UNIVERSITY OF THE AEGEAN

DOCTORAL THESIS

A Study of Time Series with Time-Varying Variance: A Practical Approach with Applications in Actuarial Science and Finance

Author

Nikolaos G. Galanopoulos

Supervisor

Alexandros E. Milionis

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

of the

Department of Statistics and Actuarial-Financial Mathematics

of the University of the Aegean

Samos, 24 October 2023

Υπεύθυνη Δήλωση

Είμαι ο αποκλειστικός συγγραφέας της υποβληθείσας Διδακτορικής Διατριβής με τίτλο «Μελέτη Χρονοσειρών με Χρονικά Μεταβαλλόμενη Διακύμανση: Μια Πρακτική Προσέγγιση με Εφαρμογές στην Αναλογιστική Επιστήμη και τα Χρηματοοικονομικά _ A Study of Time Series with Time-Varying Variance: A Practical Approach with Applications in Actuarial Science and Finance».
Η συγκεκριμένη Διδακτορική Διατριβή είναι πρωτότυπη και εκπονήθηκε αποκλειστικά για την απόκτηση του Διδακτορικού διπλώματος του Τμήματος. Κάθε βοήθεια, την οποία είχα για την προετοιμασία της, αναγνωρίζεται πλήρως και αναφέρεται επακριβώς στην εργασία. Επίσης, επακριβώς αναφέρω στην εργασία τις πηγές, τις οποίες χρησιμοποίησα, και μνημονεύω επώνυμα τα δεδομένα ή τις ιδέες που αποτελούν προϊόν πνευματικής ιδιοκτησίας άλλων, ακόμη κι εάν η συμπερίληψή τους στην παρούσα εργασία υπήρξε έμμεση ή παραφρασμένη. Γενικότερα, βεβαιώνω ότι κατά την εκπόνηση της Διδακτορικής Διατριβής έχω τηρήσει απαρέγκλιτα όσα ο νόμος ορίζει περί διανοητικής ιδιοκτησίας και έχω συμμορφωθεί πλήρως με τα προβλεπόμενα στο νόμο περί προστασίας προσωπικών δεδομένων και τις αρχές Ακαδημαϊκής Δεοντολογίας.

20/10/2023

Ο - Η Δηλ.

ΝΙΚΟΛΑΟΣ ΓΑΛΑΝΟΠΟΥΛΟΣ

ADVISING COMMITTEE

Alexandros Milionis, Supervisor

Department of Statistics and Actuarial-Financial Mathematics

University of the Aegean

Petros Hatzopoulos, Advisor

Department of Statistics and Actuarial-Financial Mathematics

University of the Aegean

Athanasios Rakitzis, Advisor

Department of Statistics and Insurance Science

University of Piraeus

University of the Aegean

October 2023

APPROVED BY THE EXAMINING COMMITTEE:

Alexandros Milionis

Associate Professor, University of the Aegean

Petros Hatzopoulos

Associate Professor, University of the Aegean

Athanasios Rakitzis

Assistant Professor, University of Piraeus

Konstantinos Katrakylidis

Professor, Aristotle University of Thessaloniki

Panagiotis Alexakis

Emeritus Professor, National and Kapodistrian University of Athens

---

Alexandros Karagrigoriou

Professor, University of the Aegean

Panagiotis Nastou

Assistant Professor, University of the Aegean

University of the Aegean

October 2023

Nikolaos G. Galanopoulos

Department of Statistics and Actuarial – Financial Mathematics

University of the Aegean

Acknowledgements

I would like to wholeheartedly thank my supervisor, Mr. Milionis, who has been by my side at every stage of my doctoral thesis, from the very first day until today, guiding and advising me. His contribution to the completion of my doctoral thesis is invaluable, as he was always willing to listen to my concerns, thoroughly discuss them, and share his knowledge on how to conduct research work.

Furthermore, I would like to thank Mr. Rakitzis, who played a crucial role in the initial steps of the doctoral thesis, particularly in the development of the code in the R programming language.

Additionally, I would like to thank Mr. Chatzopoulos, with whom we collaborated on the study of the proposed methodology in actuarial science. At this point, I also want to express my gratitude to Mrs. Sagianou and Mr. Porichis, with whom we also collaborated on the respective aspects related to actuarial science.

Moreover, I would like to thank Mr. Karagrigoriou for the feedback he provided on the presentations of various talks that were given on my behalf.

In addition, I would like to thank Mr. Nastou for the extensive discussions we had, both at the undergraduate level due to our coexistence in the Department of Mathematics at the University of the Aegean, and during my doctoral studies. His recommendations for various programming books helped me understand the philosophy of programming.

A heartfelt expression of gratitude is owed to my parents, Dimitra and George, without whom the journey I have followed at all levels of my studies would not have been possible. Their unwavering support in all areas cannot be described in words. Additionally, I thank my brother, Petros, for the encouragement he provided me during my studies.

A special chapter for me is Elena, whom I thank for her presence, understanding, and encouragement, as she has been helping me achieve my goals throughout all these years.

To my parents,

Dimitra and George…

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 General

A time series consists of a series of observations collected over time, where the observations are usually dependent on each other. The dependency between adjacent observations in a time series is an inherent characteristic. Understanding and analyzing this dependency is crucial for practical applications. Time series analysis focuses on developing stochastic and dynamic models to capture the patterns and characteristics of time series data. These models are then applied to various important fields of study.

The concept of utilizing a mathematical model to explain how a physical phenomenon behaves is widely acknowledged. A deterministic time series is a type of time series where the values can be precisely determined based on known mathematical functions. On the other hand, a stochastic time series is a type of time series where the future values can only be determined in terms of a probability distribution. If this probability distribution is constant over time, then the time series is said to be stationary (further details about the very important property of stationarity are given later in this chapter). If a time series is not stationary, then there are formal statistical procedures to transform a non-stationary series into a stationary one. A less strict condition for stationarity requires that at least the level and variance of the time series be constant over time. While researchers typically test for non-stationarity in the level of a time series using various tests, they sometimes overlook non-stationarity in the variance when conducting applied research. This may seriously affect (negatively) the quality of subsequent analysis and modeling of the series. The development of a formal statistical test for the existence of variance non-stationarity is among the cornerstones of this thesis. In the next paragraphs of this first chapter some basic concepts of time series analysis which will be useful for the subsequent analysis are reviewed emphasizing on macroeconomic, actuarial, and financial time series in conjunction with the objectives of the thesis.

1.1.1 Stationary stochastic models

One notable category of stochastic models utilized for characterizing time series, which has garnered significant interest, consists of what are commonly referred to as stationary models. Typically, the characteristics of a stationary time series can be effectively described by its mean, variance, and autocorrelation function.

Consider that observations are obtained regularly at discrete, fixed time intervals. For instance, consider a time series consisting of values $y_1, y_2, \dots, y_T$. This time series is generated by a group of random variables $Y_1, Y_2, \dots, Y_T$ which are governed by a joint probability distribution $P(Y_1, Y_2, \dots, Y_T)$. This group of random variables is formally referred to as a stochastic process. Therefore, the recorded time series represents only one of the potential results originating from the joint probability distribution $P(Y_1, Y_2, \dots, Y_T)$ and it is called as a realization or sample path of the stochastic process (Milionis, 2016).

Suppose the probability density function associated with time t is denoted as f_{Y_t}. In that case, the expected value of the element within the time series of order t can be expressed as follows:

$$\mu_t = E(Y_t) \equiv \int_{-\infty}^{+\infty} y_t f_{Y_t}(y_t)\, dy_t$$

and the equation below will provide the variance of Y_t:

$$\gamma_{0t} = E(Y_t - \mu_t)^2 \equiv \int_{-\infty}^{+\infty} (y_t - \mu_t)^2 f_{Y_t}(y_t)\, dy_t$$

When examining the time moments $t, t-1, \dots, t-j$, the jth-order autocovariance is defined in the following manner:

$$\gamma_{Jt} = E\{(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j})\}$$

$$\equiv \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} (y_t - \mu_t)(y_{t-j} - \mu_{t-j})\, f_{Y_t, Y_{t-1}, \dots, Y_{t-j}}(y_t, y_{t-1}, \dots, y_{t-j})\, dy_t dy_{t-1} \dots dy_{t-j}$$

where, $f_{Y_t, Y_{t-1}, \dots, Y_{t-j}}(y_t, y_{t-1}, \dots, y_{t-j})$ is the joint probability function of $Y_t, Y_{t-1}, \dots, Y_{t-j}$.

When the mean value, the variance, and the autocovariances (i.e. $\mu_t, \gamma_{0t}\ \gamma_{Jt}, j = 1,2, ...$ correspondingly) do not vary with the time moment t, the stochastic process is referred to as weakly stationary or second-order stationary. In other words, for a weakly stationary stochastic process, the following conditions are satisfied:

$$E(Y_t) = \mu, \forall t, E\{(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j})\} = E\{(Y_t - \mu)(Y_{t-j} - \mu)\} = \gamma_{Jt}, \forall t$$

Therefore, the autocovariance will depend only on the temporal lag j.

To qualify as strictly stationary, a stochastic process must exhibit a property wherein the entire probability distribution of $Y_t, Y_{t+1}, \dots, Y_{t+s}$ remains unaffected by the time t, indicating that it remains independent despite any temporal shift.

$$f_{Y_t, Y_{t+1}, \dots, Y_{t+s}}(y_t, y_{t+1}, \dots, y_{t+s}) = f_{Y_{t+k}, Y_{t+k+1}, \dots, Y_{t\mp k+s}}(y_{t+k}, y_{t+k+1}, \dots, y_{t+k+s})\ \forall t, k, s$$

1.1.2 Backward shift (or lag) operator

A time series operator is a tool that converts either one time series or a set of time series into another time series. Among these operators, the lag operator, represented by the symbol B, holds particular significance in time series analysis. When the lag operator is applied to a time series, it causes the series to undergo a transformation where the resulting new series is the same as the original series but shifted by a number of time periods equal to the order of the operator. In other words, it moves the values backward in time depending on the operator's order, i.e.

$$BY_t = Y_{t-1}, B^2Y_t = Y_{t-2}, \dots, B^kY_t = Y_{t-k}$$

When considering negative powers of B, denoted as B^{-k} with $k > 0$, they represent the forward operator F. The following relationships are valid:

$B^{-k}Y_t = F^{+k}Y_t = Y_{t+k}$, which can also be written as $B^{-k} = F^{+k}$.

The utilization of the operator B allows to express difference equations and stochastic models in a concise manner. Moreover, through B, we can establish differentiation operators and seasonal differentiation operators, which offer specific and efficient ways to handle time series data. More specifically:

Regular Differentiation operator $\equiv \nabla \equiv (1 - B)$, where $Y_t - Y_{t-1} = (1 - B)Y_t$

Seasonal differentiation operator $\equiv \nabla_{12} \equiv (1 - B^{12})$ where $Y_t - Y_{t-12} = (1 - B^{12})Y_t$

1.1.3 The general form of a linear stationary stochastic process

The fundamental component of discrete stochastic time series models is the white noise process, which serves as the essential building block. This time series, denoted as ε_t, has zero mean ($E(\varepsilon_t) = 0$), constant variance ($E({\varepsilon_t}^2) = \sigma^2$), and uncorrelated terms ($(\varepsilon_t \varepsilon_{t-k}) = 0$ for $k > 0$)

The Wold's decomposition theorem (1938) is a fundamental theorem in time series analysis that enables us to represent a weakly stationary stochastic process with zero mean using the following equation:

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} + K_t$$

where: $\psi_0 = 1, \varepsilon_t$ is the white noise, K_t is the causal component, and $\sum_{i=0}^{\infty}|\psi_i| < \infty$.

The symbol K_t denotes any component that can be fully predictable solely based on its past values, like an exponential function of time. If K_t equals zero, the stochastic process becomes is purely non-deterministic. Additionally, K_t is entirely independent of the values $\varepsilon_{t-i}, \forall i$ (Milionis, 2016).

The equation mentioned above can alternatively be represented through the so-called linear filter representation, which can be stated as follows: When the lag operator is applied to Y_t, where Y_t is a purely non-deterministic series, we obtain:

$$\begin{aligned} \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots &= \psi_0 \varepsilon_t + \psi_1 B \varepsilon_t + \psi_2 B^2 \varepsilon_t + \cdots \\ &= (\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)\varepsilon_t = \Psi(B)\varepsilon_t \end{aligned}$$

In other words, the time series Y_t is generated by applying the linear filter $\Psi(B)$ to the white noise ε_t. The function $\Psi(B)$ represents the transfer function of this linear filter, and its coefficients ψ_i are commonly referred to as psi-weights.

1.1.4 The autocorrelation function of the general stationary stochastic process and the sample autocorrelation function

In the case of purely non-deterministic linear stochastic processes with zero mean, the expected value will be:

$$E(Y_t) = E\{\Psi(B)\varepsilon_t\} = 0$$

and the variance is:

$$\gamma_0 = E\big(Y_t - E(Y_t)\big)^2 = E(\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots)^2$$
$$= E({\varepsilon_t}^2 + \psi_1^2\varepsilon_{t-1}^2 + \psi_2^2\varepsilon_{t-2}^2 + \cdots) = \sigma^2 + \psi_1^2\sigma^2 + \psi_2^2\sigma^2 + \cdots \Leftrightarrow$$

$$\gamma_0 = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2$$

where σ^2 is the variance of the white noise (Milionis, 2016).

It is important to highlight that as ε_t represents the white noise, the following statement remains valid: $E\big(\varepsilon_{t-i}\varepsilon_{t-j}\big) = 0 \ \forall i \neq j$, and the autocovariance will be:

$$\gamma_k = E\big\{\big(Y_t - E(Y_t)\big)\cdot\big(Y_{t-k} - E(Y_t)\big)\big\} =$$

$$E(\varepsilon_t + \psi_1\varepsilon_{t-1} + \cdots + \psi_k\varepsilon_{t-k} + \psi_{k+1}\varepsilon_{t-k+1} + \cdots)\cdot(\varepsilon_{t-k} + \psi_1\varepsilon_{t-k-1} + \psi_2\varepsilon_{t-k-2} + \cdots) =$$

$$\sigma^2\psi_k + \sigma^2\psi_1\psi_{k+1} + \sigma^2\psi_2\psi_{k+2} + \cdots \Leftrightarrow \gamma_k = \sigma^2 \sum_{i=0}^{\infty} \psi_i\psi_{i+k}$$

The role of the autocorrelation function, indicated by ρ_k, is of utmost importance in the field of applied time series analysis. In combination with the first and second moment, it offers insights into the characteristics of the stochastic process governing the evolution of the time series. Its definition is as follows:

$$\rho_k = \frac{Cov(Y_t, Y_{t-k})}{\sqrt{Var(Y_t)Var(Y_{t-k})}} = \frac{E\{(Y_t - \mu_t)(Y_{t-k} - \mu_{t-k})\}}{\sqrt{E(Y_t - \mu_t)^2 E(Y_{t-k} - \mu_{t-k})^2}}$$

In other words:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{i=0}^{\infty}\psi_i\psi_{i+k}}{\sum_{i=0}^{\infty}\psi_i^2}$$

In a stationary process, the mean at time t equals the mean at time $t-k$ ($\mu_t = \mu_{t-k}$), and the variance of Y_t is equal to the variance of Y_{t-k} ($Var(Y_t) = Var(Y_{t-k})$). Consequently, in the context of stationarity, the following can be deduced:

$$\rho_k = \frac{E\{(Y_t - \mu)(Y_{t-k} - \mu)\}}{E(Y_t - \mu_t)^2}$$

The definitions presented thus far hold greater theoretical significance, given that, in practical scenarios, obtaining multiple realizations of the stochastic process is often unfeasible. With just one realization of the stochastic process, i.e., the time series comprising our data $y_1, y_2, \dots, y_T$, the mean value can only be calculated across time:

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

and the jth-order autocovariance is calculated utilizing the following equation:

$$c_j = \frac{1}{T-j}\sum_{t=j+1}^{T} (y_t - \bar{y})(y_{t-j} - \bar{y})$$

The previously mentioned definition of ρ_k is purely theoretical since it pertains to a stochastic process for which we possess merely a finite number of terms from a single sample path. Under the assumption of stationarity, we can estimate ρ_k from the given N observations using the subsequent relationship, which yields the sample autocorrelation function (ACF):

$$\hat{\rho}_k = ACF(k) = \frac{\sum_{t=1}^{N-k}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{N}(Y_t - \bar{Y})^2}$$

Because ρ_k is symmetric, the $ACF(k)$ also exhibits symmetry, implying that $ACF(k) = ACF(-k)$. Due to this reason, when graphically depicting the ACF, we solely focus on the positive values of k.

1.1.5 The partial autocorrelation function

In the context of a time series, the partial autocorrelation of order k, represented as φ_{kk}, is described as the correlation between Y_t and Y_{t+k} while keeping the intervening terms $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$ constant (Milionis, 2016). In other words:

$$\varphi_{kk} = Correlation(Y_t, Y_{t+k} \ / \ Y_{t+1}, Y_{t+2}, \ldots, Y_{t+k-1} \ constant)$$

The partial autocorrelation function ($PACF$) provides the partial autocorrelation coefficient for various time lags, such as $k = 1,2,3$ and so on. Like the autocorrelation function, the $PACF$ serves as a valuable resource for understanding the interdependence patterns produced by a stochastic process. It proves to be a useful tool in identifying the appropriate stochastic model that best fits the data.

1.1.6 Test for autocorrelation

The Box-Pierce statistic (1970) and Ljung-Box statistic (1978) are both used in time series analysis to test the null hypothesis of no autocorrelation in the residuals of a fitted model. The Box–Pierce statistic is defined as:

$$BP = N \sum_{k=1}^{m} ACF(k)^2$$

and is a measure that quantifies the overall autocorrelation in the residuals of a time series model. The Box-Pierce statistic follows a chi-square distribution with m degrees of freedom under the null hypothesis that all autocorrelation coefficients up to order m are zero. If the computed BP statistic is found to be greater than the chi-square critical value at a chosen significance level, it suggests evidence of autocorrelation in the residuals (the null hypothesis is rejected), indicating that the model may need further refinement.

The Box–Pierce statistic does not always give accurate results, even when applied to datasets of moderate size. Ljung and Box highlighted the improved performance of the modified statistic in small samples, which is calculated as follows:

$$LBQ = N(N+2) \sum_{k=1}^{m} \frac{ACF(k)^2}{N-k}$$

Like the Box-Pierce statistic, the Ljung-Box statistic also follows a chi-square distribution with $m - s$ degrees of freedom, where s is the number of coefficients being estimated. Similarly, if the computed LBQ statistic is greater than the chi-square critical value at a chosen significance level, it indicates evidence of autocorrelation in the residuals.

In practical applications, following the approach mentioned earlier, we examine whether individual autocorrelation coefficients lie within the 95% confidence intervals and collectively assess their statistical significance using the Ljung-Box statistic to test H_0.

Nevertheless, many time series observed across diverse scientific domains, including economics, actuarial science, and finance, often exhibit characteristics that are better represented as non-stationary.

1.1.7 The model of a random walk

An example of a simple random walk is the series Y_t, where the t-th term is the sum of terms up to order t of a white noise process. In a more general context, the model of a simple random walk can be expressed as $Y_t = Y_{t-1} + \varepsilon_t$, with $E({\varepsilon_t}^2) = \sigma^2$.

The mean and variance of Y_t can be described as follows (Milionis, 2016):

$E(Y_t) = 0$ and $E(Y_t^2) = t\sigma^2$, therefore $\lim_{t\to\infty} E(Y_t^2) = \infty$.

Consequently, the series Y_t exhibits non-stationary behavior.

If we make the assumption that the initial value of a time series, described by a random walk process, is y_0 at time 0, then the general expression for the time series can be represented as:

$$Y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

If there is a deterministic drift, the random walk with a drift model is expressed by the equation:

$$Y_t = Y_{t-1} + \varepsilon_t + \xi.$$

For $\xi > 0$, the series will tend to move upwards. The opposite (downwards) holds good if $\xi < 0$. The presence of the constant term in the model creates a trend, which can be demonstrated as follows:

When $\xi > 0$, the series will tend to exhibit an upward movement, while the opposite (downward movement) holds true for $\xi < 0$. The inclusion of the constant term in the model creates a trend, which can be demonstrated as follows:

Through iterative substitution of $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k}$ into the initial equation, we obtain:

$$Y_t = Y_{t-1} + \varepsilon_t + \xi = Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t + 2\xi = \cdots = Y_{t-k} + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-k+1} + k\xi$$

Assuming that at $t = 0, Y_t = y_0$, we ultimately arrive at:

$$Y_t = y_0 + \xi t + \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

The last equation indicates that the constant term ξ represents the slope in the deterministic drift, as denoted by the term ξt. The inclusion of this term significantly improves the predictive performance of the model.

1.1.8 Difference stationary series and test for non-stationarity

A non-stationary series that can be transformed into a stationary one by differencing it d times is called a homogeneous non-stationary series of order d or an integrated of order d, and is denoted as $I(d)$. The stationary series is denoted as $I(0)$. For a homogeneous non-stationary series, the autocorrelations in the $ACF(k)$ decrease very slowly as k increases. This serves as an initial practical criterion for the presence of homogeneous non-stationarity. If the series is stationary, the autocorrelations would decrease rapidly as k increases. This criterion should be used only as a supplementary tool to the classical tests for non-stationarity, such as the test referred to below, and not as a standalone method.

Numerous tests and techniques are available for assessing non-stationarity in levels. Dickey and Fuller (1979, 1981) were the first researchers to explore unit root tests. Building upon their work, Said and Dickey (1984) extended the basic autoregressive unit root test to handle more complex $ARMA(p, q)$ models with unknown orders. This enhanced test is commonly known as the Augmented Dickey–Fuller (ADF) test. An extension of the Dickey–Fuller test is the Phillips and Perron test (Phillips and Perron, 1988). The Phillips–Perron (PP) test varies from the ADF test primarily in its treatment of serial correlation and heteroskedasticity. More specifically, the PP test does not account for any serial correlation, and the main advantage compared to the ADF test is its robustness to various forms of heteroskedasticity in the error term. The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test, developed by Kwiatkowski et al. (1992), takes a different approach by testing for stationarity rather than a unit root. It examines the null hypothesis of stationarity, which is that a time series is $I(0)$, against the alternative hypothesis of a unit root. On the other hand, the null hypothesis of the ADF and PP tests is that a time series is $I(1)$. A drawback of the ADF and PP tests is their low statistical power. Elliott, Rothenberg and Stock (1996) proposed an alternative test, namely the ERS (Elliott–Rothenberg–Stock) test, which has higher power than the ADF and PP unit root tests. In addition, the examination of autocorrelation and partial autocorrelation function patterns is a useful technique to test non-stationarity in levels.

However, non-stationarity can be present not only in the mean but also in the variance. Despite the significance of addressing non-constant variance in time series modeling, there is limited theoretical research on its detection and correction. Further, at the practical level, the treatment of non-stationary variance is insufficient, since when a particular transformation is used its selection is often arbitrary. The main objective of this Ph.D. thesis is to address this research gap.

1.2 Univariate $ARIMA(p, d, q)$ modeling

According to Wold's theorem, the approach of analyzing a stationary stochastic process as a weighted sum of an infinite number of white noise terms necessitates the determination of an infinite set of parameters ψ_i. However, in practice, this becomes practically impossible as we typically have only a finite amount of data available. Consequently, we will explore the patterns that emerge by introducing additional assumptions about the nature of $\psi_1, \psi_2, \ldots$.Specifically, we assume that the infinite-term polynomial $\Psi(B)$ can be represented as the division of two polynomials with finite degrees, as follows (Milionis, 2016):

$$\Psi(B) = \frac{1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q}{1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p} = \frac{\Theta(B)}{\Phi(B)} = \Phi^{-1}(B)\Theta(B)$$

Given the aforementioned assumption, a purely non-deterministic stationary stochastic process Y_t with a mean of zero can be represented in the following manner:

$$\Phi(B)Y_t = \Theta(B)\varepsilon_t$$

The coefficients of $\Phi(B)$ and $\Theta(B)$ are determined from the available data. In practical applications, the mathematical models we use often involve certain constants or parameters that need to be estimated based on the available data. It is crucial to strive for simplicity by using the minimum number of parameters that still provide sufficient representations. This principle of parsimony (Tukey, 1961) emphasizes the importance of keeping the models concise and efficient.

When setting $\Theta(B)$ to 1, the model can be expressed in the following manner:

$$\Phi(B)Y_t = \varepsilon_t \Leftrightarrow \left(1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p\right)Y_t = \varepsilon_t \Leftrightarrow$$

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$$

To clarify, Y_t is represented as a linear combination of past values up to lag p, along with a white noise process, which signifies the stochastic nature of the model. This specific model is referred to as an autoregressive model of order p and is denoted as $AR(p)$.

When setting $\Phi(B)$ to 1, the general model can be expressed as follows:

$$Y_t = \Theta(B)\varepsilon_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)\varepsilon_t$$
$$= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

To elaborate further, Y_t is represented as a linear combination of past values of a white noise process, along with the current value of that process. This specific model is referred to as the moving average model of order q and is denoted as $MA(q)$.

The general stationary stochastic process encompasses both the autoregressive process of order p and the moving average process of order q. This type of process is referred to as a mixed process of order p and q and is denoted as $ARMA(p, q)$. Mixed processes are useful for effectively representing an $AR(p)$ or $MA(q)$ process, particularly when either p or q is large.

When the process Y_t is derived from a homogeneous non-stationary process W_t through d successive differentiations, W_t is referred to as an integrated mixed process of order p, d, q and is denoted as $ARIMA(p, d, q)$.

1.2.1 The general autoregressive model $AR(p)$ and the general moving average model $MA(q)$

The $AR(p)$ model is considered stationary when the roots of its characteristic polynomial $1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_P B^P$ are located outside the unit circle. The characteristic polynomial is derived from the following equation:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t \Leftrightarrow Y_t - \varphi_1 B Y_t - \varphi_2 B^2 Y_t - \cdots - \varphi_P B^P Y_t = \varepsilon_t \Leftrightarrow (1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_P B^P) Y_t = \varepsilon_t$$

To calculate the variance of an autoregressive model, we compute the product of the $AR(p)$ model with Y_t and then take the expected values:

$$E[Y_t Y_t] = \varphi_1 E[Y_t Y_{t-1}] + \varphi_2 E[Y_t Y_{t-2}] + \cdots + \varphi_p E\left[Y_t Y_{t-p}\right] + E[Y_t \varepsilon_t] \Leftrightarrow$$

$$\gamma_0 = \varphi_1 \gamma_1 + \varphi_2 \gamma_2 + \cdots + \varphi_p \gamma_p + \sigma^2 \Leftrightarrow 1 = \varphi_1 \rho_1 + \varphi_2 \rho_2 + \cdots + \varphi_p \rho_p + \frac{\sigma^2}{\gamma_0} \Leftrightarrow$$

$$\gamma_0 = \frac{\sigma^2}{1 - \varphi_1 \rho_1 - \varphi_2 \rho_2 - \cdots - \varphi_p \rho_p}$$

To determine the autocovariance of an autoregressive model, we multiply the $AR(p)$ model by Y_{t-k} and then calculate the expected values:

$E[Y_t Y_{t-k}] = \varphi_1 E[Y_{t-1} Y_{t-k}] + \varphi_2 E[Y_{t-2} Y_{t-k}] + \cdots + \varphi_p E\left[Y_{t-\rho} Y_{t-k}\right] + 0$ since $k > 0$

So, $\gamma_k = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2} + \cdots + \varphi_p \gamma_{k-p}, k > 0$

Dividing γ_k by γ_0, we can deduce that $\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \cdots + \varphi_p \rho_{k-p}$

If we substitute the index t with the index k, the difference equation mentioned above is equivalent to the homogeneous part of the difference equation that describes the $AR(p)$ process (Milionis, 2016).

