior, particularly at the endpoints of the series, as discussed below.

Therefore, for $d=1$, $\mu \neq 0$ and $Var(v) = \sigma_v^2\boldsymbol{I}$, (3.1.8), (3.1.9), (3.1.11) and (3.1.12) becomes.

$$\widehat{\boldsymbol{g}} = (I_N + \lambda K_1{'}K_1)^{-1}(\boldsymbol{Y} + \lambda\mu K_1{'}\boldsymbol{1}_{N-1}) \quad (4.1.3)$$

$$\Gamma^{-1} = Var(\widehat{\boldsymbol{g}}) = \sigma_g^2(I_N + \lambda K_1{'}K_1)^{-1} \quad (4.1.4)$$

$$\widehat{\boldsymbol{g}} = (I_N + \lambda K_1{'}K_1)^{-1}(I_N + \lambda(N-1)^{-1}K_1{'}\boldsymbol{1}_{N-1}\boldsymbol{1}{'}_{N-1}K_1)\boldsymbol{Y} \quad (4.1.5)$$

$$\tilde{\sigma}_v^2 = [\boldsymbol{Y}'\boldsymbol{Y} - \widehat{\boldsymbol{g}}'(I_N + \lambda K_1{'}K_1)\widehat{\boldsymbol{g}}]/(N-1) + \lambda\mu^2 \quad (4.1.6)$$

where K_1 is the $(N\text{-}1)$xN matrix representation of the first difference operator appearing on the above formulas

$$K_1 = \begin{pmatrix} -1 & 1 & 0 & 0 \ldots 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \ldots 0 & 0 & 0 \\ & & & \ldots & & \\ 0 & 0 & 0 & 0 \ldots -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \ldots 0 & -1 & 1 \end{pmatrix} \quad (4.1.7)$$

And (3.2.6) becomes

$$\hat{\sigma}_v^2 = \left[\sum_{t=1}^{N} (Y_t - \hat{g}_t)^2 + \lambda \sum_{t=2}^{N} (\nabla \hat{g}_t - \hat{\mu})^2 \right] \Big/ (N-2) \quad (4.1.8)$$

As in the case $d = 2$, to appreciate the effect of the constant μ, it should be noticed that the array $K_1' \mathbf{1}_{N-1}$ appearing in (4.1.3) is an N-dimensional vector of zeros, except for the first and the last element, that is, $K_1' 1_{N-1} = (1,0,0,\ldots,0,0,1)'$. Therefore, the observed values of the original series $\{y_t\}$ enter the formula of the estimator $\hat{\boldsymbol{g}}$ modified in both of its extremes by the value of μ, weighted by λ. That is, (4.1.3) indicates applying the smoother matrix $(I_N + \lambda K_1' K_1)^{-1}$ to $\boldsymbol{Y} + \lambda\mu K_1' \mathbf{1}_{N-1} = (y_1 + \lambda\mu, y_2, y_3, \ldots, y_{N-2}, y_{N-1}, y_N + \lambda\mu)$, by doing this, the filter adjust the first and the last values of the series by means of $\lambda\mu$.

When analyzing daily financial data, there is typically a large number of observations compared to quarterly data analysis. As a result, calculating $\hat{\boldsymbol{g}}$ using (4.1.3) or (4.1.5) requires an NxN matrix inversion, which can lead to instability and imprecise solutions for large N. While the penalized approach clearly demonstrates the impact of λ, it is not an efficient calculation method.

To simplify the estimation procedure for the ES filter, we can rewrite the underlying minimization problem in a state-space form. This casting allows us to utilize the Kalman filter for parameter estimation with smoothing.

Kalman filtering requires the formulation of a state space model which in its general form has quite a few components: unobservable states, observable data, shocks and mapping matrices. The model is writing in the following form:

$$\boldsymbol{X}_t = A_t \boldsymbol{X}_{t-1} + \boldsymbol{Z}_t + F_t \boldsymbol{W}_t \quad (4.1.9)$$

$$Y_t = \varsigma_t + \boldsymbol{C_t}'\boldsymbol{X_t} + V_t \quad (4.1.10)$$

The $\boldsymbol{X}_t s$ are the (unobservable) state variables, $\boldsymbol{Y}_t$ are the observable data. (4.1.9) is known as the state equation while (4.1.10) is known as the measurement equation. The $\boldsymbol{Z}$, if present, are exogenous variables in the evolution of the state. The $\boldsymbol{W}_t$ are shocks to the states; the $\boldsymbol{F}_t$ matrix has the loadings from those shocks to states. The ς_t are any components in the description of the observable data which depend upon exogenous variables. The $\boldsymbol{V}_t$ are measurement errors. The $\boldsymbol{W}_t$ and $\boldsymbol{V}_t$ are assumed to be mean zero, normally distributed, independent across time and independent of each other at time *t* as well. For the ES filter, the model takes the form:

$$\boldsymbol{X}_t = g,\ A_t = 1,\ \boldsymbol{Z}_t = \mu,\ F_t = 1, \boldsymbol{W}_t = \varepsilon_t \quad and\ \boldsymbol{C}_t = 1, \varsigma_t = 0, \boldsymbol{V}_t = \nu_t$$

with $Var(\varepsilon_t) = \sigma_\varepsilon^2$ and $Var(\nu_t) = \sigma_\nu^2$. Thus, the state and measurement equations for the ES filter are:

$$g_t = g_{t-1} + \mu + \varepsilon_t \text{ and } Y_t = g_t + \nu_t.$$

Besides, λ is given by the variance ratio $\sigma_\nu^2/\sigma_\varepsilon^2$. Thus, to equate the results of the Kalman filter to those obtained with (4.1.3) or (4.1.5), it is assumed that $\sigma_\varepsilon^2 = 1$ and $\sigma_\nu^2 = \lambda$.

It's important to note that financial series may have missing values during holidays. However, we can still determine the trend using the available data. This is possible due to the use of the Kalman filter for estimation. We skip the filter step in the Kalman filter recursions to estimate trend values with missing data.

4.2. A measure of smoothness

The smoothness index (3.1.13) for the non-autocorrelated case becomes.

$$S(\lambda, N) = 1 - tr\,[(I_N + \lambda K'_1 K_1)^{-1}]/N \quad (4.2.1)$$

Just as when $d=2$, the index solely depends on the values λ and N since K_1 remains fixed. It is important to note that K_1 is a matrix of rank $N-1$. Therefore, the matrix K'_1K_1 has one eigenvalue equal to zero, while the remaining $N\text{-}1$ nonzero eigenvalues can be arranged in descending order as $e_1 \geq e_2 \geq \ldots \geq e_{N\text{-}1}$. Consequently, the expression in (4.2.1) for the trace can be expressed as:

$$tr(I_N + \lambda K'_2 K_2)^{-1} = (1 + \lambda e_1)^{-1} + \ \ldots + (1 + \lambda e_{n-1})^{-1} + 1$$

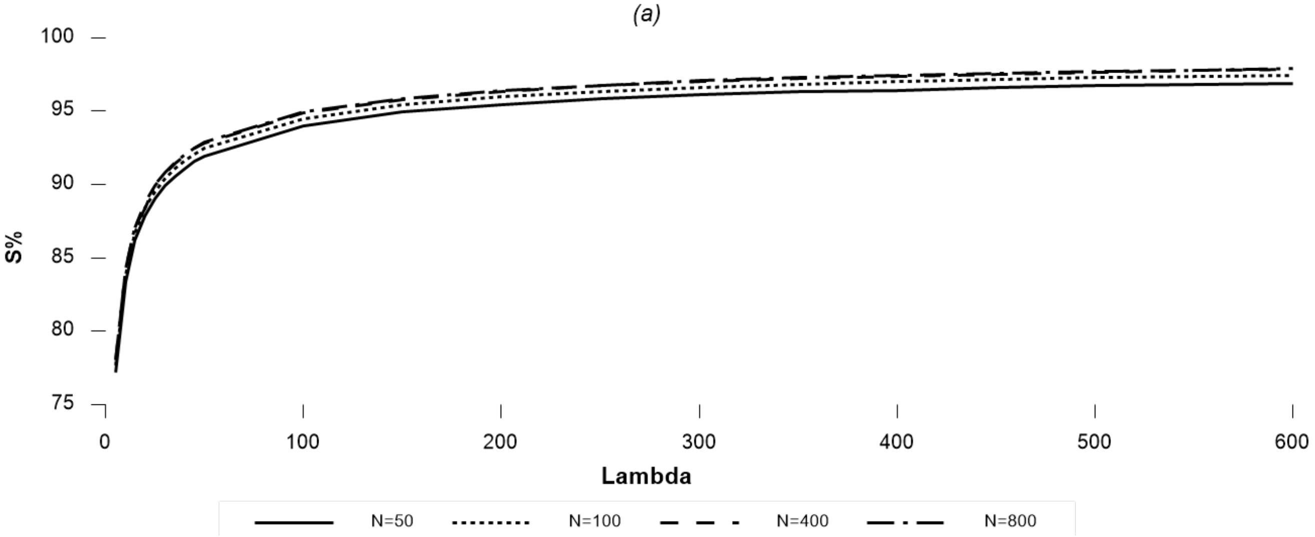
and it can be observed that $S(\lambda,N) \to 0$ as $\lambda \to 0$ and $S(\lambda,N) \to 1 - 1/N$ as $\lambda \to \infty$.

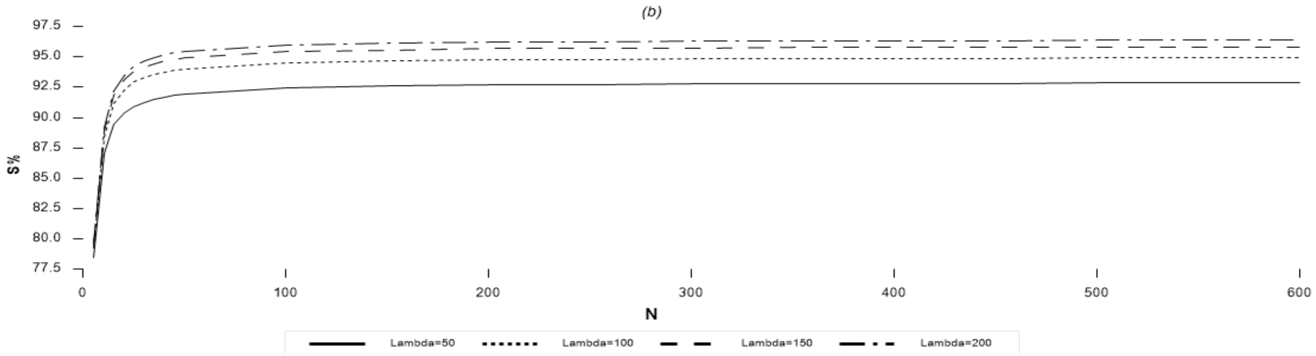
It is important to note that no matter how high the smooth constant λ is, the trend will never attain 100% smoothness. However, more smoothness can be achieved with a larger sample size (N). To find the corresponding λ for fixed values of N, we need to specify the smoothness $S(\lambda,N)$%. It is essential to remember that there is no analytical solution for λ from expression (4.2.1), so it cannot be calculated directly. Instead, the calculations are performed numerically by keeping N and $S(\lambda,N)$% fixed.

In Figure 4.2.1, the behavior of $S(\lambda,N)$% is displayed for different values of N and λ. Figure 4.2.1(a) depicts the impact of sample size on fixed λ values. For $\lambda > 50$, the percentage of smoothness is greater than 95% in all four cases shown in the figure. Additionally, Figure 4.2.1(b) indicates that $S(\lambda,N)$% grows rapidly with increasing sample sizes. When $N=50$, the percentage of smoothness is over 90% for λ values as small as 50, and the smoothness remains essentially constant when the sample size is greater than 50, regardless of the λ value.

Figure 4.2.1

Behavior of $S(\lambda,N)\%$ for: (a) N=50, 100, 400, 800, and (b) λ = 50, 100, 150, 200





Source: own estimates

Regrettably, just like in the case of d=2, it is not feasible to derive an analytical expression for λ as a function of N and $S\%$ from (4.2.1). As a result, Table 4.2.1 presents λ values corresponding to different percentages of smoothness for various sample sizes of daily series. These values were computed numerically by solving equation (4.2.1) for λ, using specific N and $S\%$ values.

Table 4.2.1

Values of λ as a function of sample size N and percentage of smoothness S% (daily series)

N	50%	52.5%	55%	57.5%	60%	62.5%	65%	67.5%	70%	72.5%	75%	75.5%	80%	82.5%	85%	85.7%	90%	92.5%	95%
4	1.366	1.626	1.955	2.384	2.962	3.779	5.012	7.079	11.230	23.715	-	-	-	-	-	-	-	-	-
8	0.984	1.142	1.333	1.565	1.851	2.209	2.666	3.260	4.056	5.159	6.756	9.226	13.446	22.042	48.143	-	-	-	-
12	0.894	1.033	1.198	1.397	1.638	1.937	2.312	2.790	3.415	4.252	5.410	7.078	9.608	13.745	21.342	38.683	109.66	-	-
16	0.854	0.984	1.138	1.323	1.546	1.820	2.161	2.594	3.153	3.893	4.903	6.329	8.437	11.744	17.380	28.309	55.37	195.79	-
20	0.832	0.957	1.105	1.281	1.494	1.755	2.078	2.486	3.010	3.700	4.633	5.937	7.842	10.778	15.653	24.657	44.47	108.49	-
28	0.807	0.927	1.068	1.236	1.438	1.685	1.989	2.371	2.859	3.496	4.351	5.533	7.236	9.817	13.999	21.445	36.71	76.67	277.48
36	0.794	0.911	1.048	1.212	1.409	1.648	1.942	2.311	2.780	3.390	4.205	5.326	6.930	9.338	13.193	19.937	33.39	66.62	195.80
44	0.785	0.901	1.036	1.197	1.391	1.625	1.913	2.273	2.731	3.325	4.116	5.201	6.746	9.052	12.716	19.060	31.51	61.40	167.83
52	0.780	0.894	1.028	1.187	1.378	1.609	1.894	2.248	2.698	3.282	4.056	5.117	6.622	8.861	12.401	18.486	30.30	58.15	153.06
60	0.776	0.889	1.022	1.180	1.369	1.598	1.879	2.230	2.674	3.250	4.013	5.056	6.534	8.725	12.177	18.082	29.46	55.94	143.68
72	0.771	0.884	1.016	1.172	1.359	1.586	1.864	2.210	2.649	3.216	3.968	4.992	6.440	8.581	11.942	17.658	28.59	53.68	134.53
84	0.768	0.880	1.011	1.167	1.353	1.578	1.853	2.197	2.631	3.193	3.935	4.947	6.374	8.481	11.778	17.365	27.99	52.14	128.54
96	0.766	0.877	1.008	1.162	1.347	1.571	1.845	2.186	2.618	3.175	3.911	4.913	6.325	8.406	11.657	17.150	27.545	51.04	124.30
108	0.764	0.875	1.005	1.159	1.343	1.566	1.839	2.179	2.608	3.161	3.893	4.887	6.288	8.349	11.564	16.985	27.212	50.202	121.15
120	0.763	0.873	1.003	1.157	1.340	1.562	1.834	2.172	2.600	3.151	3.878	4.867	6.258	8.304	11.491	16.855	26.949	49.548	118.71
136	0.761	0.871	1.001	1.154	1.337	1.558	1.829	2.166	2.591	3.139	3.863	4.845	6.227	8.257	11.414	16.719	26.676	48.869	116.21
152	0.760	0.870	0.999	1.152	1.334	1.555	1.825	2.160	2.584	3.130	3.851	4.828	6.202	8.219	11.354	16.613	26.463	48.342	114.30
168	0.759	0.869	0.998	1.150	1.332	1.552	1.822	2.156	2.579	3.123	3.841	4.815	6.183	8.190	11.305	16.528	26.292	47.923	112.78
184	0.758	0.868	0.997	1.149	1.331	1.550	1.819	2.153	2.574	3.117	3.833	4.804	6.166	8.165	11.266	16.458	26.152	47.580	111.55
200	0.758	0.867	0.996	1.148	1.329	1.548	1.817	2.150	2.570	3.112	3.826	4.794	6.153	8.144	11.232	16.400	26.036	47.295	110.53
220	0.757	0.866	0.995	1.146	1.328	1.546	1.814	2.147	2.567	3.107	3.819	4.784	6.139	8.123	11.198	16.339	25.915	47.000	109.48
240	0.756	0.866	0.994	1.145	1.326	1.545	1.812	2.144	2.563	3.103	3.813	4.776	6.127	8.105	11.169	16.289	25.814	46.756	108.61
260	0.756	0.865	0.993	1.144	1.325	1.544	1.811	2.142	2.561	3.099	3.808	4.769	6.117	8.090	11.145	16.246	25.730	46.551	107.89
280	0.755	0.865	0.992	1.144	1.324	1.542	1.809	2.140	2.558	3.096	3.804	4.764	6.109	8.077	11.124	16.210	25.658	46.377	107.28

N	50%	52.5%	55%	57.5%	60%	62.5%	65%	67.5%	70%	72.5%	75%	75.5%	80%	82.5%	85%	85.7%	90%	92.5%	95%
300	0.755	0.864	0.992	1.143	1.324	1.541	1.808	2.139	2.556	3.093	3.801	4.759	6.101	8.066	11.107	16.179	25.596	46.226	106.75
324	0.755	0.864	0.991	1.142	1.323	1.540	1.807	2.137	2.554	3.090	3.797	4.753	6.094	8.055	11.088	16.146	25.532	46.071	106.26
348	0.754	0.863	0.991	1.142	1.322	1.540	1.806	2.136	2.552	3.088	3.793	4.749	6.087	8.045	11.072	16.118	25.477	45.938	105.74
372	0.754	0.863	0.990	1.141	1.321	1.539	1.805	2.134	2.551	3.086	3.791	4.745	6.081	8.036	11.058	16.094	25.429	45.822	105.34
396	0.754	0.863	0.990	1.141	1.321	1.538	1.804	2.133	2.549	3.084	3.788	4.741	6.076	8.029	11.046	16.073	25.387	45.721	104.99
420	0.754	0.862	0.990	1.140	1.320	1.538	1.803	2.132	2.548	3.082	3.786	4.738	6.072	8.022	11.036	16.054	25.350	45.632	104.68
448	0.753	0.862	0.989	1.140	1.320	1.537	1.802	2.131	2.547	3.081	3.784	4.735	6.068	8.015	11.025	16.035	25.312	45.540	104.36
476	0.753	0.862	0.989	1.140	1.319	1.536	1.802	2.131	2.546	3.079	3.782	4.732	6.064	8.009	11.015	16.018	25.278	45.459	104.08
504	0.753	0.862	0.989	1.139	1.319	1.536	1.801	2.130	2.545	3.078	3.780	4.730	6.060	8.004	11.006	16.003	25.249	45.387	103.83
532	0.753	0.861	0.989	1.139	1.319	1.535	1.800	2.129	2.544	3.077	3.778	4.728	6.057	7.999	10.999	15.990	25.222	45.323	103.61
560	0.753	0.861	0.988	1.139	1.318	1.535	1.800	2.129	2.543	3.076	3.777	4.726	6.054	7.995	10.992	15.977	25.198	45.266	103.41

In practical applications, we simplify the selection of λ by searching for an approximating function of N and $S\%$ that fits well with the values in Table (4.2.1). We seek several regression models for each $S\%$ and show the estimation results of the best generic fitting models in Table (4.2.1). It is important to note that the value of λ produced by the regression models are just approximations of the true values in Table (4.2.1). In Table (4.2.2), it is required that $\hat{\beta}_1+\hat{\beta}_0N$ be positive for λ to be positive. The models in Table (4.2.2) are helpful in interpolating and, more importantly, extrapolating values for large sample sizes compared to those in Table (4.2.1).

Table 4.2.2

Estimation results of fitting models that relate λ with N and S% (Daly series)

	Approximating function: $\lambda = N/(\hat{\beta}_0(1/N)+\hat{\beta}_1)$		
S%	$\hat{\beta}_0$	$\hat{\beta}_1$	R^2
50	1.330926	-2.441994	0.9989
52.5	1.163152	-2.236259	0.9988
55	1.013489	-2.048598	0.9988
57.5	0.879735	-1.877245	0.9988
60	0.760042	-1.720511	0.9989
62.5	0.652910	-1.577394	0.9991
65	0.557032	-1.446773	0.9994
67.5	0.471344	-1.328060	0.9996
70	0.394926	-1.220657	0.9997
72.5	0.327008	-1.124258	0.9992
75	0.265943	-0.966738	0.9989
77.5	0.212628	-0.853266	0.9989
80	0.166080	-0.746888	0.9991
82.5	0.125913	-0.648564	0.9994
85	0.091809	-0.559848	0.9992
87.5	0.063195	-0.452528	0.9992
90	0.040246	-0.366093	0.9990
92.5	0.022526	-0.273268	0.9991
95	0.009949	-0.177600	0.9994

4.3. Daily data and extension to other frequencies

When estimating the trend of a time series with a frequency of observations different than daily, it is not sufficient to use the same λ obtained for a daily time series. This is because the sample size changes for each type of periodicity under consideration. For instance, if the observation period spans the years 2015-2022, there are either 2920 daily data or 2080 daily data by considering 5-day weeks, 416 weekly data, or 96 monthly data. Therefore, for each of these series and the same percentage of smoothness $S\%$, Table 4.2.1 would lead to different λ values. However, it is essential to remember that the long-term behavior of the series must be essentially the same, regardless of the periodicity of the time series. Additionally, it is worth noting that a time series with a lower frequency of observation is related to that with a higher frequency using some aggregation mechanism. Maravall & Rio (2007) recognized this fact and proposed different solutions to find λ values that produce equivalent results on time series with different periodicities, from a frequency domain perspective.

In cases where time series are nondaily, the selection of the smoothing constant will be determined by a time domain methodology that produces an equivalent level of smoothness as the daily series. This methodology considers the aggregation type that connects a lower-frequency series $\{Y^*_T\}$ with a higher-frequency time series $\{Y_t\}$.

The aggregation is assumed to be linear, that is,

$$Y_T^* = \sum_{j=1}^{k} \delta_j \, Y_{k(T-1)+j} \text{ for } T = 1, \dots, n \quad (4.3.1)$$

with $n=[N/k]$, where $[x]$ stands for the integer part of a real number x, and k is the number of Y_t observations between two successive observations Y^*_T. The constants δ_i determine the aggregation type. For example, $\delta_1=\delta_2= \cdots =\delta_k=1$ is used to aggregate a flow series and $\delta_1=\delta_2= \cdots =\delta_k=1/k$ is used for working with an index or an annualized flow series (which is also considered a flow series). When working with a series of stocks, the aggregated time series is generated by systematic sampling. In this case, the usual values are $\delta_1=1, \delta_2= \cdots =\delta_k=0$ or $\delta_1=\delta_2= \cdots =\delta_{k-1}=0$, $\delta_k=1$. Without loss of generality, in what follows we shall assume that $\delta_1=\delta_2= \cdots =\delta_k=1$ for a flow time series and $\delta_1=\delta_2= \cdots =\delta_{k-1}=0$, $\delta_k=1$ for a time series of stocks. Thus, let T and t represent the time sub-index for the aggregated and disaggregated time series, respectively, then,

$$Y_{T-j}^* = \begin{cases} Y_{t-jk} + Y_{t-jk-1} + \cdots + Y_{t-(j+1)k+1}, & \text{for flows} \\ Y_{t-jk}, & \text{for stocks} \end{cases}, \quad j = 0,1,\ldots,T-1$$

(4.3.2)

The model to be applied to the aggregated data preserves the form (3.1.5) and (3.1.6) for $d=1$, that is,

$$\boldsymbol{Y}^* = \boldsymbol{g}^* + \boldsymbol{\nu}^* \quad \text{with } E(\boldsymbol{\nu}^*) = \boldsymbol{0}, \quad Var(\boldsymbol{\nu}^*) = \sigma_\nu^{*2} I_n \quad (4.3.3)$$

$$K_{1n}\boldsymbol{g}^* = \mu^* \boldsymbol{1}_{n-1} + \boldsymbol{\varepsilon}^* \quad \text{with } E(\boldsymbol{\varepsilon}^*) = \boldsymbol{0}, \quad Var(\boldsymbol{\varepsilon}^*) = \sigma_\varepsilon^{*2} I_{n-1}$$

(4.3.4)

and $E(\varepsilon^* \nu^{*\prime})=0$, where * is used to denote aggregate variables. Therefore, (4.1.3), (4.1.4), (4.1.5) and (4.1.6) become

$$\hat{\boldsymbol{g}}^* = (I_n + \lambda_k^* K_{1n}' K_{1n})^{-1}(\boldsymbol{Y}^* + \lambda_k^* \mu^* K_{1n}' \boldsymbol{1}_{n-1}) \quad (4.3.5)$$

$$\Gamma^{*-1} = Var(\hat{\boldsymbol{g}}^*) = \sigma_\nu^{*2}(I_n + \lambda_k^* K_{1n}' K_{1n})^{-1} \quad (4.3.6)$$

$$\hat{\boldsymbol{g}}^* = (I_n + \lambda_k^* K_{1n}' K_{1n})^{-1}(I_n + \lambda_k^*(n-1)^{-1} K_{1n}' \boldsymbol{1}_{n-1} \boldsymbol{1}'_{n-1} K_{1n}) \boldsymbol{Y}^* \quad (4.3.7)$$

$$\tilde{\sigma}_\nu^{*2} = [\boldsymbol{Y}^{*\prime}\boldsymbol{Y}^* - \hat{\boldsymbol{g}}^{*\prime}(I_n + \lambda_k^* K_{1n}' K_{1n})\hat{\boldsymbol{g}}^*]/(n-1)\,\lambda_k^* \mu^{*2} \quad (4.3.8)$$

While expressions (4.1.5) and (4.3.5) share the same form, they produce different trends. In fact, aggregating the trend $\{\hat{g}_t\}$ estimated from the disaggregated series results in different values than those of the estimated trend $\{g^*_T\}$ acquired directly from the aggregated series. Nevertheless, as it is shown in the Appendix, it is possible to find a smoothing constant λ for a disaggregated series that is equivalent to the λ^*_k value for the aggregated data, as follows:

$$\lambda = \begin{cases} (k^2-1)/6 + k^2 \lambda_k^* & \text{for flows} \quad (4.3.9) \\ k\lambda_k^* & \text{for stocks} \end{cases}$$

It is important to note that if the smoothing constant λ for the disaggregated time series is known, we can use equation (4.3.9) to solve for the corresponding value of λ^*_k. This will be helpful in the analysis.

4.4. A simulation study

To verify numerically the performance of the suggested procedure, a simulation study was run. Two random walks were considered, one with drift and the other without drift. The random walk hypothesis (Fama, 1965) corresponds to a financial theory stating that stock market prices evolve according to a random walk with drift (so price changes are random) and thus cannot be predicted exactly. Let X_t be today's log-price of a stock. The value of X for tomorrow, X_{t+1}, will equal today's value, X_t, plus a constant value ρ_0, plus a random shock ε_t. The shock is a random variable satisfying $\varepsilon_t \sim i.i.d. N(0, \sigma_\varepsilon^2)$. Then the random walk model is the following.

$$X_t = \rho_0 + X_{t-1} + \varepsilon_t \quad (4.4.1)$$

ε_t is a random shock for each day, resulting from the log-price movements due to all news that influence the price, while ρ_0 refers to the drift of the series. If $|\rho_0|>0$ the series is a random walk with a drift. If $\rho_0>0$, then the series will have a positive trend over time, if $\rho_0<0$, the series will have a negative trend.

To simulate a random walk, the values of the following parameters are required.

- X_0, the first value of the series
- ρ_0, the drift of the series
- σ_ε^2, the volatility of the random shock.

Substituting repeatedly for lagged values of X_t gives

$$X_1 = \rho_0 + X_0 + \varepsilon_1$$

$$X_2 = \rho_0 + X_1 + \varepsilon_2$$

$$= \rho_0 + \rho_0 + X_0 + \varepsilon_1 + \varepsilon_2$$

$$= 2\rho_0 + X_0 + \varepsilon_1 + \varepsilon_2$$

Therefore, doing the same until the last T value, it follows that

$$X_T = T\rho_0 + X_0 + \sum_{i=1}^{T} \varepsilon_i \quad (4.4.2)$$

so that,

$$E(X_T) = T\rho_0 + X_0$$

from this expression, ρ_0 can be estimated as

$$\rho_0 = \frac{(X_T - X_0)}{T} \quad (4.4.3)$$

Now,

$$Var(X_T) = Var(T\rho_0) + Var(X_0) + \sum_{i=1}^{T} Var(\varepsilon_i) = T\sigma_\varepsilon^2$$

and the measure of the shocks' volatility, in relation to the series' overall volatility, is determined by:

$$\sigma_\varepsilon^2 = \frac{\sigma_X^2}{T} \quad (4.4.4)$$

To estimate the three parameters for simulating a random walk series, we will employ the S&P500 daily data from Yahoo Finance, from January 2011 to May 2023. To achieve this, we need to generate the log of the S&P500 index. Using (4.4.3), we obtain the value of ρ_0 equal to 0.000383058; from (4.4.4) we get the value of σ_ε^2 equal to 0.006913685. Additionally, the initial value is 7.148. We then generated random shocks from a N(0, 0.006913685) distribution. Finally, we simulated two random walk series. The first one, called rw1, is a random walk with drift, while the second one, named rw2, is a random walk without drift.

$$rw1_t = \rho_0 + rw1_{t-1} + \varepsilon_t$$

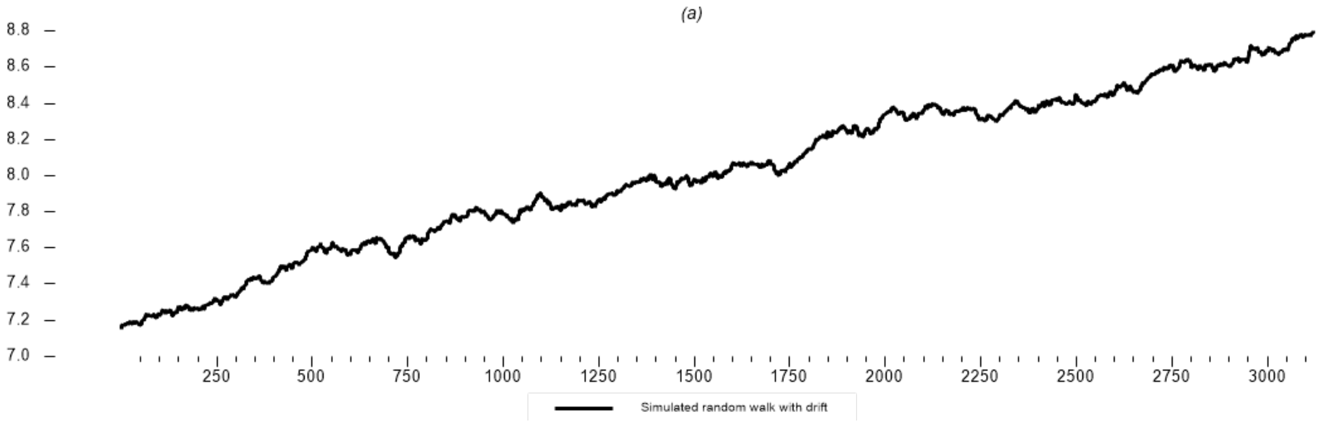
Figure 4.4.1 panel (a) shows the simulated random walk with drift. We started the random walk with the first value of the log of S&P500. Then, from day 2 we ran the simulation according to the previous formula, using the random shock jus created. In a similar way, we simulate rw2 from

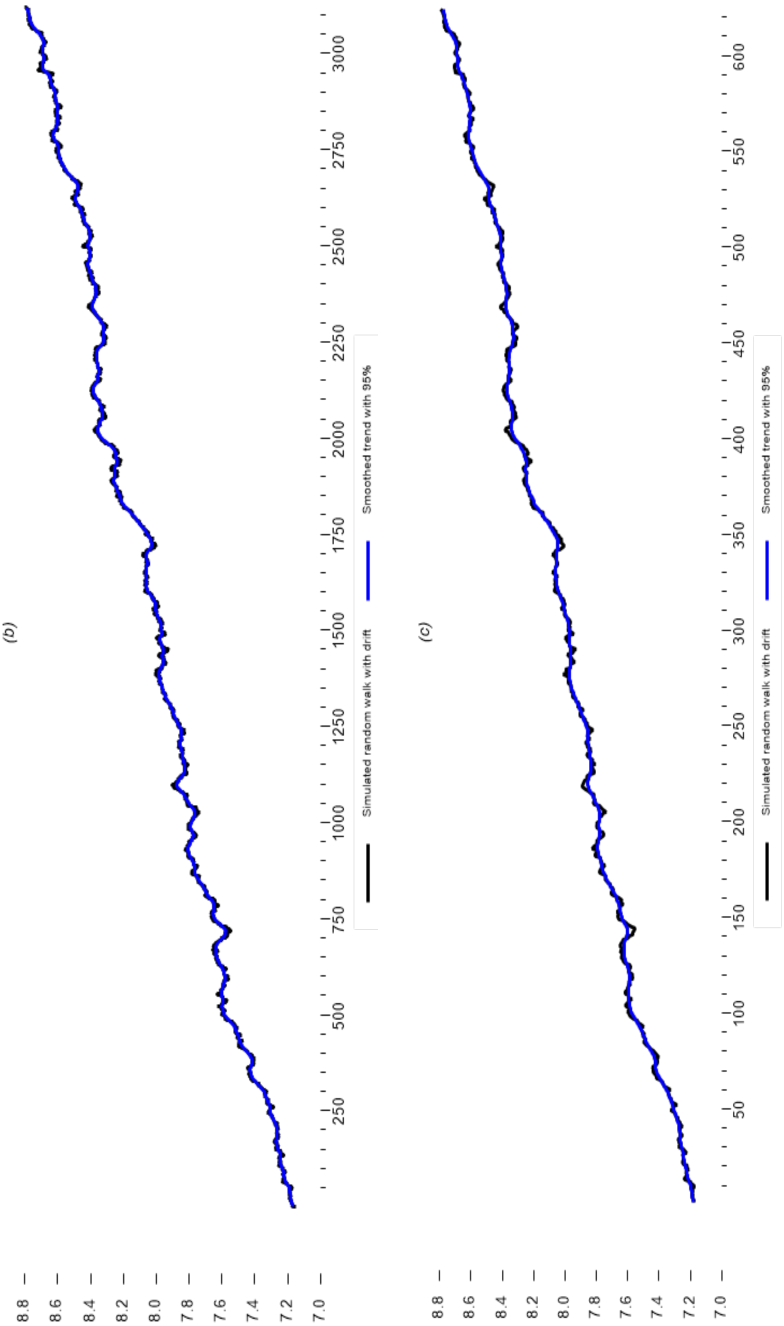
$$rw2_t = rw2_{t-1} + \varepsilon_t$$

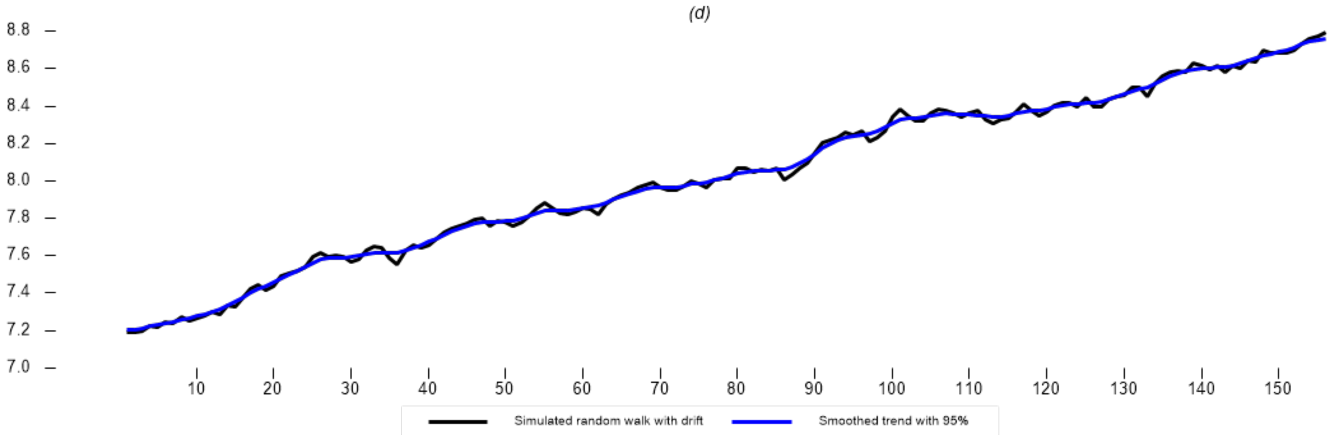
Figure 4.4.1, Panel (a) displays the simulated random walk with drift and sample size of *N=3122*. Following Guerrero's guideline (2017), a 95% percentage of smoothness was set to select the corresponding smoothing parameter from (4.2.1). For daily data, the smoothing parameter value is λ = 101.1. Panel (b) shows the daily time series and its trend. Panel (c) exhibits the weekly time series and its trend, constructed from the daily data. It is important to note that this is a series of stocks, and the sample now comprises *N=624* weekly observations, considering five-day weeks. The smoothing constant for the weekly series is λ^*_5 = 20.22 for S%=95%, as obtained from (4.3.8) by $\lambda = 5\lambda^*_5$. Lastly, Panel (d) shows the monthly time series and its trend, also built from the daily data. The sample size for this series has *N=157* observations, and the value of λ for monthly data is λ^*_{20} = 5.05 for *S%=95%,* obtained from (4.3.9) by $\lambda = 20\lambda^*_5$. By visually inspecting Panels (b), (c), and (d), we can observe that the resulting trends with the same smoothness percentage display essentially the same dynamic behavior, regardless of the frequency of data observation.

Figure 4.4.1

Panel (a), simulated random walk with drift, sample size of $N= 3123$ daily data. Panel (b), simulated daily random walk with drift and estimated trend with S%=95% (λ = 101.1). Panel (c) weekly random walk with drift and estimated trend with S%= 95% (λ^*_5 = 20.22)$N=624$ observations. Panel (d) monthly random walk with drift and estimated trend with S%= 95% (λ^*_{20}=5.05), $N=157$ observations.