The general solution can be obtained in the following manner: First, the characteristic polynomial $\Phi(B)$ is derived as $1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_P B^P$, which factorizes into

$\Phi(B)=\prod_{i=1}^{\rho}(1-g_iB)$, where $g_1^{-1},g_2^{-1},\dots,g_p^{-1}$ represent the roots of $\Phi(B)$. Consequently, the general solution will take the form $\rho_k=A_1g_1^k+A_2g_2^k+\cdots+A_pg_p^k$, with $A_1,A_2,\dots,A_p$ representing constants determined by the initial conditions.

We can categorize the following scenarios:

1) The term $A_ig_i^k$ diminishes exponentially towards zero when the corresponding g_i is a real number.
2) The term $A_1d^k cos(\omega k+A_2)$ represents a damped sine wave with decreasing amplitude when the corresponding g_i is a complex number. In this case, the aforementioned term is formed by the root and its complex conjugate.

As a result, the autocorrelation function will exhibit a mixture of exponential decays and damped sinusoidal waves with declining amplitudes.

Identifying the order of the AR process based solely on the form of the autocorrelation function (ACF) is challenging. However, the partial autocorrelation function ($PACF$) is highly beneficial in this regard and aids in determining the appropriate order. In an $AR(p)$ process, the ACF slowly decreases, but the $PACF$ displays exactly as many statistically significant autocorrelations as the order of the $AR(p)$ process.: Let's represent φ_{kj} as the j-th coefficient in an autoregressive model of order k, where φ_{kk} is the coefficient of the last term. In this context, the following equations are valid:

$\rho_j=\varphi_{k1}\rho_{j-1}+\varphi_{k2}\rho_{j-2}+\cdots+\varphi_{k(k-1)}\rho_{j-k+1}+\varphi_{kk}\rho_{j-k}$ for $j=1,2,\dots,k$

Thus, the Yule–Walker equations are derived, which can be expressed as follows:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix}\begin{bmatrix}\varphi_{k1}\\ \varphi_{k2}\\ \vdots \\ \varphi_{kk}\end{bmatrix}=\begin{bmatrix}\rho_1\\ \rho_2\\ \vdots\\ \rho_k\end{bmatrix}$$

Or $\rho_k\overrightarrow{\varphi_k}=\overrightarrow{\rho_k}$

By solving the above equations successively for $k=1,2,3,\dots$, we obtain:

$$\varphi_{11}=\rho_1,\ \varphi_{22}=\frac{\begin{vmatrix}1 & \rho_1\\ \rho_1 & \rho_2\end{vmatrix}}{\begin{vmatrix}1 & \rho_1\\ \rho_1 & 1\end{vmatrix}}=\frac{\rho_2-\rho_1^2}{1-\rho_1^2},\ \varphi_{33}=\frac{\begin{vmatrix}1 & \rho_1 & \rho_1\\ \rho_1 & 1 & \rho_2\\ \rho_2 & \rho_1 & \rho_3\end{vmatrix}}{\begin{vmatrix}1 & \rho_1 & \rho_2\\ \rho_1 & 1 & \rho_1\\ \rho_2 & \rho_1 & 1\end{vmatrix}}$$

In an $AR(p)$ process, the values of φ_{kk} will not be zero for $k \leq \rho$, whereas they will be zero for $k > \rho$. To obtain empirical estimate of the φ_{kk} coefficients, we replace ρ_k with their sample estimates, denoted as $\widehat{\rho_k}$.

The general moving average model can be represented as follows:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \Leftrightarrow$$

$$Y_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) \varepsilon_t$$

For the general moving average model to be invertible, the roots of its characteristic polynomial $1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ must be located outside the unit circle. The characteristic polynomial is derived from the following equation:

$$Y_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) \varepsilon_t \Leftrightarrow \varepsilon_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)^{-1} Y_t$$

The variance γ_0 and the autocovariance γ_k are calculated as follows:

$$\gamma_0 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 = \sigma^2 \left(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)$$

$\gamma_k = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} = \sigma^2 \left(-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \cdots + \theta_{q-k} \theta_q\right), k = 1, \ldots, q$
with $\theta_0 = 1$ and $\gamma_k = 0$ for $k > q$.

Consequently, the autocorrelation function is calculated as:

$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \cdots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2}$, with $k = 1,2, \ldots, q$ and $\rho_k = 0$ for $k > q$.

The behavior of the partial autocorrelation function ($PACF$) at lag k closely resembles the autocorrelation function (ACF) for an autoregressive process. In other words, the $PACF$ of a moving average process will exhibit a mix of exponential decays towards zero and damped sine waves with diminishing amplitudes, depending on the type of the roots of $\Theta(B) = 0$ (the principle of duality between AR and MA processes).

1.2.2 Mixed models

To achieve parsimony, it may be required to incorporate both autoregressive and moving average components. Hence, it might be necessary to utilize a mixed $ARMA$ model:

$$\Phi(B)Y_t = \Theta(B)\varepsilon_t$$

If all the roots of the characteristic equation $\Phi(B) = 0$ lie outside the unit circle, the process will be considered stationary. Similarly, for the process to be invertible, the roots of $\Theta(B) = 0$ must be located outside the unit circle.

The ACF and $PACF$ of an $ARMA(p, q)$ process will begin to decay either exponentially towards zero or with damped sine waves after lag q and p, respectively.

1.3 Time series decomposition and seasonal $ARIMA(p, d, q)(P, D, Q)_s$ models

In practical applications, it is beneficial to assume that an observed time series can be decomposed into unobservable components using the following equation:

$$Y_t = S_t + P_t + C_t + U_t$$

where:

Y_t is the observed time series or some transformation of it,

S_t is the seasonal component,

P_t is the long-term trend,

C_t is the cyclical component,

U_t is the irregular component.

The seasonal component of time series decomposition refers to the recurring patterns that occur within a time series at regular intervals, typically over the course of a year. These patterns can be influenced by various factors such as weather, holidays, or cultural events. Two examples illustrating the seasonal component are tourist arrivals (e.g. May to September in Greece) and retail sales (e.g. a retail business that sells winter apparel).

The long-term trend signifies the general change in the level of the time series. In non-stationary time series, the long-term trend can either remain constant or vary over time. In instances where the trend is time-varying, linear regression models often lead to highly inaccurate forecasts in many cases.

The cyclical component of time series decomposition refers to the medium-term fluctuations in a time series. Unlike the seasonal component, which has regular and predictable patterns, cyclical movements are irregular and can vary in duration and amplitude. These fluctuations are typically linked to economic or business cycles and can span multiple years.

The irregular component, also known as noise, encompasses the cumulative effects of non-systematic factors. It represents the random or unpredictable variations that are not accounted for by the systematic components of the time series. The irregular component can include various sources of randomness, measurement errors, outliers, or unexplained fluctuations that cannot be attributed to the underlying patterns or components of the series.

By drawing an analogy with the non-seasonal model $ARIMA(p, d, q)$, it is possible to formulate a similar model to capture the correlations between observations for the same month in different years, in the following format:

$$\Phi(B^{12})\nabla_{12}Y_t = \Theta(B^{12})\varepsilon_t$$

The polynomials $\Phi(B^{12})$ and $\Theta(B^{12})$ are of degree P and Q, respectively, in relation to the lag operator B raised to the 12th power. In other words:

$$\Phi(B^{12}) = 1 - \varphi_1 B^{12} - \varphi_2 B^{24} - \cdots - \varphi_P B^{P\cdot 12}$$

$$\Theta(B^{12}) = 1 - \theta_1 B^{12} - \theta_2 B^{24} - \cdots - \theta_Q B^{Q\cdot 12}$$

If we combine the seasonal and non-seasonal model, we arrive the composite model:

$$\Phi(B)\Phi(B^{12})\nabla\nabla_{12}Y_t = \Theta(B)\Theta(B^{12})\varepsilon_t$$

In other words, the composite model takes the form of a multiplicative model denoted as $ARIMA(p, 1, q)(P, 1, Q)_{12}$. The prevalent type of seasonal model is referred to as the airline model, which can be expressed as $ARIMA(0,1,1)(0,1,1)_s$, where s indicates the seasonality.

1.3.1 The variable seasonal pattern

The process of removing the seasonal component from time series data involves intricate statistical procedures, and it is commonly executed using specialized statistical software like X-12-REG-ARIMA, TRAMO-SEATS, and JDemetra+. More specifically:

X-12-REG-ARIMA is a software package developed by the U.S. Census Bureau and the Statistical Office of Canada (Findley et al., 1998), designed for seasonal adjustment and time series analysis. It is widely used by statistical agencies, researchers, and economists to remove seasonal variations from economic time series data.

TRAMO-SEATS is widely regarded as a robust and sophisticated tool for seasonal adjustment, particularly for complex time series with irregular seasonality. This program was developed by Gómez and Maravall (1996) at Banco de España, with the support of Eurostat. The latest version operates in a Windows environment under the name TSW. TRAMO stands for "Time series Regression with ARIMA noise Missing observations and Outliers" and SEATS stands for "Signal Extraction in ARIMA Time Series".

JDemetra+: is an advanced software tool developed by the National Bank of Belgium and Eurostat since February 2015. It has gained popularity among statistical practitioners due to its powerful capabilities, comprehensive reporting, and user-friendly interface. It is continually updated and maintained to incorporate the latest developments in seasonal adjustment methods, ensuring the accuracy and reliability of the results.

However, users should have a solid understanding of seasonal adjustment principles and the specific characteristics of their time series data to make appropriate and informed decisions when using the aforementioned software.

1.4 Box–Jenkins univariate stochastic models

Box-Jenkins univariate stochastic models, also known as ARIMA models, are a powerful class of time series models used for analyzing and forecasting single-variable time series data. Developed by Box and Jenkins (1976), these models have become widely popular and essential in time series analysis due to their flexibility, simplicity, and effectiveness in handling a wide range of time series patterns. The Box-Jenkins procedure consists of four main stages for developing an ARIMA model These stages are as follows:

(i) The identification of the model: The first stage involves identifying the appropriate orders of the AR and MA components, as well as the order of differencing needed to achieve stationarity. This is done by analyzing the autocorrelation function and partial autocorrelation function of the time series data. Additionally, if the original data is non-stationary, the order of differencing required to make the data stationary is determined.

(ii) The estimation of the model: The coefficients of the ARIMA model are estimated using various techniques, with maximum likelihood estimation being one of the most commonly used methods. The coefficients of the model should fall within the bounds of invertibility, stationarity, and demonstrate statistical significance.

(iii) The diagnosis of the model: The third stage involves performing diagnostic tests to assess the adequacy of the model. The diagnosis of the models involve analyzing the residuals to ensure that they meet certain assumptions. (i.e. the null hypothesis that the residuals of the model are white noise). To avoid rejecting the null hypothesis, two conditions must be met: a) there should be no significant correlation up to the initial lags and b) the LBQ test's value should not exhibit statistical significance.

(iv) The metadiagnosis of the model: The proposed model is evaluated against other competing models, with the key criteria being the model's parsimony and the residual mean square (RMS) value. The RMS is calculated as follows:

$$RMS = \frac{1}{N}\sqrt{\sum_{t=1}^{N} \hat{\varepsilon}_t^2}$$

For the parsimony of the model with the best fit, various statistical criteria can be used such as:

a) The Bayesian information criterion (BIC) of Schwarz (1978):

$BIC(p, q) = ln\hat{\sigma}^2 + (p + q)N^{-1}lnN$, where $\hat{\sigma}^2 = \frac{\sum_{t=1}^{T} \varepsilon_t^2}{N}$ is the estimation of the variance of the residuals and N is the number of terms.

b) The Akaike (1974) criterion: $AIC(p,q) = ln\hat{\sigma}^2 + 2(p+q)N^{-1}$

BIC has a stronger penalty for model complexity compared to AIC. This means that BIC tends to prefer simpler models with fewer parameters. If the primary goal is to obtain a simpler model and the sample size is relatively large, BIC may be preferred as it tends to favor more parsimonious models. When choosing between models, the lowest value of the above tests indicates a better fit.

1.5 Time series linearization

The $ARIMA$ models provide a practical approach to capturing the features and patterns present in time series. Building an $ARIMA$ model may require pre-adjustments for the following reasons:

a) Outliers: Outliers are data points in a time series that deviate significantly from the overall pattern or trend. They are extreme values that lie far away from the majority of data points and can have a substantial impact on the statistical properties of the time series. $ARIMA$ models assume that the data are generated from a stationary stochastic process with no significant outliers. If outliers are present in the data and left untreated, they can lead to inaccurate model parameter estimates, affect the model's ability to capture the underlying patterns, and result in unreliable forecasts. Outliers are typically linked to three primary types of effects:
   i) Additive Outlier (AO): This effect only impacts a single isolated observation.
   ii) Transitory Change (TC): This resembles an additive outlier, but its effect does not immediately fade away but rather persists over several periods.
   iii) Level Shift (LS): This implies a change in the mean level of the series.

b) Calendar effects: Calendar effects refer to regular patterns that occur at specific time intervals within a year. Some common calendar effects include holiday effect and day of the week effect. Typically, these effects are integrated into the model by using regression variables.
c) Intervention variables: Time series data can be influenced by extraordinary or uncommon events that are difficult to incorporate into an ARIMA model.

Consequently, it becomes necessary to "intervene" in the series to account for the effects of these exceptional events. Some examples of such events are policy changes, technological advancements, labor strikes, major events and celebrations (like the Olympics) and so on. These particular effects can be included in the model as regression variables, commonly known as intervention variables according to Box and Tiao (1975).

The general framework of linearization (Kaiser and Maravall, 2001) for the original series can be written as

$$y_t = w_t'b + C_t'\eta + \sum_{j=1}^{m} \alpha_j \mu_j(B) I_t(t_j) + x_t$$

Where: $y_t = f(z_t), f$ is some transformation of the raw series z_t, which may be necessary to stabilize the variance;

$b = (b_1, \dots, b_n)$ is a vector of regression coefficients;

$w_t' = (w_{1t}, \dots, w_{nt})$ denotes n regression or intervention variables;

C_t' denotes the matrix whose columns represent possible calendar effect variables (e.g., trading day);

η is the vector of associated coefficients;

$I_t(t_j)$ is an indicator variable for the possible presence of an outlier at period t_j;

$\mu_j(B)$ captures the transmission of the j-th effect;

α_j denotes the coefficient of an outlier in a multiple regression model with m outliers;

x_t, in general, follows a multiplicative seasonal ARIMA(p,d,q)(P,D,Q)s model:

$$\varphi(B)\Phi(B^s)\nabla^d\nabla_s^D x_t = \theta(B)\Theta(B^s)\varepsilon_t$$

where:

$\varphi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is the so-called autoregressive polynomial of order p;

$\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ is the so-called moving average polynomial of order q;

$\nabla^d \equiv (1 - B)^d$ is an arithmetic difference operator of order d;

$\nabla_s^D \equiv (1 - B^s)^D$ is a seasonal arithmetic difference operator of order D and seasonality s;

$\Phi(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{P \cdot s}$ is the so-called seasonal autoregressive polynomial of order P and seasonality s;

$\Theta(B^s) = 1 - \Theta_1 B^s - \cdots - \Theta_Q B^{Q \cdot s}$ is the so-called moving average polynomial of order Q and seasonality s;

ε_t is the stochastic disturbance.

There are several software programs (such as X-12-REG-ARIMA, TRAMO-SEATS, and JDemetra+) that implement these procedures to estimate the parameters of the general framework of linearization.

1.6 Thesis motivation and main objectives

1.6.1 Further discussion on stationarity in the second moment

As far as variance stabilization is concerned, if the variance is functionally related to the level and the level is non-stationary (which is often the case with economic-financial time series), the variance is neither conditionally nor unconditionally constant. Hence, the process is non-homogeneously non-stationary in the sense of Box and Jenkins and cannot be made stationary by simply differencing it. One way to tackle variance non-stationarity is to employ power transformations, such as the well-known class of the Box and Cox transformations (Box and Cox, 1964). For instance, the following transformation is very often used:

$$f(z_t) = \begin{Bmatrix} \dfrac{z_t^{\lambda} - 1}{\lambda} \; if \; \lambda \neq 0 \\ ln z_t \; if \; \lambda = 0 \end{Bmatrix}$$

Regarding the arithmetic values of the exponent λ of the above equation, for practical purposes, Makridakis et al. (1998) state that there is no merit in using arithmetic values with several decimal points, as nearby values will produce very similar results. Simple arithmetic values of λ are easier to interpret and, hence, more meaningful. Furthermore, Kalligeris et al. (2019) acknowledge that when $\lambda < 0$, an alternative model is necessary, which diminishes the attractiveness of the methodology. To address this limitation, they propose a comprehensive model selection approach that remains applicable even in cases involving negative values.

At the practical level, the treatment of non-stationary variance is occasionally biased towards over-rejection of the null hypothesis of unconditionally constant variance, as is argued in subsequent chapters.

More specifically, the existing statistical approaches for the detection and correction of variance non-stationarity appear to have several disadvantages, viz.: (i) although they detect variance non-stationarity, the correction they suggest is not formally and rigorously documented (e.g., Hay and McLeay, 1979; Milionis and Davies, 1994); (ii) they usually suffer from subjectivity (see, for instance, Mills, 1990 chapter 4, for a short review); and (iii) although they detect variance non-stationarity and are formally suggestive of a solution, they lack robustness (Milionis, 2003; Milionis, 2004).

Failure to account for non-stationarity in variance leads to distortions in both the variance itself and the autocovariances of the time series. This can result in a correlation between terms that are lags apart, leading to the presence of artificially statistically significant coefficients at certain lags. These issues contribute to the problem of overparameterized models. Furthermore, variance stationarity is a critical requirement for outlier analysis. Without it, the identification of outliers, their types, and any economic significance associated with them become invalid.

It is important to distinguish between two concepts: variance non-stationarity, also known as heteroscedasticity, and conditional heteroscedasticity. Heteroscedasticity implies a functional relation between the variance of a series which is non-stationary in its level and its mean level. This entails non-stationarity in the variance (Milionis, 2004). On the other hand, conditional heteroscedasticity, usually described by ARCH or GARCH models (Engle, 1982; Bollerslev, 1986), means that while the conditional variance changes over time, the unconditional variance remains constant. Therefore, the series is stationary in the second moment.

Indeed, considering, without loss of generality, the simple ARCH(1) model, i.e., $f(X_t) = f(X_{t-1}; b) + e_t$, where X_t is a stochastic process, b represents a parameter vector, $e_t = v_t\sqrt{\omega + a_1 e_{t-1}^2}$, $\omega > 0$, $0 < a_1 < 1$, and v_t is a white noise with unit variance. For this model it is easily proved that (Enders (1995), Milionis (2004)): i) the unconditional mean of e_t is equal to zero, ii) the conditional variance of e_t is equal to $\omega + a_1 e_{t-1}^2$, meaning it varies with time, and iii) the unconditional variance of e_t is equal to $\frac{\omega}{1-a_1}$, i.e. it is a constant.

A specific case is that of an integrated GARCH (IGARCH) process, in which volatility exhibits persistence, and may require additional examination. Research by Nelson

(1990) has demonstrated that, unlike the typical random walk, even this process is strictly stationary, yet its unconditional variance is unbounded.

In the aforementioned models, the parameters governing the conditional variance are concurrently estimated alongside the parameters related to the series' level. From a methodological perspective, it is essential to address variance non-stationarity before addressing non-stationarity in levels.

1.6.2 Thesis aim

The aim of this Ph.D. thesis is to present a formal econometric approach that not only identifies non-stationary variance and suggests appropriate transformations for correction but is also robust to the specific partitioning of a time series, which is a necessary step for conducting the test, and the possible presence of outliers. The importance of the application of this methodology in macroeconomics, actuarial science and finance is thoroughly examined and evidenced in subsequent chapters. A brief outline is given below.

1.6.2a) Applications in macroeconomics

A univariate ARIMA model is a concise quantitative summary of the internal dynamics of a time series in a linear framework. It is therefore useful for several reasons, including for forecasting. More specifically, univariate forecasts usually serve either as short-term or benchmark forecasts.

The effect of the application of this methodology to some crucial elements of macroeconomic time series modeling, such as forecasting and outlier detection, will be examined. It is of much interest to investigate how variance non-stationarity could potentially affect the specification of the univariate ARIMA model and the detection of outliers.

For instance, examining the time series on monthly external trade statistics from the Balance of Payments for Greece (see Chapter 2), the presence of variance non-stationarity leads to seriously mis-specified univariate ARIMA models, a result that is in accordance with that of Milionis (2004). Also, in properly transformed data, the pattern of detected outliers is clearly different, a conclusion that is also in accordance with that of Milionis (2004).

It will also be shown that the TSW program, a specialized software for time series analysis, occasionally appears to be biased, favoring the log transformation of the data. Furthermore, the results obtained using simulated data show a bias in TSW that depends on the initial conditions. Moreover, it will be established that the consequences of falsely transforming a time series which is originally variance stationary do exist but are less severe than the consequences of falsely not transforming an originally variance non-stationary series.

In addition, utilizing 20 of the most important time series for the Greek economy, the empirical findings show a significant improvement in the confidence intervals of forecasts but no substantial improvement in point forecasts (see Chapter 3). Furthermore, the combined transformation–linearization procedure improves substantially the non-normality problem encountered in many macroeconomic time series.

1.6.2b) Applications in actuarial science

Longevity is a threat for insurance companies or pension funds. Longevity risk is considered the possibility that life expectancy, or actual survival rates, will exceed the "expected". If indeed this happens, then the outflow of money from the funds will be greater and, as a result, the risk now lurks for the company or the pension fund. This risk exists—in the last 50–60 years there has been a trend of increasing life expectancy—therefore insured persons and pensioners have to receive proceeds for more time. On the other hand, the number of people reaching retirement age is constantly increasing. The combination of the two results in higher payout levels than originally thought. The types of plans exposed to the highest levels of longevity risk are pension plans and life annuities. Figures for average life expectancy are increasing, and even a small change in life expectancy can create serious solvency issues for pension plans and insurance companies. Therefore, it is very crucial to predict mortality rates as accurately as possible. Aiming at possible improvements of such forecasts, the effect of data transformation–linearization on the quality of time series forecasts of mortality is examined, and results indicate a clear improvement for interval forecasts of mortality (see Chapter 4). The documented improvement in interval forecasts can significantly affect the Solvency Capital Requirement, giving some pension providers a competitive advantage.

1.6.3c) Applications in finance

The predictability of stock returns and the concept of market efficiency have earned significant interest among researchers. This is evident from the extensive body of research published on these topics. The efficient market hypothesis posits that "security prices fully reflect all available information" (Fama, 1970).

The traditional classification of available information (Roberts, 1959) categorizes market efficiency into three forms: weak-form efficiency, which considers past prices as the information set; semi-strong efficiency, which contains all publicly available information in the information set; and strong-form efficiency, which encompasses both publicly and privately available information in the information set. In an efficient market, tests of return predictability should fail to reject the null hypothesis of no predictability.

To empirically test the weak-form market efficiency hypothesis (WFME), it is initially assumed that conditions of equilibrium can be described in terms of expected returns. This can be written as:

$$E(\tilde{P}_{j,t+1} \,/\, \Phi_t) = [1 + E(\tilde{R}_{j,t+1} \,/\, \Phi_t)]P_{j,t}$$

where $P_{j,t}$ represents the price of security j at time t, $R_{j,t+1}$ represents the percentage return of security j between t and $t+1$, and Φ_t represents the information set that is fully reflected in $P_{j,t}$. The use of tildes denotes random variables at time t.

These expected returns are determined using a pricing model, making the test of WFME a joint test of WFME and the pricing model. When the adopted model assumes constant expected returns in a risk-unadjusted framework, it is common to apply tests for autocorrelation in security returns. However, it is important to note that statistically significant correlations alone do not necessarily imply the rejection of the WFME hypothesis, as the joint hypothesis being tested includes both WFME and the pricing model with constant expected returns.

Elton et al. (2014) provide a comprehensive compilation of autocorrelation test results conducted by multiple researchers. These tests analyze the first differences of prices or the logarithms of prices. It is important to note that these tests are valid only when the series of first differences of prices exhibits variance stationarity. If this condition is not met, the significance testing of autocorrelation coefficients becomes invalid. So, to elucidate the conditions under which autocorrelation tests or similar tests for market efficiency using returns are significant, the following proof offers a comprehensive statistical rationale that establishes the compatibility between interdependence in returns and market efficiency (Milionis, 2007).

Proof: $COV(\tilde{R}_{j,t}, \tilde{R}_{j,t+1}) =$

$$\iint_{R_t R_{t+1}} \left(\tilde{R}_{j,t} - E(\tilde{R}_{j,t})\right) \cdot \left(\tilde{R}_{j,t+1} - E(\tilde{R}_{j,t+1})\right) f(R_{j,t}, R_{j,t+1}) dR_t dR_{t+1} =$$

$$\iint_{R_t R_{t+1}} \left(\tilde{R}_{j,t} - E(\tilde{R}_{j,t})\right) \cdot \left(\tilde{R}_{j,t+1} - E(\tilde{R}_{j,t+1})\right) f(R_{j,t}) f(R_{j,t+1}/R_t) dR_t dR_{t+1}$$

as $f(R_t, R_{t+1}) = f(R_t) f(R_{t+1}/R_t)$.

Based on the definition of conditional expected value, it holds that:

$$\int_{R_t} \left(\tilde{R}_{j,t+1} f(R_{t+1}/R_t)\right) dR_{t+1} = E(\tilde{R}_{j,t+1}/R_t)$$

Hence:

$$COV(\tilde{R}_t, \tilde{R}_{t+1}) = \int_{R_t} \{\tilde{R}_t - E(\tilde{R}_t)\} \cdot \{E(\tilde{R}_{t+1}/R_t) - E(\tilde{R}_{j,t+1})\} \cdot f(R_t) dR_t$$

The above equation is used by Fama. As $E(\tilde{R}_{t+1}) = E\{E(\tilde{R}_{t+1}/R_t)\}$, the equation provided above can be alternatively expressed as:

$$COV(\tilde{R}_t, \tilde{R}_{t+1}) = \int_{R_t} \{\tilde{R}_t - E(\tilde{R}_t)\} \cdot \left\{E(\tilde{R}_{t+1}/R_t) - E\{E(\tilde{R}_{t+1}/R_t)\}\right\} \cdot f(R_t) dR_t$$

Based on the analysis presented above, it is clear that both of the previously mentioned equations are equivalent to the definition of autocovariance, without any specific reference to $\tilde{R}_{j,t}, \tilde{R}_{j,t+1}$. These equations hold true for any random variables $\tilde{X}, \tilde{Y}$.

Therefore, in general, $COV(\tilde{R}_t, \tilde{R}_{t+1}) \neq 0$ as $E(\tilde{R}_{t+1}/R_t) - E\{E(\tilde{R}_{t+1}/R_t)\} \neq 0$, and it is evident that for the integral in the last equation to vanish, an additional assumption needs to be made. This assumption is known as the constant expected returns, and when it holds good, it follows that $E(\tilde{R}_{t+1}/R_t) = E\{E(\tilde{R}_{t+1}/R_t)\}$ and therefore $COV(\tilde{R}_t, \tilde{R}_{t+1}) = 0$.

Autocorrelation tests can only be meaningful in cases where the assumption of constant expected returns and, in extension, of variance stationarity holds true. In this scenario, the presence of autocorrelation in stock returns can be considered as evidence for rejecting market efficiency. Specifically, it indicates the rejection of the joint hypothesis that assumes both market efficiency and constant expected returns.

In a risk-adjusted framework, when risk fluctuates over time instead of remaining constant, it follows that expected returns, considering risk aversion, should also vary over time. One of the most apparent quantitative expressions of this temporal relationship between risk and return is found in GARCH-M models, where the conditional variance can serve as a predictor for returns (Milionis, 2016). These models can be broadly formulated as:

$$\Delta log P_t = f(h_t^2, \Phi_{t-1}, \beta) + \varepsilon_t$$

The equation presented above represents returns as a function f of three components: the conditional variance (h_t^2) which reflects the risk, the information available up to time t-1 (Φ_{t-1}), and the parameter vector β. According to Milionis and Moschos (2000), two scenarios can be illustrated based on this equation:

1) When $\frac{\partial f}{\partial h_t^2} < 0$, if h_t^2 increases, this will result in a decrease in expected returns. In that scenario, if the model is correctly specified, it would lead to the rejection of the WFME.

2) When $\frac{\partial f}{\partial h_t^2} > 0$, if h_t^2 increases, this will result in an increase in expected returns. This is not incompatible with the overall concept of WFME, if investors anticipate a positive return.

1.7 Thesis outline

The structure of this thesis is as follows. In Chapter 2, the theory of the proposed methodology for non-stationarity in the second moment will be developed, and it will be demonstrated that the proposed methodology outperforms other methods. Through simulations, the superiority of the proposed methodology is highlighted, while also exposing the biases in alternative approaches. In Chapter 3, the proposed methodology is applied to macroeconomic time series, demonstrating the improvement in prediction confidence intervals that arises from its application. In Chapter 4, the aforementioned methodology will be used for longevity forecasting, and its advantages over existing approaches will be presented. In the last chapter (Chapter 5), the developed methodology will contribute to the improvement of the framework of econometric assumptions and tests in finance, aiming to determine the rejection or non-rejection of the hypothesis of weak-form market efficiency.

SUMMARY OF CHAPTER 1

While researchers typically test for non-stationarity in the level of a time series using various tests, they sometimes overlook non-stationarity in the variance when conducting applied research. Indeed, regarding time series variance, the research focus is mainly directed in modeling conditional heteroscedasticity through a plethora of ARCH-GARCH type of models. However, as argued earlier in this chapter, such models are variance stationary because although the conditional variance is time-varying the unconditional variance is constant. The implications of disregarding non-stationarity in the variance in macroeconomic time series, in actuarial science and finance are examined. A formal econometric approach is proposed to test and address non-stationarity in the variance. The existence of non-stationarity in the variance results in an inaccurate specification of univariate autoregressive integrated moving average (ARIMA) models, and the identification and analysis of outliers are then influenced by the existence of non-stationarity in the variance. The consequences of testing the hypothesis of weak-form market efficiency (WFME) are considered, particularly highlighting the inadequacy of conventional autocorrelation tests when applied to the differences in asset prices.

CHAPTER 2

DEVELOPMENT OF THE STATISTICAL METHODOLOGY AND EVALUATION OF ITS MERIT

2.1 Introduction

Over the last five decades a vast volume of research work, at both the theoretical and the applied level, has been devoted to time series with time-varying second moment. This non-constancy in the second moment may be due to various reasons. For the purposes of this work, it is methodologically useful to distinguish between type (i): series with conditionally non-constant, but unconditionally constant variance, and type (ii): series with non-constant variance both conditionally and unconditionally. In the present the focus is on the latter.

If the variance is functionally related to the level and the latter is non-stationary, the variance is not constant both conditionally and unconditionally [this is the typical case of type (ii)]. Therefore, the procedure is non-homogeneously non-stationary in the sense of Box and Jenkins (1976) and cannot be made stationary by merely taking differences. To address the issue of non-constant variance, one approach is to utilize power transformations, such as the widely recognized Box and Cox transformations (refer to section 1.6.1).

In spite of its importance for time series modelling, there is not much work at the theoretical level on the detection and correction of non-constancy in the variance owing to its dependence on a non-stationary mean level. Additionally, at the practical level, dealing with non-constant variance is not only inadequate (in fact, the choice of a specific transformation is often arbitrary) but also, occasionally, tends to show a bias towards rejecting the null hypothesis of unconditionally constant variance. This argument is discussed later in this chapter.

More specifically, the existing statistical methods used to detect and correct variance non-stationarity seem to have various drawbacks, as outlined in section 1.6.1. The aim of this work is to develop a formal econometric approach, which not only allows the detection of non-stationary variance and is suggestive of the transformation necessary to correct for it but also it is robust to the particular partition of a time series –a procedure necessary for the test- and the possible existence of outliers. Further, the possible advantages of the application of this methodology on some crucial elements of time series modelling such as outlier detection and seasonal adjustment, as compared to existing methods, are examined.

The chapter is structured as follows: In section 2.2 the statistical methodology (testing procedure) is developed and the framework is set upon which the possible advantages of the new testing procedure on univariate time series modelling are evaluated. In section 2.3 the testing procedure is applied on Greek real data, as well as on artificial time series created by statistical simulation. This aims to serve two purposes: (i) to evaluate the usefulness of the methodology in analyzing and forecasting time series from the real world; (ii) to identify biases in the algorithms incorporated in existing specialized statistical software for variance non stationarity testing. Section 2.4 concludes the chapter.

2.2 Development of the statistical methodology

As in most other similar studies (Mills, 1990; Milionis and Davies 1994; Milionis 2004), for the statistical testing approach used in this work time series are partitioned into segments (subsamples) of equal length. For each subsample the (local) mean (LM) as well as the (local) standard deviation (LSD) are calculated. Local Standard Deviation is assumed to be functionally dependent on Local Mean in a non-linear fashion as follows:

$$LSD = aLM^{\beta} e^{u} \quad (1)$$

where a, β are model parameters, e is the base of natural logarithms and u the stochastic disturbance. Model parameters a, β are estimated via Ordinary Least Squares (henceforth OLS) using the corresponding log-log model. The estimated value of β ($\hat{\beta}$) provides the necessary information for the existence (or non-existence) and the type of data transformation needed to ensure variance stationarity (e.g. for the most popular transformations, namely the log-transformation and the square root one, correspond to $\beta = 1$, and $\beta = 0.5$, respectively). This is formally stated and tested by hypothesis H_a below.

To ensure robustness with respect to a particular partition and the possible existence of outliers, as this procedure should precede the detection of outliers, the procedure is repeated for different partitions. The number of different partitions is at least equal to the number of divisors of the series' length, giving a quotient (series length over divisor) ≥ 5 and restricting the size of subsamples to be ≥ 5[1]. Robustness is formally stated and tested by hypothesis H_b below.

Finally, the previous steps are repeated with the already transformed data. The purpose of this last step is to test, whether or not, the suggested transformation is sufficient to stabilize the series variance. This is formally stated and tested by hypothesis H_c below.

2.2.1 Notation and equations

Before the description of the testing procedure some explanation on the notation and definition of the various symbols is necessary.

- Index (k) indicates the ascending number of a subsample in a partition.
- Index (j) indicates the ascending number of the particular partition, $j = 1,2, \dots, j_{max}$.
- Index i_j represents the maximum value of k (number of subsamples) in partition j.
- N is the total length (size) of the initial time series.
- n_{ij} represents the size of subsamples in partition j.
- $\hat{\beta}_j$ is the estimate of the exponent β using subsamples derived from partition with ascending number j.
- $\hat{u}_{jk}$, $\hat{\varepsilon}_j$, $\hat{u}^*_{jk}$ are independent of each other regression residuals.

An asterisk (*) over a symbol denotes the corresponding transformed data, or the corresponding parameter estimate derived from the transformed data.

- $i_j = (N/n_{ij})$, if (N/n_{ij}) is an integer; $n_{ij} \geq 5$,
- $i_j = int(N/n_{ij}) + 1$, if (N/n_{ij}) is not an integer, $n_{ij} \geq 5$ and the residual of the division is ≥ 5,

[1]Five (5) was selected as a reasonable lower limit for both the size of a subsample, as well as the number of subsamples in any partition of the original series.

- $i_j = int(N/n_{ij})$, if (N/n_{ij}) is not an integer, $n_{ij} \geq 5$ and the residual of the division is < 5,
- $\hat{\beta}_j$ is estimated for each partition j, $j = 1,2, \ldots, j_{max}$ via OLS from the model (First stage regression):

$$ln(LSD_{jk}) = ln(\alpha_j) + \hat{\beta}_j\, ln(LM_{jk}) + \hat{u}_{jk} \quad (2)$$

- $\hat{\beta}$ is estimated via OLS as the constant term of the model (Second stage regression):

$$\hat{\beta}_j = \hat{\beta} + \hat{d}j + \hat{\varepsilon}_j \quad (3)$$

- Model using the transformed data (Third stage regression):

$$ln(LSD^*_{jk}) = ln(\alpha^*_j) + \hat{\beta}^*_j\, ln(LM^*_{jk}) + \hat{u}^*_{jk} \quad (4)$$

2.2.2 Statistical Hypotheses and comments

Applying the procedure described above, it can be made possible to state and test the following statistical hypotheses:

1) $\boldsymbol{H_a}$: $\boldsymbol{\beta_j = 0}\ \forall \boldsymbol{j}$ (or at least the majority of $\boldsymbol{\beta_j s}$).

This hypothesis can be tested from the first stage regression (Equation 2) and is utilized to ensure that indeed there exists a dependence of local standard deviation on local mean. Failure to reject $\boldsymbol{H_a}$ means that there is no such dependence, hence, no variance instability of type II exists, and therefore, the algorithm stops.

2) $\boldsymbol{H_b}$: $\boldsymbol{d = 0}$[2] **(Robustness test)**

The dependent variable in Equation 3 (second stage regression) is the estimate of β derived from the partition of ascending number j $(\hat{\beta}_j)$, while the independent variable is the ascending number of the partition itself (j). Therefore, $\hat{d}$ is the estimate of the slope of the regression. This hypothesis states that the slope $\hat{d}$ should not be statistically significant, and non-rejection of it, means that $\hat{\beta}$ is robust to any particular partition of the series, or outliers. Additionally, non-rejection of $\boldsymbol{H_b}$ also ensures a better estimate of β by making more efficient use of information available in all partitions.

[2]Typically, an additional Hypothesis: **H'$_b$:** β=**0** should also be tested, but practically its rejection is ensured by the previous rejection of **H$_a$.**

3) $\boldsymbol{H_c}: \boldsymbol{\beta_j^*} = \mathbf{0}\ \forall \boldsymbol{j}$ **(Under-transformation test)**

where β_j^* are the corresponding β_j for the transformed data estimated from the third stage regression (Equation 4). The third hypothesis is a kind of an "under-transformation" test. If $\hat{\beta}_j^*$ are statistically significant it means that there will still be a dependence between local mean and local standard deviation in the already transformed data, consequently the chosen transformation did not succeed in removing type II variance instability. Therefore, $\boldsymbol{H_c}$ states that there is no remaining dependence of local mean on local standard deviation in the transformed data. Non-rejection of this hypothesis ensures that the chosen transformation is sufficient and has adequately rendered an unconditionally stable variance.

The addition of robustness and under-transformation tests to the above methodology offers advantages over existing methods, as the latter are devoid of these features.

2.3 Evaluation of methodology's merit

Real data will be used to evaluate the proposed statistical methodology against existing statistical software programs in time series analysis. More specifically, the advantages arising from univariate ARIMA modelling, outlier detection, seasonal adjustment, and forecasting performance of the univariate models will be presented through the application of the proposed methodology.

Furthermore, by utilizing real data, the bias which is present in existing statistical tests concerning the rejection of the null hypothesis of unconditionally constant variance will be highlighted. Additionally, simulated time series will be used to identify one of the sources for the bias in rejecting the null hypothesis of constancy in the second moment.

2.3.1 Application on Greek real data

The Bank of Greece produces routinely, for internal use, seasonally adjusted data, as well as purely statistical (atheoretical) univariate forecasts for several Balance of Payments(BOP) series of monthly observations. Such benchmark forecasts are useful, amongst others, for the comparison with actual values, when the latter become available. Seasonal adjustment and forecasts are produced in conjunction with outlier detection and use is made of the algorithms of the specialized statistical software TSW for this purpose. TSW stands for TRAMO-SEATS for Windows, a Windows version of the DOS programmes TRAMO and SEATS (see Gómez and Maravall, 1996 and section 1.3.1). TSW is freely available by the provider (Bank of Spain) and is currently used by many NCBs, NSIs as well as many other academic and international institutions (universities, ECB, EUROSTAT, etc.)[3]. Using TSW in the course of the routine time series analysis within the Bank of Greece it was observed that TSW suggested the logarithmic transformation of the original data in order to stabilize the variance far too often. This observation was the starting point of this work.

The time series examined are monthly external trade statistics from the Balance of Payments and prices of consumer goods and services for Greece. The particular data were selected due to their obvious importance given the continuing economic crisis in the country and the initially large current account deficit of Greece at the beginning of the economic crisis. This current account deficit is attributed primarily to the deficit of the balance of Goods (see press releases on the web site of the Bank of Greece). It is apparent that proper statistical modelling is vital for the short-term monitoring and forecasting of such series.

More specifically, the time series from external trade statistics from the Balance of Payments for Greece are Total Imports, and Total Exports of Goods excluding fuels and ships (source: Bank of Greece). Those series are of special interest for re-analysis, not least because they have recently undergone adjustments in several ways. More specifically, the International Monetary Fund (IMF henceforth) in its 6th Manual on Balance of Payments (IMF, 2009) redefined the item “Total Goods” so as the new definition be firmly based on the “change of ownership” principle. In that sense, the sub-items “goods for processing” as well as “repairs on goods”, which were included in “goods” before the new definition of IMF, are now classified as services, since no change of ownership takes place. By contrast, transactions under “merchanting”, which used to be classified as services, with the new definition are included in goods, again because according to the change of ownership principle such a change does occur in the merchanting process. A further adjustment in the series of imports and exports of goods occurred owing to the need for harmonization with the external trade statistics produced by the Hellenic Statistical Authority (ELSTAT). The recording of goods in the latter is based on customs data. Additionally, ELSTAT data include estimates about

[3]TSW routines are also incorporated in other econometric software such as E-VIEWS.

the volume of illegal trade. The objective of this adjustment is to make BOP statistics on external trade on goods fully compatible with the relative notions of the wider framework of the System on National Accounts (SNA, 2008).

The data cover the period from January 2003 to June 2015 and consist of one hundred and fifty (150) monthly observations of Total Imports, and Total Exports of Goods excluding fuels and ships (source: Bank of Greece). The dates of some of the important events that occurred during the crisis are noted in Table 1, while graphical representations of the time series are shown in Figures 1 and 2.

Table 1. Dates of important events during period

Event	Date
Lehman Brothers' bankruptcy	15/09/2008
Commencement of the first economic adjustment programme for Greece	06/05/2010
Commencement of the second economic adjustment programme for Greece	13/02/2012

Figure 1. Imports of Goods (in million euro)

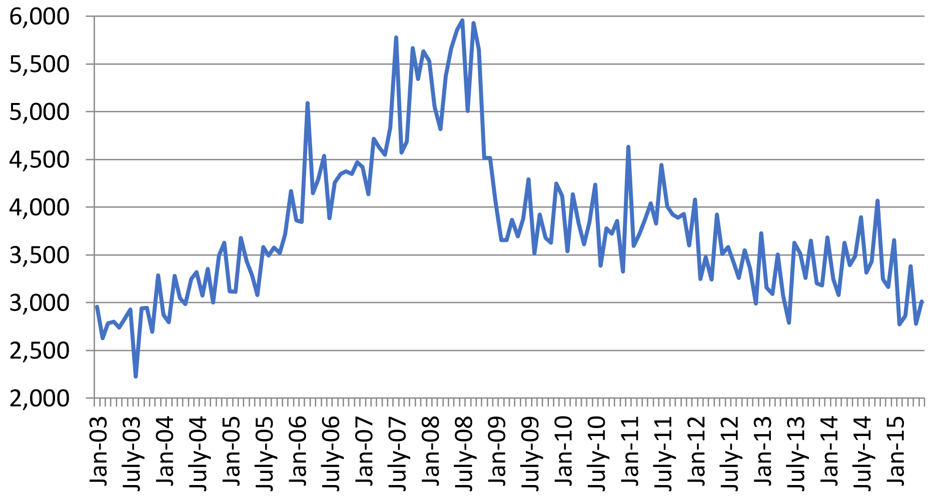
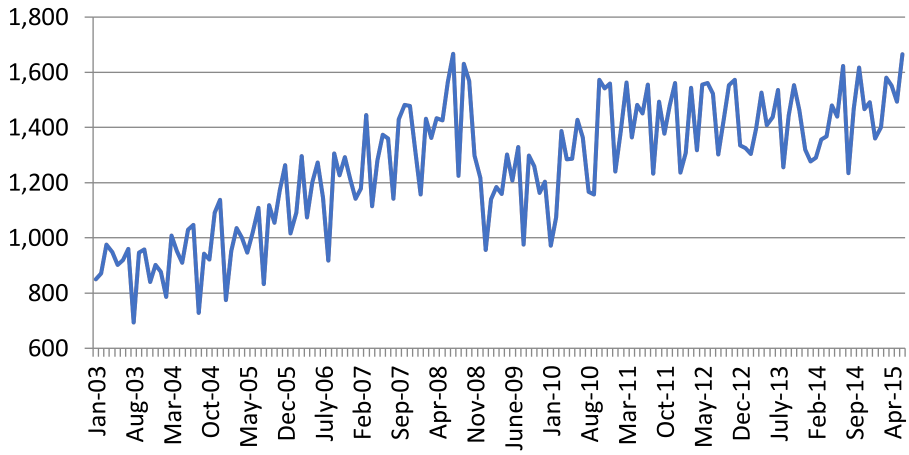


Figure 2. Exports of Goods without fuels and ship (in million euro)



The sharp decline in both series at end 2008, which is conspicuous from the visual inspection of those figures, may be attributed, amongst others, to the Lehman Brothers' bankruptcy and the subsequent sharp reduction in economic activity.

In addition to the real BOP data, time series of Greek Consumer Price Index (CPI) and Harmonized Index of Consumer Price (HICP) covering the same time period (January 2003 to June 2015, source: Hellenic Statistical Authority) and simulated time series will also be used in order to provide further supporting evidence for the conclusions drawn from the BOP data.

2.3.1.1 Results – Discussion

At first the possible need for a data transformation will be examined using the new method, as well as the TSW routine. Further, each decision derived from the new methodology and the corresponding one derived from the TSW routine will be compared. This task is of course of interest in its own right. However, in terms of applied time series analysis in general, it is of crucial importance to examine also the extent to which that decision affects some crucial elements of time series modelling such as outlier detection and seasonal adjustment. Once a decision about the proper data transformation is made by the two methods, TSW will be used for both cases for this further analysis.

As far as outlier detection is concerned, outliers are automatically detected, classified and corrected using the Chen and Liu (1993) approach. It is noted that in TSW framework, outliers are classified into three types, according to their effect on a time

series (see section 1.5), as follows (for further details as well as the theoretical background good references are Hilmer et al., 1983; Tsay, 1984; Tsay, 1986):
Additive outliers (AO), Transitory Change outliers (TC), Level shifts (LS).
Seasonal adjustment with TSW is made after the series are "linearized" (see section 1.5 for a description on "linearization"). Seasonal adjustment itself is based on ARIMA model-based signal extraction. This method uses the Burman-Wilson algorithm (Burman, 1980) to decompose a time series into unobserved components [for further details see Maravall, (1995); Gómez and Maravall, (1996)].

2.3.1.2 Comparative analysis of the time series of "Imports of goods"

i. Analysis using of the new statistical testing methodology

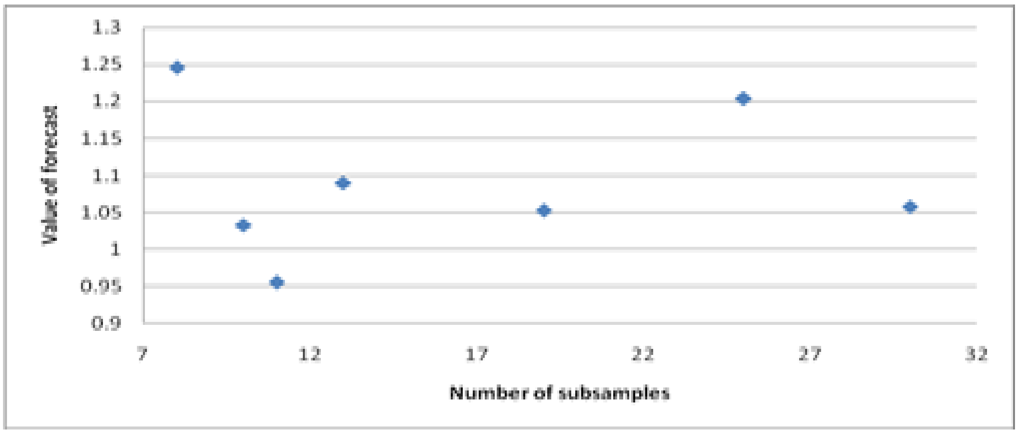
The results for the first stage regressions for the time series "Imports of goods" are presented in Table 2. It can be seen that in all subsample pairs the β_j^* estimates are statistically significant at the 10% significance level and in almost all subsample pairs at the 5% significance level (exceptions only for the partitions with subsample size 10 and 18, where the estimates are "marginally" significant for the 5% significance level). Hence, H_a is clearly rejected.

Table 2. Estimates of $\hat{\beta}_j$ for the various partitions for the first stage regression

Subsample size (n_{ij})	5	6	8	10	12	14	16	18	20
Number of subsample pairs (i_j)	30	25	19	15	13	11	10	9	8
$\hat{\beta}_j$	1.058	1.203	1.052	0.863	1.090	0.955	1.032	0.791	1.246
t-statistic	2.642	3.230	3.211	2.096	4.204	2.297	3.321	2.278	2.758
p-value	0.013	0.004	0.005	0.056	0.001	0.047	0.010	0.057	0.033

The results for the first stage regressions are also depicted in Figure 3, where the x-axis represents the number of subsamples and the y-axis the value of exponent. From the visual inspection of Figure 3 it is apparent that no systematic association between estimates and the sample size seems to exist.

Figure 3. Graphical representation of the first stage regressions results



This is further supported by the results of the second stage regression with the aid of which the Hypothesis H_b: $d = 0$ can be formally tested. Those results are presented in Table 3. As is evident, the constant $\hat{\beta}$ is statistically significant at the 5% level and equal to 1.082, whereas the slope $\hat{d}$ is not statistically significant. Hence, H_b cannot be rejected.

Table 3. Results for the second stage regression

	Estimate	t-statistic	standard error
$\hat{\beta}$	1.082	10.956	0.099
$\hat{d}$	0.0005	0.101	0.005

The above results clearly suggest that: (i) the original data series is variance non-stationary; (ii) the estimated value of $\hat{\beta}$ suggest that the data should be log-transformed. To examine, whether or not, the chosen logarithmic transformation is indeed sufficient to stabilize the variance the "under-transformation" test is performed. To this end the logarithms of the original data are subjected to the logarithmic transformation once more and the parameters of the third stage regression are estimated. The results of the "under-transformation" test are presented in Table 4. The results of the "under-transformation" test are presented in Table 4.

As is evident from the results of Table 4 none of the $\hat{\beta}_j^*$s is statistically significant, therefore the hypothesis H_c: β_j^*=0 ∀j is not rejected.

The above results clearly suggest that the log-transformation of the original data makes the series variance stationary.

Table 4. Results for "under-transformation" test

Subsample size (n_{ij})	5	6	8	10	12	14	16	18	20
Number of subsample pairs (i_j)	30	25	19	15	13	11	10	9	8
β_j^*	0.305	1.655	0.222	-1.59	0.558	-0.75	5.882	-1.61	6.470
t-statistic	0.092	0.540	0.085	-0.50	0.258	-0.23	1.325	-0.56	1.350
p-value	0.927	0.594	0.933	0.623	0.801	0.821	0.222	0.592	0.220

ii. Analysis using exclusively the TSW testing approach

The same series were reanalyzed following the standard TSW procedure[4]. The way TSW tests whether or not the data need to be transformed in order to stabilize the variance is based on a variant of the so-called range-mean regression (see Gómez and Maravall, 1996). For range-mean regression, the series is divided into subsamples and the range and mean for each subsample are calculated. Then a regression model using the subsamples' ranges and means is estimated. If the regression slope is found to be significant the data are log-transformed.

Using the TSW procedure, TSW also suggested the logarithmic transformation of the original data, as was the conclusion using the new approach. However, when the TSW procedure was repeated with the log-transformed data ("under-transformation" test), TSW suggested a logarithmic transformation again(!). Indeed, TSW output states that: «LOG-LEVEL PRETEST: SSlevels/(SSlog*Gmean(levels)^2)=1.0078847 LOGS ARE SELECTED».

Therefore, for the "under-transformation" test, TSW is biased towards rejection of the null hypothesis of no transformation.

iii. The effect of data transformation on univariate modelling and outlier detection

It is of interest to further investigate how variance non-stationarity could potentially affect the specification of the univariate ARIMA model and the detection of outliers.

[4]It is noted that the only alternatives with TSW is either the log-transformation, or no transformation.

Table 5 below presents the results of univariate ARIMA modelling with and without the log-transformation, while the estimation details are quoted in Tables 6-7.