(b)
Simulated random walk with drift
Smoothed trend with 95%
8.8
8.6
8.4
8.2
8.0
7.8
7.6
7.4
7.2
7.0
250
500
750
1000
1250
1500
1750
2000
2250
2500
2750
3000
(c)
Simulated random walk with drift
Smoothed trend with 95%
8.8
8.6
8.4
8.2
8.0
7.8
7.6
7.4
7.2
7.0
50
100
150
200
250
300
350
400
450
500
550
600

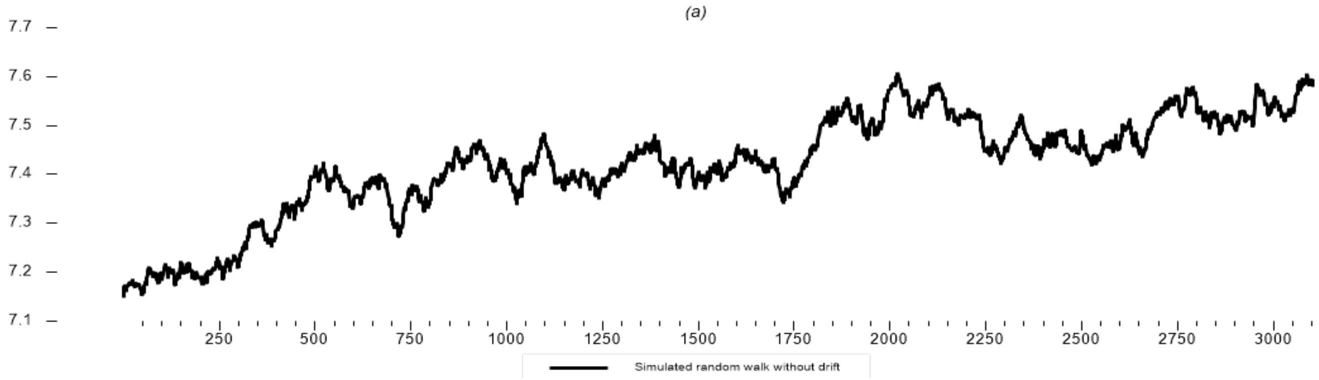


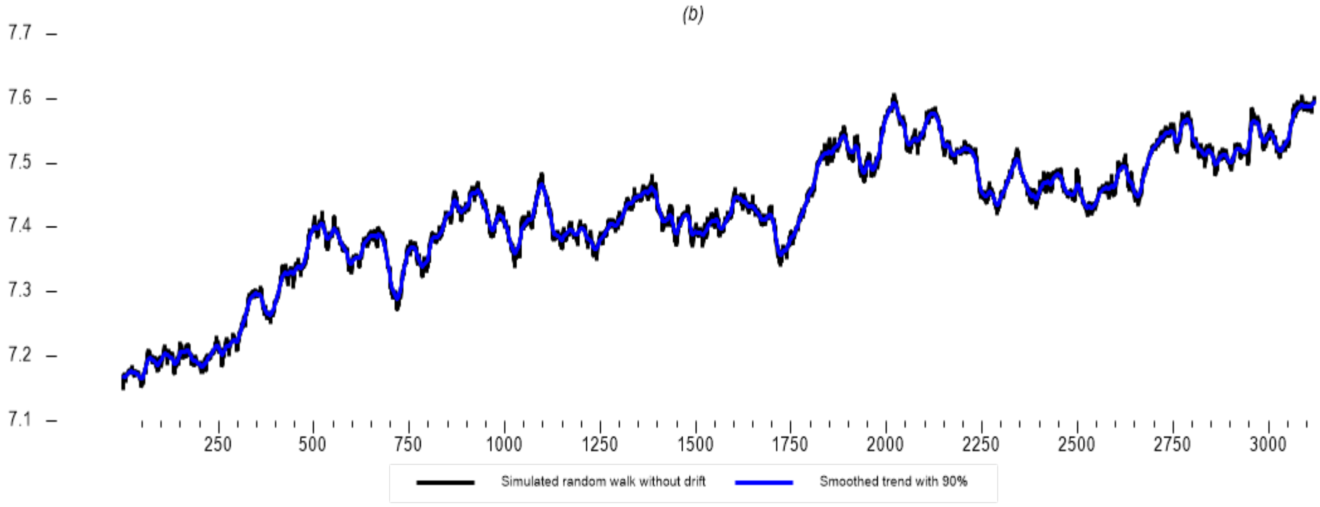
Source: own estimates

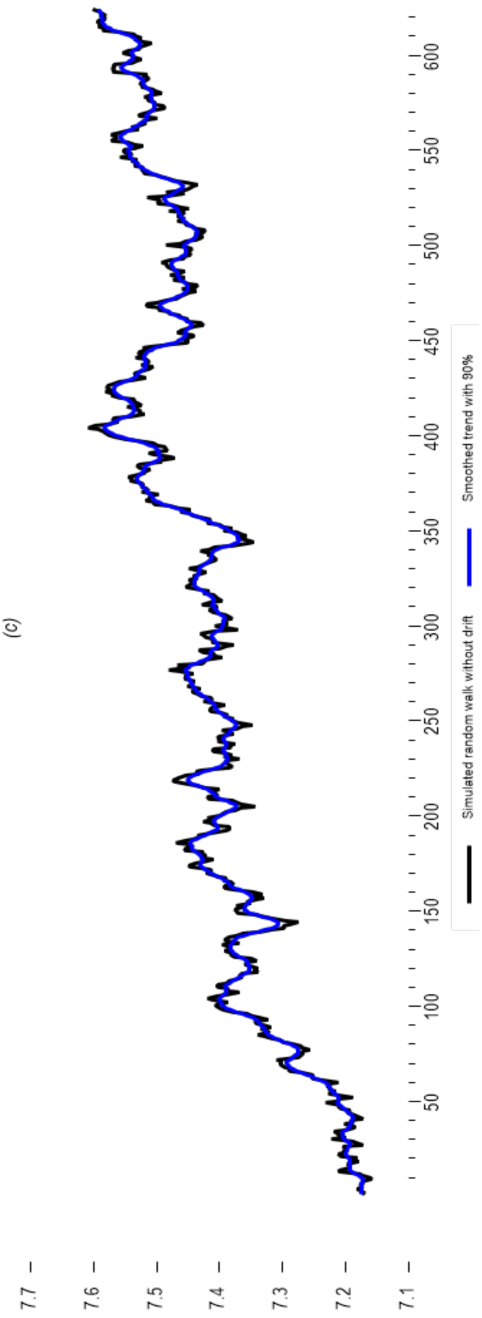
On the other hand, Figure 4.4.2 Panel (a) shows the simulated random walk without drift and sample size $N=3122$. Then, following Guerrero´s (2017) guideline the percentage of smoothness suggested is 90% to choose the corresponding smoothing parameter from (4.2.1), which takes the value $\lambda= 24.9$ for daily data, Panel (b) shows the daily time series without drift and its trend. Panel (c) shows the weekly time series without drift and its trend built from the daily data. As in the random walk with drift, it should be stressed that this is a series of stocks (it is chosen as every fifth value of the sample under consideration). The sample now consists of $N=624$ weekly observations, considering five-day weeks. The value of λ is obtained from (4.3.8) as follows $\lambda = 5\lambda^{*}_{5}$, therefore, the smoothing constant for the weekly series is $\lambda^{*}_{5} = 4.98$ for S% = 90%. Finally, Panel (d) shows the monthly time series without drift and its trend built, also, from the daily data. The sample for this series is $N=157$ observations. The value of λ for the monthly data is also obtained from (4.3.8) as $\lambda = 20\lambda^{*}_{5}$, and takes the value $\lambda^{*}_{20} = 1.245$ for S%=90%. In the same way, by a visual inspection of Panels (b), (c), and (d), the resulting trends with the same percentage of smoothness show essentially the same dynamic behavior, no matter what the frequency of observation of the data is.

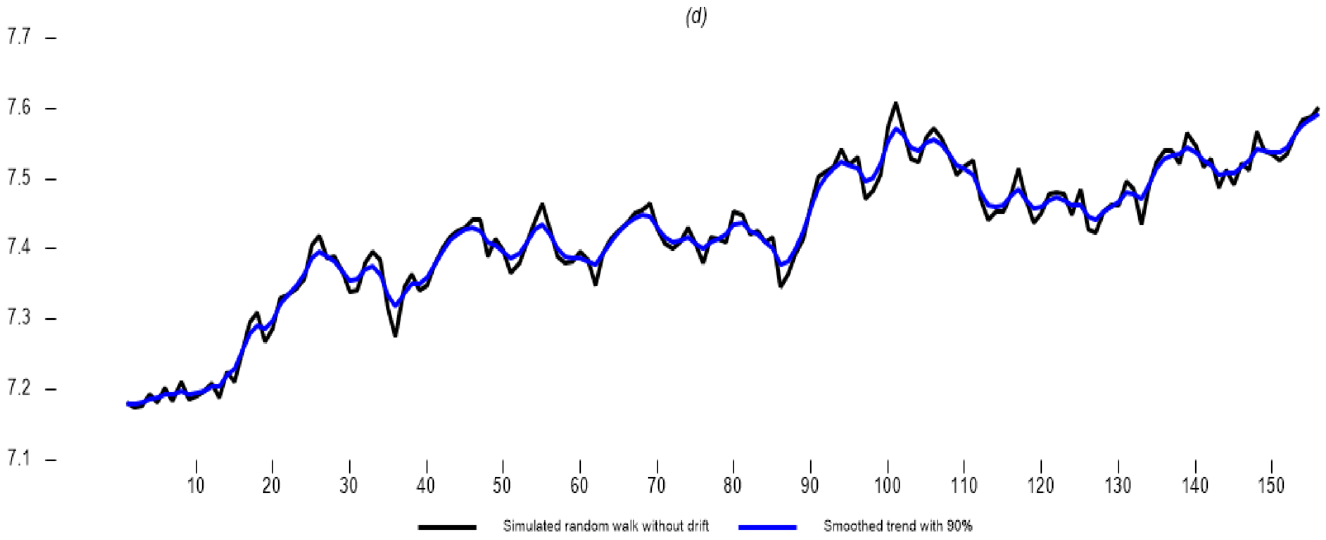
Figure 4.4.2

Panel (a), simulated random walk without drift, sample size *N=3123* daily data. Panel (b), simulated daily random walk without drift and estimated trend with *S*%=90% (λ = 24.9). Panel (c) weekly random walk without drift and estimated trend with *S*%= 90% (λ^*_5 = 4.98), *N=624* observations. Panel (d) monthly random walk without drift and estimated trend with *S*%= 90% (λ^*_{20} = 1.245), *N=157* observations.




(b)
7.7
7.6
7.5
7.4
7.3
7.2
7.1
250
500
750
1000
1250
1500
1750
2000
2250
2500
2750
3000
Simulated random walk without drift
Smoothed trend with 90%


(c)
7.7
7.6
7.5
7.4
7.3
7.2
7.1
50
100
150
200
250
300
350
400
450
500
550
600
Simulated random walk without drift
Smoothed trend with 90%



Source: own estimates

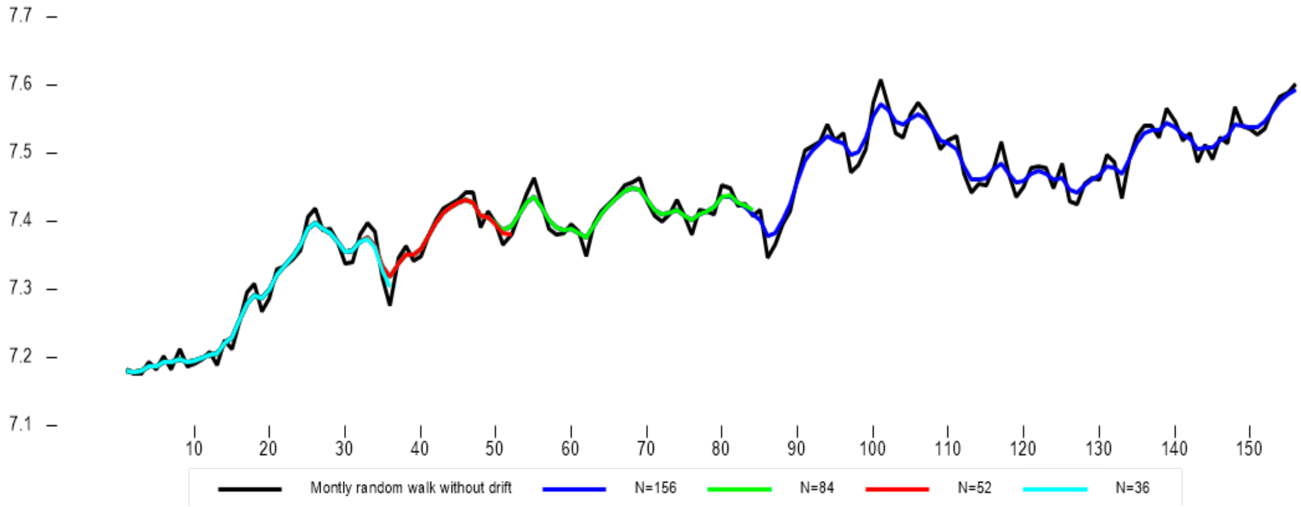
4.4.1. Same time series with different sample sizes

It is important to note that trends obtained with the same percentage of smoothness are comparable even if the sample sizes differ, just like in the case of $d = 2$. To demonstrate this, we estimated the trend of the monthly random walk without drift, previously shown in Figure (4.4.2) Panel (d), with different sample sizes. To achieve this goal, we need to obtain the λ^*_{20} value of the monthly series, which is equivalent to the smoothing parameter λ used with the daily series. In the case of a stock series, the equivalent smoothing constants are related by $\lambda = 20\lambda^*_{20}$. For sample sizes of daily data, N=84, 52, and 36, the values of lambda that produce 90% of the smoothness of the trend are λ=24.982, λ=25.066, and λ=25.165, respectively. Therefore, the corresponding smoothed parameters for the monthly series are $\lambda^*_{20} = 1.249$, $\lambda^*_{20} = 1.253$ and $\lambda^*_{20} = 1.258$. The resulting trends are depicted in Figure 4.4.3, and it can be observed that the trend estimates are reasonably close to each other.

Based on the examples given, it is evident that the λ values in Table 4.2.1 and the values derived from the fitting models in Table 4.2.2 work best for time series comprising daily observations. The process for selecting the smoothing constant of a daily time series can also be applied to other types of time series. The formula for determining the equivalent smoothing constant is simple for flow and stock series. Moreover, using the daily observation frequency as a standard reference point offers empirical benefits. Thus, we address the issue of finding the trend for the same time series with different periodicities. Our findings suggest that trends with the same degree of smoothness, but different periodicities, display similar dynamic behavior.

Figure 4.4.3

Selected trend estimates with 90% smoothness and different sample sizes for the simulate monthly random walk without drift series.



Source: own estimates

Chapter 5

Empirical Applications

In this chapter, we showcase how our developed UCS web-tool can be used to obtain trend and cyclical components of observed time series. We present three examples, two of which come from economics and one is financial. The first problem we address is related to the high levels of crime rates faced by Mexican society for almost three decades. The second example measures the impact of monetary policy decisions made by Mexico's Central Bank on economic activities in different regions of Mexico. The third example studies the long-run evolution of oil prices.

5.1 Case 1: Economic growth and crime

Let us start with the problem of public insecurity. According to Mexico's Federal Penal Code (FPC) crimes can be classified into two types: Federal Crimes (FC) and Common Crimes (CC). In general, FC are those that affect people's health, heritage, and the nation's security. For instance, offenses like treason, espionage, sedition, rebellion, and terrorism among others. CC on the other hand, are those that affect directly to people as individuals. These are organized by the type of legal right affected; that is, life and bodily integrity, personal liberty, sexual liberty and security, property, family and society among others. Examples are robbery, extortion, rape, etc.

By its nature, CC represent the most serious challenge to Mexican authorities since almost 95 percent of total crimes are common crimes. To the extent that they involve a wide range of felonies, it has been very difficult to pinpoint their main determinants. However, an increasing number of studies have been able to identify key characteristics of their behavior over time. This identification process shall eventually help us define and implement a public security policy that would attenuate and even reduce the level of crime significatively.

Since the mid-90s crime analysis has been a growing area within criminology that uses not only different theories of crime but methods of analysis as well. For instance, analyses of crime by geographical areas, -led by Place Theories of Crime-, have allowed researchers to identify crime hotspots, i.e., areas with high crime intensity. Moreover, for Crime Pattern Theory, crime is not a random phenomenon since it is either planned or opportunistic (Brantingham & Brantingham, 2008). In effect, this theory sustains that those criminals commit their crime in specific areas and times that respond to victims' and offenders' patterns of behavior.

Within criminology there are competing theories about how to analyze and identify the key determinants of crime. Environmental criminology, for instance, is a family of theories that focus on crime events, the circumstances in which they occur and their changing nature, while crime economics focuses on identifying the main determinants of individual's criminal behavior and evaluating their economic impact. In any event, all theories need to test their hypothesis so that they can make predictions about emerging problems or future trends. This information will help develop strategies that might be employed to prevent crime.

Nowadays, there are different statistical instruments that help analyze crime data. These instruments can help us identify spatial as well as temporal patterns of crime. Spatial-temporal patterns of crime can be analyzed using spatial econometrics, while temporal patterns can be studied through time series econometrics. As already discussed in the previous two chapters, a first step is the decomposition of the series into its trend and cyclical components.

Time series of crime can refer to any type of crime (burglary, kidnapping, fraud, rape, robbery, etc.) and its scope can be a neighborhood, county, city, region, or country. Therefore, the use of time series techniques is flexible and can be applied to any type of crime or to any level of aggregation. Study of the trend component is more associated to the study of the long-term behavior of the series; that is, the crime behavior that is persistent over time.

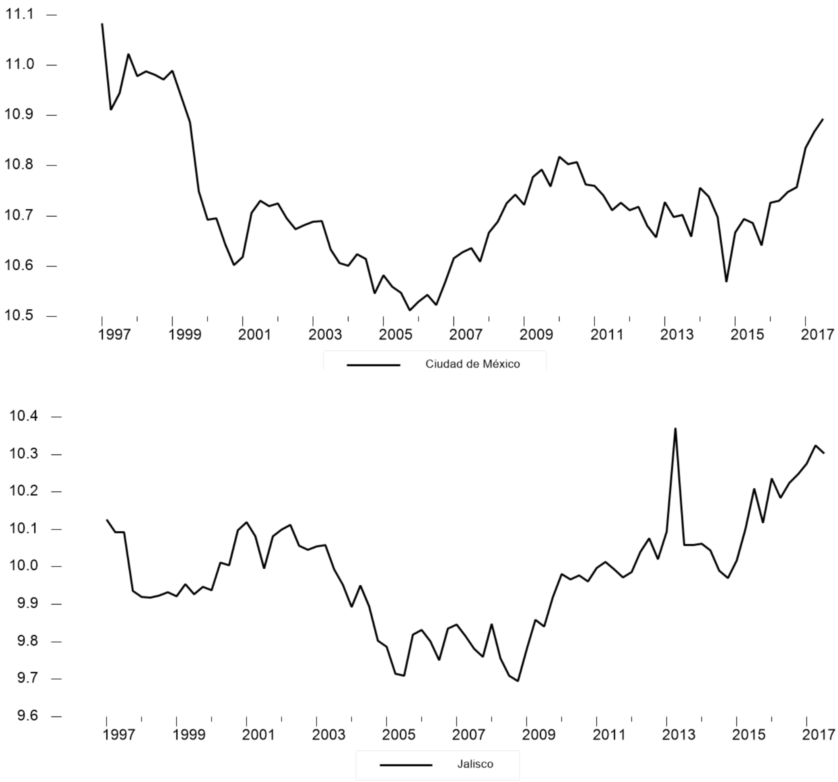
There are several explanations about the persistence of crime. From the economics of crime's perspective, once a person has committed a crime, his/her criminal human capital relative to his/her legal human capital has increased. This higher (relative) criminal human capital makes the individual not to give up easily criminal activities. This effect might be reinforced by the fall of learning costs associated to criminal activities, which will increase crime's rate of return. Another explanation is based on the effect that crime has upon the interaction between moral costs and learning of crime technology; that is, before a person commits any crime, both high moral costs and low knowledge of criminal technology act as crime deterrent factors. But once this person starts committing criminal activities, these deterrent factors start disappearing, i.e., moral costs fall while learning crime technologies increase.

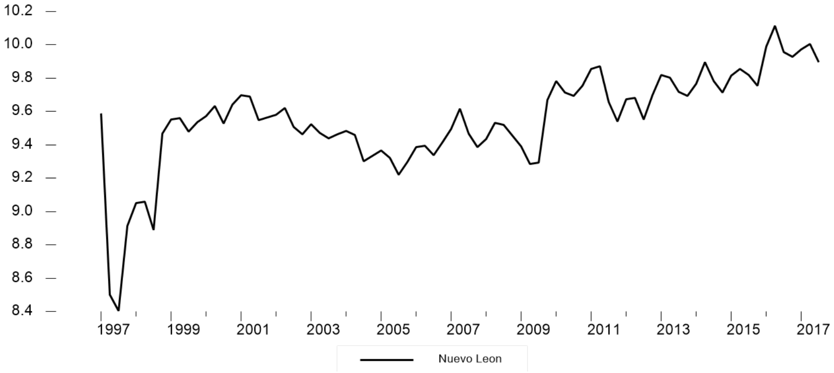
These arguments help us understand why we observe crime's long-term behavior. But the analysis of the long-term component would also tell us whether there were changes in the underlying factors that explain it and in what direction. In addition, we can also use the long-term component to predict long-term behavior.

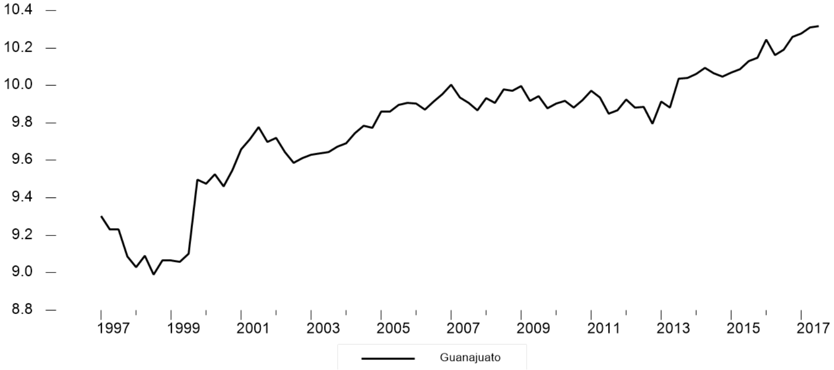
In what follows we analyze the behavior of total crime in four Mexican States: Nuevo Leon, Mexico City, Jalisco, and Guanajuato. Mexico City, Nuevo Leon and Jalisco are three of the largest states in terms of population. Guanajuato, on the other hand, is one of the fastest growing States in manufacturing. We are interested in comparing the evolution of their crime levels' trend component.