Table 5. Univariate ARIMA modelling

ARIMA model with linearized original data [Model (1)]
ARIMA(2,1,0) $(1,0,1)_{12}$ MODEL SPECIFICATION
$(1-0.583B-0.401B^2)\ (1+0.857B^{12})\ (1\text{-}B)Y_t=(1+0.594B^{12})\varepsilon_t$
ARIMA model with linearized transformed data (TSW and new method) [Model (2)]
ARIMA(0,1,1) $(0,1,1)_{12}$ MODEL SPECIFICATION
$(1-B)\ (1\text{-}B^{12})\ log\,Y_t=(1+0.609B)\ (1+0.592B^{12})\varepsilon_t$

From the results of Table 5 it is apparent that, when variance non-stationarity is taken into account, the univariate model is the so-called “airline” model, often encountered in data with seasonality (see Box and Jenkins, 1976). In contrast, without considering variance non stationarity a much more complicated ARIMA model is selected. Hence, the presence of variance non stationarity leads to seriously mis-specified univariate ARIMA models, a result that is in accordance with that of other studies (e.g. Milionis, 2004).

Table 6. Parameter estimation of model (1)

Parameter	**Value**	**t-statistic**
ϕ_1	0.583	7.480
φ_2	0.401	5.290
Φ_1	-0.857	-10.890
Θ_1	-0.594	-4.810

Table 7. Parameter estimation of model (2)

Parameter	Value	t-statistic
θ_1	-0.609	-8.750
Θ_1	-0.592	-8.360

The results on the detection of outliers are presented in Table 8. As is evident a TC outlier at period 39 in the original data does not exist in the transformed data. Further, the LS outlier at period 71 (November 2008) in the original data, which could be related to the Lehman bankruptcy, has been shifted forward one period in the transformed data, while the AO outlier in period 97 in the original data has been shifted backward one period in the transformed data. Hence, in properly transformed data the pattern of detected outliers is clearly different, a conclusion that is also in accordance with that of Milionis (2004).

Table 8. Outlier Detection (series of Imports of goods)

Outliers with original data	Outliers with log-transformed data (TSW, new method)
39 TC (3 2006), 71 LS (11 2008), 97 A0 (1 2011)	72 LS (12 2008), 96 A0 (12 2010)

2.3.1.3 Comparative analysis of the time series of "Exports of goods excluding fuels and ships"

i. Testing for variance non-stationarity

Table 9 presents the results of the first stage regressions for the exports of goods excluding fuels and ships. From these results it is evident that $\hat{\beta}_j$ is not statistically significant at the 5% significance level, except for the partition with subsample size 6. Thus, according to the new testing approach, the series variance is (unconditionally) stationary and no transformation of the original data is required.

However, the conclusion is different when the approach of TSW is followed. Indeed, TSW log-transforms the data as a consequence of the range-mean regression. Indeed, TSW output states that:

«LOG-LEVEL PRETEST: SSlevels/(SSlog*Gmean(levels)^2)=1.0106056 LOGS ARE SELECTED».

Table 9. Estimates of $\hat{\beta}_j$ for the various partitions for the first stage regression

Subsample size (n_{ij})	5	6	8	10	12	14	16	18	20
Number of subsample pairs (i_j)	30	25	19	15	13	11	10	9	8
$\hat{\beta}_j$	0.654	0.723	0.530	0.467	0.455	0.440	0.595	0.466	0.246
t-statistic	1.812	2.110	2.015	1.713	1.233	1.111	1.167	1.279	0.392
p-value	0.081	0.046	0.060	0.110	0.243	0.295	0.277	0.241	0.709

Therefore, once again, TSW is biased towards the logarithmic transformation, whereas no transformation of the original data needs to be performed.

Even worse than that, when the TSW procedure was repeated once again with the already log-transformed data (the "under-transformation" test), which were supposed to be variance stable, TSW suggested the logarithmic transformation again. Indeed, TSW output states that:

«LOG-LEVEL PRETEST: SSlevels/(SSlog*Gmean(levels)^2)=0.98497426 LOGS ARE SELECTED».

Apparently, that suggestion is seriously biased and misleading.

ii. Univariate ARIMA modelling, outlier detection and seasonal adjustments

Table 10 below presents the results of univariate ARIMA modelling with and without the log-transformation for exports of goods excluding fuels and ships, while the estimation details are quoted in Tables 11-12.

From the results of Table 10 it is apparent that in contrast to the univariate models for imports of goods here the differences in the two univariate models are of minor character, as in both cases the univariate model is of the same type i.e., the so-called "airline" model. The differences are confined only to the estimated values of the parameters of the two models. This is not surprising as with exports of goods excluding fuels and ships, the original, as well as the log-transformed series, are both variance stationary. Indeed, this result advocates our previous conclusion for imports of goods where the pronounced difference in the character of the univariate ARIMA model for the original and the log-transformed data was attributed to the existence of non-stationary variance in the original data series.

Table 10. Univariate ARIMA modelling for the series of Exports of goods excluding fuels and ships

ARIMA model with linearized original data, (new method) [Model (3)]
ARIMA(0,1,1) (0,1,1)$_{12}$ MODEL SPECIFICATION
$(1-B)$ (1-B^{12})$Y_t = (1+0.584B)\ (1+0.7.64B^{12})\varepsilon_t$
ARIMA model with linearized transformed data (TSW) [Model (4)]
ARIMA(0,1,1) (0,1,1)$_{12}$ MODEL SPECIFICATION
$(1-B)$ (1-B^{12}) $log\, Y_t = (1+0.566B)\ (1+0.821B^{12})\varepsilon_t$

Table 11. Parameter estimation of model (3)

Parameter	**Value**	**t-statistic**
θ_1	-0.584	-8.190
Θ_1	-0.764	-13.490

Table 12. Parameter estimation of model (4)

Parameter	**Value**	**t-statistic**
θ_1	-0.566	-7.850
Θ_1	-0.821	-16.430

The results for the detection of outliers for both the original and the log-transformed data for exports of goods excluding fuels and ships are quoted in Table 13. As is evident the AO outlier is the same in both cases, while the level shift, which could be related, amongst other things, to the Lehman bankruptcy and its repercussions, has only been moved forward by one time-period in the log-transformed data.

Table 13. Outlier Detection (series of Exports of goods excluding fuels and ships)

Outliers with original data (new method)	Outliers with log-transformed data (TSW)
71 LS (11 2008), 93 AO (9 2010)	72 LS (12 2008), 93 AO (9 2010)

Our next task is to examine if and to what extent the data transformation affects the forecasting performance of the univariate models, as well as the seasonally adjusted series. The former was evaluated using the Mean Absolute Percentage Error (MAPE). The MAPE statistic is given by: $MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$, where A_t is the actual value and F_t is the forecast value.

Table 14 presents the results. As is evident both one-step-ahead and twelve-step- ahead forecasts with no transformation (as suggested by the new method) are superior in terms of the MAPE value, as compared to the corresponding forecasts with the data log-transformed, as suggested by TSW.

Table 14. Values of Mean Absolute Percentage Error of forecasts

	MAPE (%)	
Forecasts	Original Data (New Method)	Log-transformed data (TSW)
Twelve step ahead	3.618	4.501
One step ahead	3.265	3.778

A statistic akin to MAPE, which is used as a means to evaluate the quality of a forecast, will be employed in order to assess the difference in the seasonally adjusted data produced with and without the log-transformation.

To this end, the Mean Absolute Percentage Difference (MAPD) statistic was employed to assess the differences in the seasonally adjusted series produced from original data versus transformed data. MAPD is calculated as: $MAPD = \frac{1}{J}\sum_{j=1}^{N}\left(\frac{\left|\tilde{X}_j^{ln} - \tilde{X}_j^{lev}\right|}{\tilde{X}_j^{ln}} \cdot 100\right)$, where $\tilde{X}_j^{ln}$ is the seasonally adjusted value using log-transformed data and $\tilde{X}_j^{lev}$ the corresponding seasonally adjusted data using the original data themselves.

The results are presented in Table 15. It should be noted that in the results of Table 15 quite substantial differences are observed. Indeed, although the value of the MAPD between the two seasonally adjusted series is not so high (approximately 1.3%), the minimum percentage difference is as high as -4.3% and the maximum percentage difference is equal to 5.5%.

Table 15. Differences in Seasonally Adjusted Series produced from original data versus transformed data

MAPD (%)	Minimum Percentage Error (%)	Maximum Percentage Error (%)
1.297	-4.314	5.515

2.3.1.4 Comparative analysis of the time series of "CPI" and "HICP"

i. Analysis using of the new statistical testing approach

The results for the first stage regressions for the time series "CPI" are presented in Table 16. It can be seen that in almost all subsample pairs the estimates are statistically insignificant at 5% significance level (exceptions exist only for the partitions with subsample size 5 and 8). Hence, H_a is not rejected and therefore no transformation of the original data is suggested.

Table 16. Estimates of $\hat{\beta}_j$ for the various partitions for the first stage regression (CPI)

Subsample size (n_{ij})	5	6	8	10	12	14	16	18	20
Number of subsample pairs (i_j)	30	25	19	15	13	11	10	9	8
$\hat{\beta}_j$	1.604	0.681	1.148	0.933	0.898	0.565	0.377	-0.629	0.011
t-statistic	2.112	1.390	2.166	1.315	1.656	0.891	0.325	-0.667	0.010
p-value	0.047	0.178	0.048	0.211	0.126	0.396	0.753	0.526	0.992

Table 17 presents the results of the first stage regressions for the series of HICP. From these results it is evident that is not statistically significant in the 5% significance level.

Table 17. Estimates of $\hat{\beta}_j$ for the various partitions for the first stage regression (HICP)

Subsample size (n_{ij})	5	6	8	10	12	14	16	18	20
Number of subsample pairs (i_j)	30	25	19	15	13	11	10	9	8
$\hat{\beta}_j$	1.003	0.076	0.639	0.363	0.399	0.039	0.018	-0.957	-0.503
t-statistic	1.383	0.146	1.266	0.529	0.689	0.057	0.016	-1.010	-0.460
p-value	0.178	0.885	0.223	0.606	0.505	0.956	0.988	0.346	0.662

Thus, according to the new testing approach, in both cases no transformation of the original data is suggested as the series variance is (unconditionally) stationary.

ii. Analysis using exclusively the TSW testing approach

Using the TSW procedure for both series, TSW suggested the logarithmic transformation of the original data for CPI as well as HICP as the output from TSW that refers to this test stated that «LOG-LEVEL PRETEST: SSlevels/(SSlog*Gmean(levels)^2)=1.1253168 LOGS ARE SELECTED» and «LOG-LEVEL PRETEST: SSlevels/(SSlog*Gmean(levels)^2)=1.1096478 LOGS ARE SELECTED», respectively.

Therefore, TSW falsely suggests the logarithmic transformation of the data in both cases and so there is a bias in favor of the logarithmic transformation, whereas no transformation of the original data needs to be performed as shown using the new method.

iii. The effect of data transformation on univariate modelling and outlier detection

The possible effect of log transforming an already variance stationary time series on the specification of the univariate ARIMA model and the detection of outliers is examined through Tables 18-25.

For the time series of CPI, Table 18 presents the results on univariate ARIMA modelling with and without the log-transformation, while the estimation details are quoted in the

Tables 19-20. From the results of Table 18 it is obvious that different ARIMA models are proposed for the linearized transformed data and linearized original data.

Table 18. Univariate ARIMA modelling (CPI)

ARIMA model with linearized original data (new method) [Model (5)]
ARIMA(0,1,1) $(1,1,0)_{12}$ MODEL SPECIFICATION
$(1-0.221B^{12})\,(1\text{-}B)(1\text{-}B^{12})\,Y_t = (1-0.208B)\varepsilon_t$
ARIMA model with linearized transformed data (TSW) [Model (6)]
ARIMA(0,1,1) $(0,1,1)_{12}$ MODEL SPECIFICATION
$(1-B)\,(1\text{-}B^{12})\,log\,Y_t = (1-0.206B)\,(1+0.223B^{12})\varepsilon_t$

Table 19. Parameter estimation of model (5)

Parameter	**Value**	**t-statistic**
θ_1	0.208	2.49
Φ_1	0.221	2.65

Table 20. Parameter estimation of model (6)

Parameter	**Value**	**t-statistic**
θ_1	0.206	2.46
Θ_1	-0.223	-2.68

The results on the detection of outliers are presented in Table 21. As is evident both approaches identify the same outliers in terms of both time and type.

Table 21. Outlier Detection (CPI)

Outliers with original data (new method)	Outliers with log-transformed data (TSW)
105 LS (9 2011), 118 TC (10 2012), 131 A0 (11 2013)	105 LS (9 2011), 118 TC (10 2012), 131 A0 (11 2013)

For the series of HICP, Table 22 presents the results on univariate ARIMA modelling with and without the log-transformation (for estimation details see Tables 23-24). From the results of Table 22 it is apparent that the differences in the two univariate models are of minor character, as in both cases the univariate model is the so-called "airline model" of Box and Jenkins (1976). The differences, weak as they are, are confined only to the estimated values of the parameters of the two models.

Table 22. Univariate ARIMA modelling (HICP)

ARIMA model with linearized original data (new method) [Model (7)]
ARIMA(0,1,1) $(0,1,1)_{12}$ MODEL SPECIFICATION
$(1-B)(1-B^{12})Y_t = (1-0.188B)(1+0.203B^{12})\varepsilon_t$
ARIMA model with linearized transformed data (TSW) [Model (8)]
ARIMA(0,1,1) $(0,1,1)_{12}$ MODEL SPECIFICATION
$(1-B)(1-B^{12})\log Y_t = (1-0.213B)(1+0.175B^{12})\varepsilon_t$

Table 23. Parameter estimation of model (7)

Parameter	**Value**	**t-statistic**
θ_1	0.188	2.23
Θ_1	-0.203	-2.42

Table 24. Parameter estimation of model (8)

Parameter	**Value**	**t-statistic**
θ_1	0.213	2.55
Θ_1	-0.175	-2.08

The results on the detection of outliers with both the original and the log-transformed data are quoted in Table 25. As is evident, TSW falsely identifies one more outlier (25 AO) due to the scale "squizing" caused by the logarithmic transformation.

Table 25. Outlier Detection (HICP)

Outliers with original data (new method)	Outliers with log-transformed (TSW)
105 LS (9 2011), 131 AO (11 2013)	25 AO (1 2005), 105 LS (9 2011), 131 AO (11 2013)

These results should be contrasted to those of the time series of BOP statistics (see sections 2.3.1.2 and 2.3.1.3) in which it was documented that the consequences of not transforming a variance non-stationary time series were indeed severe, at least as far as univariate time series modelling is concerned (Milionis and Galanopoulos, 2017). The above conclusions, regarding the severity of the consequences of a wrong decision about the transformation of the original values of a time series, may be backed by theoretical argumentation. Indeed, in the case of unnecessarily (over) transforming an already variance stationary series, the original and the over-transformed series are both variance stationary. In contrast, when the original series is variance non-stationary and is not transformed (as it should) it is evident that the usual univariate analysis, which legitimately can be applied strictly to variance stationary series only, is falsely applied to a variance non-stationary series. Hence, it is natural for sharp differences with the analysis of the properly transformed series to appear. Obviously, this conclusion is of much practical importance.

2.3.1.5 Further Analysis

The examination of the forecasting performance of the univariate models accordingly with the data transformation is assessed using the MAPE statistic. Table 26 presents the results and as is evident twelve-step-ahead forecasts with no transformation (as suggested by the new method) are superior in terms of the MAPE value, as compared to the corresponding forecasts with the data log-transformed, as suggested by TSW. As regards the one-step-ahead forecasts, the forecasting performance of CPI is better with the new method in terms of the MAPE value, while the MAPE value is exactly the same for HICP.

Table 26. Values of Mean Absolute Percentage Error of forecasts

		MAPE (%)	
Time series	Forecasts	Original Data (new method)	Log-transformed data (TSW)
CPI	Twelve step ahead	1.294	1.457
HICP	Twelve step ahead	0.848	0.865
CPI	One step ahead	0.395	0.467
HICP	One step ahead	0.452	0.452

To evaluate the differences in the seasonally adjusted series produced from original data versus transformed data, the MAPD statistic was employed, and the results are presented in Table 27. It is remarked that, as expected according to the previous conclusion argumentation, the values of MAPD themselves are small in the statistical sense, although with possibly higher economic importance as they refer to consumer price indices.

Table 27. Differences in Seasonally Adjusted Series produced from original data versus transformed data

Time series	MAPD (%)	Minimum Percentage Error (%)	Maximum Percentage Error (%)
CPI	0.018	-0.057	0.037
HICP	0.017	-0.316	0.035

2.3.2. Application on time series created by statistical simulation

It should be recalled that in the normal course of an analysis of a time series, the test for the possible need to transform the original data in order to stabilize the variance precedes all other actions such as the creation of the univariate ARIMA model, the seasonal adjustment etc. Therefore, it is obvious that the outcome of those actions is affected by the decision on data transformation. That lends even more importance to that decision and thus further evidence was sought in favor of our previous finding that TSW suggests the log-transformation far too often due to bias. To this end, use was made of simulated data. After some experimentation, it was observed that with series following $ARIMA(p,1,q)$ processes, the initial conditions influence (although they should not) the decision of TSW regarding the log-transformation. In sequence, 40 time series were artificially created. The first 20 were built simulating a simple Gaussian random walk model:

$$X_t = X_{t-1} + e_t \text{ with } e_t \approx N(0,1) \text{ and initial condition } X_o = 0.$$

The other 20 were created by the same manner, except that X_o was set equal to 1000. All simulated data are available in Milionis and Galanopoulos (2018) working paper. All these time series are by construction homogeneously stationary, so no data transformation is needed.

Applying the new testing procedure, the outcome was indeed no transformation in all 40 cases (Milionis and Galanopoulos, 2018). In the simulated series with initial condition $X_o = 0$ the presence of negative values was observed in every single series. Therefore, before log-transforming the series the usual practice of shifting them upwards by adding a constant so that the minimum value of each series becomes slightly positive was adopted.

Applying TSW, the outcome was no transformation only in the cases with $X_o = 0$. In all the cases with $X_o = 1000$ TSW, falsely, suggested the log transformation (see Table 28). Indeed, the results of TSW are, once again, biased in favor of data transformation where no transformation is needed and in the particular case at hand this bias is clearly related to the initial condition of the simulated series.

Table 28. Test results for data transformation using TSW

LOG-LEVEL PRETEST :	
1st simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0004275 LOGS ARE SELECTED
2nd simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0004200 LOGS ARE SELECTED
3rd simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99937392 LOGS ARE SELECTED
4th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99821483 LOGS ARE SELECTED
5th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0029646 LOGS ARE SELECTED
6th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99932366 LOGS ARE SELECTED
7th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99998323 LOGS ARE SELECTED

8th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99976469 LOGS ARE SELECTED
9th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99899563 LOGS ARE SELECTED
10th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0009120 LOGS ARE SELECTED
11th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99940611 LOGS ARE SELECTED
12th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99783903 LOGS ARE SELECTED
13th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99956026 LOGS ARE SELECTED
14th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99996242 LOGS ARE SELECTED
15th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0009673 LOGS ARE SELECTED
16th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99844882 LOGS ARE SELECTED
17th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99780523 LOGS ARE SELECTED
18th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0011250 LOGS ARE SELECTED
19th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0010450 LOGS ARE SELECTED
20th simulated series with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0001285 LOGS ARE SELECTED
LOG-LEVEL PRETEST :	
21st simulated series with initial value 0	LEVELS are Selected
22nd simulated series with initial value 0	LEVELS are Selected
23rd simulated series with initial value 0	LEVELS are Selected
24th simulated series with initial value 0	LEVELS are Selected

25th simulated series with initial value 0	LEVELS are Selected
26th simulated series with initial value 0	LEVELS are Selected
27th simulated series with initial value 0	LEVELS are Selected
28th simulated series with initial value 0	LEVELS are Selected
29th simulated series with initial value 0	LEVELS are Selected
30th simulated series with initial value 0	LEVELS are Selected
31st simulated series with initial value 0	LEVELS are Selected
32nd simulated series with initial value 0	LEVELS are Selected
33rd simulated series with initial value 0	LEVELS are Selected
34th simulated series with initial value 0	LEVELS are Selected
35th simulated series with initial value 0	LEVELS are Selected
36th simulated series with initial value 0	LEVELS are Selected
37th simulated series with initial value 0	LEVELS are Selected
38th simulated series with initial value 0	LEVELS are Selected
39th simulated series with initial value 0	LEVELS are Selected
40th simulated series with initial value 0	LEVELS are Selected

Moreover, even after the log-transformation, when the 20 log-transformed series with $X_o = 1000$ were tested again by TSW, the result was biased again (suggestion to log-transform the (already log-transformed) series, as was case with the time series of exports of goods excluding fuels and ships). Those results are presented in Table 29.

Table 29. Test results for data transformation using TSW and log-transformed series

LOG-LEVEL PRETEST :	
1st simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0000550 LOGS ARE SELECTED
2nd simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0000529 LOGS ARE SELECTED
3rd simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99989907 LOGS ARE SELECTED
4th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99973569 LOGS ARE SELECTED
5th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0003951 LOGS ARE SELECTED
6th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99989544 LOGS ARE SELECTED
7th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99999495 LOGS ARE SELECTED
8th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99996343 LOGS ARE SELECTED
9th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99984226 LOGS ARE SELECTED
10th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0001277 LOGS ARE SELECTED
11th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99990892 LOGS ARE SELECTED
12th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99967945 LOGS ARE SELECTED
13th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99993525 LOGS ARE SELECTED
14th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99999001 LOGS ARE SELECTED

15th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0001311 LOGS ARE SELECTED
16th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99976445 LOGS ARE SELECTED
17th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 0.99966162 LOGS ARE SELECTED
18th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0001710 LOGS ARE SELECTED
19th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0001444 LOGS ARE SELECTED
20th simulated series in log-transformed data with initial value 1000	SSlevels/(SSlog*Gmean(levels)^2)= 1.0000163 LOGS ARE SELECTED

2.4. Conclusions

In this work a new statistical testing procedure for variance non-stationary time series is proposed. This procedure improves on existing ones as it combines detection, correction and robustness. This is of value in its own right as it results in better univariate time series modelling.

In addition, it was shown empirically using real data (balance of payments and prices of consumer goods and services) for Greece, as well as simulated data, that an existing test, namely the widely used algorithm of TSW software, provides, occasionally, biased results. As a matter of fact, with the aid of the simulated data it was possible to identify one of the possible sources responsible for this bias. More specifically, with simulated homogeneously non-stationary processes, it was possible to identify that the bias of TSW depends on the initial conditions.

Further, on the basis of the empirical evidence presented it is argued that the type of data transformation and the entailed correction for variance–non stationarity is also crucial for the detection of outliers and the seasonal adjustment of the original time series. In addition, the empirical results provide evidence of an improved forecasting performance by the proper use of a data transformation, a result that needs further backing by additional empirical evidence.

It was also established that the consequences of falsely transforming a time series, which is originally variance stationary, do exist, but are less severe than the consequences of falsely not transforming an originally variance non-stationary series, a conclusion of much practical importance.

Overall, the proposed statistical testing methodology, placed in a more general framework, seems to be a promising tool in applied time series analysis.

SUMMARY OF CHAPTER 2

This chapter aims to fill an existing gap in the literature regarding the statistical testing for the existence and the identification of the character of time-varying second moment in its dependence on a non-constant mean level in time series. To this end a new statistical testing procedure is introduced with some considerable advantages over the existing ones. Amongst others it is argued that the existing statistical tests are insufficient and sometimes lead to biased results. Further the effect of the application of this methodology on some crucial elements of time series modelling such as outlier detection and seasonal adjustment is examined, through case studies conducted on a comparative basis using both the new methodology and an established one. It is established that the consequences of falsely transforming a time series, which is originally variance stationary, do exist, but are less severe than the consequences of falsely not transforming an originally variance non-stationary series. This empirical evidence is supported by theoretical arguments. The data set comprises time series on monthly external trade statistics and prices of consumer goods and services for Greece. Overall, the resulting empirical evidence favors the new approach. Further supporting evidence is provided by the application of the new methodology to simulated data.

CHAPTER 3

FORECASTING MACROECONOMIC TIME SERIES IN THE PRESENCE OF VARIANCE INSTABILITY AND OUTLIERS

3.1 Introduction

A univariate ARIMA model is a concise quantitative summary of the internal dynamics of a time series in a linear framework. As such, is useful for several reasons, amongst others for forecasting and model-based time series decomposition in unobserved components. This work will deal with the former and, in particular, with univariate forecasts, which usually serve either as short-term, or benchmark forecasts. However, economic time series from the real world are not usually «ready» to be used for forecasting purposes and they need to undergo some statistical preparation and pre-adjustment. This is because in time series of raw data variance non-stationarity may be present. Furthermore, very often there exist causes that disrupt the underlying stochastic process (existence of outliers, calendar effects, etc.). Their treatment is known as «linearization».

Within that line of reasoning, statistical forecasts can be made after a series itself, or some variance stabilizing transformation of it, is «linearized» according to the general framework that is described in section 1.5.

As far as variance stabilization is concerned, if variance is somehow functionally related to the mean level it is possible to select a transformation to stabilize the variance. Widely used transformations to tackle this problem belong to the class of the power Box and Cox transformation (see section 1.6.1).

So, there are two effects with potential influence on forecasting: transformation and «linearization», each of which separately, as well as in combination, may play an important role on time series forecasting.

At the empirical level, studies which have considered the merits of mathematical transformations on forecasting have demonstrated that a data transformation generally does not have a positive effect on forecast accuracy (Nelson and Granger 1979; Makridakis and Hibon, 1979; Makridakis et. al, 1998; Meese and Geweke, 1984).

On the other hand, at the theoretical level, Granger and Newbold (1976) found that such forecasts are not optimal in terms of minimization of Mean Square Forecast Error (MSFE). More specifically, for instance for the most popular transformation, namely the logarithmic one, they showed that the minimum MSFE h-step ahead forecast is not equal to $\hat{z}_{T+h} = exp(\hat{y}_{T+h})$, as implied by the previous discussion, but is given by the expression $\hat{z}_{T+h} = exp\left(\hat{y}_{T+h} + \frac{1}{2}\sigma_h^2\right)$, where σ_h^2 is the h -step ahead forecast error variance. Pankratz and Dudley (1987), building up further on the work of Granger and Newbold (1976), relate the bias in using simply the inversely transformed value of the forecasts on the transformed time series (as compared to the minimum MSFE forecast) amongst others to the value of the exponent λ of the power transformation. The two most frequent transformations, namely the logarithmic and the square root ones, under certain conditions may be associated with serious biases (Pankratz and Dudley, 1987).

Regarding time series linearization, such a procedure is utilized thus far mainly as a preadjustment task for seasonal adjustment (Kaiser and Maravall, 2001), so its effect on forecasting has not been examined systematically, but only indirectly and fragmentally. It is also remarked that even in studies coping with forecasting with transformed data the attention focuses almost exclusively on point forecasts, by and large disregarding interval forecasts.

Aiming at covering this research gap in the literature the objective in this chapter of the thesis is in fact twofold: (a) to examine the effect of «linearization» and transformation separately, as well as in combination, on both point forecasts and confidence interval forecasts; (b) to use two algorithms specializing in testing, whether or not, a transformation of the original data is necessary, namely the algorithm of TSW and the algorithm developed in Chapter 2 and, compare the derived results from both (see also Milionis and Galanopoulos 2018a). Hereafter the latter will be called M-G algorithm for convenience. As a further application, we rank main economic indicators of the Greek economy in terms of statistical «forecastability». The intended approach will be practical.

The structure of the chapter is as follows: In section 3.2 details about the data to be used for the empirical analysis are given; section 3.3 presents the empirical results and relevant comments; section 3.4 concludes the chapter.

3.2 Data

The data set comprises some of the most important macroeconomic time series for the Greek economy, which refer to: GDP; unemployment; prices of consumer goods and services; monetary aggregates; and balance of payments statistics. More specifically, the time series from the balance of payments statistics are imports-exports excluding fuels and ships and imports-exports including them. Particularly, in the balance of payments, a distinction is made between imports – exports of all goods and imports - exports of goods without fuels and ships for several reasons. More specifically: i) the IMF in its 6th Manual on Balance of Payments (IMF, 2009), revised the definition of the item "Total Goods" to firmly align with the principle of "change of ownership" (for more details see section 2.3.1), ii) to ensure consistency with the external trade statistics generated by the Hellenic Statistical Authority, iii) according to a study by the Bank of Greece (Oikonomou et al., 2010), the dependence of the Greek economy on oil was high and was rising at the fastest pace among the euro area countries. For these reasons, the time series of Imports of Goods and Exports of Goods without fuels and ships from Table 1 (referred to as Total Imports and Total Exports of Goods excluding fuels and ships in Chapter 2) are being re-examined. Furthermore, from the same study it is noted that the balance of payment of sea transport is significant in the Greek balance of current transactions (4% of GPD in 2008) and will be considered separately from other BOP transactions on transport.

Of the twenty economic time series that are used, nineteen are monthly time series, one is a quarterly time series (sources: Bank of Greece (BoG) and Hellenic Statistical Authority (ELSTAT)). The list of time series used is given in Table 1.

The monthly time series data cover the period from January 2004 to August 2018 and consist of one hundred and seventy-six (176) observations, except for Industrial Production Index, where available data existed from January 2010 to August 2018 (104 observations). The quarterly time series is that of Gross Domestic Product and covers the period from 1995 Quarter 1 to 2018 Quarter 3 (95 observations).

Table 1. Data

Time Series	**Observation frequency**	**Source**
Gross Domestic Product (GDP)	Quarterly	ELSTAT
Industrial Production Index (IPI)	Monthly	ELSTAT
Consumer Price Index (CPI)	Monthly	ELSTAT
Harmonised Index of Consumer Prices (HICP)	Monthly	ELSTAT
Unemployment – thousands	Monthly	ELSTAT
Unemployment – percentage	Monthly	ELSTAT
Retail sales	Monthly	ELSTAT
M1	Monthly	BoG
M2	Monthly	BoG
M3	Monthly	BoG
Balance of payments (BOP) – Transport – Payments	Monthly	BoG
Balance of payments (BOP) – Transport – Receipts	Monthly	BoG
Balance of payments (BOP) – Travelling – Payments	Monthly	BoG
Balance of payments (BOP) – Travelling – Receipts	Monthly	BoG
Balance of payments (BOP) – Sea transport – Payments	Monthly	BoG
Balance of payments (BOP) – Sea transport – Receipts	Monthly	BoG
Exports of Goods	Monthly	BoG
Exports of Goods without fuels and ships	Monthly	BoG
Imports of Goods	Monthly	BoG

Imports of Goods without fuels and ships	Monthly	BoG

3.3 Empirical results and comments

As mentioned in section 3.1, the effect of transformation and the effect of linearization on forecasting will be examined at first each one separately and, subsequently, in combination.

The aforementioned effects will be studied on a comparative basis utilizing both the TSW and the M-G algorithms. In that way, together with those effects themselves, it will also be possible to evaluate the performance of each methodology.

Typical statistics to be used for the assessment of the quality of point forecasts are the following:

i) the Mean Absolute Percentage Error (MAPE) statistic given by:

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|,$$

ii) the Mean Square Forecast Error (MSFE) statistic given by:

$$MSFE = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2,$$ and

iii) the Mean Absolute Error (MAE) statistic given by:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t|,$$

where A_t is the actual value and F_t is the forecast value.

Furthermore, as far as interval forecasts are concerned, the width of the forecast confidence interval (CI), or the forecast standard error, will be considered.

Best forecast will obviously be perceived the one with the minimum value of each time utilized statistic from the ones mentioned above.

3.3.1 The effect of «linearization» on forecast quality

We will investigate how time series linearization affects the quality of both point forecasts and confidence interval forecasts. Here linearization will not be considered in its generality, as described in section 1.5, but will be confined to outliers' detection and adjustment. (Calendar effects such as the trading day and leap effects were considered and indeed were found to be statistically significant on some occasions. All series were properly adjusted for calendar effects before further analysis.). Table 2 presents the number of best forecasts with data in levels. Auxiliary Table 3 presents the number of best forecasts with log-transformed data indistinguishably for all time series, as it is often the case to use log-transformed data in econometric analyses. It is noted that in one time series with levels (that of unemployment expressed in percentages) and one time series in logs (that of industrial production index) no outliers were detected, hence, the total number of time series considered reduced to nineteen for each case.

Table 2. Summary table - Number of best forecasts (levels)[5]

Point Forecasts	With detected Outliers	Without detection of Outliers
MAPE	10/19	9/19
MSFE	8/19	11/19
MAE	9/19	10/19
Interval Forecasts	**With detected Outliers**	**Without detection of Outliers**
Forecast Standard Error (SE)	19/19	0/19

From the results of Tables 2 and 3 it is apparent that, when outliers are considered, forecasts are better in every single case in terms of the width of the forecast confidence interval. In contrast, there is no obvious improvement in point forecasts. One point that should be stressed is that such results are in general dependent upon the specific characteristics of each time series, especially upon whether an outlier lays among the first, the middle or the last observations. For this reason, it would be desirable to use a large number of time series, so as to draw conclusions of indisputable confidence. Although the number of time series used in this work is relatively small (though comparable to that of other similar works, see for instance Nelson and Granger, 1979)

[5] As the usual practice, the original data set was split up into the estimation sample, over which model estimation is performed, and the holdout (test) sample. In all cases the holdout sample for ex-post forecasts was originally set to twelve time periods for the monthly series and ten time periods for GDP. Presented results are based on one-step-ahead forecasts. Results for longer forecasting periods (not presented) are very similar and are available by the author.

the evidence that led to the above conclusions, in particular regarding the width of the forecast confidence interval, is so convincing that it really stands far and beyond any concern related to micronumerosity.

Table 3. Summary table – Number of best forecasts (log-data)

Point Forecasts	**With detected Outliers**	**Without detection of Outliers**
MAPE	9/19	10/19
MSFE	11/19	8/19
MAE	10/19	9/19
Interval Forecasts	**With detected Outliers**	**Without detection of Outliers**
Forecast Standard Error (SE)	19/19	0/19

3.3.2 The effect of Level Shifts (LS), in particular, on forecast quality

After a level shift outlier, all observations subsequent to the outlier move to a new level. In contrast to additive and transitory outliers a level shift outlier reflects a major change in the stochastic process and affects many observations, as it has a permanent effect. For this reason, the case with only additive and transitory outliers (i.e. excluding level shifts) was considered, and their effect on forecasts was examined separately, performing the same analysis as in section 3.3.1. It is noted that this time only fifteen time series were considered, i.e. those including all types of outliers. The results are presented in Tables 4 and 5.

From the results below it is obvious that there is a trade-off: confidence interval forecasts are better with level shift outliers included and, conversely, point forecasts are better excluding level shifts. Given the influence of the level shift outliers it would be desirable to possibly consider stricter identification criteria for them relative to the other two types of outliers. It is noted that in existing statistical software specializing on time series analysis there is no such an option, and a purpose-built routine should be created by the researcher.

Table 4. Summary table - Number of best forecasts (levels)

Point Forecasts	All Outliers	Outliers without LS
MAPE	6/15	9/15
MSFE	5/15	10/15
MAE	6/15	9/15
Interval Forecasts	**All Outliers**	**Outliers without LS**
Forecast Standard Error (SE)	14/15	1/15

Table 5. Summary table - Number of best forecasts (log-data)

Point Forecasts	All Outliers	Outliers without LS
MAPE	6/15	9/15
MSFE	5/15	10/15
MAE	6/15	9/15
Interval Forecasts	**All Outliers**	**Outliers without LS**
Forecast Standard Error (SE)	13/15	2/15

3.3.3 The effect of a data transformation on forecast quality

As far as the effect of data transformation is concerned, at first it is important to note that the effect of a transformation is meant in two ways: 1) direct and 2) indirect (through its influence on outlier detection). Indeed, regarding the latter, it has been shown that data transformation affects the number and the character of outliers in a time series (Milionis 2003; Milionis, 2004).

The possible need for a data transformation of the original time series data will be examined using both the algorithms of TSW and M-G. Furthermore, each decision derived from the Milionis and Galanopoulos methodology and the corresponding one derived from the TSW routine will be compared. Once a decision about the proper data transformation is made, TSW will be used for both cases for further analysis on statistical forecasting.

Regarding the arithmetic values of the exponent λ (see section 1.6.1) and the closely related parameter $\hat{\beta}$, as is estimated by Equation (3) of section 2.2.1, for practical purposes Makridakis et al. (1998) mention that it is of no merit to use arithmetic values with several decimal points, as nearby values will produce very similar results. Simple arithmetic values of λ are easier to interpret, hence, are more meaningful.

In line with that argument, nearby arithmetic values of $\hat{\beta}$ will be grouped together, so as to create two sub-logarithmic transformations, namely the square root and cubic root ones, the logarithmic itself, and one over-logarithmic, namely the negative inverse transformation. More specifically, following some experimentation, the grouping is as follows (it is noted that no case with negative value of $\hat{\beta}$ was encountered):

(a) $\hat{\beta}$ not statistically significant, then $\lambda = 1$;
(b) $\hat{\beta}$ statistically significant and $0 < \hat{\beta} \leq \hat{\beta} + 1.96se(\hat{\beta})$ or 0.65, whichever is lower, then $\lambda = 1/2$;
(c) $\hat{\beta} - 1.96se(\hat{\beta})$, or 0.65, whichever is higher $< \hat{\beta} \leq \hat{\beta} + 1.96se(\hat{\beta})$ or 0.80, whichever is lower, then $\lambda = 1/3$;
(d) $\hat{\beta} - 1.96se(\hat{\beta})$ or 0.80, whichever is higher $< \hat{\beta} \leq \hat{\beta} + 1.96se(\hat{\beta})$, then $\lambda = 0$;
(e) $\hat{\beta} - 1.96se(\hat{\beta}) > 1$, then $\lambda = -1$.

Table 6 presents the results on the decision about, transforming or not, the original time series data. From these results it is evident that, according to the M-G algorithm, no transformation of the original data is suggested in fifteen out of the twenty cases, the negative inverse transformation is suggested in four cases and the logarithmic transformation in only one case.

The same series were reanalyzed following the standard TSW procedure. It is noted that the only alternatives available with TSW are either the log-transformation, or no transformation. Using the TSW routine for these twenty cases, TSW suggested the logarithmic transformation of the original data for eighteen cases. It is remarkable that only for the two series of unemployment TSW suggests no transformation, as does the Milionis Galanopoulos method as well, for the particular two series. It should be stressed, however, that as shown by Milionis and Galanopoulos (2018a, 2018b), the TSW routine is biased towards suggesting the log-transformation.

Table 6. Decision about data transformation

TIME SERIES	METHOD OF TRANSFORMATION	
	LOG-LEVEL PRETEST (Output from TSW)	**M-G**
Gross Domestic Product (GDP)	SSlevels/(SSlog*Gmean(levels)^2)= 1.1380170 LOGS ARE SELECTED	Levels
Consumer Price Index (CPI)	SSlevels/(SSlog*Gmean(levels)^2)= 1.0781750 LOGS ARE SELECTED	Levels
Harmonised Index of Consumer Prices (HICP)	SSlevels/(SSlog*Gmean(levels)^2)= 1.0954455 LOGS ARE SELECTED	Levels
Industrial Production Index (IPI)	SSlevels/(SSlog*Gmean(levels)^2)= 1.0224433 LOGS ARE SELECTED	Levels
Unemployment – thousands	SSlevels/(SSlog*Gmean(levels)^2)= 0.87725642 LEVELS ARE SELECTED	Levels
Unemployment – percentage	SSlevels/(SSlog*Gmean(levels)^2)= 0.86356273 LEVELS ARE SELECTED	Levels
Retail sales	SSlevels/(SSlog*Gmean(levels)^2)= 1.2755206 LOGS ARE SELECTED	Negative Inverse
M1	SSlevels/(SSlog*Gmean(levels)^2)= 0.98393639 LOGS ARE SELECTED	Levels
M2	SSlevels/(SSlog*Gmean(levels)^2)= 1.0714007 LOGS ARE SELECTED	Levels
M3	SSlevels/(SSlog*Gmean(levels)^2)= 1.0422806 LOGS ARE SELECTED	Levels
Balance of payments (BOP) – Transport – Payments	SSlevels/(SSlog*Gmean(levels)^2)= 1.0351033 LOGS ARE SELECTED	Levels

Balance of payments (BOP) – Transport – Receipts	SSlevels/(SSlog*Gmean(levels)^2)= 1.1641507 LOGS ARE SELECTED	Negative Inverse
Balance of payments (BOP) – Travelling – Payments	SSlevels/(SSlog*Gmean(levels)^2)= 1.1509645 LOGS ARE SELECTED	Levels
Balance of payments (BOP) – Travelling – Receipts	SSlevels/(SSlog*Gmean(levels)^2)= 4.3996100 LOGS ARE SELECTED	Logarithmic
Balance of payments (BOP) – Sea transport – Payments	SSlevels/(SSlog*Gmean(levels)^2)= 0.98863656 LOGS ARE SELECTED	Levels
Balance of payments (BOP) – Sea transport – Receipts	SSlevels/(SSlog*Gmean(levels)^2)= 1.1948699 LOGS ARE SELECTED	Negative Inverse
Exports of Goods	SSlevels/(SSlog*Gmean(levels)^2)= 0.95751942 LOGS ARE SELECTED	Levels
Exports of Goods without fuels and ships	SSlevels/(SSlog*Gmean(levels)^2)= 0.96487436 LOGS ARE SELECTED	Levels
Imports of Goods	SSlevels/(SSlog*Gmean(levels)^2)= 1.1118244 LOGS ARE SELECTED	Levels
Imports of Goods without fuels and ships	SSlevels/(SSlog*Gmean(levels)^2)= 1.2957291 LOGS ARE SELECTED	Negative Inverse

The possible effect of transforming time series on forecasting quality is examined through Tables 7a and 7b. From the results below it is concluded that point forecasts with either transformation method are slightly better than with no transformation in terms of MAPE and MAE, but not in terms of MSFE. As already explained, forecasts on transformed variables are not optimal in terms of MSFE. On the other hand, confidence interval forecasts are shorter in four of the five cases using transformations with the M-G approach. In contrast this happens in only eight out of the eighteen cases using the TSW approach. Though it seems that M-G approach leads to shorter confidence interval forecasts, obviously there are very few cases available. Further empirical evidence with a larger dataset is needed on that point so as to draw safer conclusions.

Table 7a. Summary table - Number of best forecasts (M-G versus benchmark)

Point Forecasts	**M-G - no outliers**	**Levels-no outliers (Benchmark)**
MAPE	3/5	2/5
MSFE	2/5	3/5
MAE	3/5	2/5
Interval Forecasts	**M-G - no outliers**	**Levels-no outliers (Benchmark)**
Forecast Standard Error (SE)	4/5	1/5

Table 7b. Summary table - Number of best forecasts (TSW versus benchmark)

Point Forecasts	**TSW - no outliers**	**Levels-no outliers (Benchmark)**
MAPE	9/18	9/18
MSFE	7/18	11/18
MAE	9/18	9/18
Interval Forecasts	**TSW - no outliers**	**Levels-no outliers (Benchmark)**
Forecast Standard Error (SE)	8/18	10/18

3.3.4 The combined effect of linearization and data transformation

The results of the examination of the forecasting performance combining both linearization and data transformation are presented in Tables 8a and 8b. The conclusion that is derived is that, by and large, the combined effect does not lead to better point forecasts but leads to improved confidence interval forecasts with better performance for the M-G approach. The conclusion about the forecast confidence interval is reasonable and, to a large extent, expected, as with the transformation of the original time series data and the adjustment for outliers the process variance is reduced. It is possible to exploit this reduction in obtaining forecasts with increased confidence.

Table 8a. Summary table - Number of best forecasts (M-G versus benchmark)

Point Forecasts	M-G - All outliers	Levels-no outliers (Benchmark)
MAPE	2/5	3/5
MSFE	2/5	3/5
MAE	2/5	3/5
Interval Forecasts	**M-G - All outliers**	**Levels-no outliers (Benchmark)**
Forecast Standard Error (SE)	4/5	1/5

Table 8b. Summary table - Number of best forecasts (TSW versus benchmark)

Point Forecasts	TSW - All outliers	Levels-no outliers (Benchmark)
MAPE	8/18	10/18
MSFE	8/18	10/18
MAE	8/18	10/18
Interval Forecasts	**TSW - All outliers**	**Levels-no outliers (Benchmark)**
Forecast Standard Error (SE)	12/18	6/18

Table 9 presents the ARIMA models for the benchmark model and the combination of Milionis-Galanopoulos variance stabilizing method - linearization. It is noted that the differences in the ARIMA models for the time series where no transformation was needed should be attributed to the existence of outliers adjusted by linearization.

Table 9. Univariate ARIMA models with and without transformation-linearization

Time series	**Benchmark**	**M-G**
Gross Domestic Product (GDP)	ARIMA (0,1,1) $(0,1,1)_4$	ARIMA (1,0,0) $(1,1,0)_4$
	$\nabla\nabla_4 Y_t = (1 + 0.118B)(1 + 0.425B^4)\varepsilon_t$	$(1 + 0.953B)(1 - 0.335B^4)\nabla_4 Y_t = \varepsilon_t$
Industrial Production Index (IPI)	ARIMA (2,0,0) $(0,1,1)_{12}$	ARIMA (2,0,0) $(0,1,1)_{12}$
	$(1 + 0.379B + 0.547B^2)\nabla_{12} Y_t = (1 + 0.950B^{12})\varepsilon_t$	$(1 + 0.379B + 0.547B^2)\nabla_{12} Y_t = (1 + 0.950B^{12})\varepsilon_t$
Consumer Price Index (CPI)	ARIMA (0,1,0) $(0,1,1)_{12}$	ARIMA (1,1,0) $(0,1,0)_{12}$
	$\nabla\nabla_{12} Y_t = (1 + 0.260B^{12})\varepsilon_t$	$(1 + 0.134B)\nabla\nabla_{12} Y_t = \varepsilon_t$
Harmonised Index of Consumer Prices (HICP)	ARIMA (0,1,0) $(0,1,1)_{12}$	ARIMA (0,1,0) $(0,1,1)_{12}$
	$\nabla\nabla_{12} Y_t = (1 + 0.347B^{12})\varepsilon_t$	$\nabla\nabla_{12} Y_t = (1 + 0.287B^{12})\varepsilon_t$
Unemployment – thousands	ARIMA (3,2,1) $(0,1,1)_{12}$	ARIMA (3,2,1) $(0,1,1)_{12}$
	$(1 - 0.681B - 0.674B^2 + 0.062B^3)\nabla^2\nabla_{12} Y_t = (1 + 0.758B)(1 + 0.938B^{12})\varepsilon_t$	$\begin{pmatrix} 1 - 1.153B - 1.123B^2 \\ -0.340B^3 \end{pmatrix}\nabla^2\nabla_{12} Y_t = (1 + 0.614B)(1 + 0.907B^{12})\varepsilon_t$
Unemployment – percentage	ARIMA (2,2,1) $(0,1,1)_{12}$	ARIMA (2,2,1) $(0,1,1)_{12}$
	$(1 - 0.726B - 0.715B^2)\nabla^2\nabla_{12} Y_t = (1 + 0.734B)(1 + 0.816B^{12})\varepsilon_t$	$(1 - 0.726B - 0.715B^2)\nabla^2\nabla_{12} Y_t = (1 + 0.734B)(1 + 0.816B^{12})\varepsilon_t$
Retail sales	ARIMA (0,1,1) $(0,1,1)_{12}$	ARIMA (0,1,1) $(0,1,1)_{12}$
	$\nabla\nabla_{12} Y_t = (1 + 0.364B)(1 + 0.566B^{12})\varepsilon_t$	$\nabla\nabla_{12} \frac{-1}{Y_t} = (1 + 0.316B)(1 + 0.586B^{12})\varepsilon_t$
M1	ARIMA (0,2,1) $(0,1,1)_{12}$	ARIMA (3,1,0) $(0,1,1)_{12}$

	$\nabla^2\nabla_{12}Y_t = (1 + 0.838B)(1 + 0.682B^{12})\varepsilon_t$	$\begin{pmatrix}1 + 0.007B + 0.156B^2 \\ +0.420B^3\end{pmatrix}\nabla\nabla_{12}Y_t = (1 + 0.668B^{12})\varepsilon_t$
M2	ARIMA (3,1,0) (1,0,1)12	ARIMA (3,1,0) (0,1,1)12
	$(1 + 0.328B + 0.040B^2 + 0.307B^3)(1 + 0.868B^{12})\nabla Y_t = (1 + 0.656B^{12})\varepsilon_t$	$\begin{pmatrix}1 + 0.375B + 0.087B^2 \\ +0.311B^3\end{pmatrix}\nabla\nabla_{12}Y_t = (1 + 0.822B^{12})\varepsilon_t$
M3	ARIMA (0,2,1) (0,1,1)12	ARIMA (0,2,1) (0,1,1)12
	$\nabla^2\nabla_{12}Y_t = (1 + 0.695B)(1 + 0.824B^{12})\varepsilon_t$	$\nabla^2\nabla_{12}Y_t = (1 + 0.683B)(1 + 0.859B^{12})\varepsilon_t$
Balance of payments (BOP) – Transport – Payments	ARIMA (0,1,1) (0,1,1)12	ARIMA (0,1,1) (0,1,1)12
	$\nabla\nabla_{12}Y_t = (1 + 0.188B)(1 + 0.847B^{12})\varepsilon_t$	$\nabla\nabla_{12}Y_t = (1 + 0.312B)(1 + 0.859B^{12})\varepsilon_t$
Balance of payments (BOP) – Transport – Receipts	ARIMA (3,1,1) (0,1,1)12	ARIMA (0,1,1) (0,1,1)12
	$(1 - 0.393B - 0.050B^2 + 0.264B^3)\nabla\nabla_{12}Y_t = (1 - 0.288B)(1 + 0.950B^{12})\varepsilon_t$	$\nabla\nabla_{12}\frac{-1}{Y_t} = (1 + 0.180B)(1 + 0.829B^{12})\varepsilon_t$
Balance of payments (BOP) – Travelling – Payments	ARIMA (1,0,0) (0,1,1)12	ARIMA (1,0,0) (1,0,0)12
	$(1 + 0.339B)\nabla_{12}Y_t = (1 + 0.506B^{12})\varepsilon_t$	$(1 + 0.314B)(1 + 0.613B^{12})Y_t = \varepsilon_t$
Balance of payments (BOP) – Travelling – Receipts	ARIMA (1,0,0) (1,1,0)12	ARIMA (1,0,0) (1,1,0)12
	$(1 + 0.731B)(1 - 0.371B^{12})\nabla_{12}Y_t = \varepsilon_t$	$(1 + 0.598B)(1 - 0.422B^{12})\nabla_{12}\mathit{ln}Y_t = \varepsilon_t$
	ARIMA (0,1,1) (0,0,0)12	ARIMA (0,1,1) (0,1,1)12

Balance of payments (BOP) – Sea transport – Payments	$\nabla Y_t = (1 + 0202B)\varepsilon_t$	$\nabla\nabla_{12}Y_t = (1 + 0.290B)(1 + 0.846B^{12})\varepsilon_t$
Balance of payments (BOP) – Sea transport – Receipts	ARIMA (3,1,1) (0,1,1)12	ARIMA (3,1,1) (0,1,1)12
	$(1 - 0.388B - 0.020B^2 + 0.281B^3)\nabla\nabla_{12}Y_t = (1 - 0.262B)(1 + 0.848B^{12})\varepsilon_t$	$\begin{pmatrix}1 - 0.533B - 0.125B^2 \\ +0.217B^3\end{pmatrix}\nabla\nabla_{12}\frac{-1}{Y_t} = (1 - 0.414B)(1 + 0.826B^{12})\varepsilon_t$
Exports of Goods	ARIMA (0,1,1) (0,1,1)12	ARIMA (0,1,1) (0,1,1)12
	$\nabla\nabla_{12}Y_t = (1 + 0.414B)(1 + 0.950B^{12})\varepsilon_t$	$\nabla\nabla_{12}Y_t = (1 + 0.387B)(1 + 0.950B^{12})\varepsilon_t$
Exports of Goods without fuels and ships	ARIMA (0,1,1) (0,1,1)12	ARIMA (0,1,1) (0,1,1)12
	$\nabla\nabla_{12}Y_t = (1 + 0.485B)(1 + 0.922B^{12})\varepsilon_t$	$\nabla\nabla_{12}Y_t = (1 + 0.587B)(1 + 0.785B^{12})\varepsilon_t$
Imports of Goods	ARIMA (0,1,1) (0,1,1)12	ARIMA (0,1,1) (0,1,1)12
	$\nabla\nabla_{12}Y_t = (1 + 0.495B)(1 + 0.950B^{12})\varepsilon_t$	$\nabla\nabla_{12}Y_t = (1 + 0.655B)(1 + 0.934B^{12})\varepsilon_t$
Imports of Goods without fuels and ships	ARIMA (0,1,1) (0,1,1)12	ARIMA (0,1,1) (0,1,1)12
	$\nabla\nabla_{12}Y_t = (1 + 0.434B)(1 + 0.785B^{12})\varepsilon_t$	$\nabla\nabla_{12}\frac{-1}{Y_t} = (1 + 0.313B)(1 + 0.799B^{12})\varepsilon_t$

3.3.5 Sensitivity analysis - Outliers (dependence of outlier detection on the parameter τ)

Let $\hat{Y}_{T+1}/\Phi_T$ denote the optimal one-step-ahead linear forecast of Y_{T+1} given the information set Φ_T, which includes information up to time T, $e_{T+1} = Y_{T+1} - \hat{Y}_{T+1}/\Phi_T$ denote the associated forecast error, and $\sigma_{T+1}^2 = [Y_{T+1} - \hat{Y}_{T+1}/\Phi_T]^2$ denote the associated variance. The observation Y_{T+1} is considered as an outlier if the null Hypothesis: $H_0: e_{T+1} = 0$ is rejected. The appropriate statistic to test H_0 is: $\tau = \frac{e_{T+1}}{\sigma_{T+1}}$.