Figure 5.1.1 displays quarterly total crime data from 1997-I to 2017-IV in Nuevo Leon, Mexico City, Jalisco, and Guanajuato. The data shows non-linear behavior with no overall trend, and the concept of stochastic local trend seems better to describe the underlying dynamics of the trend, so we chose $\mu \neq 0$.

Figure 5.1.1
Quarterly total crime data from 1997-I to 2017-IV
in Mexico City, Jalisco, Nuevo Leon, and Guanajuato



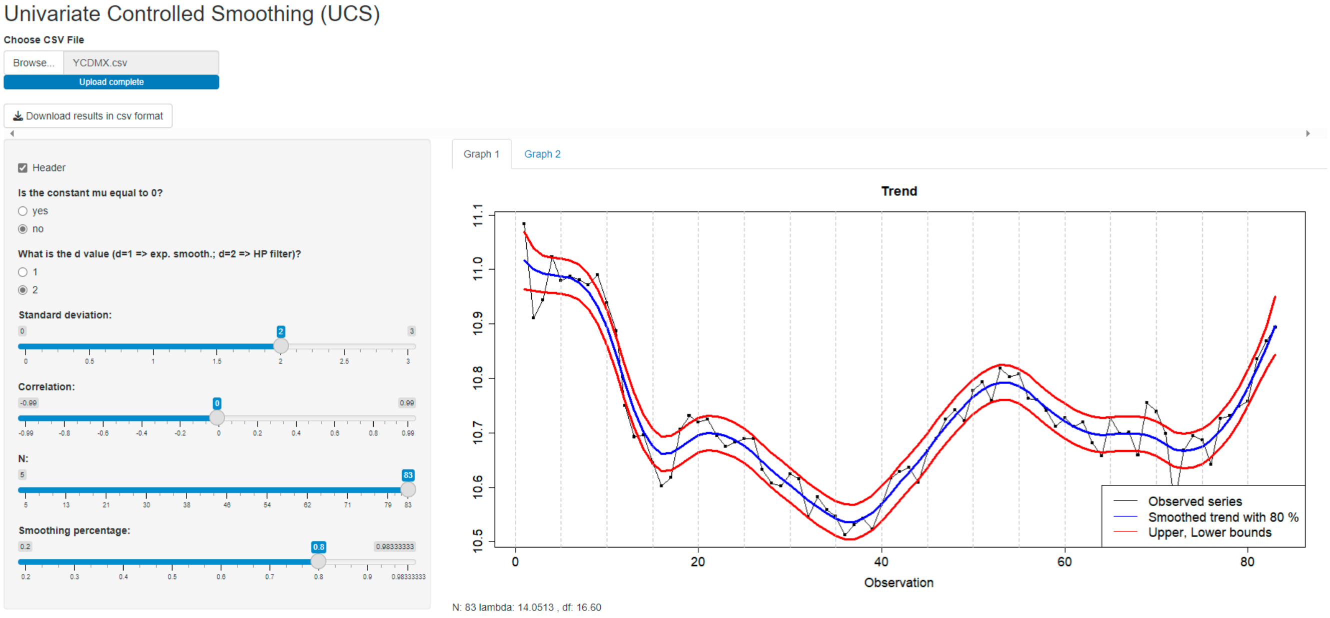




The suggested percentage of smoothness of 80% was used to select the smoothing parameter, based on Guerrero's (2017) guideline, resulting in $\lambda=14.05$ for a sample size of N=83 quarters. The smoothness percentage was kept constant among the four series under study to enable adequate trend comparison. To illustrate the use the UCS web-tool to estimate of the trend and cycle's components we use the time series of Mexico City's total crime.

Figure 5.1.2

Defining the Model's parameters



Univariate Controlled Smoothing (UCS)

Choose CSV File

Browse... YCDMX.csv

Upload complete

Download results in csv format

Graph 1 Graph 2

☑ Header

Is the constant mu equal to 0?

○ yes

● no

What is the d value (d=1 => exp. smooth.; d=2 => HP filter)?

○ 1

● 2

Standard deviation: 2

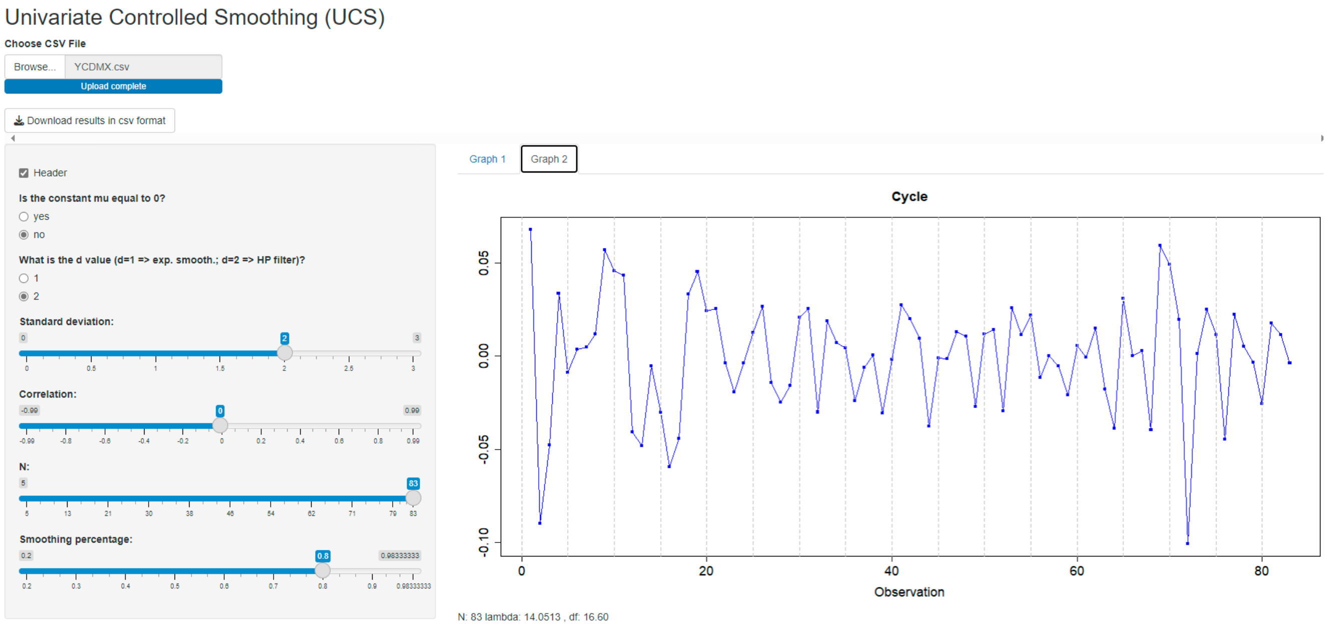
Correlation: 0

N: 83

Smoothing percentage: 0.8

N: 83 lambda: 14.0513 , df: 16.60

0
0.5
1
1.5
2
2.5
3
-0.99
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
0.99
5
13
21
30
38
46
54
62
71
79
83
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.98333333
Trend
11.1
11.0
10.9
10.8
10.7
10.6
10.5
0
20
40
60
80
Observation
Observed series
Smoothed trend with 80 %
Upper, Lower bounds


Browse...
Download results in csv format
Graph 1
Graph 2
Cycle
0.05
0.00
-0.05
-0.10
0
20
40
60
80
Observation
0
0.5
1
1.5
2
2.5
3
-0.99
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
0.99
5
13
21
30
38
46
54
62
71
79
83
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.98333333

Univariate Controlled Smoothing (UCS)

Choose CSV File

YCDMX.csv

Upload complete

☑ Header

Is the constant mu equal to 0?

○ yes

● no

What is the d value (d=1 => exp. smooth.; d=2 => HP filter)?

○ 1

● 2

Standard deviation: 2

Correlation: 0

N: 83

Smoothing percentage: 0.8

N: 83 lambda: 14.0513 , df: 16.60

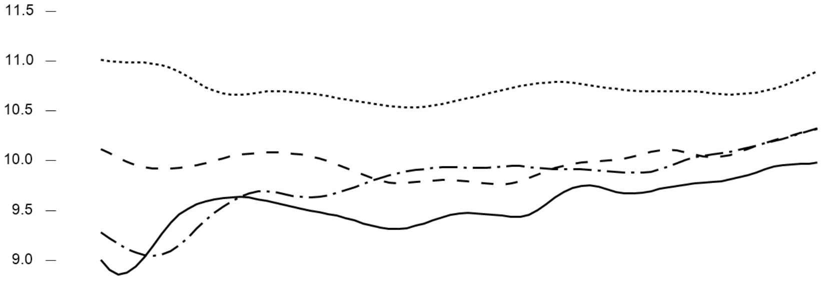
Source: Own estimates

To the left of the UCS tool (in Figure 5.1.2), we select parameters related to the filter we want to use; in this case, we chose $\mu \neq 0$, a difference order (d) equals to 2, two standard deviations (SD) for the estimated error bands of the trend, noncorrelation, $\rho=0$ (see section 3.4), and percentage of smoothing S%= 80%. Figure 5.1.2 displays trend estimates (Graph 1), as well as cycle estimates (Graph 2) once data in .csv format for the univariate crime time series was uploaded. The estimated trend component ($\hat{g}$) is shown with ∓ 2 standard error bands, while the estimated cycle component ($y_t - \hat{g}_t$) is shown below it. Our results were downloaded in .csv format for additional analyses, by just pushing the top-right button ("Download results in .csv format"). We estimated the trend and cycle of the total crime series for the other States, i.e., Jalisco, Nuevo León and Guanajuato in a similar manner.

Figure 5.1.3 shows total crime's long-term evolution during the period of analysis. Except for Mexico City, total crime shows an upward trend in Nuevo Leon, Jalisco, and Guanajuato. The graph also shows that crime in Nuevo Leon and Jalisco evolves in a parallel fashion since the late 90s. A third characteristic is that among these States, Guanajuato shows higher long-term growth of total crime.

Figure 5.1.3

Trend of Total Crime in selected states: 1997-I to 2017-IV



Source: Own estimates

We now evaluate the relationship between the Index of Physical Volume's and Total Crimes' cyclical components following the methodology proposed by Kydland & Prescott (1990). Table 5.1.1 presents the estimated cross correlations between output and crime's cyclical components for the selected entities. Each cell represents the value of the correlation between the cyclical components of the Index of Physical Volume (IPV) and Total Crime (TC) at different lags for each of the selected states,

Table 5.1.1

Cross Correlation between the cyclical components
of the state's Index of Physical Volume and Total Crime

	State			
	Nuevo Leon	Jalisco	Cd Mexico	Guanajato
X(t-8)	-0.078	-0.068	0.133	-0.063
X(t-7)	-0.149	-0.188	-0.193	-0.262
X(t-6)	-0.064	-0.152	0.032	0.091
X(t-5)	0.204	0.23	-0.018	0.047
X(t-4)	0.078	-0.078	0.211	0.104
X(t-3)	-0.066	-0.088	-0.129	-0.071
X(t-2)	-0.061	-0.078	0.042	0.106
X(t-1)	0.115	0.197	-0.158	-0.096
X(t)	0.035	-0.034	0.14	-0.034
X(t+1)	0.111	-0.155	-0.198	-0.203
X(t+2)	0.052	-0.129	-0.015	0.094
X(t+3)	0.149	0.241	-0.118	0.04
X(t+4)	-0.034	0.001	0.174	0.123
X(t+5)	-0.103	-0.153	-0.137	-0.061
X(t+6)	-0.102	-0.029	0.058	0.164
X(t+7)	0.048	0.145	-0.085	-0.003
X(t+8)	-0.057	0.095	0.224	0.05

Source: Own Estimates

In the case of Nuevo Leon, we observe that there is a slight indication that it may be procyclical; however, crime might anticipate economic activity by about five quarters. Crime in Jalisco, on the other hand, seems to be procyclical as well; however, it appears to lag economic activity by about three quarters, that is, crime is procyclical. Crime in Mexico City also seems to be procyclical, but the strongest correlation occurs when crime lags to economicactivity by eight quarters. Unlike the previous cases, the crime and economic activity seems to be countercyclical in Guanajuato.

The cross correlation is strongest in the seventh quarter when crime antecedes economic activity.

In short, Table 5.1.1 present some evidence about the relationship between the cyclical components of economic activity and crime. The results, however, are not as conclusive as one might have expected. A possible explanation is that the crime variable may need to be more specific. In our analysis we have used total crime, probably we would need to use economic crimes instead. Another element that we need to consider is whether the Index of Physical Volume is the proper indicator of economic activity. Perhaps a more precise indicator could be an income indicator.

5.2. Case 2: Regional Effects of Monetary Policy in Mexico

For this example, we also use Kydland & Prescott´s (1990) proposed methodology to analyze business cycles to explore the regional impact of monetary policy in Mexico. To recall, the latter is based on the estimation of cross correlations between the cyclical components of monetary and real sector variables against the GDP's cyclical component. Our interest is to somehow evaluate the effect of monetary policy upon output fluctuations in a selected group of Mexican States. As is well known, Mexico's Central Bank, Banco de México, is the sole authority that determines the country's monetary policy. Since 1994, it has been an autonomous monetary authority that officially adopted inflation targeting as its strategy to maintain price stability in 2002; albeit it had begun implementing early measures towards that goal since 1995 (Ramos-Francia & Torres-Garcia, 2005).

Banco de México's decision was in tune with other Central Banks' decisions around the world about implementing inflationary targeting as a framework to conduct their monetary policy. The main goal of such a strategy is price stability, achieved through the setting of a nominal anchor to tie down the price level. It is argued that inflation stability is a necessary condition for sustainable growth for several reasons. First, it induces low inflation expectations which affects positively private sector's investment decisions. Second, it contributes to an efficient allocation of resources since relative prices are likely to remain stable. Third, low and stable inflation implies low social and economic costs since the purchasing power of money remains stable; thus, consumers enjoy higher welfare. An additional positive outcome is that there is no redistribution of income from loaners to borrowers, which would induce more savings.

To accomplish its goal, Banco de Mexico (Banxico) sets the equilibrium interbank interest rate (TIIE) based on quotes presented by credit's institutions. The initial change in interest rate affects inflation through several channels. These channels, also known as transmission mechanisms of monetary policy, affect both aggregate demand and supply to different degrees, both of which induce price changes. In other words, to achieve its objective of price control, monetary policy induces changes in the economy's real sector.

Let us see what these transmission mechanisms are and how they work. Banxico in its statement about the effects of monetary policy (Banco de México, 2023b)[1] reports five different transmission channels: i) interest rates, ii) credits, iii) exchange rates, iv) other assets' prices, and v) expectations.

In the first case, Banxico decides to change short-term interest rates when inflation expectations are off its expected values. In general, an increase in the interest rate increases capital´s costs to finance projects as well as the opportunity cost of current consumption. These two factors constrain demand, and thus price growth. Credits, in turn, are also affected by the increase in interest rates since now it is more expensive to finance investments and purchases. Financial intermediaries may also decide to restrict the amount of available credit since the risk of default is greater. All in all, aggregate demand is negatively affected.

The exchange rate channel kicks in when the domestic interest rate makes more attractive domestic financial assets compared to the foreign ones. In this case, there is an influx of foreign capital which appreciates the domestic currency. As a result, net exports tend to decline, i.e., exports decline while imports increase. An additional effect of the depreciation of foreign currency is that it affects overall domestic prices directly since prices in domestic currency of imported goods decline. In short, the exchange rate channel affects not only the real sector of the economy, but also the price level directly.

The fourth channel operates when the increase of the interest rate makes Bonds more attractive than other financial assets, in particular stock prices, which, in turn, may have as an unintended consequence the decline in the market value of some private firms. The latter may further complicate these firms´ access to finance new investment projects, which will result in a decline in aggregate demand.

---

[1] A further description can be found in Banco de México (2023).

The last channel, expectations, -and in particular inflation expectations-, have become a very important channel since depending on whether they are off targets, can induce Banxico to modify its monetary policy. However, as Tobias (2023) points out there are different expectations; for example, near- and long run-time expectations. There are also households', markets', and forecasters' expectations, and Banco de México must consider the behavior of all of them in its decision-making process. This is so because they affect economic agents' consumption and investment decisions. Long-term expectations are also considered indicative of the Central Bank's credibility in controlling inflation.

It is well known that Mexico's economy is characterized by having great regional contrasts. The regional differences include not only differences in endowments, climate, but also in productive structure and stage of economic development. Given these differences across states or regions we are interested in exploring how different are the effects of a given monetary policy across a small group of Mexican States. To do that we propose to use Guerrero's filter to obtain the cyclical components of an indicator of monetary conditions as a proxy of monetary policy and of the economic sectors for a group of five Mexican States.

Therefore, we estimate a Monetary Conditions Index (MCI). We argue that MCI can be a good alternative indicator of monetary conditions when we are dealing with a small open economy like Mexico. In an open economy with free capital mobility and exchange rate market, interest rates´ increases might induce an influx of financial capital which, in turn, induces an appreciation of the domestic currency; that is, a contractive monetary policy is reinforced by an appreciation of the domestic currency which further contracts aggregate demand. There are other instances when monetary conditions change without explicit intervention of the Central Bank. For instance, foreign banks may decide to change their relevant interest rates or when there are external shocks that affect trade balances that, in turn, pressure exchange rates to either devaluate or appreciate the domestic currency.

However, the use of MCI as an indicator of monetary policy is not without controversy. As Stevens (1998) has pointed out there are some problems with its use as an indicator of monetary policy because monetary conditions may change even when the Central Bank has not modified any of its instruments as is the case when there is an unforeseen external shock to the foreign exchange market. There are other issues that we shall not discuss here because it will deviate us from our main objective. Let us just say that MCI is not an instrument of monetary policy, it is just an alternative indicator of monetary conditions that affects the economy's real sector. In that sense, it tells us whether the monetary conditions are favorable to aggregate demand or not. An increase of the MCI is seen as tighter monetary conditions, whereas a reduction of the MCI is associated to looser monetary conditions, thus favoring aggregate demand (Stevens, 1998; Costa, 2000).

We follow Stevens' formulation to estimate the MCI. We start with the estimation of the interest rate and exchange rate weights, respectively. Those weights are estimated from the following equation,

$$y_t = -\alpha r_t - \beta e_t + other\ var's \quad (5.2.1)$$

Where y is output, r is the interest rate, e is the exchange rate (US dollars per peso). The coefficients, α, β are the interest rate and exchange rate's weights, respectively. Equation (5.2.1) is a simplified version of a more general model that can include lags or other variables. The MCI, on the other hand, is defined as,

$$MCI_t = (r_t - r_0) + \frac{\beta}{\alpha}(e_t - e_0) + 100 \quad (5.2.2)$$

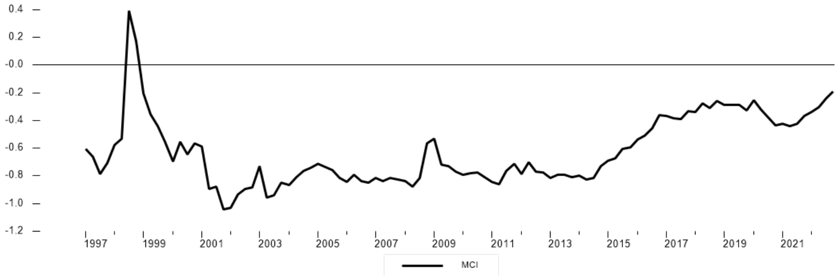
To estimate equation (5.2.1) we use first differences of: Gross Domestic Product, nominal interest rates (29 days) and exchange rate[2]. We included two dummy variables to control for the decline in GDP in 1999 and 2013. The resulting weights were: $\alpha = -0.0193245$; $\beta = -0.0310272$. We use the first quarter of 1997 as the reference period for both the interest rates and exchange rates. We obtained the following MCI,

---

[2] First difference of the exchange rates was estimated as $\nabla e = \ln(e)_t - \ln(e)_{t-1}$. Where e is the exchange rate (dollar per peso).