However, theory cannot predict the critical value of $\boldsymbol{\tau}$ above which the corresponding observation can be considered as an outlier. A usual practice is to relate the critical value of $\boldsymbol{\tau}$ with the length of a time series. The default values of TSW for $\boldsymbol{\tau}$ are presented in Table 10[6]. In the course of our experimentation, it was observed that outlier detection (as well as ARIMA models for the linearized-transformed series), were very sensitive to the value of parameter $\boldsymbol{\tau}$. In order to examine, whether or not, the critical $\boldsymbol{\tau}$ values could have any noticeable effect on our final conclusions, as an alternative set of critical values for $\boldsymbol{\tau}$ we used those suggested by Fischer and Planas (2000), who examined a very large number of time series. Their critical values for $\boldsymbol{\tau}$ were set at 3.5, 3.7 and 4.0 for series lengths of less than 130 observations, between 131 and 180, and more than 180 observations, respectively.

Table 10. Critical values for τ

Observations	**Default values for $\boldsymbol{\tau}$ in TSW**
164	0.358E+01
165 – 168	0.359E+01
169 – 172	0.360E+01
173 – 175	0.361E+01

The comparison of the results based on default critical $\boldsymbol{\tau}$ values, as well as on Fischer – Planas recommendations are presented in Table 11, while the detected outliers for each time series and each set of values for the parameter $\boldsymbol{\tau}$ are presented in Table 12. Looking at Table 12 it is observed that the detection of outliers is indeed sensitive even to the examined small changes in the value of $\boldsymbol{\tau}$. On the other hand, however, from the results of Table 11, it is apparent that using the Fischer and Planas critical values for $\boldsymbol{\tau}$ leads to mixed results regarding the effect on forecast quality.

By and large, there is only very weak evidence of improvement using the Fischer – Planas recommendations[7].

[6] In the TSW framework the subroutine TERROR is designed especially for outlier detection. Incoming data volume in institutions like EUROSTAT, ECB, OECD, NCBs, NSOs etc. may be enormous. Such data may be contaminated by errors of various types and origins. Using TERROR is a convenient, yet formal way to spot aberrant observations (outliers). It is highly possible that if erroneous data do exist, they will be included in the set of observations characterized as outliers by TERROR, hence, in a second stage, their possible identification is focused exclusively on that data set. In this work we used the first stage only.

[7] Indeed, setting the Fischer –Planas critical values instead of the default ones, the results are identical regarding those of Table 8a, while the results pertaining to those of Table 8b they are identical in terms

Table 11. Results based on Fischer – Planas recommendations

Time series	Improvement of forecast quality	Same forecast quality	Deterioration of forecast quality
Gross Domestic Product (GDP)		MAPE, MSFE, MAE, SE (TSW)	MAPE, MSFE, MAE, SE (M-G)
Consumer Price Index (CPI)	MAPE, MSFE, MAE (TSW)	MAPE, MSFE, MAE, SE (M-G)	SE (TSW)
Harmonised Index of Consumer Prices (HICP)	MAPE, MSFE, MAE (M-G) MAPE, MSFE, MAE, SE (TSW),		SE (M-G)
Industrial Production Index (IPI)	MAPE, MSFE, MAE (M-G)	MAPE, MSFE, MAE, SE (TSW)	SE (M-G)
Unemployment – thousands	MSFE (M-G, and TSW)		MAPE, MAE, SE (M-G, and TSW)
Unemployment – percentage	MAPE, MSFE, MAE (M-G, and TSW)	SE (M-G, and TSW)	
Retail sales			MAPE, MSFE, MAE, SE (M-G, and TSW)
M1		MAPE, MSFE, MAE, SE (M-G, and TSW)	
M2	MAPE, MAE (M-G)		MAPE, MSFE, MAE, SE (TSW), MSFE, SE (M-G)
M3	MAPE, MSFE, MAE (M-G)		MAPE, MSFE, MAE, SE (TSW),

of the standard error, and 8/18 for MAPE, MAE and MSFE with TSW, as compared to 7/18 using the default critical values).

			SE (M-G)
Balance of payments (BOP) – Transport – Payments	MAPE, MSFE, MAE (TSW)		MAPE, MSFE, MAE, SE (M-G) SE (TSW)
Balance of payments (BOP) – Transport – Receipts	MAPE, MSFE, MAE, SE (M-G, and TSW)		
Balance of payments (BOP) – Travelling – Payments		MAPE, MAE, SE (TSW)	MSFE (TSW) MAPE, MSFE, MAE, SE (M-G)
Balance of payments (BOP) – Travelling – Receipts		SE (M-G, and TSW)	MAPE, MSFE, MAE (M-G, and TSW)
Balance of payments (BOP) – Sea transport – Payments	MAPE, MSFE, MAE (M-G, and TSW)		SE (M-G, and TSW)
Balance of payments (BOP) – Sea transport – Receipts	MAPE, MSFE, MAE (M-G, and TSW)		SE (M-G, and TSW)
Exports of Goods	MAPE, MSFE, MAE (M-G, and TSW)		SE (M-G, and TSW)
Exports of Goods without fuels and ships			MAPE, MSFE, MAE, SE (M-G, and TSW)
Imports of Goods		MAPE, MSFE, MAE, SE (M-G)	MAPE, MSFE, MAE, SE (TSW)

Imports of Goods without fuels and ships	MAPE, MSFE, MAE (TSW)	MAPE, MSFE, MAE, SE (M-G)	SE (TSW)

Table 12. Detected outliers for the different values of parameter τ (the first number indicate the serial number of the corresponding observation, then follows the type of outlier and within the parentheses the corresponding month, or quarter, and year*)*

Time series	τ -default TSW critical values	τ -Fischer-Planas
Gross Domestic Product (GDP)	OUTLIERS: 57 AO (1 2009)	OUTLIERS: 57 AO (1 2009)
Industrial Production Index (IPI)	OUTLIERS: NO OUTLIERS DETECTED	OUTLIERS: NO OUTLIERS DETECTED
Consumer Price Index (CPI)	OUTLIERS: 93 LS (9 2011), 119 AO (11 2013)	OUTLIERS: 119 AO (11 2013)
Harmonised Index of Consumer Prices (HICP)	OUTLIERS: 119 AO (11 2013)	OUTLIERS: 119 AO (11 2013)
Unemployment – thousands	OUTLIERS: 60 LS (12 2008), 95 LS (11 2011), 98 TC (2 2012), 126 LS (6 2014), 148 TC (4 2016), 156 TC (12 2016)	OUTLIERS: 60 LS (12 2008), 95 LS (11 2011), 98 TC (2 2012), 126 LS (6 2014), 148 TC (4 2016), 156 TC (12 2016)
Unemployment – percentage	OUTLIERS: NO OUTLIERS DETECTED	OUTLIERS: NO OUTLIERS DETECTED
Unemployment – thousands	OUTLIERS: NO OUTLIERS DETECTED	OUTLIERS: 113 AO (5 2013), 139 AO (7 2015)
M1	OUTLIERS: 139 LS (7 2015)	OUTLIERS: 139 LS (7 2015)

M2	OUTLIERS: 100 AO (4 2012), 102 AO (6 2012), 133 LS (1 2015), 138 TC (6 2015)	OUTLIERS: 100 AO (4 2012), 102 AO (6 2012), 138 TC (6 2015)
M3	OUTLIERS: 100 AO (4 2012), 102 AO (6 2012), 133 LS (1 2015), 138 TC (6 2015)	OUTLIERS: 100 AO (4 2012), 102 AO (6 2012), 133 LS (1 2015), 138 TC (6 2015)
Balance of payments (BOP) – Transport – Payments	OUTLIERS: 60 LS (12 2008), 133 LS (1 2015)	OUTLIERS: 60 LS (12 2008), 133 LS (1 2015)
Balance of payments (BOP) – Transport – Receipts	OUTLIERS: 59 LS (11 2008)	OUTLIERS: 36 TC (12 2006), 59 LS (11 2008)
Balance of payments (BOP) – Travelling – Payments	OUTLIERS: 92 AO (8 2011)	OUTLIERS: 92 AO (8 2011)
Balance of payments (BOP) – Travelling – Receipts	OUTLIERS: 2 AO (2 2004), 113 LS (5 2013)	OUTLIERS: 2 AO (2 2004), 113 LS (5 2013)
Balance of payments (BOP) – Sea transport – Payments	OUTLIERS: 60 LS (12 2008), 113 LS (5 2013), 133 LS (1 2015)	OUTLIERS: 59 LS (11 2008), 113 LS (5 2013), 133 LS (1 2015)
Balance of payments (BOP) – Sea transport – Receipts	OUTLIERS: 36 TC (12 2006), 59 LS (11 2008), 129 AO (9 2014)	OUTLIERS: 36 TC (12 2006), 59 LS (11 2008), 129 AO (9 2014)
Exports of Goods	OUTLIERS: 81 AO (9 2010)	NO OUTLIERS DETECTED

Exports of Goods without fuels and ships	OUTLIERS: 60 LS (12 2008), 81 AO (9 2010)	OUTLIERS: 60 LS (12 2008), 81 AO (9 2010)
Imports of Goods	OUTLIERS: NO OUTLIERS DETECTED	OUTLIERS: NO OUTLIERS DETECTED
Imports of Goods without fuels and ships	OUTLIERS: 39 AO (3 2007), 59 LS (11 2008), 75 AO (3 2010), 82 AO (10 2010), 139 TC (7 2015)	OUTLIERS: 39 AO (3 2007), 59 LS (11 2008), 75 AO (3 2010), 82 AO (10 2010), 139 TC (7 2015)

3.3.6 Evaluation of models' forecasting performance

The skill of a forecast can be assessed by comparing the relative proximity of both the forecast and a benchmark to the observations. The presence of a benchmark makes it easier to compare approaches and for this reason a benchmark is proposed to establish a common ground for comparison. In the present case an obvious benchmark is to use the univariate ARIMA forecasts of the twenty-time series described in section 3.2, non-linearized and non-transformed. This benchmark forecasts will be used together with the forecasts from the TSW and M-G approaches as the three alternatives, the performances of which are to be evaluated and compared. Forecasts' evaluation for each model will be based on both point and interval forecasts. A simple and transparent ad-hoc approach will be used for this purpose. More specifically, for the point forecasts for each time series and for each model an arithmetic value is assigned in ascending order based on the corresponding value of the MSFE statistic (i.e. 1 for the minimum MSFE value, 2 for the mid- MSFE value, 3 for the maximum MSFE value). Then, adding up the arithmetic values for all series for a particular model their sum will represent the performance of the model. Models will be ranked according to the value of the corresponding sum. Apparently, the model with the lowest sum will be considered as the best one. For interval forecasts the same procedure will be followed replacing the value of the MSFE statistic with the value of the corresponding standard error around the point forecasts.

From the above, it is apparent that use will be made repetitively of the same data set. This could potentially make the whole process susceptible to the data snooping trap

(White, 2000)[8]. Such a case is quite common for instance in developing trading strategies in financial markets. A well-known tool used by the developers of such strategies is the so-called reality check with its refinements and extensions (White 2000; Romano and Wolf, 2005; Hansen et. al 2011). In the present case however, the possibility that the forecasting performance of one of the three models to be used (namely the benchmark model, TSW and M-G) is superior than that of the other two simply due to chance is reduced by the fact that the number of models is much lower than the number of time series (three against twenty). Therefore, it is unlikely that one and the same model would obtain superior performance in all, or at least in most of the twenty-time series, just as a result of pure chance. For this reason, the usage of the reality check, bearing in mind also its weaknesses (Hansen, 2005; Hansen et. al 2011), is not deemed as necessary.

The results are shown in Tables 13 and 14[9] and more detailed results are quoted in Table 15. It is clarified that both the TSW and M-G transformation approaches are coupled with the outlier detection-adjustment approach.

Table 13. Ranking of forecasting performance according to MSFE (point forecasts)

Time series	Benchmark	TSW	M-G
Consumer Price Index (CPI)	1	2	3
Harmonised Index of Consumer Prices (HICP)	1	3	2
M3	1	3	2
M2	2	3	1
Gross Domestic Product (GDP)	3	1	2
M1	3	1	2

[8] Halbert White in his seminal paper (White, 2000) states that: "data snooping occurs when a given set of data is used more than once for purposes of inference or model selection. When such data reuse occurs, there is always the possibility that any satisfactory results obtained may simply be due to chance rather than to any merit inherent in the method yielding the results. This problem is practically unavoidable in the analysis of time series data…"
[9] If for two models the value of MSFE or SE is exactly the same, the mid-point will be used for both.

Industrial Production Index (IPI)	1	3	2
Retail sales	2	3	1
Unemployment – thousands	1	2.5	2.5
Balance of payments (BOP) – Transport – Receipts	1	2	3
Balance of payments (BOP) – Sea transport – Receipts	1	3	2
Unemployment – percentage	2	2	2
Balance of payments (BOP) – Transport – Payments	1	3	2
Imports of Goods without fuels and ships	2	1	3
Exports of Goods without fuels and ships	3	1	2
Exports of Goods	2	1	3
Balance of payments (BOP) – Sea transport – Payments	3	1	2
Imports of Goods	3	1	2

Balance of payments (BOP) – Travelling – Receipts	3	1.5	1.5
Balance of payments (BOP) – Travelling – Payments	2	1	3
SUM	**38**	**39**	**43**

Table 14. Ranking of forecasting performance according to SE (interval forecasts)

Time series	Benchmark	TSW	M-G
Harmonised Index of Consumer Prices (HICP)	3	2	1
Consumer Price Index (CPI)	3	2	1
M1	3	2	1
M3	3	1	2
M2	3	1	2
Gross Domestic Product (GDP)	3	1	2
Unemployment – percentage	2	2	2
Industrial Production Index (IPI)	2	3	1
Unemployment – thousands	3	1.5	1.5
Exports of Goods without fuels and ships	2	3	1

Retail sales	3	2	1
Exports of Goods	2	3	1
Balance of payments (BOP) – Transport – Receipts	3	2	1
Balance of payments (BOP) – Transport – Payments	2	3	1
Balance of payments (BOP) – Sea transport – Receipts	3	2	1
Imports of Goods without fuels and ships	3	1	2
Balance of payments (BOP) – Sea transport – Payments	2	3	1
Imports of Goods	3	1	2
Balance of payments (BOP) – Travelling – Payments	3	1	2
Balance of payments (BOP) – Travelling – Receipts	1	2.5	2.5
SUM	**52**	**39**	**29**

From the results of Tables 13 and 14 it is evident that the performance of neither TSW nor M-G approach for point forecasts is better than that of the benchmark model (as a matter of fact both are slightly worse). On the other hand, for the forecast confidence intervals M-G has a better performance than TSW and the benchmark model. Furthermore, TSW outperforms the benchmark model. A rather crude way to procced to an overall evaluation of the three models is to add up their performances in the two categories (i.e. point and interval forecasts). The addition gives the values of 90, 78 and 72 for the benchmark model, TSW and M-G respectively, which means that both TSW and M-G perform clearly better than the benchmark model and further the performance of M-G is better than that of TSW.

Table 15. Detailed forecast quality statistics: MSFE, MAE and Forecast Standard Error

Time series	Benchmark	TSW	M-G
Consumer Price Index (CPI)	0.074 0.241 0.461	0.123 0.293 0.450	0.163 0.332 0.426
Harmonised Index of Consumer Prices (HICP)	0.100 0.255 0.466	0.114 0.272 0.452	0.107 0.267 0.448
M3	1,551,599 947 2,448	2,166,840 1,100 1,709	1,947,577 1,116 1,989
M2	2,410,091 1,048 2,440	2,479,304 1,094 1,831	2,224,942 1,165 2,046
Gross Domestic	252,244	212,606	230,028

Product (GDP)	371 1,004	354 819	363 869
M1	1,318,053 908 1,490	849,764 752 1,470	1,138,385 815 1,319
Industrial Production Index (IPI)	1.618 0.955 2.665	1.639 1.049 2.751	1.619 0.963 2.663
Retail sales	3.159 1.423 5.111	4.389 1.671 3.646	3.048 1.480 2.821
Unemployment – thousands	546.2 20.8 26.6	819.2 24.8 24.4	819.2 24.8 24.4
Balance of payments (BOP) – Transport – Receipts	1,919 36.0 70.3	2,225 38.5 68.0	3,828 50.8 66.4
Balance of payments (BOP) – Sea transport – Receipts	1,215 31.2 70.2	1,574 32.6 58.5	1,352 30.0 58.1
Unemployment – percentage	0.399 0.584	0.399 0.584	0.399 0.584

	0.544	0.544	0.544
Balance of payments (BOP) – Transport – Payments	1,002 25.4 49.7	1,217 29.0 51.8	1,106 27.3 37.9
Imports of Goods without fuels and ships	12,479 98.1 224.7	12,246 96.5 152.1	14,772 102.5 175.1
Exports of Goods without fuels and ships	6,020 67.3 81.3	2,793 45.5 97.7	4,520 58.4 71.0
Exports of Goods	20,174 130.6 138.8	16,877 108.5 192.8	20,562 133.4 133.9
Balance of payments (BOP) – Sea transport – Payments	1,276 31.0 39.8	711.8 21.8 42.9	1,095 28.7 31.8
Imports of Goods	97,620 263.4 345.1	93,330 252.6 319.3	93,509 250.3 324.0
Balance of payments (BOP) –	19,885 87.7	13,120 78.6	13,120 78.6

Travelling – Receipts	96.6	98.9	98.9
Balance of payments (BOP) – Travelling – Payments	1,563 24.4 28.8	1,560 23.4 23.0	1,687 267.0 25.7

Nelson and Granger (1979) utilized the Box-Cox transformations, amongst others, for forecasting purposes (point forecasts) using twenty-one actual economic time series. As they failed in getting superior forecasts, they reached to the rather pessimistic conclusion that it is not worthwhile to make use of these transformations bearing in mind the extra inconvenience, effort, and cost. Their point of view was subsequently adopted by other researchers as well, as already mentioned in the introductory section. Lest to get too disappointed, despite the fact that cost and effort are much lower nowadays than what they were at that time, we further note that Nelson and Granger did not associate forecasts on transformed time series with an outlier detection-adjustment approach. Furthermore, their conclusion was based only on point forecasts, disregarding forecast confidence intervals. The latter are of much importance especially in cases where the focus is on best-worst forecast scenarios. For instance, such is the case with actuarial time series on mortality rates, which may be used further for the construction of pension plans. As shown above, the combination of transformation-linearization leads to shorter forecast confidence intervals.

It should also be stressed that neither in the existing research works thus far, nor in the present one, the treatment of the effect of data transformation on time series forecasting is complete for the simple reason that no work extends the analysis in a bivariate (in general multivariate) framework. Indeed, the existence of variance non-stationarity in time series could potentially contaminate the pre-whitening process (for details about the pre-whitening process see Box and Jenkins, 1976), consequently the sample cross correlation function, so it will mask the true dynamic relationship between two series, one of which is supposed to be the leading indicator, thus affecting negatively the conditional (in this case) forecasts.

3.3.7 The shift towards normality

Another serious concern expressed by Nelson and Granger (1979) was the fact that the problem of acute non-normal distributions they found in most macroeconomic time series they analyzed was restored only very little by their use of data transformations. Table 16 presents the results for the Jarque-Bera statistic for normality (Jarque and Bera, 1980). This statistic is distributed as chi-square with two degrees of freedom. An asterisk right next to an arithmetic value of Table 16 indicates a rejection of the null hypothesis of normality at the 5% significance level (critical value = 5.99).

Table 16. Values of the Jarque –Bera statistic (statistically significant values are indicated with an asterisk)

Time series	Benchmark	TSW	M-G
Consumer Price Index (CPI)	2.889	0.999	0.423
Harmonised Index of Consumer Prices (HICP)	6.289*	5.850	8.263*
M3	19.78*	14.72*	12.44*
M2	16.71*	7.519*	16.31*
Gross Domestic Product (GDP)	14.17*	0.541	3.699
M1	152.6*	2.879	3.597
Industrial Production Index (IPI)	1.118	0.996	1.118
Retail sales	2.328	0.771	0.545
Unemployment – thousands	9.745*	7.613*	7.613*
Balance of payments (BOP)	5.526	0.563	3.587

– Transport – Receipts			
Balance of payments (BOP) – Sea transport – Receipts	7.447*	0.9231E-01	0.7904E-01
Unemployment – percentage	7.584*	7.584*	7.584*
Balance of payments (BOP) – Transport – Payments	137.5*	1.651	5.289
Imports of Goods without fuels and ships	7.938*	0.928	0.266
Exports of Goods without fuels and ships	28.26*	0.473	0.593
Exports of Goods	0.404	0.380	0.180
Balance of payments (BOP) – Sea transport – Payments	210.5*	4.633	4.598
Imports of Goods	1.589	4.115	0.924
Balance of payments (BOP) – Travelling – Receipts	15.31*	4.696	4.696
Balance of payments (BOP)	2.286	1.978	2.013

– Travelling – Payments			

The results of Table 16 allow, again, for a more optimistic view, inasmuch as it is evident that there is a general shift towards normality from the benchmark model to either TSW, or M-G transformation-linearization procedure. The phenomenon on some occasions is really very pronounced indeed (e.g. in the series of M1 and Balance of Payments–transport-payments). This allows for computational algorithms such as maximum likelihood estimation, as well as standard statistical tests, to be legitimately employed with transformed-linearized data.

3.3.8 Statistical benchmark forecasting

Seizing the opportunity of the above analysis, it is useful to assess the forecastability of the twenty time series of the Greek economy. Here forecastability will be perceived in both point and confidence interval forecasts. For the former the MAPE statistic will be employed. For the latter the percentage standard error statistic will be introduced as the mean average of the ratio of the forecasts' standard error over the corresponding actual value, so as to make forecasts of the various series mutually comparable. In all cases one-step-ahead forecasts will be performed[10]. It is stressed that although these forecasts are technically perfectly acceptable, nevertheless they are purely statistical, hence, a-theoretical, and they can only serve as benchmark forecasts in order to evaluate the merit of more structural econometric forecasts. Tables 17-18 show the results in descending order in terms of statistical forecastability according to the Milionis - Galanopoulos method.

[10] Two-(or more)-step-ahead forecasts are available from the author.

Table 17. Forecastability of main economic indicators. Greece. Point forecasts

MAPE			
Time series	**Benchmark**	**TSW**	**M-G**
Harmonised Index of Consumer Prices (HICP)	0.241%	0.257%	0.252%
Consumer Price Index (CPI)	0.238%	0.289%	0.328%
M3	0.561%	0.653%	0.661%
M2	0.625%	0.650%	0.697%
M1	0.786%	0.652%	0.706%
Gross Domestic Product (GDP)	0.760%	0.729%	0.745%
Industrial Production Index (IPI)	1.011%	1.111%	1.019%
Retail sales	1.424%	1.666%	1.458%
Unemployment – thousands	2.170%	2.608%	2.608%
Balance of payments (BOP) – Sea transport – Receipts	2.789%	2.902%	2.640%
Unemployment – percentage	2.917%	2.917%	2.917%
Balance of payments (BOP) – Transport – Payments	2.929%	3.309%	3.134%
Exports of Goods without fuels and ships	3.718%	2.517%	3.208%
Imports of Goods without fuels and ships	3.032%	3.026%	3.258%
Balance of payments (BOP) – Transport – Receipts	2.748%	2.922%	3.835%
Balance of payments (BOP) – Sea transport – Payments	5.515%	3.883%	5.077%
Exports of Goods	5.021%	4.238%	5.129%
Imports of Goods	6.027%	5.750%	5.705%
Balance of payments (BOP) – Travelling – Receipts	12.194%	7.729%	7.729%
Balance of payments (BOP) – Travelling – Payments	12.553%	11.775%	13.994%

Table 18. Forecastability of main economic indicators. Greece. Interval forecasts

Percentage Standard Error			
Time series	**Benchmark**	**TSW**	**M-G**
Consumer Price Index (CPI)	0.454%	0.443%	0.419%
Harmonised Index of Consumer Prices (HICP)	0.439%	0.427%	0.423%
M1	1.290%	1.272%	1.142%
M3	1.451%	1.013%	1.180%
M2	1.455%	1.090%	1.219%
Gross Domestic Product (GDP)	2.145%	1.745%	1.855%
Unemployment – thousands	2.809%	2.572%	2.572%
Unemployment – percentage	2.737%	2.737%	2.737%
Retail sales	5.110%	3.636%	2.803%
Industrial Production Index (IPI)	2.808%	2.890%	2.805%
Exports of Goods without fuels and ships	4.582%	5.483%	4.008%
Balance of payments (BOP) – Transport – Payments	5.817%	6.101%	4.469%
Exports of Goods	5.495%	7.552%	5.294%
Balance of payments (BOP) – Sea transport – Receipts	6.514%	5.394%	5.332%
Imports of Goods without fuels and ships	7.151%	4.833%	5.528%
Balance of payments (BOP) – Sea transport – Payments	7.234%	7.779%	5.807%
Balance of payments (BOP) – Transport – Receipts	5.565%	5.348%	6.317%
Balance of payments (BOP) – Travelling – Receipts	24.967%	7.679%	7.679%

Imports of Goods	8.170%	7.559%	7.700%
Balance of payments (BOP) – Travelling – Payments	17.607%	14.157%	16.151%

From the results of the Tables 17 – 18, it is observed that although there are many similarities in the two Tables, the ordering is not exactly the same. For this reason, the linear correlation coefficient between orderings based on MSFE and the percentage standard error was used. In all cases there is a strong positive correlation (see Table 19). The method of Milionis-Galanopoulos has the highest correlation, while TSW has the lowest.

From Tables 17 and 18 it is also noticeable that the BOP series are the least forecastable in both Tables. Regarding the imports-exports time series it is noted that the former is less forecastable than the latter. Furthermore, imports-exports excluding fuels and ships are clearly more forecastable than imports-exports including them. This justifies, here form the statistics point of view, the separate recording and usage of the imports-exports without the inclusion of fuels and ships for further economic analysis.

Table 19. Linear correlation coefficient between MSFE and percentage SE ordering

Method	Correlation
Benchmark	95.40%
TSW	93.05%
M-G	97.23%

3.4 Conclusions

This work dealt with the effect of data transformation for variance stabilization and linearization for outlier adjustment on the quality of univariate time series forecasts, using two methods for data transformation, those of TSW and Milionis Galanopoulos, and following a practical approach.