Figure 5.2.1

Mexico's MCI, 1997-I-2022-IV



Source: own estimates

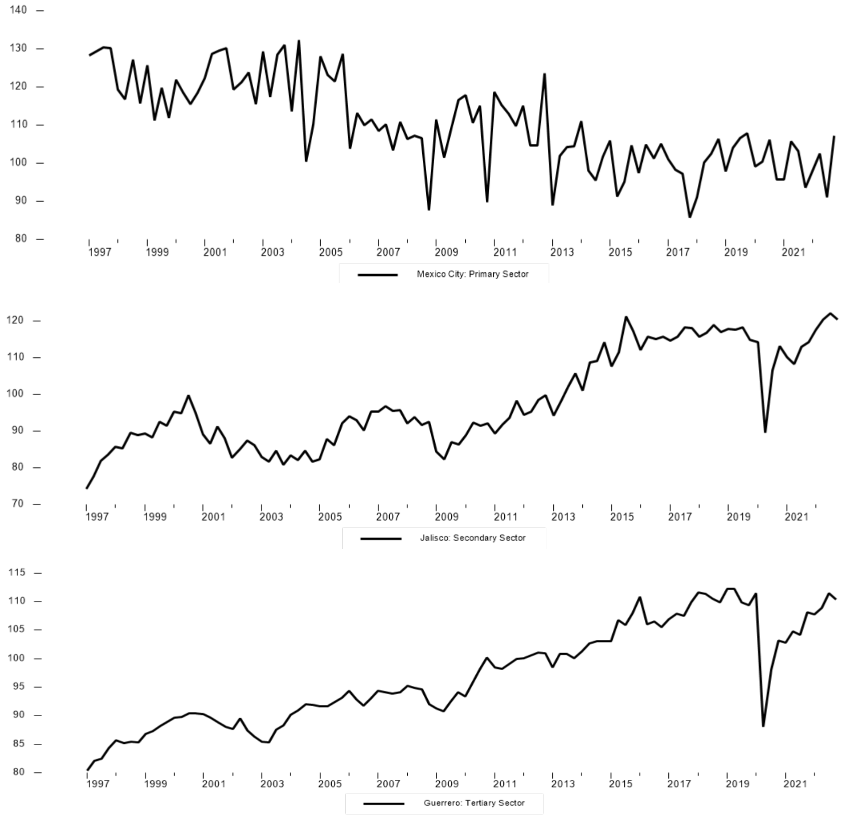
Figure 5.2.1 shows the resulting MCI during the period 1997-I-2022-IV. Compared to 1997-I, the monetary conditions in the Mexican economy can be characterized as loose until 2001-III. After an initial short period of stringent monetary condition (2001-I-2005-I), the conditions remained stable with some fluctuations until 2015-I. Beginning 2015-I, monetary conditions were stringent until the pandemic. The last two years 2020-2022, monetary conditions were somewhat looser.

We estimate the cyclical component of this indicator to evaluate its relationship with the cyclical components of the index of physical volume of five Mexican States: Mexico City, Jalisco, Guerrero, Chihuahua, and Nuevo Leon. One might expect that the relationship between the cyclical components of the MCI and economic activity may depend on the type of economic activity we are analyzing; thus, we use three indicators of economic activity, one for each economic sector: primary, secondary, and tertiary economic sector indexes. This will give us an idea about the type of relationship that there may be between monetary conditions and real sector.

To illustrate the selection of parameters in the filter, Figure 5.2.2 displays the quarterly Index of Physical Volume from 1997-I to 2022-IV of the Primary Sector for Mexico City, Secondary Sector for Jalisco, and Tertiary Sector for Guerrero. The data in Figures 5.2.1 and 5.2.2 shows non-linear behavior with no overall trend, and the concept of stochastic local trend seems better to describe the underlying dynamics of the trend, so we chose $\mu \neq 0$.

Figure 5.2.2

Quarterly Index of Physical Volume from 1997-I to 2022-IV in the Primary Sector for Mexico City, Secondary Sector for Jalisco, and Tertiary Sector for Guerrero



The suggested percentage of smoothness of 85% was used to select the smoothing parameter, based on Guerrero's (2017) guideline, resulting in $\lambda=43.05$ for a sample size of $N=104$ quarters. The smoothness percentage was kept constant among the six series which would enable adequate trend comparison.

Figure 5.2.3 shows the UCS output displays trend estimates of the MCI, $\hat{\mathbf{g}}$, (Graph 1), and cycle estimates (Graph 2). On the left side of the UCS tool, we selected the parameters to filter the series, we chose $\mu \neq 0$, a difference order (d) equals to 2, two standard deviations (SD) for the estimated error bands of the trend, noncorrelation, $\rho=0$, (see section 3.3), and percentage of smoothing S%= 85%. Our results were downloaded in .csv format for additional analyses, by just pushing the top-right button ("Download results in .csv format"). We estimated the trend and cycle of the remaining series in a similar manner.

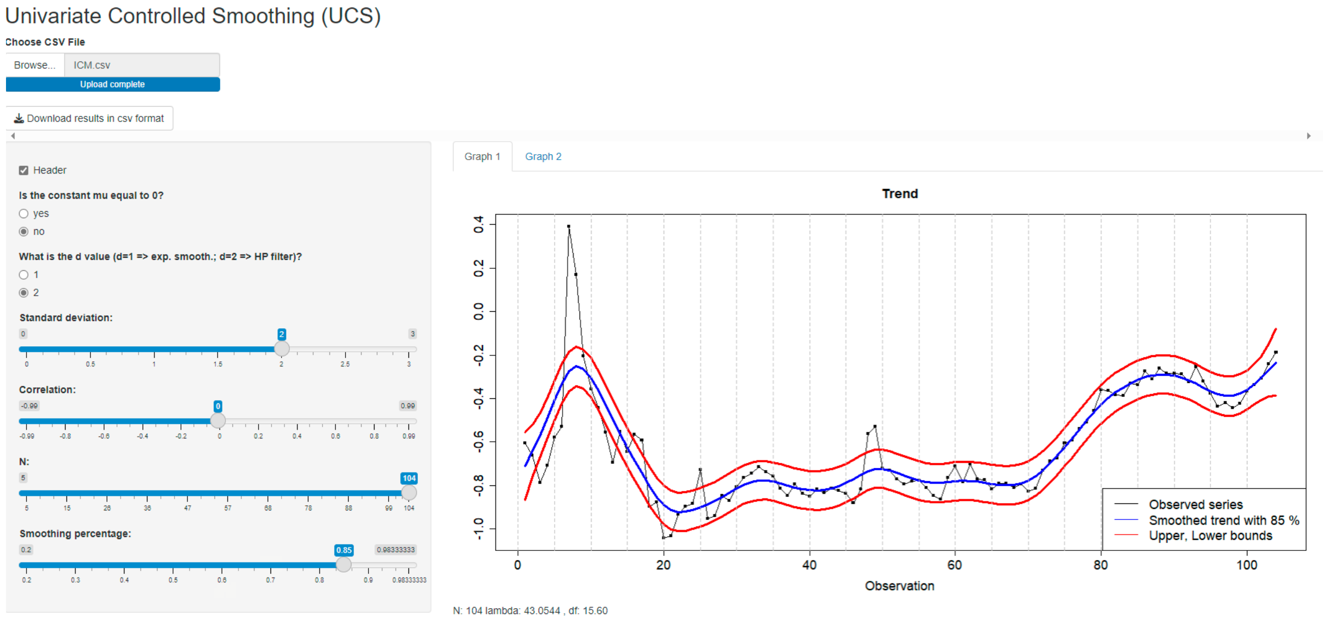
Tables 5.2.1, 5.2.2. and 5.2.3 present the estimated cross correlations between the cyclical components of the MCI and the sectors' index of physical volume for each of the five selected States.

The results suggest that in general the secondary sector is countercyclical and lags between one quarter (Nuevo Leon and Jalisco), two quarters (Mexico City and Guerrero) and six quarters (Chihuahua), although in this latter case output seems to be procyclical. The evidence in the tertiary sector points to a mixed relationship, i.e., it is procyclical for some states (Nuevo Leon, Jalisco, and Chihuahua), whereas is countercyclical in Mexico City and Guerrero. In this latter case, the economic activity lags the change in the MCI one quarter. An additional comment is that in the case of Nuevo Leon, Jalisco, and Chihuahua their tertiary sectors are antecede the change in MCI.

In the case of the primary sector, the relationship between the sector and the MCI is interesting for in all five states the change in the primary sector antecedes the change in the MCI. Furthermore, we found evidence that there is a positive relationship between the sector's evolution and MCI (Jalisco, Chihuahua) and a negative relationship in states like Mexico City, Guerrero.

Figure 5.2.3

Defining the Model's parameters



Univariate Controlled Smoothing (UCS)

Choose CSV File

Browse... ICM.csv

Upload complete

Download results in csv format

☑ Header

Is the constant mu equal to 0?

○ yes

● no

What is the d value (d=1 => exp. smooth.; d=2 => HP filter)?

○ 1

● 2

Standard deviation: 2

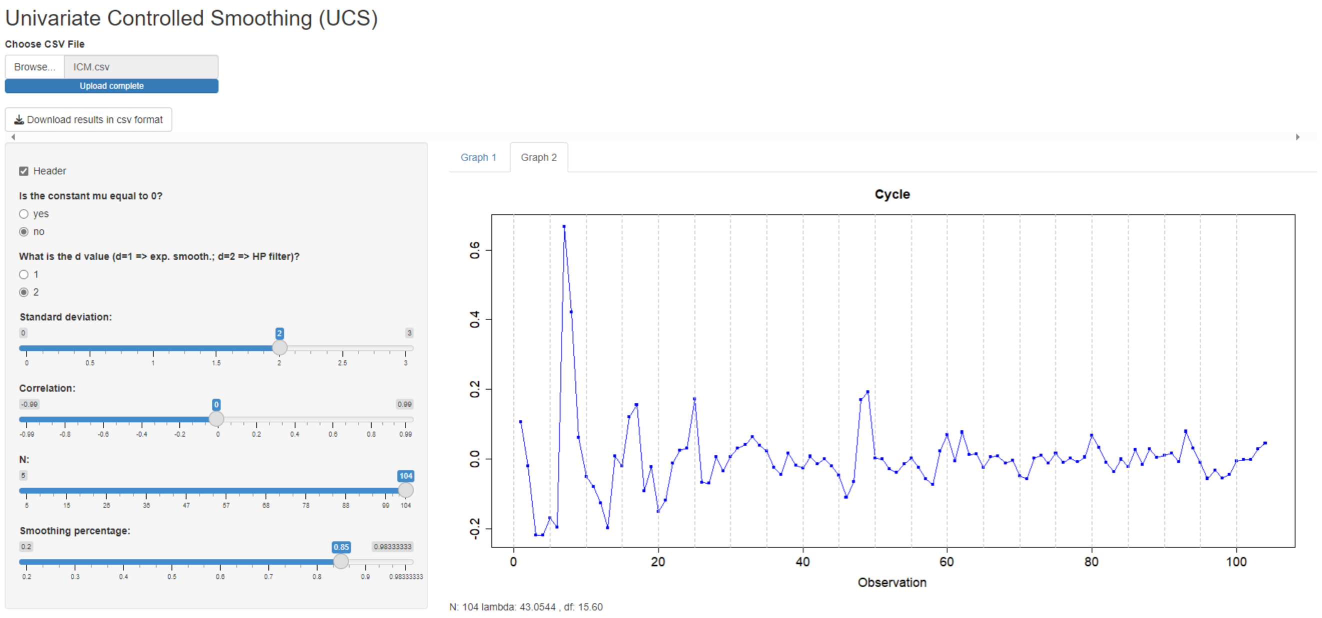
Correlation: 0

N: 104

Smoothing percentage: 0.85

N: 104 lambda: 43.0544 , df: 15.60

Graph 1
Graph 2
0
3
0
0.5
1
1.5
2
2.5
3
-0.99
0.99
-0.99
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
0.99
5
5
15
26
36
47
57
68
78
88
99
104
0.2
0.98333333
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.98333333
Trend
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
0
20
40
60
80
100
Observation
Observed series
Smoothed trend with 85 %
Upper, Lower bounds



Univariate Controlled Smoothing (UCS)

Choose CSV File

Browse... ICM.csv

Upload complete

Download results in csv format

☑ Header

Is the constant mu equal to 0?

○ yes
● no

What is the d value (d=1 => exp. smooth.; d=2 => HP filter)?

○ 1
● 2

Standard deviation: 2

Correlation: 0

N: 104

Smoothing percentage: 0.85

N: 104 lambda: 43.0544 , df: 15.60

Graph 1
Graph 2
0
3
0
0.5
1
1.5
2
2.5
3
-0.99
0.99
-0.99
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
0.99
5
104
5
15
26
36
47
57
68
78
88
99
104
0.2
0.98333333
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.98333333
Cycle
0.6
0.4
0.2
0.0
-0.2
0
20
40
60
80
100
Observation

Source: own estimates

Table 5.2.1

Cross Correlation between the cyclical components of Monetary Condition Index and IVF of Primary sector

State	X(t-8)	X(t-7)	X(t-6)	X(t-5)	X(t-4)	X(t-3)	X(t-2)	X(t-1)	X(t)	X(t+1)	X(t+2)	X(t+3)	X(t+4)	X(t+5)	X(t+6)	X(t+7)	X(t+8)
NL	0.05	-0.01	-0.09	-0.06	-0.04	0.09	0.04	0.09	0.13	0.01	0.00	-0.11	-0.12	-0.04	-0.07	-0.07	0.01
Jal	0.06	0.01	0.01	0.01	-0.14	0.01	0.04	0.14	0.12	0.03	-0.08	-0.20	0.00	-0.04	0.04	0.05	-0.03
Cdmx	0.03	-0.02	-0.03	0.12	0.01	0.03	-0.09	-0.14	0.02	0.01	0.15	0.02	0.05	-0.02	-0.04	-0.21	-0.08
Gro	0.00	-0.02	0.00	0.08	-0.09	0.03	0.03	-0.03	0.05	-0.05	0.03	-0.06	0.03	0.08	0.01	-0.04	-0.02
Chih	0.09	0.17	0.13	0.08	-0.13	-0.13	-0.14	-0.07	0.09	0.09	0.12	-0.01	-0.04	-0.10	-0.08	-0.01	0.04

Source: Own estimates

Table 5.2.2

Cross Correlation between the cyclical components of Monetary Condition Index and IVF of Secondary sector

State	X(t-8)	X(t-7)	X(t-6)	X(t-5)	X(t-4)	X(t-3)	X(t-2)	X(t-1)	X(t)	X(t+1)	X(t+2)	X(t+3)	X(t+4)	X(t+5)	X(t+6)	X(t+7)	X(t+8)
NL	0.00	-0.02	0.00	0.04	0.10	0.06	0.06	0.00	-0.02	-0.12	-0.04	-0.05	-0.02	-0.03	0.02	0.09	0.07
Jal	0.00	-0.07	-0.06	0.02	0.07	0.07	0.13	0.10	-0.01	-0.20	-0.15	-0.09	-0.03	0.01	0.13	0.14	0.16
Cdmx	-0.02	-0.16	-0.13	0.05	0.15	0.05	-0.01	0.09	0.06	-0.12	-0.16	-0.02	0.01	-0.02	0.03	0.12	0.10
Gro	0.00	-0.07	-0.08	0.08	0.14	0.08	0.09	0.02	-0.03	-0.14	-0.16	0.02	0.06	0.03	-0.01	-0.02	-0.08
Chih	-0.01	-0.15	-0.11	-0.04	0.00	0.04	0.08	0.08	0.03	-0.04	-0.02	-0.11	-0.15	0.02	0.22	0.18	0.20

Source: Own estimates

Table 5.2.3

Cross Correlation between the cyclical components of Monetary Condition Index and IVF of Tertiary sector

State	**X(t-8)**	**X(t-7)**	**X(t-6)**	**X(t-5)**	**X(t-4)**	**X(t-3)**	**X(t-2)**	**X(t-1)**	**X(t)**	**X(t+1)**	**X(t+2)**	**X(t+3)**	**X(t+4)**	**X(t+5)**	**X(t+6)**	**X(t+7)**	**X(t+8)**
NL	-0.01	-0.02	-0.02	0.03	0.05	0.13	0.15	0.06	-0.07	-0.21	-0.11	-0.04	0.01	0.05	0.05	0.07	0.05
Jal	-0.03	-0.02	-0.03	0.03	0.04	0.13	0.15	0.04	-0.10	-0.18	-0.08	-0.01	0.04	0.04	0.02	0.01	0.03
Cdmx	-0.04	-0.03	-0.01	0.01	0.02	0.10	0.14	0.06	-0.07	-0.18	-0.10	-0.05	0.03	0.07	0.06	0.09	0.04
Gro	-0.05	-0.04	-0.05	0.02	0.10	0.16	0.13	0.02	-0.08	-0.18	-0.08	-0.03	0.01	0.04	0.03	0.08	0.03
Chih	-0.04	-0.07	-0.07	0.00	0.05	0.17	0.23	0.11	-0.05	-0.22	-0.17	-0.12	-0.06	0.04	0.16	0.13	0.09

Source: Own estimates

5.3. Case 3. The long-run evolution of oil prices

Energy producers and consumers often analyze the causes of price volatility of oil, coal, and other resources. Producers are interested in price volatility for strategic planning and assessing investment decisions, such as resource exploration, reserve development, and production. Industrial consumers, like petrochemical companies or electric utilities, also share this interest because oil, coal, and natural gas are essential input costs that can influence investment choices. For instance, they may need to decide between building an oil- or coal-powered plant for an electric utility or selecting products to manufacture.

The growth rate of prices may reflect a depletion of resources or technological changes. Some may assume that prices will continue to grow from their current level, following a random walk with a drift. Others may assume that prices will revert to a trend line that either increases or decreases. This behavior would be consistent with the idea that the resource is produced and sold in a competitive market, so the price will eventually return to the long-run marginal cost. This cost is likely to change slowly, indicating that price shocks are temporary. Over a long enough period, prices are more likely to be mean-reverting rather than random.

In this example, we focus on oil prices. The literature on the sources of oil price fluctuations agrees on the role of global demand conditions. Elekdag & Laxton (2007), Elekdag et al. (2008), and Kilian (2008) emphasize the importance of aggregate demand. Barsky & Kilian (2002) and Kilian (2009) argue that global demand played a significant role in the recent oil price episode and was also influenced by supply conditions. Hamilton (2009) mentioned that disruptions in crude oil production, caused mostly by geopolitical events, played a large part in determining oil price dynamics; additionally, he pointed out the role of speculation on the oil market as an extra source of volatility.

Based on the above discussion, numerous empirical studies have tried to understand the fluctuating nature of oil prices. Sadorsky (1999) suggests that a stochastic trend should be taken into account due to the rise in the volatility of oil prices after the mid-1980s. Similarly, Morana (2001) demonstrates alternating periods of high volatility in oil price changes, which can best be analyzed using a stochastic local trend.

In this example, we use daily Brent oil spot prices ranging from January 3, 2000– June 23, 2023. Throughout the day, the price of Brent crude in the

Spot market fluctuates due to various factors, such as geopolitical and economic events and changes in supply and demand. It is worth noting that the daily prices published by the *Energy Information Administration* represent the market closing price, which is considered the day's benchmark. The price of Brent crude is measured in dollars per barrel (USD/B) in the Spot market.

Figure 5.3.1 displays the plotted oil price level and its absolute rate of change. Two significant observations can be made. Firstly, the oil price level does not follow a global trend, and the concept of stochastic local trend seems better to describe the underlying dynamics of the oil price series. Secondly, the absolute rate of oil price change shows alternating periods of high volatility followed by relative tranquility. These two features are commonly found in speculative asset prices.

Therefore, we will examine how oil prices change over time by breaking them down into its trend and cycle components. The trend component follows a stochastic trend with a drift as follows,

$$price_{oil,t} = trend_{oil,t} + cycle_{oil.t}$$

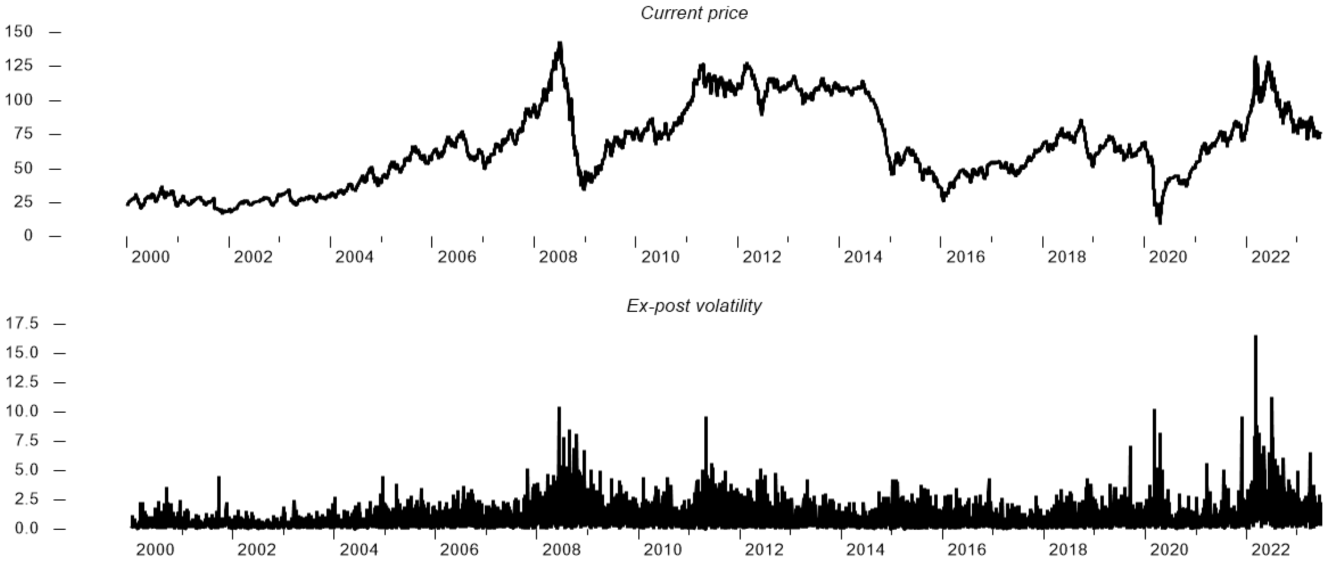
With a time series specification for the trend as:

$$trend_{oil,t} = \mu + trend_{oil,t-1} + \varepsilon_t \quad \text{or} \quad \nabla trend_{oil,t} = \mu + \varepsilon_t \quad \text{for } t = 2, \ldots, N$$

Following Guerrero's (2017) guideline, a smoothness percentage of 90% was set to select the corresponding smoothing parameter from the information in Table 4.2.2. For a sample size of *N= 6122* daily data, the smoothing parameter value is obtained from $\lambda = 6122/(-0.366093+0.040246 * 6122) = 24.9$, as indicated in Table 4.2.2. Figure 5.3.2 Panel (a) shows the daily time series and its trend. Panel (b) exhibits the weekly time series and its trend, constructed from the daily data. It is important to note that this is a series of stocks, and the sample now comprises *N=1224* weekly observations, considering five-day weeks. The smoothing constant for the weekly series is $\lambda^*_5 = 4.98$ for S%=90%, as obtained from (4.3.8) by $\lambda = 5\lambda^*_5$. Lastly, Panel (c) shows the monthly time series and its trend, also built from the daily data. The sample size for this series is *N=281* observations, and the value of λ for monthly data is $\lambda^*_{20} = 1.24$ for *S%=90%,* obtained from (4.3.8) by $\lambda = 20\lambda^*_5$. By visually inspecting Panels (a), (b), and (c), we can observe that the resulting trends with the same smoothness percentage display essentially the same dynamic behavior, regardless of the frequency of data observation.

Figure 5.3.1

Level and Rate of Change of Oil Price



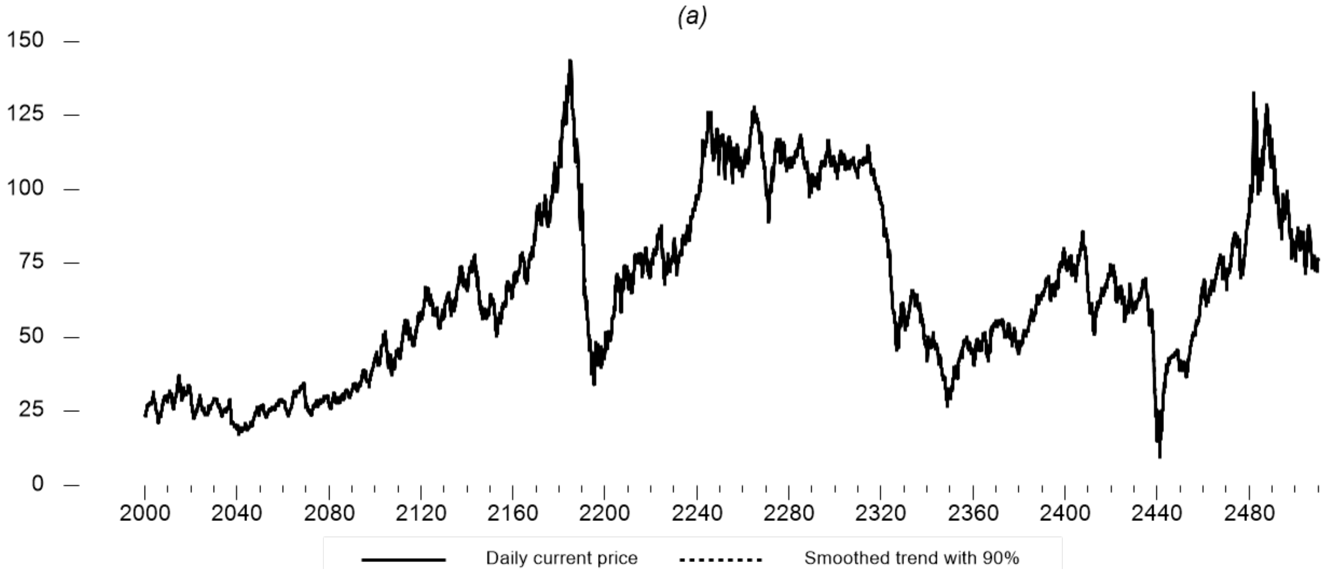
Source: own estimates

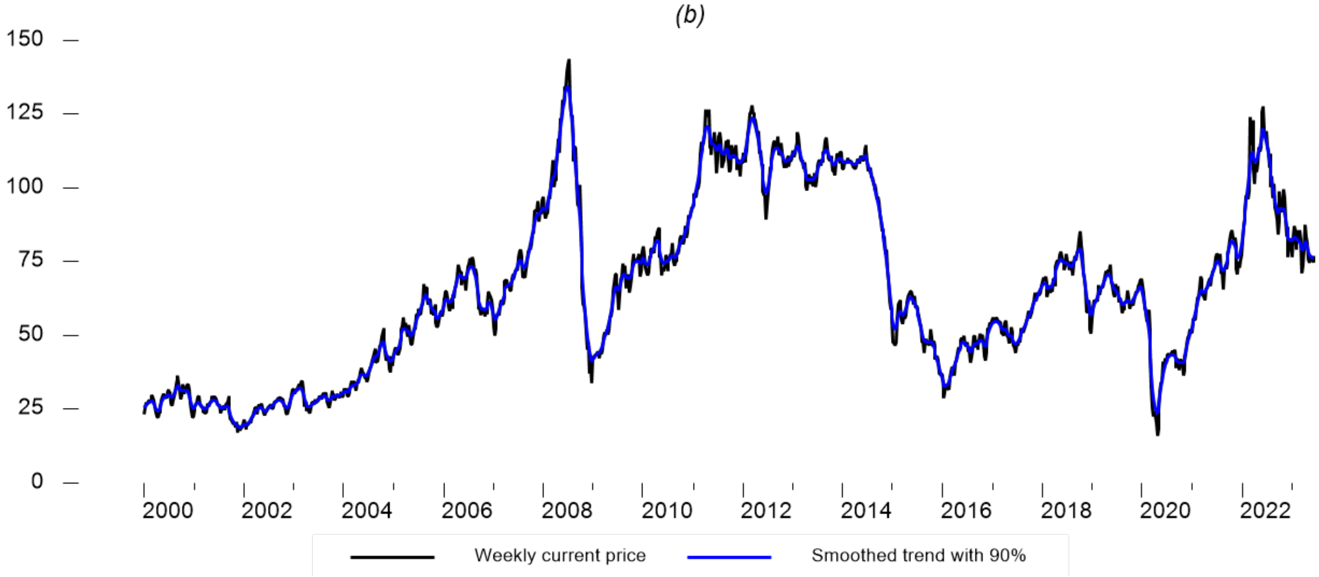
Figure 5.3.2. Panel (a), daily current prices and estimated trend with S%=90% (λ = 24.9), sample size *N= 6122.* Panel (b), weekly current prices and estimated trend with S%= 90% (λ^*_5 = 4.98, *N=1224* observations. Panel (c), monthly current prices and estimated trend with S%= 90% (λ^*_{20} = 1.24, *N=281* observations.

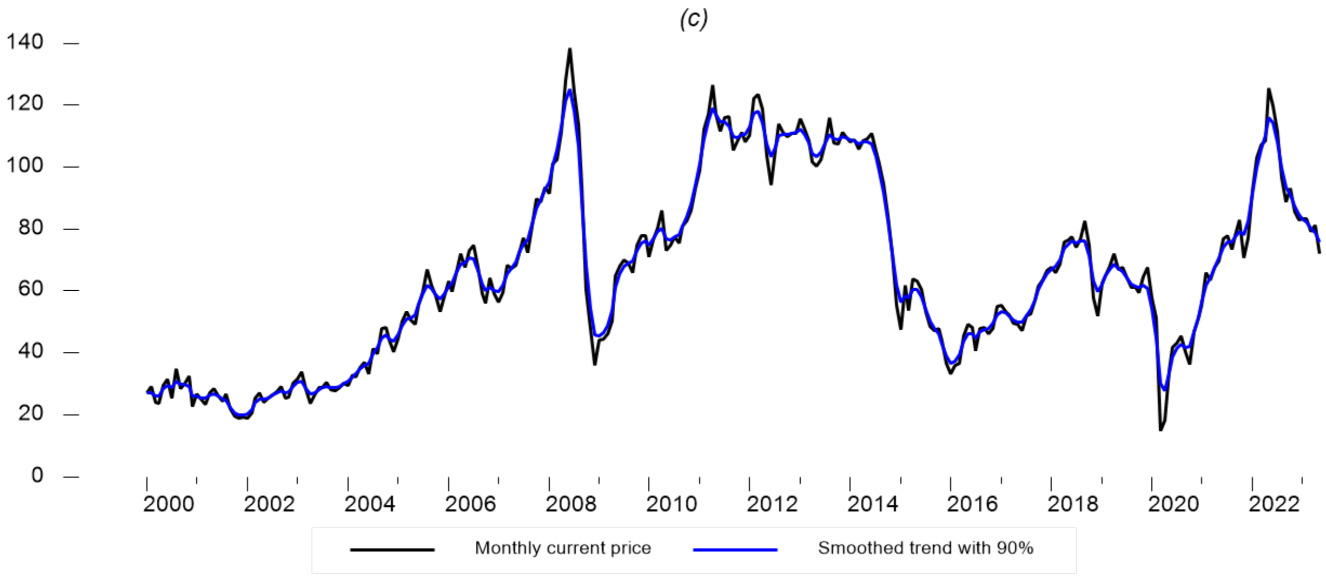
In his analysis of oil prices between 2007 and 2008, Hamilton (2009) suggests that three factors played a significant role in shaping the dynamics of the market. Firstly, the global supply of crude oil was affected by the failed attempt to increase productive capacity between 2005 and 2007. The pressure in each oil field eventually reduces over time, leading to a decrease in production. Secondly, the demand for crude oil continued to increase globally, particularly in China, which went from importing 800,000 barrels per day in 1998 to 3.7 million barrels in 2007. As a result, the market became imbalanced by 2007, leading to a significant reduction in crude oil inventories worldwide. Finally, ccording to Hamilton's research, a speculative bubble was formed during those years that eventually caused prices to skyrocket. In other words, speculation in financial markets was one of the reasons that led to the historical maximum price of oil. On the other hand, Masteres (2008) stated that the rise in price since 2007 was due to investors who started purchasing oil as a financial asset rather than a consumer good.

In early July 2008, oil prices reached their all-time high, but then they fell sharply to $40 per barrel in December of the same year. This was due to the global financial crisis, which caused a decrease in economic activity worldwide. As a result, oil demand also decreased significantly (OPEC, 2008).

According to Figure 5.3.2, the exponential filter accurately identifies the trend dynamics between 2007 and 2009. This is because during this period, there was a significant episode of price growth from February 2007 to June 2008, followed by a period of falling prices from July 2008 to February 2009. These findings are consistent with what we can observe in the historical price series.


(a)
150
125
100
75
50
25
0
2000
2040
2080
2120
2160
2200
2240
2280
2320
2360
2400
2440
2480
Daily current price
Smoothed trend with 90%


(b)
150
125
100
75
50
25
0
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Weekly current price
Smoothed trend with 90%



Source: own estimates

The drop in oil prices that occurred in mid-2014 is believed to be due to the advancements in oil production technology. The United States' oil supply grew significantly due to the development of hydraulic fracturing (fracking), enabling it to exploit vast reserves in Texas and New Mexico in the Permian Basin. This event is known as the fracking revolution (The USA Shale Oil Revolution). By 2014, crude oil production in the United States had increased by 80% compared to 2008 (Bordoff & Losz, 2015).

The exponential filter results suggest that from June 2014 to December 2015, the state of trend decline dominated the process. This trend's behavior is consistent with the information mentioned above.

In 2020, the price of crude oil fell considerably due to two primary reasons. Firstly, the COVID-19 pandemic caused a reduction in economic activity worldwide. Secondly, a price war erupted between Russia and Saudi Arabia, as they could not agree on reducing the volume of crude oil produced. As a result, both countries increased their production, which led to a market flood and a subsequent price drop. The prices of various futures even reached negative values, indicating the lack of storage capacity and the urgent need to balance the market (Johnston, 2022). The exponential filter accurately recognizes the impact of the COVID-19 pandemic on the market, as Figure 5.3.2 shows the oil trend price became dominated by falling prices from December 2019 to May 2020, consistent with the historical series of prices. The lowest price was recorded in March 2020.

To summarize, the exponential filter accurately represents the significant events that have impacted the long-term trends in oil prices, regardless of the data's periodicity. For instance, during the price slump, we can observe that the trend reduced during the 2008 financial crisis, the emergence of fracking in 2014, the COVID-19 pandemic, and the start of the Russian invasion of Ukraine. Graph 5.3.2 demonstrates that the trend estimated with the exponential filter encompasses all the relevant price drops since 2000. It also correctly describes the behavior of price stability and growth throughout the analysis period.

Appendix
Technical details

Proofs of (3.1.8)

By applying GLS to (3.1.7) we obtain the following estimation equation.

$$(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_v^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{Y} \\ \mu \mathbf{1}_{N-d} \end{pmatrix} = (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_v^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} \begin{pmatrix} I_N \\ K_d \end{pmatrix} \widehat{\boldsymbol{g}}$$

We can solve for $\hat{\boldsymbol{g}}$ as follow.

$$\widehat{\boldsymbol{g}} = \left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_v^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} \begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_v^2 V_{N\text{x}N} & 0_{N\text{x}N-d} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{Y} \\ \mu \mathbf{1}_{N-d} \end{pmatrix}$$

$$= (\sigma_v^{-2} V_{N\text{x}N}^{-1} + \sigma_\varepsilon^{-2} K_d{}' K_d)^{-1} (\sigma_v^{-2} V_{N\text{x}N}^{-1} \boldsymbol{Y} + \sigma_\varepsilon^{-2} K_d{}' \mu \mathbf{1}_{N-d})$$

$$= \sigma_v^2 (V_{N\text{x}N}^{-1} + \sigma_v^2 \sigma_\varepsilon^{-2} K_d{}' K_d)^{-1} \sigma_v^{-2} (V_{N\text{x}\text{X}N}^{-1} \boldsymbol{Y} + \sigma_v^2 \sigma_\varepsilon^{-2} K_d{}' \mu \mathbf{1}_{N-d})$$

$$= (V_{N\text{x}N}^{-1} + \lambda K_d{}' K_d)^{-1} (V_{N\text{x}N}^{-1} \boldsymbol{Y} + \lambda \mu K_d{}' \mathbf{1}_{N-d}),$$

where the smoothing parameter is defined as $\lambda = \sigma_v^2 \sigma_\varepsilon^{-2} = \sigma_v^2 / \sigma_\varepsilon^2$.
The GLS estimator $\hat{\boldsymbol{g}}$ is unbiased, that is,

$$\widehat{\boldsymbol{g}} = \left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{N-dxN} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{Y} \\ \mu\mathbf{1}_{N-d} \end{pmatrix}$$

$$= \left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\left(\begin{pmatrix} I_N \\ K_d \end{pmatrix}\boldsymbol{g} + \begin{pmatrix} \boldsymbol{\nu} \\ -\boldsymbol{\varepsilon} \end{pmatrix}\right)$$

$$= \left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\boldsymbol{g}$$

$$+ \left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{\nu} \\ -\boldsymbol{\varepsilon} \end{pmatrix}$$

$$= \boldsymbol{g} + \left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{\nu} \\ -\boldsymbol{\varepsilon} \end{pmatrix}$$

Therefore,

$$E(\widehat{\boldsymbol{g}}) = \boldsymbol{g} + \left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} E\left[\begin{pmatrix} \boldsymbol{\nu} \\ -\boldsymbol{\varepsilon} \end{pmatrix}\right] = \boldsymbol{g}$$

Proofs of (3.1.9)

The variance-covariance matrix of $\hat{g}$ is given by

$$Var(\widehat{\boldsymbol{g}}) = E[(\widehat{\boldsymbol{g}} - \boldsymbol{g})(\widehat{\boldsymbol{g}} - \boldsymbol{g})'] = E\left\{\left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{\nu} \\ -\boldsymbol{\varepsilon} \end{pmatrix}\right\}$$

$$\left\{\left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{\nu} \\ -\boldsymbol{\varepsilon} \end{pmatrix}\right\}'$$

$$= \left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} E\left\{\begin{pmatrix} \boldsymbol{\nu} \\ -\boldsymbol{\varepsilon} \end{pmatrix}(\boldsymbol{\nu}' \quad -\boldsymbol{\varepsilon}')\right\}$$

$$\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}}$$

$$= \left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}$$

$$\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}}$$

$$= \left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}} (I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}$$

$$\left[(I_N{}' \quad K_d{}')\begin{pmatrix} \sigma_\nu^2 V_{N\mathrm{x}N} & 0_{N\mathrm{x}(N-d)} \\ 0_{(N-d)\mathrm{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}\begin{pmatrix} I_N \\ K_d \end{pmatrix}\right]^{\mathbf{-1}}$$

$$= \left[\begin{pmatrix} I_N' & K_d' \end{pmatrix} \begin{pmatrix} \sigma_\nu^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} \begin{pmatrix} I_N \\ K_d \end{pmatrix} \right]^{\mathbf{-1}} = (\sigma_\nu^{-2} V_{N\text{x}N}^{-1} + \sigma_\varepsilon^{-2} K_d' K_d)^{-1}$$

$$= \sigma_\nu^2 (V_{N\text{x}N}^{-1} + \lambda K_d' K_d)^{-1}$$

Proof of (3.1.12)