There is clear evidence that linearization improves the forecasts' confidence intervals and some evidence that data transformation acts likewise. However, the effect of the latter needs to be reconfirmed using a larger dataset. In contrast no evidence was found that either transformation or linearization lead to better point forecasts. The combined effect of transformation-linearization improves further the forecasts confidence intervals but worsens point forecasts. Furthermore, there is also evidence that the overall forecasting performance using the Milionis Galanopoulos data transformation procedure is somewhat better than the one using the data transformation procedure of TSW.

One field that the documented in this chapter improvement in forecast confidence intervals may be employed with some considerable advantages, is that of the actuarial science and more specifically the longevity risk. This risk is caused by the uncertainty surrounding the future trend of mortality rates of pensioners, as advancements in science and medicine make the prediction of mortality rates a difficult task. One method of addressing the aforementioned issue is to utilize mortality models to forecast the trend of mortality rates and its associated uncertainty in the future. The latter is directly associated with forecast confidence intervals. The whole upcoming chapter (Chapter 4) is exclusively devoted to this topic.

Last, but certainly not least, the combined transformation-linearization procedure improves substantially the non-normality problem encountered in many macroeconomic time series.

SUMMARY OF CHAPTER 3

Very often in actual macroeconomic time series there are causes that disrupt the underlying stochastic process and their treatment is known as «linearization». In addition, variance non-stationarity is in many cases also present in such series and is removed by proper data transformation. The impact of either of them (data transformation - linearization) on the quality of forecasts has not been adequately studied to date. This work examines their effect on univariate forecasting considering each one separately, as well as in combination, using twenty of the most important time series for the Greek economy. Empirical findings show a significant improvement in forecasts' confidence intervals, but no substantial improvement in point forecasts. Furthermore, the combined transformation-linearization procedure improves substantially the non-normality problem encountered in many macroeconomic time series.

CHAPTER 4

MODELLING LONGEVITY RISK: A PRACTICAL STUDY OF THE EFFECT OF STATISTICAL PRE-ADJUSTMENTS ON MORTALITY TREND FORECASTS

4.1 Introduction

The utilization of Chapter 3's findings from analyzing macroeconomic time series in the field of modeling longevity risk is highly advantageous, especially in terms of dealing with variability. In Chapter 3, it's demonstrated that by applying statistical pre-adjustments like transformation and linearization, forecast confidence intervals can be made shorter. This is achieved through transforming the original time series data and adjusting for outliers, which reduces the process variance. This reduction can be employed to generate forecasts with higher levels of confidence.

The significance of forecasting confidence intervals becomes particularly pronounced when considering scenarios that encompass both the most optimistic and the most pessimistic forecasts. This is exemplified in instances like actuarial time series concerning mortality rates, which have the potential for extended application in developing pension plans.

Additionally, managing outliers representing rare real-world events in actuarial data could enhance forecasts. A comparable situation that could exert substantial influence on mortality data and forecasting is the Covid-19 pandemic, characterized by its escalating death toll.

In this chapter use will be made of the findings of previous chapters in modeling actuarial time series, in particular in forecasting longevity risk. The structure of the chapter is as follows: In the next section a skeletal review of the subject is provided, emphasizing on the link of the findings of Chapter 3 with modeling longevity risk. In section 4.3 we describe the data and the software to be employed. In section 4.4 we present and comment upon our results. In section 4.5 we conclude.

4.2 Skeletal review of the subject

As time passes, the average lifespan is getting longer, presenting difficulties to both the insurance sector and the academic community. The rise in life expectancy along with the simultaneous decrease in fertility rates are placing noteworthy financial strain on retirement income programs (Dowd et al., 2010; Oeppen and Vaupel, 2002). The growing quantity of retirement plans and payouts resulting from longer lifespans creates a potential risk of exceeding the budget of pension funds and life insurance companies. Thus, financial organizations including pension funds, governments and life insurance companies must confront the longevity risk. To address this, various regulations have been implemented to ensure the stability of an institution's reserve funds and manage the associated risks. The Solvency II (Directive 2009/138/EC) establishes a standardized method for determining capital requirements across all EU member states with the goal of maintaining the financial stability (solvency) and risk management capabilities of organizations. This capital requirement is called Solvency Capital Requirement (SCR) and covers all the potential risks that an insurance company may encounter. One of the most substantial non-diversifiable risks, among others, is the longevity risk. The longevity risk is meant to be a composition of several components. For the non-familiar reader, a short description of these components is given in Appendix. This risk is caused by the uncertainty surrounding the future trend of mortality rates of those receiving annuities (Dowd et al., 2010; Kleinow and Richards, 2017), as advancements in science and medicine make the prediction of mortality rates a difficult task. In other terms, pensioners are living longer than anticipated causing life insurance policies and retirement plans to pay out compensations for an extended period of time. As a result, profits are decreasing and there is a risk of insolvency. Considering the aforementioned points, Solvency II mandates that insurers maintain sufficient reserves to cover 99.5% of potential scenarios that could arise within a one-year period. Nevertheless, the longevity risk is associated with the prolonged trend of mortality rates over the long term. The aforementioned trend develops over numerous years as a result of the accumulation of minor alterations. Although many insurance risks can be easily incorporated into a one-year value-at-risk framework, not all risks can be treated in the same manner. Demanding that the risk associated with the trend of longevity be evaluated solely over a one-year period would be excessively rigid (Richards and Currie, 2009). This view of longevity trend risk is sometimes called the run-off approach, and it does not correspond with the one-year view demanded by a pure value-at-risk methodology. Therefore, assessing the risk associated with longevity requires predicting data related to longevity over a time horizon of multiple years.

One method of addressing the aforementioned issue is to utilize mortality models from existing literature to forecast the trend of mortality rates and its associated uncertainty in the future. By employing this method, an insurance company can strengthen the process of determining capital requirements. Accurately and methodically predicting the mortality rates is of paramount importance in managing longevity risk.

Numerous mortality models have been proposed over time, starting from the Gompertz law of mortality in 1825, in order to achieve this goal (Cairns et al., 2006; Currie, 2006; Hatzopoulos and Haberman, 2011; Hatzopoulos and Sagianou, 2020; Hyndman and Ullah, 2007; Lee and Carter, 1992; Plat, 2009; Renshaw and Haberman, 2006). Recent advancements in mortality modeling have tended to be extrapolative in nature, with the principal components (PC) approach gaining significant attention. Thus, Bell and Monsell (1991) expanded on the Ledermann and Breas (1959) method by utilizing a PC approach to predict age-specific mortality rates. Lee and Carter (1992) conducted a fundamental study on this method by investigating a modified version of it for the purpose of predicting mortality rates. The primary statistical technique employed was least-squares estimation via singular value decomposition (SVD) of the matrix of the log age specific observed forces of mortality. Improvements to the LC model occur when the model is adjusted by fitting a Poisson regression model to the number of deaths at each age (Brillinger, 1986). Renshaw and Haberman (2003) incorporate age differential effects, introducing a double bilinear predictor structure into the LC forecasting methodology, and optimize the Poisson likelihood. Also, Hyndman and Ullah (2005) use several PCs in order to capture the differential movements in age-specific mortality rates, using functional PCA. A number of recent studies have suggested new approaches to forecasting mortality rates, which involve (nonparametric) smoothing. Thus, Currie et al (2004) use bivariate penalized B-splines to smooth the mortality surface in both the time and age dimensions within a penalized GLM framework. Hyndman and Ullah (2005) smooth the observed log-mortality rates with constrained and weighted penalized regression splines. De Jong and Tickle (2006) introduce a state space framework using B-spline smoothing. Gao and Hu (2009) introduce a Generalized Dynamic Factor method and multivariate BEKK GARCH model to describe mortality dynamics under conditional heteroskedasticity. Lazar and

Denuit (2009) utilize dynamic factor analysis and the methodology of Johansen cointegration to project mortality through a linear state space representation which assumes that common factors can be modelled as a multivariate random walk with drift. Further, in many developed countries (including UK, USA, Japan and Germany), there is evidence of a cohort effect – thus, in the UK, generations born between 1925 and 1945 approximately seem to have experienced more rapid mortality decreases than earlier or later generations. Renshaw and Haberman (2006) incorporate this effect by developing an age-period-cohort version of the LC model which provides an improved fit to the data compared to the basic LC model.

A stochastic mortality model can be used to analyze historical mortality data and gain insights into mortality dynamics, including the trend of mortality rates. The mortality rates obtained from a stochastic mortality model using historical data can be used with the intention to predict the future behavior of mortality trends.

In this chapter, we use the multiple-component stochastic mortality model Hatzopoulos-Sagianou (hereafter called HS) to model the mortality dynamics. The HS model uses a semi parametric estimation method. This method adopts Generalized Linear Models (GLMs) and Sparse Principal Components Analysis (SPCA). A sparsity factor (s value) is necessary for the SPCA to identify the optimal and most informative age–period and age–cohort components. To achieve this, the definition of the sparsity factor is based on a methodology tailored for the HS model and is able to measure the Unexplained Variance (UVR) of each of the age–period and age–cohort components that are incorporated in the proposed model. For more details about the novel dynamic structure and estimation method of the HS model see Hatzopoulos and Sagianou (2020).

In the family of age-period-cohort stochastic mortality models the dynamics of mortality are driven by the period and the cohort indices. Therefore, the forecasting of mortality rates requires the modeling of these indices using time series techniques. We adopt the random walk with drift (henceforth RWD) model, as the standard approach in the actuarial literature, for modelling the period indices (Cairns et al., 2006; 2011; Haberman and Renshaw, 2011; Lee and Carter, 1992; Lovász, 2011; Pitacco et al., 2009; Villegas et al., 2018): $Y_t = d + Y_{t-1} + u_t$ where Y_t is a stochastic time series, u_t is a white noise process, and d is a constant. Nevertheless, researchers have attempted to use other types of stochastic models to improve the accuracy of mortality forecasts (Hatzopoulos and Sagianou, 2020; Lee and Miller, 2001; Plat, 2009; Villegas et al., 2018). A set of models that belong to this category are the AutoRegressive Integrated Moving Average (henceforth ARIMA) models. Before using time series for forecasting purposes, they typically require some statistical preparation and pre-adjustment, as they are not usually suitable in their raw form. For instance, a time series of raw data may have variance non-stationarity. Moreover, it is common to find outliers and other factors, like calendar effects, that disturb the inherent stochastic process. Their treatment is known as "linearization". Variance non-stationarity and outliers not only affect the variance of time series data, but also have an impact on the nature of the ARIMA model and the identification and character of outliers (Milionis, 2003; 2004; Milionis and Galanopoulos, 2019). So, both variance non-stationarity and outliers have an impact on the accuracy of point and interval forecasts. Therefore, the presence of variance non-stationarity and outliers in time series data can negatively affect the accuracy of forecasts, leading to wider confidence intervals, which can in turn adversely impact the management of longevity risk. In the actuarial field, the potential presence of variance instability and outliers in longevity data can lead to an increase in time series variance which can impact the uncertainty surrounding the solvency capital requirements of a pension fund or insurance institution, among other factors. An increase in time series variance may lead some insurers at a competitive disadvantage as they have more capital locked in than the risk profile of the company would imply. On the other hand, if the outliers in the actuarial time series data represent rare events in the real world, such as world wars or pandemics like the Spanish influenza (1917), it may be possible to enhance the accuracy of forecasts by appropriately managing their impact. A similar phenomenon that may have a significant impact in mortality data and forecasting is the Covid-19 pandemic, with the increasing number of deaths attributed to it. In a possible future study of mortality rates, we need to take into account the presence of the Covid-19 pandemic and its influence on mortality, in order to possibly improve the accuracy and quality of predictions. Recent evidence is conducive to such an approach. Indicatively, the New York City Department of Health and Mental Hygiene in a recent analysis finds that life expectancy in the city of New York has been decreased by as much as 4.6 years as a result of the COVID-19 pandemic (Department of Health and Mental Hygiene, 2023).

Despite the importance of both point and interval forecasts on actuarial time series, particularly in mortality rates, the potential presence and nature of variance instability and outliers, their importance and impact on such forecasts, and the potential consequences for the performance of actuarial models have not been thoroughly investigated to date. This is indeed the scope of this work. To this end, the RWD model will be used as a benchmark, as the choice of this particular model is strongly backed by the existing literature. Indeed, Lee and Carter (1992) found that a simple random walk with drift was an appropriate model for the U.S. data they studied, and although they highlighted the possibility of more general models, the random walk with drift is typically used in applications. Lee and Carter developed their approach specifically for U.S. mortality data, 1933-1987. In fact, the method is now being applied to all-cause and cause-specific mortality data from many countries and time periods, all well beyond the application for which it was designed. So, the method proposed in Lee and Carter (1992) has become the "leading statistical model of mortality forecasting in the demographic literature" (Deaton and Paxson, 2004). In addition, stochastic models will be utilized with and without statistical pre-adjustments to evaluate the impact of such adjustments on forecast accuracy. The intention is clearly towards a practical approach.

4.3 Data and software-computational details

In this chapter, we use the HS multiple-component stochastic mortality model in order to model the mortality dynamics. By utilizing mortality models we estimate the death rates and, in turn, the mortality trends in terms of time series, which reveal the behavior of mortality over time. In the family of age-period-cohort stochastic mortality models the dynamics of mortality are driven by the period and the cohort indices. The data used, in order to estimate the time series of the period and cohort indices, consist of the number of deaths, $D_{t,x}$, and the corresponding central exposures to risk, $E_{t,x}$, which are defined in rectangular arrangement (t, x) over a unit range of individual calendar years $t(t_1, \ldots, t_n)$, and individual ages, x, last birthday $(x_1, \ldots, x_n)$. Thus, we calculate the crude (unsmoothed) central death rate for any age x and calendar year t, as $m_{t,x} = D_{t,x}/E_{t,x}$. $E_{t,x}$ is usually approximated by an estimate of the population aged x last birthday in the middle of the calendar year t or by an estimate of the average population aged x last birthday of the beginning and the end of the calendar year t. We model the number of deaths as independent Poisson realizations; that is, $D_{t,x}$ follow Poisson distribution with mean $E_{t,x} \cdot m_{t,x}$ (Brillinger, 1986; Brouhns et al., 2002).

Hatzopoulos and Sagianou (2020) proposed a dynamic multiple-component model that includes δ_1 age-period and δ_2 age-cohort effects. The HS model can be represented by the following generic formula:

$$log(\tilde{m}_{t,x}) = a_x + \sum_{i=1}^{\delta_1} \beta_x^{(i)} \kappa_t^{(i)} + \sum_{j=1}^{\delta_2} \beta_x^{c(j)} \gamma_c^{(j)} \qquad (5)$$

In equation 5 the tilt above $m_{t,x}$ indicates expected value, the term a_x reflects the main age profile of mortality by age, $\beta_x^{(i)}$ and $\beta_x^{c(j)}$ represent the age effect for each period and cohort component, respectively. The terms $\kappa_t^{(i)}$ reflect period-related effects and determine the mortality trend. The terms $\gamma_c^{(j)}$ represent the cohort-related effects, where $c = t - x$. The parameters δ_1 (>=1) and δ_2 (>=0) are indices for the number of period and cohort components included in the model structure, respectively. The number of period and cohort components vary depending on the experimental dataset, i.e., the intrinsic mortality peculiarities of the examined population in a given time frame. For the England and Wales dataset, for the period 1841-2006, δ_1 = 5 and δ_2 = 2 and for the period 1961-2006, δ_1 = 4 and δ_2 = 1. For full details of the Estimation Methodology, see Hatzopoulos and Sagianou (2020).

Therefore, these κ values must be projected. These period, $\kappa_t^{(i)}$, indices reveal the mortality trends of unique age clusters and can be used by a time series analysis technique in order to forecast future mortality trends.

In this spirit, the approach adopted in this paper is the traditional two-stage process: firstly, we fit the stochastic mortality model in order to estimate κ values (see Hatzopoulos and Sagianou, 2020) and then we fit a projection model to the estimated κ values for forecasting.

Therefore, considering the aforementioned and according to Hatzopoulos and Sagianou (2020) results, the dataset for the time series analysis consists of nine annual time series of period indices $\kappa_t^{(i)}$ for England and Wales dataset, of which five are "long" time series, while four are "short" time series. The long time series data cover the period from 1841 to 2016 and consist of one hundred and seventy-six (176) observations. The short time series cover the period from 1961 to 2016 (55 observations). The graphical representations of the nine time series are shown in Figures 4-12 (row data with the arithmetic values available on request by the author).