To proof (3.1.12), first note that

$$Var\begin{pmatrix} \boldsymbol{\nu} \\ -\boldsymbol{\varepsilon} \end{pmatrix} = \begin{pmatrix} \sigma_\nu^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix} = \sigma_\nu^2 \begin{pmatrix} V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & (\sigma_\varepsilon^2/\sigma_\nu^2)\, I_{N-d} \end{pmatrix}$$

$$= \sigma_\nu^2 \begin{pmatrix} V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \lambda^{-1} I_{N-d} \end{pmatrix} = \sigma_\nu^2 \Omega,$$

and define the residual vector as

$$\begin{pmatrix} \hat{\boldsymbol{\nu}} \\ -\hat{\boldsymbol{\varepsilon}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y} \\ \mu \mathbf{1}_{N-d} \end{pmatrix} - \begin{pmatrix} \widehat{\boldsymbol{Y}} \\ \mu \mathbf{1}_{N-d} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y} \\ \mu \mathbf{1}_{N-d} \end{pmatrix} - \begin{pmatrix} I_N \\ K_d \end{pmatrix} \widehat{\boldsymbol{g}} \text{ with } \hat{\boldsymbol{\nu}} = \boldsymbol{Y} - \widehat{\boldsymbol{g}}$$

and $-\hat{\boldsymbol{\varepsilon}} = \mu \mathbf{1}_{N-d} - K_d \widehat{\boldsymbol{g}}$

$$= \begin{pmatrix} \boldsymbol{Y} \\ \mu \mathbf{1}_{N-d} \end{pmatrix} - \begin{pmatrix} I_N \\ K_d \end{pmatrix} \left\{ \left[\begin{pmatrix} I_N' & K_d' \end{pmatrix} \begin{pmatrix} \sigma_\nu^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} \begin{pmatrix} I_N \\ K_d \end{pmatrix} \right]^{\mathbf{-1}} \begin{pmatrix} I_N' & K_d' \end{pmatrix} \begin{pmatrix} \sigma_\nu^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{Y} \\ \mu \mathbf{1}_{N-d} \end{pmatrix} \right\}$$

$$= \left\{ I_{2N-d} - \begin{pmatrix} I_N \\ K_d \end{pmatrix} \left[\begin{pmatrix} I_N' & K_d' \end{pmatrix} \begin{pmatrix} \sigma_\nu^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} \begin{pmatrix} I_N \\ K_d \end{pmatrix} \right]^{\mathbf{-1}} \begin{pmatrix} I_N' & K_d' \end{pmatrix} \begin{pmatrix} \sigma_\nu^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} \right\} \begin{pmatrix} \boldsymbol{Y} \\ \mu \mathbf{1}_{N-d} \end{pmatrix}$$

$$= (I_{2N-d} - M) \begin{pmatrix} \boldsymbol{Y} \\ \mu \mathbf{1}_{N-d} \end{pmatrix}$$

with

$$M = \begin{pmatrix} I_N \\ K_d \end{pmatrix} \left[\begin{pmatrix} I_N' & K_d' \end{pmatrix} \begin{pmatrix} \sigma_\nu^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1} \begin{pmatrix} I_N \\ K_d \end{pmatrix} \right]^{\mathbf{-1}} \begin{pmatrix} I_N' & K_d' \end{pmatrix} \begin{pmatrix} \sigma_\nu^2 V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \sigma_\varepsilon^2 I_{N-d} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} I_N \\ K_d \end{pmatrix} (\sigma_\nu^{-2} V_{N\text{x}N}^{-1} + \sigma_\varepsilon^{-2} K_d' K_d)^{-1} (\sigma_\nu^{-2} V_{N\text{x}N}^{-1} + \sigma_\varepsilon^{-2} K_d')$$

$$= \begin{pmatrix} I_N \\ K_d \end{pmatrix} \sigma_\nu^2 (V_{N\text{x}N}^{-1} + (\sigma_\nu^2/\sigma_\varepsilon^2)\, K_d' K_d)^{-1} \sigma_\nu^{-2} (V_{N\text{x}N}^{-1} \quad (\sigma_\nu^2/\sigma_\varepsilon^2)\, K_d')$$

$$= \begin{pmatrix} I_N \\ K_d \end{pmatrix} (V_{N\text{x}N}^{-1} + \lambda K_d' K_d)^{-1} (V_{N\text{x}N}^{-1} \quad \lambda K_d').$$

This matrix has the important property of idempotency:

$$MM = \binom{I_N}{K_d}(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d)^{-1}(V_{N\text{x}N}^{-1} \quad \lambda K_d{'})\binom{I_N}{K_d}(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d)^{-1}(V_{N\text{x}N}^{-1} \quad \lambda K_d{'})$$

$$= \binom{I_N}{K_d}(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d)^{-1}(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d)(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d)^{-1}(V_{N\text{x}N}^{-1} \quad \lambda K_d{'})$$

$$= \binom{I_N}{K_d}(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d)^{-1}(V_{N\text{x}N}^{-1} \quad \lambda K_d{'}) = M.$$

Moreover, M is a projection matrix, that is

$$M\binom{I_N}{K_d} = \binom{I_N}{K_d}(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d)^{-1}(V_{N\text{x}N}^{-1} \quad \lambda K_d{'})\binom{I_N}{K_d}$$

$$= \binom{I_N}{K_d}(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d)^{-1}(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d) = \binom{I_N}{K_d},$$

therefore, the residual vector can be expressed as follows:

$$\boldsymbol{e}* = \binom{\hat{\boldsymbol{\nu}}}{-\hat{\boldsymbol{\varepsilon}}} = (I_{2N-d} - M)\binom{\boldsymbol{Y}}{\mu\mathbf{1}_{N-d}} = (I_{2N-d} - M)\left[\binom{I_N}{K_d}\boldsymbol{g} + \binom{\boldsymbol{\nu}}{-\boldsymbol{\varepsilon}}\right]$$

$$= (I_{2N-d} - M)\binom{I_N}{K_d}\boldsymbol{g} + (I_{2N-d} - M)\binom{\boldsymbol{\nu}}{-\boldsymbol{\varepsilon}} = \left[\binom{I_N}{K_d} - M\binom{I_N}{K_d}\right]\boldsymbol{g} + (I_{2N-d} - M)\binom{\boldsymbol{\nu}}{-\boldsymbol{\varepsilon}}$$

$$= (I_{2N-d} - M)\binom{\boldsymbol{\nu}}{-\boldsymbol{\varepsilon}}.$$

(A.1)

On the other hand, given that *tr*(*AB*)=*tr*(*BA*), where *tr* stands for the trace of a matrix, we have that

$$tr(M) = tr\left(\binom{I_N}{K_d}\left(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d\right)^{-1}(V_{N\text{x}N}^{-1} \quad \lambda K_d{'})\right) = tr(V_{N\text{x}N}^{-1} \quad \lambda K_d{'})\binom{I_N}{K_d}\left(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d\right)^{-1}$$

$$= tr(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d)(V_{N\text{x}N}^{-1} + \lambda K_d{'}K_d)^{-1} = tr(I_n) = N$$

(A.2)

Now, use (A.1), (A.2) and the fact that a scalar is its own trace to obtain:

$$E(\boldsymbol{e}*'\Omega^{-1}\boldsymbol{e}*) = E\left[(\boldsymbol{\nu}' \quad -\boldsymbol{\varepsilon}')(I_{2N-d} - M')\Omega^{-1}(I_{2N-d} - M)\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}\right]$$

$$= E\left\{(\boldsymbol{\nu}' \quad -\boldsymbol{\varepsilon}')\left[I_{2N-d} - \begin{pmatrix}V_{N\mathrm{x}N}^{-1}\\ \lambda K_d\end{pmatrix}(V_{N\mathrm{x}N}^{-1} + \lambda K_d'K_d)^{-1}(I_N \quad K_d')\right]\Omega^{-1}(I_{2N-d} - M)\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}\right\}$$

$$= E\left\{(\boldsymbol{\nu}' \quad -\boldsymbol{\varepsilon}')\left[\Omega^{-1} - \begin{pmatrix}I_N\\ \lambda K_d\end{pmatrix}(I_N + \lambda K_d'K_d)^{-1}(I_N \quad K_d')\Omega^{-1}\right](I_{2N-d} - M)\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}\right\}$$

$$= E\left\{(\boldsymbol{\nu}' \quad -\boldsymbol{\varepsilon}')\left[\Omega^{-1} - \begin{pmatrix}V_{N\mathrm{x}N}^{-1}\\ \lambda K_d\end{pmatrix}\left(V_{N\mathrm{x}N}^{-1} + \lambda K_d'K_d\right)^{-1}(I_N \quad K_d')\begin{pmatrix}V_{N\mathrm{x}N} & 0_{N\mathrm{x}N-d}\\ 0_{(N-d)xN} & \lambda^{-1}I_{N-d}\end{pmatrix}^{-1}\right](I_{2N-d} - M)\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}\right\}$$

$$= E\left\{(\boldsymbol{\nu}' \quad -\boldsymbol{\varepsilon}')\left[\Omega^{-1} - \begin{pmatrix}V_{N\mathrm{x}N}^{-1}\\ \lambda K_d\end{pmatrix}(V_{N\mathrm{x}N}^{-1} + \lambda K_d'K_d)^{-1}(V_{N\mathrm{x}N}^{-1} \quad \lambda K_d')\right](I_{2N-d} - M)\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}\right\}$$

$$= E\left\{tr\left\{(\boldsymbol{\nu}' \quad -\boldsymbol{\varepsilon}')\left[\Omega^{-1} - \begin{pmatrix}V_{N\mathrm{x}XN}^{-1}\\ \lambda K_d\end{pmatrix}(V_{N\mathrm{x}N}^{-1} + \lambda K_d'K_d)^{-1}(V_{N\mathrm{x}N}^{-1} \quad \lambda K_d')\right](I_{2N-d} - M)\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}\right\}\right\}$$

$$= E\left\{tr\left\{\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}(\boldsymbol{\nu}' \quad -\boldsymbol{\varepsilon}')\left[\Omega^{-1} - \begin{pmatrix}V_{N\mathrm{x}XN}^{-1}\\ \lambda K_d\end{pmatrix}(V_{N\mathrm{x}N}^{-1} + \lambda K_d'K_d)^{-1}(V_{N\mathrm{x}XN}^{-1} \quad \lambda K_d')\right](I_{2N-d} - M)\right\}\right\}$$

$$= \left\{tr\left\{E\begin{pmatrix}\boldsymbol{\nu}\\ -\boldsymbol{\varepsilon}\end{pmatrix}(\boldsymbol{\nu}' \quad -\boldsymbol{\varepsilon}')\left[\Omega^{-1} - \begin{pmatrix}V_{N\mathrm{x}XN}^{-1}\\ \lambda K_d\end{pmatrix}(V_{N\mathrm{x}N}^{-1} + \lambda K_d'K_d)^{-1}(V_{N\mathrm{x}N}^{-1} \quad \lambda K_d')\right](I_{2N-d} - M)\right\}\right\}$$

$$= tr\left\{\sigma_\nu^2\Omega\left[\Omega^{-1} - \begin{pmatrix}V_{N\mathrm{x}XN}^{-1}\\ \lambda K_d\end{pmatrix}(V_{N\mathrm{x}N}^{-1} + \lambda K_d'K_d)^{-1}(V_{N\mathrm{x}N}^{-1} \quad \lambda K_d')\right](I_{2N-d} - M)\right\}$$

$$= \sigma_\nu^2 tr\left\{\left[\Omega\Omega^{-1} - \Omega\begin{pmatrix}V_{N\mathrm{x}XN}^{-1}\\ \lambda K_d\end{pmatrix}(V_{N\mathrm{x}N}^{-1} + \lambda K_d'K_d)^{-1}(V_{N\mathrm{x}N}^{-1} \quad \lambda K_d')\right](I_{2N-d} - M)\right\}$$

$$= \sigma_\nu^2 tr\left\{\left[I_{2N-d} - \begin{pmatrix}V_{N\mathrm{x}N} & 0_{Nx(N-d)}\\ 0_{(N-d)\mathrm{x}N} & \lambda^{-1}I_{N-d}\end{pmatrix}\begin{pmatrix}V_{N\mathrm{x}XN}^{-1}\\ \lambda K_d\end{pmatrix}(V_{N\mathrm{x}N}^{-1} + \lambda K_d'K_d)^{-1}(V_{N\mathrm{x}N}^{-1} \quad \lambda K_d')\right](I_{2N-d} - M)\right\}$$

$$= \sigma_\nu^2 tr\left\{\left[I_{2N-d} - \begin{pmatrix}I_N\\ K_d\end{pmatrix}(V_{N\mathrm{x}N}^{-1} + \lambda K_d'K_d)^{-1}(V_{N\mathrm{x}N}^{-1} \quad \lambda K_d')\right](I_{2N-d} - M)\right\}$$

$$= \sigma_\nu^2 tr\{(I_{2N-d} - M)(I_{2N-d} - M)\} = \sigma_\nu^2 tr(I_{2N-d} - M) = \sigma_\nu^2(N - d)$$

(A.3)

On the other hand, the generalized sum of squares can also be expressed as follows:

$$\boldsymbol{e} * '\Omega^{-1}\boldsymbol{e} *= (\hat{\boldsymbol{\nu}}' \quad -\hat{\boldsymbol{\varepsilon}}')\begin{pmatrix} V_{N\text{x}N} & 0_{N\text{x}(N-d)} \\ 0_{(N-d)\text{x}N} & \lambda I_{N-d} \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\nu}} \\ -\hat{\boldsymbol{\varepsilon}} \end{pmatrix} = \hat{\boldsymbol{\nu}}' V_{N\text{x}N}^{-1}\hat{\boldsymbol{\nu}} + \boldsymbol{\lambda}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$$

$$= (\boldsymbol{Y} - \widehat{\boldsymbol{g}})' V_{N\text{x}N}^{-1}(\boldsymbol{Y} - \widehat{\boldsymbol{g}}) + \lambda(K_d\widehat{\boldsymbol{g}} - \mu\mathbf{1}_{N-d})'(K_d\widehat{\boldsymbol{g}} - \mu\mathbf{1}_{N-d})$$

$$= \boldsymbol{Y}' V_{N\text{x}\text{X}N}^{-1}\boldsymbol{Y} + \widehat{\boldsymbol{g}}' V_{N\text{x}N}^{-1}\widehat{\boldsymbol{g}} + \lambda(\widehat{\boldsymbol{g}}' K_d{}' K_d\widehat{\boldsymbol{g}} + \mu^2\mathbf{1}'_{N-d}\mathbf{1}_{N-d})$$

$$= \boldsymbol{Y}' V_{N\text{x}N}^{-1}\boldsymbol{Y} + \widehat{\boldsymbol{g}}'(V_{N\text{x}N}^{-1} + \lambda K_d{}' K_d)\widehat{\boldsymbol{g}} + (N-d)\lambda\mu^2 \tag{A.4}$$

Then, combining (A.3) and (A.4) we show that (3.1.12) is an unbiased estimator of σ_ν^2.

Proof of (3.1.23)

This proof is completely based on the following work ofTheil (1963). Lemma: Let P and Q be two positive-definite or positive semi-definite matrices of size HxH and let $\Lambda(P, P+Q)$ be a scalar function which measure the share of P in the matrix $(P+Q)^{-1}$. Let us suppose that:

i. $\Lambda(P, P+Q) + \Lambda(Q, P+Q) = 1$
ii. $\Lambda(0, Q) = 0$ and $\Lambda(P, P) = 1$
iii. $\Lambda(K'PK, K'(P+Q)K) = \Lambda(P, P+Q)$ for all squared and nonsingular K-matrices.
iv. $\Lambda(\alpha P_1 + (1-\alpha)P_2, \alpha(P_1+Q_1) + (1-\alpha)(P_2+Q_2)) = \alpha\Lambda(P_1, P_1+Q_1) + (1-\alpha)\Lambda(P_2, P_2+Q_2)$ with P_1, Q_1, P_2 and Q_2 positive-definite or positive semi-definite matrices, such that $P_1 + Q_1 = P_2 + Q_2$ and $0 \leq \alpha \leq 1$.

The only Λ which satisfies these four criteria is

$$\Lambda(P, P+Q) = Tr\,[P(P+Q)^{-1}]/H \tag{A.5}$$

Expression (3.1.23) is deduced from property (i) of Λ

$$S(\lambda, N) = \Lambda(\sigma_{\varepsilon}^{-2}K_d{}'K_d, \Gamma) = \Lambda\left(\sigma_{\varepsilon}^{-2}K_d{}'K_d, (\sigma_{\nu}^{-2}V_{N\mathrm{x}N}^{-1} + \sigma_{\varepsilon}^{-2}K_d{}'K_d)\right)$$

$$= 1 - \Lambda\left(\sigma_{\nu}^{-2}V_{N\mathrm{x}N}^{-1}, (\sigma_{\nu}^{-2}V_{N\mathrm{x}N}^{-1} + \sigma_{\varepsilon}^{-2}K_d{}'K_d)\right)$$

$$= 1 - Tr\left[\sigma_{\eta}^{-2}V_{N\mathrm{x}N}^{-1}\left(\sigma_{\eta}^{-2}V_{N\mathrm{x}N}^{-1} + \sigma_{\varepsilon}^{-2}K_d{}'K_d\right)^{-1}\right]/N$$

$$= 1 - Tr[V_{NxN}^{-1}(V_{N\mathrm{x}N}^{-1} + \lambda K_d{}'K_d)^{-1}]/N$$

Proof of (3.3.2)

The noise component follows the AR(1) process
$\nu_t = \rho\nu_{t-1} + \zeta_t$ for $t = 1, 2, \dots, N$ and $\nu_1 = \zeta_1$, where $|\rho| \leq 1$.

Although the time series is observed at time $t = 1$, the process is regarded as having started at some time in the remote past. Substituting repeatedly for lagged values of ν_t gives

$$\nu_t = \sum_{j=0}^{J-1} \rho^j \zeta_{t-j} + \rho^J \nu_{t-J} \tag{A.6}$$

Now, since $|\rho| \leq 1$, the component ρ^J is negligible, since for J large, that is, as $J \rightarrow \infty$, it effectively disappears and so if the process is regarded as having started at one point in the remote past, it is possible to write (A.6) in the form

$$\nu_t = \sum_{j=0}^{\infty} \rho^j \zeta_{t-j}\,, \qquad t = 1, 2, \dots, N, \tag{A.7}$$

and, since summing the squared coefficients as a geometric progression yields

$$\sum_{j=0}^{\infty} \rho^{2j} = 1/(1-\rho^2)$$

Therefore,

$$E(\nu_t) = \sum_{j=0}^{J-1} \rho^j E(\nu_{t-j}) = 0,$$

$$\gamma(0) = E(\nu_t^2) = E\left(\sum_{j=0}^{\infty} \rho^j \zeta_{t-j}\right)^2 = \sigma_\zeta^2 \sum_{j=0}^{\infty} \rho^{2j} = \sigma_\zeta^2/(1-\rho^2)$$

$$\gamma(k) = E(\nu_t \nu_{t+k}) = \sigma_\zeta^2 \left(\sum_{j=1}^{\infty} \rho^j \rho^{k+j} + \rho^k\right) = \sigma_\zeta^2 \rho^k \sum_{j=0}^{\infty} \rho^{2j} = \sigma_\zeta^2 \rho^k/(1-\rho^2), k = 1, 2, \dots .$$

Then, assuming σ_ζ^2=1, it follows that (3.3.2)

$$Var(\boldsymbol{\nu}) = \sigma_\nu^2 V = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \dots & \rho^{N-2} \\ \rho & 1 & \dots & \rho^{N-2} \\ & \dots & & \\ \rho^{N-1} & \rho^{N-2} & \dots & 1 \end{pmatrix}$$

Proof of (4.3.8)

Without loss of generality assume μ^* = 0 in (4.3.3) and from (4.3.4) it follows that

$$\nabla^* \boldsymbol{Y}_T^* = \varepsilon_T^* + \nabla^* \boldsymbol{\nu}_T^* \quad \text{(A.8)}$$

where ∇^* is the first order difference operator for the aggregated series, so that $\{\boldsymbol{Y}_T^*\}$ is represented by an Integrated Moving Average model of order (1,1) (IMA(1,1)), whose variance γ_0^* and autocovariance γ_1^* are given by

$$\gamma_0^* = \sigma_\varepsilon^{*2} + 2\sigma_\nu^{*2} \text{ and } \gamma_1^* = -\sigma_\nu^{*2}$$

Similarly, assuming μ = 0 and d = 1 in (3.1.6) and from (3.1.5), the disaggregate series follows the IMA(1,1) model

$$\nabla Y_t = \varepsilon_t + \nabla\nu_t \quad \text{(A.9)}$$

whose variance γ_0 and covariance γ_1 are

$$\gamma_0 = \sigma_\varepsilon^2 + 2\sigma_\nu^2 \quad \text{and} \quad \gamma_1 = -\sigma_\nu^2$$

To see how the two IMA(1,1) models (A.7) and (A.8) relate to each other, consider a time series of flows, that is, let

$$Y_{T-j}^* = S_k Y_{t-jk} \text{ with } S_k = 1 + B + \cdots + B^k.$$

Since

$$\nabla_k = 1 - B^k = S_k\nabla \text{ then } \nabla_k Y_t = S_k(\varepsilon_t + \nabla\nu_t),$$

so that

$$\nabla_k S_k Y_t = S_k^2\varepsilon_t + S_k\nabla_k\nu_t.$$

Then, as

$$\nabla^* Y_T^* = \nabla_k S_k Y_t,$$

it follows that

$$\nabla^* Y_T^* = S_k^2\varepsilon_t + S_k\nabla_k\nu_t \quad \text{for flows.} \quad \text{(A.10)}$$

Now, for a stock time series, it is known that $Y^*_{T-j} = Y_{t-jk}$ and $\nabla^* Y^*_T = \nabla_k Y_t$, then, from the previous derivation, it follows that

$$\nabla^* Y_T^* = S_k\varepsilon_t + \nabla_k\nu_t \quad \text{for stocks.} \quad \text{(A.11)}$$

Therefore, the autocovariance generating function of the disaggregated series $\gamma(B) = \Sigma_j^\infty = -\infty\gamma_j B_j$ is given by

$$\gamma(B) = \begin{cases} S_k^2\overline{S}_k^2\sigma_\varepsilon^2 + S_k\nabla_k\overline{S}_k\overline{\nabla}_k\sigma_\nu^2 & \text{for flows} \\ S_k\overline{S}_k\sigma_\varepsilon^2 + \nabla_k\overline{\nabla}_k\sigma_\nu^2 & \text{for stocks} \end{cases} \quad \text{(A.12)}$$

where an upper bar denotes the corresponding polynomial with B replaced by B^{-1} (that is, $\overline{\nabla}_k = (1\text{-}B^{-k})$.

The values $\sigma_\varepsilon^2, \sigma_v^2, \sigma_\varepsilon^{*2}$ and σ_v^{*2} that make the results of both exponential filters to become equivalent are obtained by equating γ_0^* and γ_1^* to γ_0 and γ_{1k}. That is, by solving the following system of equations

$$\begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \sigma_\varepsilon^{*2} \\ \sigma_v^{*2} \end{pmatrix} = \begin{pmatrix} a_{11,k} & a_{12,k} \\ a_{21,k} & a_{22,k} \end{pmatrix} \begin{pmatrix} \sigma_\varepsilon^2 \\ \sigma_v^2 \end{pmatrix} \quad \text{(A.13)}$$

Where $a_{11,k}$ and $a_{21,k}$ are the coefficients of B^0 and B^k in the polynomial $S_k^2\bar{S}_k^2$ for the flow time series or in the polynomial $S_k\bar{S}_k$ for a stock time series. Similarly, $a_{12,k}$ and $a_{22,k}$ are the coefficient of B^0 and B^k in the polynomial $S_k\nabla_k\bar{S}_k\bar{\nabla}_k$ for a flow time series or in the polynomial $\nabla_k\bar{\nabla}_k$ in the time series of stocks. System (A.12) has a unique solution for σ_ε^{*2} and σ_v^{*2}, or for σ_ε^2 and σ_v^2, depending on which pair of variances is known. If σ_ε^2 and σ_v^{*2} are known, so that $\lambda^*_k = \sigma_v^{*2}/\sigma_\varepsilon^{*2}$ is known, solving (A.13) yields

$$\sigma_\varepsilon^2 = \frac{a_{22,k} + (a_{12,k} + 2a_{22,k})\lambda_k^*}{a_{11,k}a_{22,k} - a_{12,k}a_{21,k}} \quad and\ \sigma_v^2 = \frac{-a_{21,k} - (a_{11,k} + 2a_{21,k})\lambda_k^*}{a_{11,k}a_{22,k} - a_{12,k}a_{21,k}}$$

from which it follows that $\lambda = \sigma_v^2/\sigma_\varepsilon^2$ is given by

$$\lambda = \frac{-a_{21,k} - (a_{11,k} + 2a_{21,k})\lambda_k^*}{a_{22,k} + (a_{12,k} + 2a_{22,k})\lambda_k^*}. \quad \text{(A.14)}$$

Now, since $\nabla_k\bar{\nabla}_k = 2 - (B^k + B^{-k})$,

$$S_k\nabla_k\bar{S}_k\bar{\nabla}_k = 2k - k(B^k + B^{-k}) + P_{1,k}(B, B^{-1}),\ S_k\bar{S}_k = k + P_{2,k}(B, B^{-1}),$$

and $S_k^2\bar{S}_k^2 = 2\sum_{i=1}^{k-1} i^2 + k^2(k^3 - k)(B^k + B^{-k})/6 + P_{3,k}(B, B^{-1})$,

where $P_{1,k}(B, B^{-1})$, $P_{2,k}(B, B^{-1})$ and $P_{3,k}(B, B^{-1})$ are polynomials in B and B^{-1} that have no powers of type B^{ik}, for $i = 0,1$. By inspection of the polynomials involved, it follows that for time series of flows, $a_{11,k} = (2k^2+k)/3$, $a_{21,k} = (k^3 - k)/6$, $a_{12,k} = 2k$and $a_{22,k} = -k$, whereas for a time series of stocks, $a_{11,k} = k$, $a_{21,k} = 0$, $a_{12,k} = 2$ and $a_{22,k} = -1$. Therefore, from (A.14) follows (4.3.9), so that

$$\lambda = \begin{cases} (k^2 - 1)/6 + k^2\lambda_k^* & \text{for flows} \\ k\lambda_k^* & \text{for stocks} \end{cases}$$

References

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60:255–265.

Backus, D. K., and Kehoe, P. J. (1992). International Evidence on the Historical Properties of Business Cycles, The American Economic Review. *82*, pp. 864-888.

Baillie, R. T., and Bollerslev, T. (1989). Common stochastic trends in a system of exchange rates. *The Journal of Finance*, 44:167–181.

Banco de México (2023a) "History of monetary policy" taken from Banxico's page https://www.banxico.org.mx/getting-to-know-banco-de-mexico/history-hierarchical-history-.html

Banco de México (2023b) https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwj_ppSkoK2BAxXUN0QIHcRDAWoQFnoECCEQAQ&url=https%3A%2F%2Fwww.banxico.org.mx%2Fpolitica-monetaria%2Fd%2F%257BCE7DEA10-0015-1138-4A2F-F3580416D34F%257D.pdf&usg=AOvVaw2ERhHf2cf_6QQyRONKXfBU&opi=89978449 reviewed on (15/09/2023).

Barberis, N., and Thaler, T. (2002). A survey of behavioral finance. *NBER Working Paper, 9222*.

Barsky, R. B., and Kilian, L. (2002). Do we really know that oil caused the great stagflation? a monetary alternative. In: Bernanke, Ben S., Rogoff, Kenneth (Eds.), NBER Macroeconomics Annual 2001, pp. 137–183.

Baxter, M., and King, R. G. (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *Rev. Econ. Stat.*, 81, 575–593.

Björnland, H. C. (2000). Detrending Methods and Stylised Facts of Business Cycles in Norway – An International Comparison. *Empirical Economics,* (*25*), 369-392.

Bordoff, J.; Losz, A. (2015). Oil Shock – Decoding the Causes and Consequences of the 2014 Oil Price Drop", CIRSD. Available in: HYPERLINK "https://www.cirsd.org/en/horizons/horizons-spring-2015--issue-no3/oil-shock-

%E2%80%94-decoding-the-causes-and%20consequences-of-the-2014-oil-price-drop#:~:text=The%20primary%20factor%20driving%20the,of%20 U.S.%20tight%20oil%20production"https://www.cirsd.org/en/horizons/horizons-spring-2015--issue-no3/oil-shock-%E2%80%94-decoding-the-causes-and consequences-of-the-2014-oil-price-drop#:~:text=The%20primary%20factor%20driving%20the,of%20U.S.%20tight%20oil%20production.

Brantingham, P., and Brantingham, P. (2008). "Crime Pattern Theory" in Wortley, Richard; Mazerolle, Lorraine (2008) Environmental Criminology and Crime Analysis, Routledge, Taylos and Francis Group, London and New York.

Bruder, B., Dao, T.-L., Richard, J.-C., and Roncalli, T. (2011, December 1). Trend Filtering Methods for Momentum Strategies. Retrieved from SSRN: https://ssrn.com/abstract=2289097 or http://dx.doi.org/10.2139/ssrn.2289097

Cleveland, W. P., and Tiao, G. C. (1976). Decomposition of seasonal time series: A model for the X-11 program. *J. Amer. Statist. Assoc*, 71, 581–587.

Cochrane, J. (2001). Asset pricing, Princeton University Press.

Costa, Sonia (2000) "Monetary Conditions Index", Economic Bulletin, Banco de Portugal, September.

Craven, P., and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numer Math*, 31:377–403.

Dolado, J. J., Sebastián, M., and Vallés, J. (1993). Cyclical Patterns of the Spanish Economy. *Investigaciones Económicas, XVII*, pp. 445-473.

Elekdag, S. A., and Laxton, D. (2007). Understanding the link between oil prices and the economy, Box 1.1 in IMF World Economic Outlook, Washington, DC.

Elekdag, S. A., Lalonde, R., Laxton, D., Muir, D., and Pesenti, P. A. (2008). Oil price movements and the global economy: a model-based assessment. NBER Working Paper No. 13792.

European Central Bank. (2000). Monthly Bulletin, October.

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance, 25*(*2*), pp. 383-417.

Giorno, C., Richardson, P., Roseveare, D., and Van den Noord, P. (1995). Estimating Potential Output, Output Gaps and Structural Budget Balances, OECD Economics Department Working Papers, 152.

Guerrero, V. M. (2007). Time series smoothing by penalized least squares. *Stat Probab Lett.*, 77:1225–1234.

Guerrero, V. M. (2008). Estimating trends with percentage of smoothness chosen by the user. *Int Stat Rev*, 76:187–202.

Guerrero, V. M. (2009). *Análisis Estadístico y Pronóstico de Series de Tiempo Económicas*. 3ª edición, México, Jit Press.

Guerrero, V. M., and Galicia-Vazquez, A. (2010). Trend estimation of financial time series. *Appl. Stochastic Models in Business and Industry*, 26:205-223.

Guerrero, V. M., Cortés, T. H., and Cervantes, J. R. (2018). Effect of autocorrelation when estimating the trend of a time series via penalized least squares with controlled smoothness. *Stat Methods Appl*, 27:109-130.

Guerrero, V. M., Islas-Camargo, A., and Ramirez-Ramirez, L. (2017). Trend estimation of multivariate time series with controlled smoothness. *Communication in Statistics – Theory and Methods, 46:13*, 6704-6726. doi:10.1080/03610926.2015.1133826

Hamilton, J. H. (2009). Causes and consequences of the oil shock of 2007–2008. NBER Working Paper No. 15002.

Harvey, A. C. (1985). Trends and cycles in macroeconomic time series. *Journal of Business and Economic Statistics, 3*, 216-227.

Harvey, A. C. (1989). Forecasting Structural Time Series Models and the Kalman Filter. Cambridge: Cambridge University Press.

Harvey, A. C., and Jaeger, A. (1993). Detrending, stylized facts and the business cycle. J. *Appl. Econometrics, 8*, 231– 247.

Henderson, R. (1924). A new method of graduation. *Trans. Actuarial Soc. Amer. 25*, 29–40.

Hodrick, R. J., and Prescott, E. C. (1997). Postwar U.S. business cycles: an empirical investigation. *J. Money Credit and Banking, 29*, 1-16.

Hong, H., and Stein, J. C. (1977). A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets, NBER Working Paper, 6324.

Hurvich, C., Simonoff, J. S., and Tsai, C. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J R Stat Soc Ser B*, 60:271–293.

Johnston, M. (2022). "What happened to Oil Prices in 2020", Investopedia, Available in: https://www.investopedia.com/articles/investing/100615/will-oil-prices-go-2017.asp

Kaiser, R., and Maravall, A. (2001). Measuring Business Cycles in Economic Time Series. *Lecture Notes in Statistics, 154*, New York: Springer –Verlag.

Kilian, L. (2008). The economic effects of energy price shocks. *Journal of Economic Literature, 46* (*4*), 871–909.

Kilian, L. (2009). Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. *American Economic Review, 99* (*3*), 1053–1069.

King, R. G., and Rebelo, S. T. (1993). Low frequency filtering and real business cycles. *J. Econom. Dynamics Control, 17*, 207–231.

Kitagawa, G., and Gersch, W. (1996). Smoothness Priors Analysis of Time Series. Lecture Notes in Statistics, 116, New York: Springer –Verlag.

Kohn, R., Ansley, C. F., and Wong, C. M. (1992). Non-parametric spline regression with autoregressive moving average errors. *Biometrika, 79*, 335–346.

Krivobokova, T., and Kauermann, G. (2007). A note on penalized spline smoothing with correlated errors. *J Am Stat Assoc*, 102:1328–1337.

Kydland, F., and Prescott, E. (1990). Business Cycles: Real Facts and a Monetary Myth. *Federal Reserve Bank of Minneapolis Quarterly Review, 14* (*2*), 3-18.

Lee, T. C. (2003). Smoothing parameter selection for smoothing splines: A simulation study. *Comput. Statist. Data Anal, 42*, 139–148.

Maravall, A., and del Río, A. (2007). Temporal aggregation, systematic sampling, and the Hodrick–Prescott filter. *Computational Statistics and Data Analysis*, 52:975–998.

Masters, M. W. (2008). Testimony of Michael W. Masters, Managing Member / Portfolio Manager, Masters Capital Management, LLC. Testimony before the U.S. Senate Committee on Homeland Security and Governmental Affairs

Mohr, M. (2001). Ein disaggregierter Ansatz zur Berechnung konjunkturbereinigter Budgetsalden für Deutschland: Methoden und Ergebnisse", Diskussionspapier 13/01, Economics Research Centre, Deutsche Bundesbank, Frankfurt am Main.

Morana, C. (2001). A semiparametric approach to short-term oil price forecasting. *Energy Economics, 23* (*3*), 325–338.

Narayan, P. K., and Smyth, R. (2005). Are OECD stock prices characterized by a random walk? Evidence from sequential trend break and panel data models. *Applied Financial Economics*, 15:547–556.

Organization of the Petroleum Exporting Countries (OPEC). (2008). Monthly Oil Market Report – December 2008

Pedersen, T. M. (2001). The Hodrick–Prescott filter, the Slutzky effect, and the distortionary effect of filters. *J. Econom. Dynam. Control, 25*, 1081–1101.

Ramos-Francia, Manuel; Torres-García, Alberto (2005) "Reducing Inflation Through Inflation Targeting: The Mexican Experience", Banco de México, Working Papers, N° 2005-01, July.

Reeves, J. J., Blyth, C. A., Triggs, C. M., and Small, J. P. (2000). The Hodrick–Prescott filter a generalization and a new procedure for extracting an empirical cycle from a series. *Stud. Nonlinear Dynamics and Econom*, 4 (1) Article 1.

Sadorsky, P. (1999). Oil price shocks and stock market activity. *Energy Economics, 21* (*5*), 449–469.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Ann Stat*, 6:461–464.

Stevens, G. R. (1998) "Pitfalls in the Use of Monetary Condition Indexes", Reserve Bank of Australia Bulletin, August.

Theil, H. (1963). On the use of incomplete prior information in regression analysis. *J Am Stat Assoc*, 58:401– 414.

Tobias, A. (2023) "The Role of Inflation Expectations in Monetary Policy", https://www.imf.org/en/News/Articles/2023/05/15/sp-role-inflation-expectations-monetary-policy-tobias-adrian

Tsay, R. S. (2002). *Analysis of Financial Time Series*. Wiley: New York.

Whittaker, E. T. (1923). On a new method of graduation. *Proc. Edinburgh Math. Soc, 41*, 63–75.

Young, P. (1994). Time-variable parameter and trend estimation in non-stationary economic time series. *J. Forecast, 13*, 179–210.

Tables Content

Table 3.2.1. Values of λ as a function of sample size N and percentage of smoothness S% (quarterly series) 49
Table 3.2.1. Values of λ as a function of sample size N and percentage of smoothness S% (quarterly series, continuation) 51
Table 3.2.2. Estimation of an approximating function that relates λ to *S*% and N, N > 120 (quarterly series). 52
Table 3.2.3. Parameter values for the nonlinear trends. 54
Table 3.3.1. S% values as a function of λ for *N*=20 and different values of ρ 74
Table 3.3.2. S% values as a function of λ for *N*=50 and different values of ρ 74
Table 3.3.3. S% values as a function of λ for *N*=100 and different values of ρ 75
Table 3.3.4. S% values as a function of λ for *N*=200 and different values of ρ 75
Table 3.3.5. S% values as a function of λ for *N*=300 and different values of ρ 76
Table 3.3.6. S% values as a function of λ for *N*=500 and different values of ρ 76
Table 4.2.1. Values of λ as a function of sample size N and percentage of smoothness S% (daily series). 101
Table 4.2.2. Estimation results of fitting models that relate λ with N and S% (Daly series) 103
Table 5.1.1. Cross Correlation between the cyclical components of the state's Index of Physical Volume and Total Crime. . . . 127
Table 5.2.1. Cross Correlation between the cyclical components of Monetary Condition Index and IVF of Primary sector . . . 137

Table 5.2.2. Cross Correlation between the cyclical components of Monetary Condition Index and IVF of Secondary sector. . . . 137
Table 5.2.3. Cross Correlation between the cyclical components of Monetary Condition Index and IVF of Tertiary sector . . . 138

Figures Content

Figure 1.1. The logarithm of Jalisco´s remittances 18-19
Figure 3.2.1. Behavior of $S(\lambda, 0)\%$ for: (a) N=50, 100, 150, 200, and (b) λ = 400, 1600, 6400 46-47
Figure 3.2.2. Examples of simulated time series with estimated trends 55-56
Figure 3.2.3. Examples of simulated time series with estimated trends based on Guerrero's (2017) guideline for fixing a smoothness percentage. Panels (a) and (b) show the nonlinear time series, while panel (c) shows the linear time series. . 62-64
Figure 3.2.4. Selected trend estimates with 85% smoothness and different sample sizes for the simulated $Nonlinear_1$ series . . . 66
Figure 3.2.5. Trend estimates for $\lambda = 1600$ and different sample sizes 68
Figure 3.2.6. Trend with 85% smoothness 70
Figure 3.3.1. Time series simulated with theoretical piecewise quadratic trend and smoothed with 85% smoothness and respective autocorrelations: (*a*) ρ=0.4, (*b*) ρ=-0.4, (c) ρ=0.9, (*d*) ρ=-0.9 78-80
Figure 3.3.2. Time series simulated with theoretical piecewise quadratic trend and smoothed with 90% smoothness and respective autocorrelations: (*a*) ρ=0.4, (*b*) ρ=-0.4, (c) ρ=0.9, (*d*) ρ=-0.9 81-84
Figure 3.3.3. Time series simulated with theoretical piecewise linear trend and smoothed with 85% smoothness and respective autocorrelations: (*a*) ρ=0.4, (*b*) ρ=-0.4, (c) ρ=0.9, (*d*) ρ=-0.9 . . 85-87

Figure 3.3.4. Time series simulated with theoretical piecewise linear trend and smoothed with 90% smoothness and respective autocorrelations: (*a*) ρ=0.4, (*b*) ρ=-0.4, (c) ρ=0.9, (*d*) ρ=-0.9 . . 88-90
Figure 4.2.1. Behavior of $S(\lambda,N)$% for: (a) N=50, 100, 400, 800, and (b) λ = 50, 100, 150, 200 98-99
Figure 4.4.1. Panel (a), simulated random walk with drift, sample size of *N= 3123* daily data. Panel (b), simulated daily random walk with drift and estimated trend with S%=95% (λ = 101.1). Panel (c) weekly random walk with drift and estimated trend with S%= 95% (λ^*_5 = 20.22)*N=624* observations. Panel (d) monthly random walk with drift and estimated trend with S%= 95% ($\lambda^*_{20=5.05}$), *N=157* observations. 109-111
Figure 4.4.2. Panel (a), simulated random walk without drift, sample size *N=3123* daily data. Panel (b), simulated daily random walk without drift and estimated trend with *S*%=90% (λ = 24.9). Panel (c) weekly random walk without drift and estimated trend with *S*%= 90% (λ^*_5 = 4.98), *N=624* observations. Panel (d) monthly random walk without drift and estimated trend with *S*%= 90% (λ^*_{20} = 1.245), *N=157* observations. 113-116
Figure 4.4.3. Selected trend estimates with 90% smoothness and different sample sizes for the simulate monthly random walk without drift series. 118
Figure 5.1.1. Quarterly total crime data from 1997-I to 2017-IV in Mexico City, Jalisco, Nuevo Leon, and Guanajuato . . . 122-123
Figure 5.1.2. Defining the Model's parameters 124-125
Figure 5.1.3. Trend of Total Crime in selected states: 1997-I to 2017-IV 126
Figure 5.2.1. Mexico's MCI, 1997-I-2022-IV 132
Figure 5.2.2. Quarterly Index of Physical Volume from 1997-I to 2022-IV in the Primary Sector for Mexico City, Secondary Sector for Jalisco, and Tertiary Sector for Guerrero 133
Figure 5.2.3. Defining the Model's parameters. 135-136
Figure 5.3.1. Level and Rate of Change of Oil Price 141
Figure 5.3.2. Panel (a), daily current prices and estimated trend with S%=90% (λ = 24.9), sample size N= 6122. Panel (b), weekly current prices and estimated trend with S%= 90% (λ^*_5 =4.98), N=1224 observations. Panel (c) monthly current prices and estimated trend with S%= 90% (λ^*_{20} =1.24, N=281 observations 143-145

Time Series Smoothing by Penalized
Least Squares with Applications
was finished printing in May 2024
in the graphic workshops of Ediciones de la Noche
Francisco I. Madero #687, Zona Centro
44100, Guadalajara, Jalisco, Mexico.

The circulation was 150 copies.

www.edicionesdelanoche.com