Figure 4

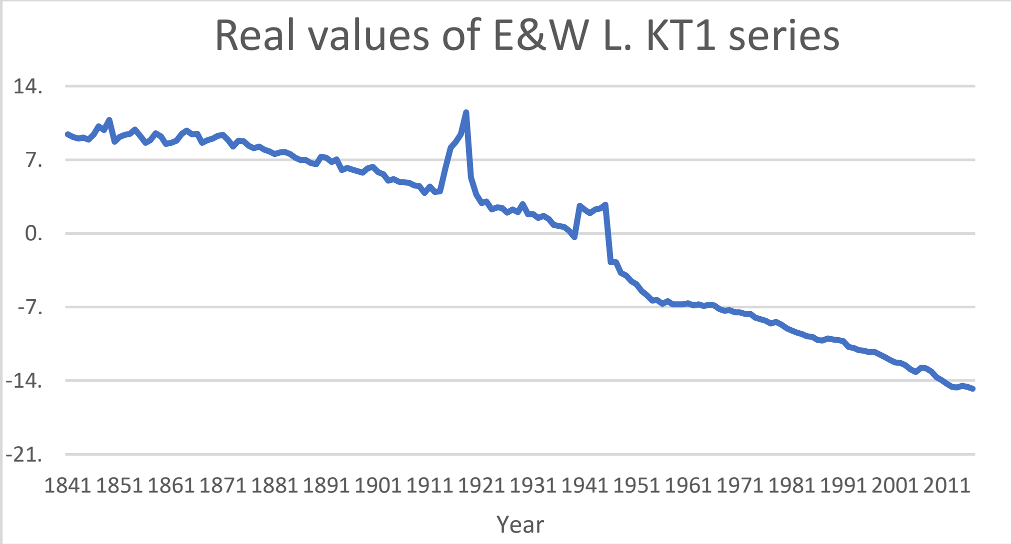


Figure 5

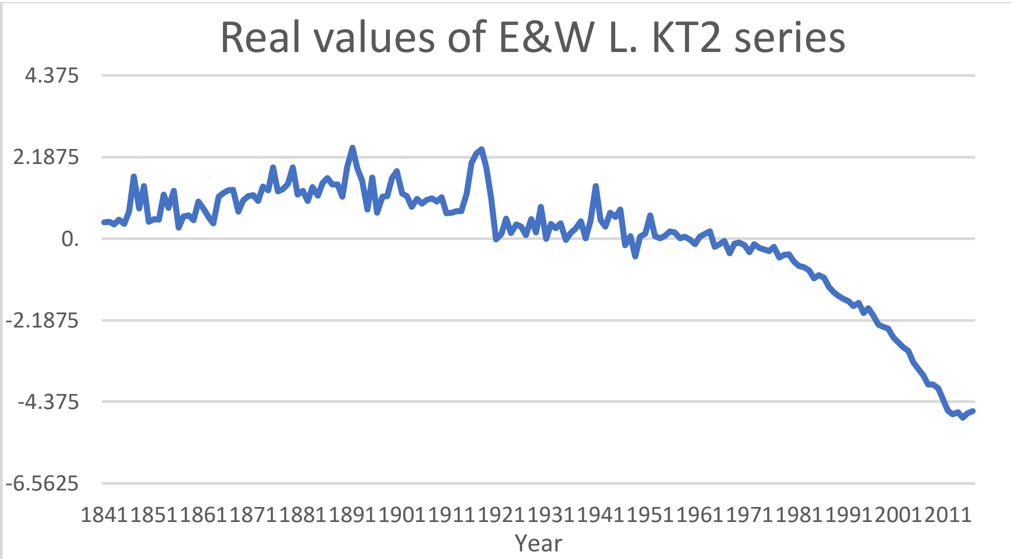


Figure 6

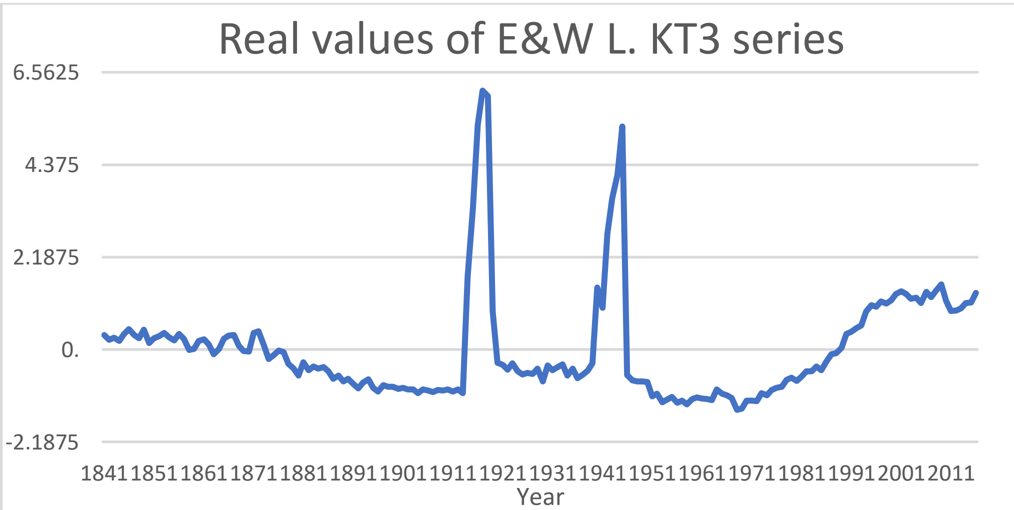


Figure 7

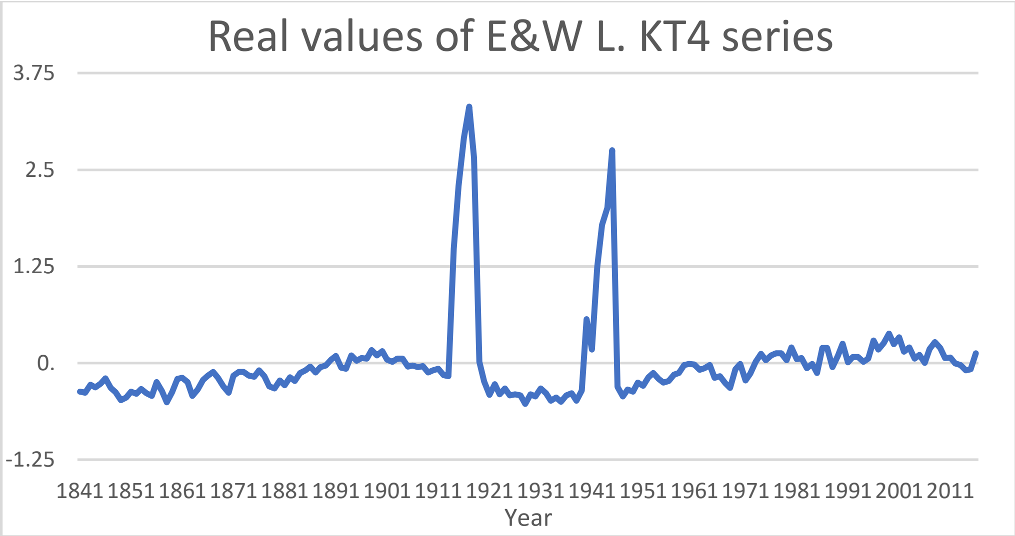


Figure 8

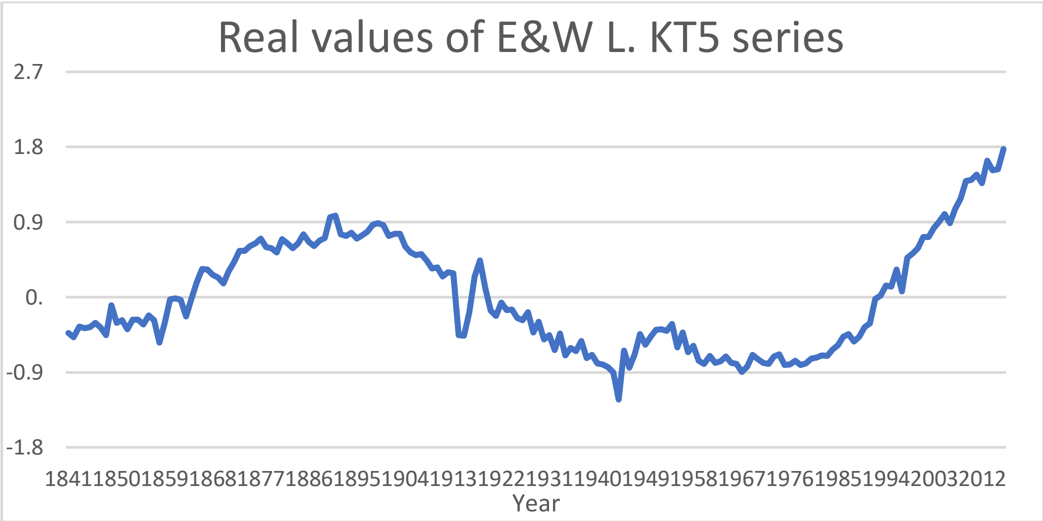


Figure 9

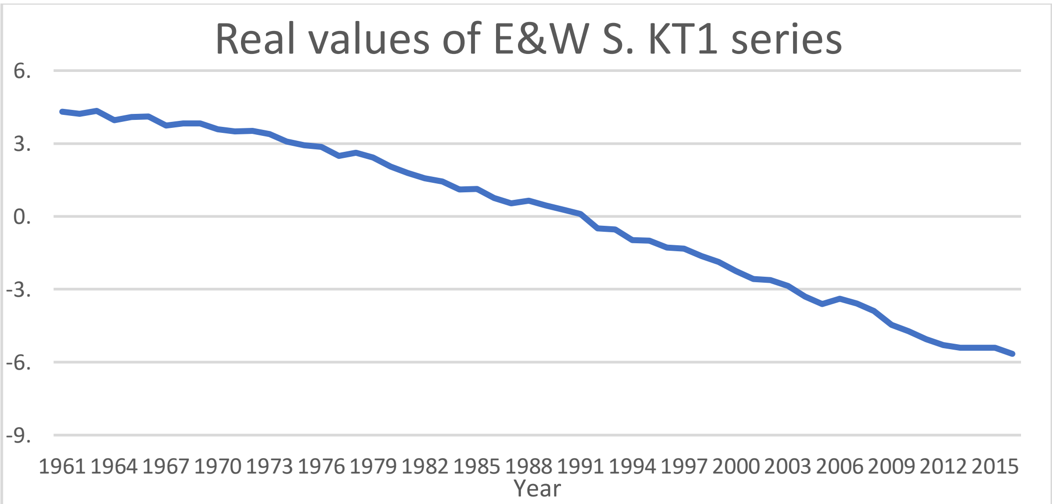


Figure 10

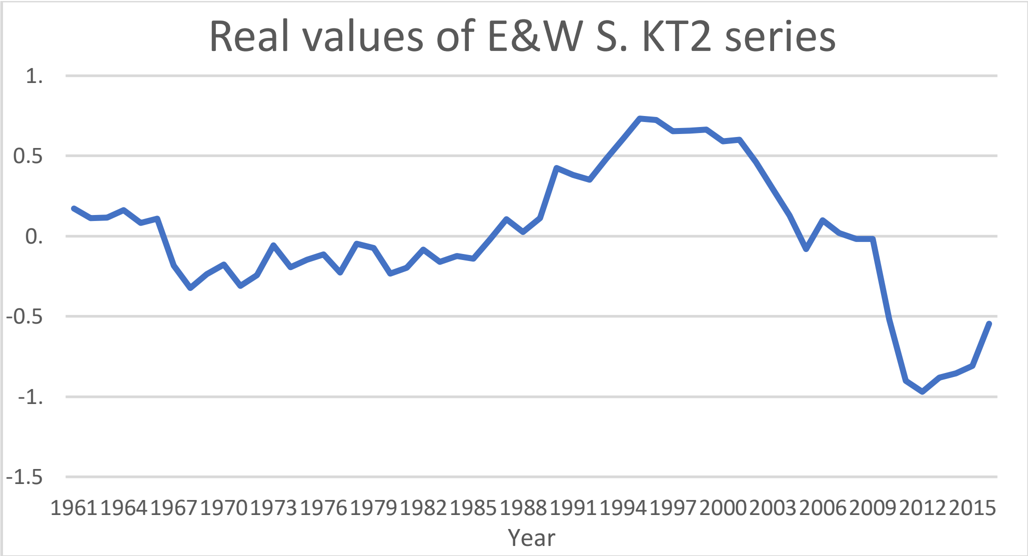


Figure 11

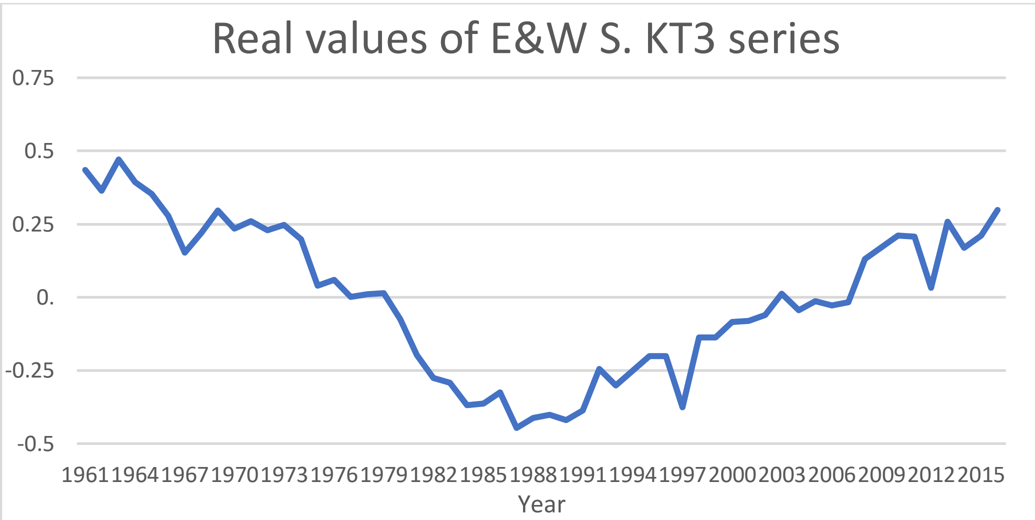
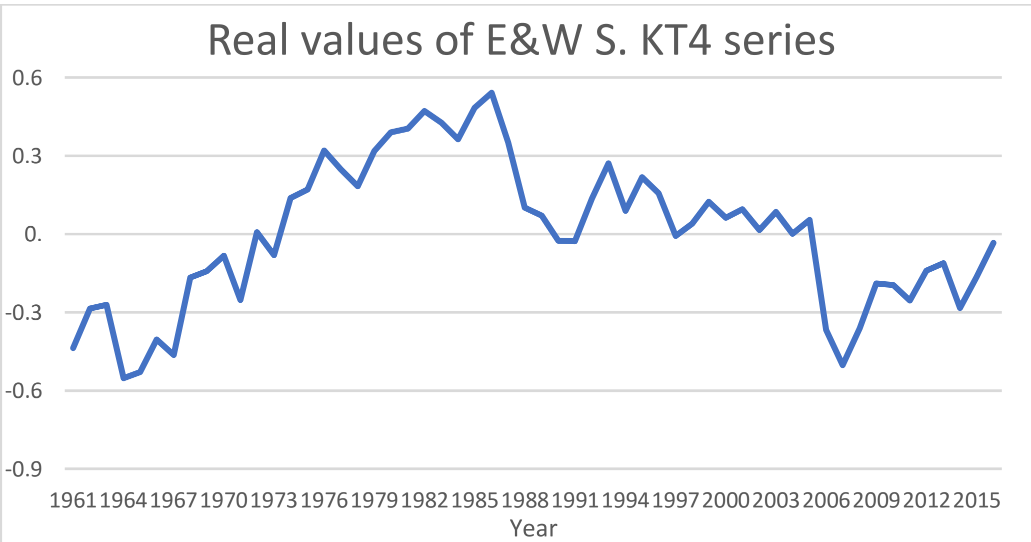


Figure 12



To assess the effect of statistical pre-adjustments on forecasts two statistical software approaches will be employed, namely the "AUTOARIMA" command and its extensions of the well-known programming software "R" and the module TRAMO of the TSW statistical package.

The "AUTOARIMA" command of "R" allows for the automatic selection of an ARIMA model. Moreover, forecasts based on the selected model may be obtained. On the other hand, TRAMO pre-tests for time series transformation to tackle with variance non-stationarity. Moreover, it offers several options for the treatment of outliers within the frames of the more general pre-adjustment procedure known as "linearization" (see section 1.5). Nevertheless, TSW only allows for logarithmic transformation, limiting the options for transformation. Therefore, to have a wider range of transformations, such as the square root transformation, the statistical approach and recommendations suggested by Milionis will also be utilized (Milionis, 2003; 2004; Milionis and Galanopoulos, 2019). An observation is classified as an outlier based on the critical value of a suitable statistic $\boldsymbol{\tau}$, which is described in Gómez and Maravall (1996); Caporello and Maravall (2004). Since the critical value of $\boldsymbol{\tau}$ cannot be predicted by theory, it is commonly related to the length of the time series (Fischer and Planas, 2000). In this study, the default options of TSW for identifying outliers will be utilized.

4.4 Results and discussion

4.4.1 Data transformation

Initially, it is crucial to acknowledge that the impact of a transformation is twofold: direct and indirect. The direct effect is evident and pertains to the transformation itself. The indirect effect concerns the influence of the transformation on detecting outliers. Studies have demonstrated that data transformation has an impact on both the number and the character of outliers in a time series (Milionis, 2003; 2004; Milionis and Galanopoulos, 2019).

Table 1 displays the results of deciding whether to transform the original time series data using TSW. From the analysis of the nine-time series examined, it was found that in seven cases, no transformation of the initial data required, while in only two cases, log transformation was deemed necessary. These results were identical when the alternative approach of Milionis was applied (2004).

Table 1. Decision about data transformation

Time series	TSW	Milionis (2004)
E&W L.KT1	Levels	Levels
E&W L.KT2	Levels	Levels
E&W L.KT3	Logs	Logs
E&W L.KT4	Logs	Logs
E&W L.KT5	Levels	Levels
E&W S.KT1	Levels	Levels
E&W S.KT2	Levels	Levels
E&W S.KT3	Levels	Levels
E&W S.KT4	Levels	Levels

To conduct a more detailed investigation about statistical forecasting, three different methods will be considered. These methods are the following: (a) The random walk with drift model, which is a commonly employed model in actuarial research due to its simplicity, as previously noted, and will be used as benchmark. (b) The "AUTOARIMA" command of the programming software "R" for automatic selection and forecasting, as in Hatzopoulos and Sagianou (2020). (c) ARIMA models following statistical pre-adjustments. The latter implies Variance Reduction and will be called "VR" forecasts henceforth.

Table 2 presents the ARIMA models utilized in methodologies (b) and (c). In our case, seasonality is out of context, as annual data will be used. Hence, a non-seasonal ARIMA model will be sought for. Moreover, $b_1, \ldots, b_n$, as well as $C_t' \eta$ in equation of the general framework of linearization (see section 1.5) will all be set equal to zero.

It should be noted that differences in the ARIMA models, for time series where no transformation was required, may be due to the presence of outliers adjusted by linearization and possible differences in the computational algorithms between the two software products employed.

Table 2. ARIMA models

Time series	"AUTOARIMA"	VR
E&W L.KT1	(0,1,0) WITH MEAN	(0,1,1) WITH MEAN
E&W L.KT2	(0,1,3) WITH MEAN	(3,1,0) WITHOUT MEAN
E&W L.KT3	(3,0,0) WITHOUT MEAN	(0,1,0) WITHOUT MEAN
E&W L.KT4	(1,0,3) WITHOUT MEAN	(0,1,1) WITHOUT MEAN
E&W L.KT5	(1,1,4) WITHOUT MEAN	(1,2,1) WITHOUT MEAN
E&W S.KT1	(0,2,2) WITHOUT MEAN	(1,1,0) WITH MEAN
E&W S.KT2	(0,1,0) WITHOUT MEAN	(0,1,0) WITHOUT MEAN
E&W S.KT3	(0,1,0) WITHOUT MEAN	(0,1,0) WITHOUT MEAN
E&W S.KT4	(0,1,0) WITHOUT MEAN	(1,0,0) WITH MEAN

4.4.2 The effect of "Linearization"

Outliers are significant fluctuations in values that are noticeable in time series. Upon visually analyzing the time series included in our dataset (*E&W L.KT3* and *E&W L.KT4* in natural logarithms), it is apparent that in certain instances the amplified variance can be attributed to outliers. To detect outliers in all time series, TSW was utilized with default settings.

Table 3 outlines the type of outlier and the order of observation in which they appear. The first number refers to the order of observation followed by the type of outlier. For instance, 80 LS in E&W L.KT2 time series (see second row of Table 3) shows that the order (80) of the observations (years) of detected outlier is the year 1921, as the initial observation is the year 1841, and the type of outlier (LS) is Level Shift.

Table 3. Detected outliers and their type

Time series	Temporal Position and Type of outliers
E&W L.KT1	9 AO, 74 LS, 75 LS, 78 TC, 79 LS, 100 LS, 106 LS
E&W L.KT2	80 LS, 100 AO
E&W L.KT3	74 LS, 79 LS, 80 LS, 89 AO, 100 LS, 102 LS, 106 LS, 111 LS, 113 TC, 116 TC, 118 AO, 124 TC, 128 TC, 130 TC, 133 LS
E&W L.KT4	9 TC, 18 AO, 74 LS, 79 LS, 88 AO, 95 AO, 100 LS, 106 LS
E&W L.KT5	9 AO, 18 AO, 23 TC, 50 TC, 74 TC, 77 TC, 78 AO, 104 AO, 157 AO
E&W S.KT1	NO OUTLIERS DETECTED
E&W S.KT2	NO OUTLIERS DETECTED
E&W S.KT3	37 AO
E&W S.KT4	NO OUTLIERS DETECTED

4.4.3 The combined effect of Data Transformation and Linearization

To evaluate the combined effect of data transformation and linearization on the quality of point forecasts some typical statistics will be used. Primarily, the Mean Square Forecast Error (MSFE) which measures the average squared difference between the forecasting values (F_t) and the actual values (A_t), i.e. $MSFE = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2$. It is well known that optimal forecasts are those with the minimum MSFE (Hamilton, 1994). Auxiliary, the following statistics will also be used:

i) the Mean Absolute Percentage Error (MAPE) statistic given by:

$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$, and

ii) the Mean Absolute Error (MAE) statistic given by:

$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t|$.

Furthermore, when evaluating interval forecasts, the forecast standard error will be taken into consideration.

In addition, the Akaike Information Criterion (AIC) will be utilized as a probabilistic statistical measure to assess the model's performance on the training dataset in conjunction with the complexity of the model.

Best forecast will obviously be perceived the one with the minimum value of the each time utilized statistic from the ones mentioned above.

Table 4 displays the count of forecasts that performed better based on the minimization of each statistic mentioned above, when comparing the VR model to the RWD model. According to the results presented in Table 4, it is evident that the VR methodology outperforms the RWD model in every single case in terms of the width of the forecast standard error. Additionally, based on the minimum value of the Akaike information criterion, the VR methodology is superior. The point forecasts generated with the VR methodology are slightly better in terms of the three statistics (MSFE, MAPE, MAE).

The results of the examination of the forecasting performance between VR methodology and "AUTOARIMA" are presented in Table 5.

Table 5 is read in the same manner as Table 4, explaining further that when the calculated values of a statistic are found to be equal, then the arithmetic value 0.5 is assigned in both methodologies. For instance, the AIC values 2.5/9 and 6.5/9 of the fourth row of the Table 5 indicate that in two out of the nine time series the corresponding statistic value is minimum with the "AUTOARIMA" methodology, in six out of the nine time series the corresponding statistic value is minimum with TSW methodology, and in one time series the estimated statistic value is equal in both methodologies.

Table 4. Summary table - Number of best forecasts (VR versus RWD)

(Table is read as follows: for each statistic, in the second and third column the cases with the minimum value of the statistic (i.e. the best forecasts) out of the total number of cases (i.e. the nine time series of the dataset) are presented).

Point Forecasts	**RWD**	**VR**
MSFE	3/9	6/9
MAPE	4/9	5/9
MAE	3/9	6/9
AIC	1/9	8/9
Interval Forecasts	**RWD**	**VR**
Forecast Standard Error (SE)	0/9	9/9

From the results of Table 5 it is seen that the VR methodology outperforms "AUTOARIMA" in terms of the interval forecasts and is better in terms of the Akaike information criterion. Additionally, it is concluded that point forecasts generated by the VR methodology are slightly better in terms of MSFE and MAE compared to those of "AUTOARIMA", and are equal in terms of MAPE.

Table 5. Summary table - Number of best forecasts (VR versus "AUTOARIMA")

Point Forecasts	**"AUTOARIMA" with further analysis in TSW**	**VR**
MSFE	3/9	6/9
MAPE	4/9	5/9
MAE	3/9	6/9
AIC	2.5/9	6.5/9
Interval Forecasts	**"AUTOARIMA" with analysis further in TSW**	**VR**
Forecast Standard Error (SE)	1.5/9	7.5/9

4.4.4 An Ad-Hoc Evaluation of the overall Models' Forecasting Performance

The skill of a forecast can be assessed by comparing the relative proximity of both the forecast and a benchmark to the observations. The use of a benchmark allows for easier comparison between different forecasting methods and for this reason a benchmark is proposed to establish a common ground for comparison. In this study an obvious benchmark is the Random Walk Model with Drift (RWD) as already mentioned.

A crude, yet very simple and transparent ad-hoc forecasting evaluation for both point and interval forecasts will be used. More specifically, for the point forecasts for each time series and for each model an arithmetic value is assigned in ascending order based on the corresponding value of the MSFE statistic (i.e., 1 for the best (minimum) MSFE value, 2 for the second best MSFE value, 3 for the worst (maximum) MSFE value). Then, adding up the arithmetic values for all series for a particular model their sum will represent the performance of the model. Models will be ranked according to the value of the corresponding sum. Apparently, the model with the lowest sum will be considered as the best one. For interval forecasts the same procedure will be followed replacing the value of the MSFE statistic with the value of the corresponding standard error around point forecasts. The results are presented in Tables 6 and 7.

Table 6. Ranking of forecasting performance according to MSFE (points forecasts)

Time series	RWD	"AUTOARIMA"	VR
E&W L.KT1	*1.5*	*1.5*	*3*
E&W L.KT2	*1*	*3*	*2*
E&W L.KT3	*2*	*3*	*1*
E&W L.KT4	*2*	*3*	*1*
E&W L.KT5	*2*	*3*	*1*
E&W S.KT1	*3*	*1*	*2*
E&W S.KT2	*1*	*2.5*	*2.5*
E&W S.KT3	*3*	*1.5*	*1.5*
E&W S.KT4	*2*	*3*	*1*
Total	*17.5*	*21.5*	*15*

From the results of Tables 6 it is evident that the performance of VR methodology for point forecast is better than that of RWD model and "AUTOARIMA". It should be noted that the RWD model performs better than the "AUTOARIMA".

Regarding interval forecasts, the findings presented in Table 7 indicate that the VR methodology has a clearly superior performance compared to both the RWD model and "AUTOARIMA". In this case "AUTOARIMA" clearly outperforms RWD model.

Table 7. Ranking of forecasting performance according to SE (intervals forecasts)

Time series	RWD	"AUTOARIMA"	VR
E&W L.KT1	*2.5*	*2.5*	*1*
E&W L.KT2	*3*	*2*	*1*
E&W L.KT3	*3*	*2*	*1*
E&W L.KT4	*3*	*2*	*1*
E&W L.KT5	*3*	*2*	*1*
E&W S.KT1	*3*	*1*	*2*
E&W S.KT2	*3*	*1.5*	*1.5*
E&W S.KT3	*3*	*2*	*1*
E&W S.KT4	*3*	*2*	*1*
Total	*26.5*	*17*	*10.5*

4.4.5 Analysis of the E&W L.KT5 time series

E&W L.KT5 is a time series which deserves special attention. The series was not found to be variance non-stationary in the sense that its variance was not found to be functionally related to a non-stationary level, either using the purpose-built TSW subroutine, or the methodology suggested by Milionis (2003; 2004). Yet, visual inspection of the series (see Figure 8) reveals a clearly non constant behavior in terms of its variance. To deal with such cases, existing bibliography suggests a (logarithmic) data transformation (Gujarati, 2003, chapter 7). However, we have some reservations in using this recommendation as a general rule.

To examine it further, we perform a forecasting experiment exclusively for the E&W L.KT5 time series. More specifically, three forecasting approaches were used: i) outlier adjustment without data transformation, ii) "AUTOARIMA", iii) a combination of outlier adjustment and one of the three most commonly used transformations, namely the logarithmic, the squared root and the negative inverse.

For the implementation of this comparison, the forecast values from the ARIMA model that derived from "AUTOARIMA" were derived with the programming language R, and the forecast values from both the proposed methods of outlier adjustment without data transformation and the outlier adjustment with data transformation, were derived exclusively with TSW.

The ARIMA models from all these different procedures, their ARMA parameters estimates, and their corresponding standard errors are presented in Table 8, while the relevant forecast evaluation statistics are presented in Table 9. The results of Table 9 indicate that both point, and intervals forecasts are better with the proposed VR methodology of solely outlier adjustment, without any data transformation. Hence, the above results counterevidence the existing recommendation in the literature regarding the treatment of variance instability. Indeed, a case-by-case treatment seems to be more reasonable that the blind application of the logarithmic transformation.

Table 8. ARIMA models and ARMA parameter estimates for E&W L.KT5 series

	Parameter Estimates					
"AUTOARIMA"	**AR(1)**	**MA(1)**	**MA(2)**	**MA(3)**	**MA(4)**	**Integration order**
Coefficients	0.8557	-1.1044	0.2115	-0.1415	0.2189	1
s.e.	0.1202	0.1280	0.1310	0.1333	0.0762	
Outlier adjustment without data transformation						
Coefficients	0.41651	-0.88345				2
s.e.	0.74705E-01	0.38500E-01				
Outlier adjustment and Logarithmic transformation						
Coefficients	0.30645					1
s.e.	0.74104E-01					
Outlier adjustment and Squared root transformation						
Coefficients	-0.24422					1
s.e.	0.75493E-01					
Outlier adjustment and Negative inverse transformation						
Coefficients	0.61122					

s.e.	0.61615E-01					1

Table 9. Summary table – forecast evaluation statistics for E&W L.KT5 (best forecasts in bold)

Point Forecasts	**"AUTOARIMA" with further analysis in R**	**TSW – levels (VR)**	**TSW - logs**	**TSW-Negative inverse**	**TSW-Squared root**
MSFE	0.1401	**0.034**	0.3102	0.2950	0.3161
MAPE	23.33%	**11.01%**	34.99%	33.86%	35.43%
MAE	0.3475	**0.1612**	0.5206	0.5051	0.5268
Interval Forecasts	**"AUTOARIMA" with further analysis in R**	**TSW – levels (VR)**	**TSW - logs**	**TSW-Negative inverse**	**TSW-Squared root**
Forecast Standard Error (SE)	0.2631	**0.2273**	0.5751	1.5162	0.3497

4.4.6 Further illustrative and detailed analysis

It is worthy to present a more detailed analysis for each series. This is done with the aid of Figures 13-21 and Tables 10-17. More specifically, Figures 13 and 14 refer to the E&W L.KT3 and E&W L.KT4 series respectively. The full potential of the VR methodology is realized in these two series where the series are log-transformed and there are multiple outliers, as seen in Table 3. Figures 13 and 14 demonstrate that the VR method substantially narrows the forecast confidence interval and leads to a noticeable improvement in point forecasts. The accuracy of these forecasts is supported by the forecast evaluation statistics in Tables 10 and 11.

Figure 13. Forecasts and Confidence intervals with both methods for the series E&W L.KT3

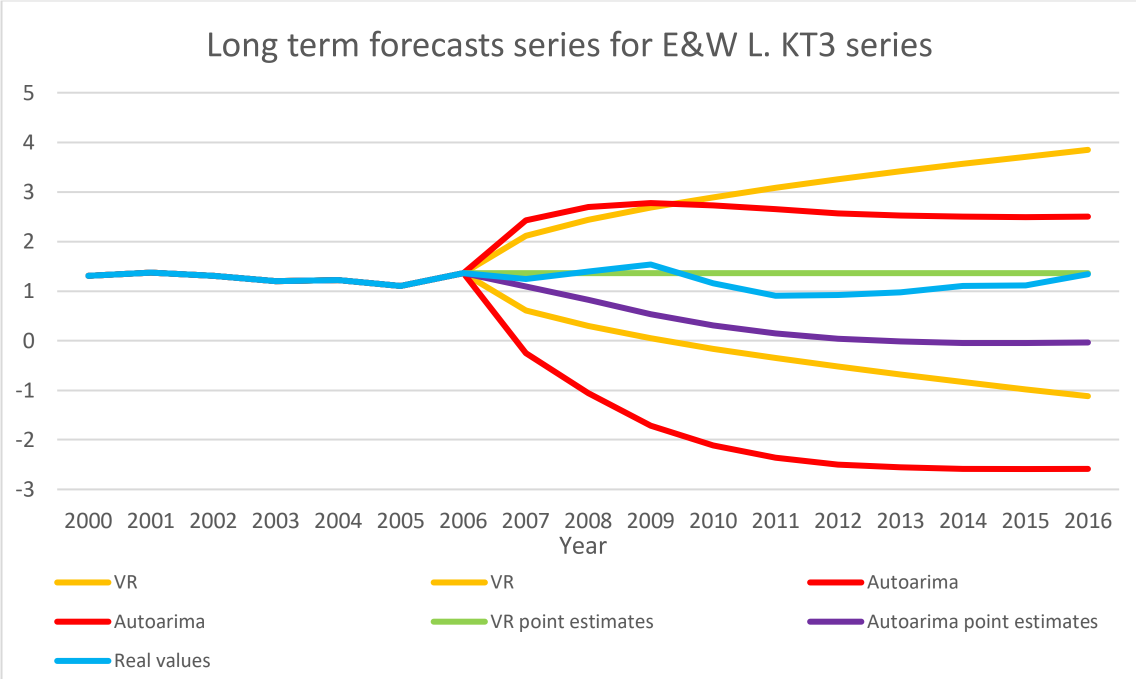


Figure 14. Forecasts and Confidence intervals with both methods for the series E&W L.KT4

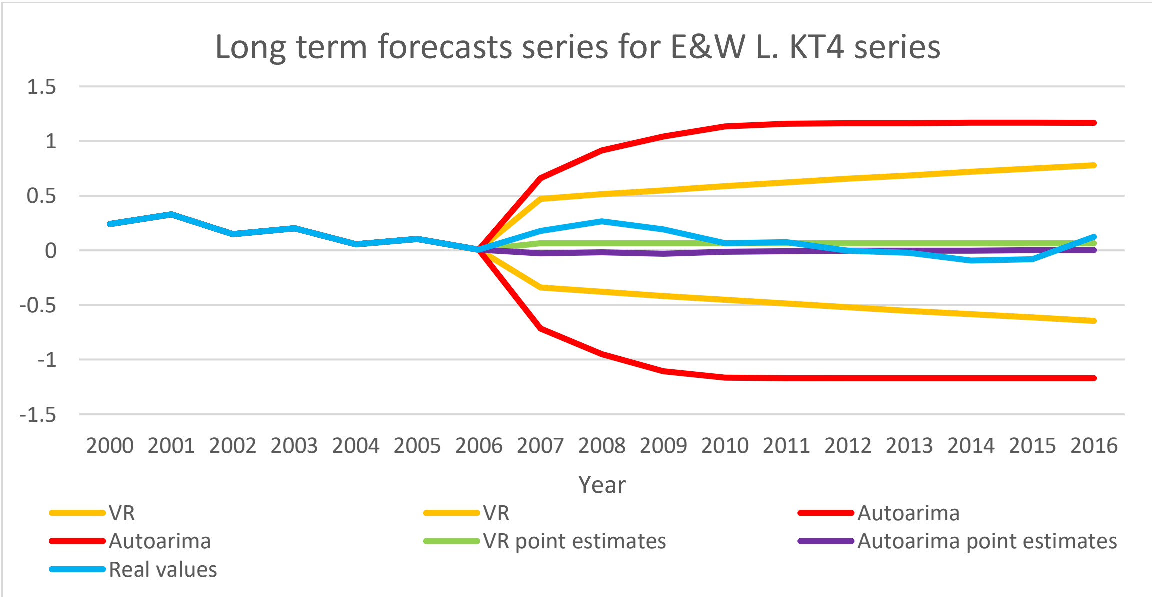


Table 10. Forecast Evaluation Statistics for the series E&W L.KT3

Point Forecasts	"AUTOARIMA"	VR
MSFE	0.90	**0.08**
MAPE	78.57%	**22.90%**
MAE	0.89	**0.24**
Interval Forecasts	**"AUTOARIMA"**	**VR**
Forecast Standard Error (SE)	1.15	**0.89**

Table 11. Forecast Evaluation Statistics for the series E&W L.KT4

Point Forecasts	"AUTOARIMA"	VR
MSFE	0.02	**0.01**
MAPE	**100.69%**	253.11%
MAE	0.12	**0.10**
Interval Forecasts	**"AUTOARIMA"**	**VR**
Forecast Standard Error (SE)	0.55	**0.29**

In passing, it is also worthy to pay some attention on the value of the MAPE statistic in Table 11. Observing Figure 14, as well as the values of MSFE and MAE statistic in Table 11 it is obvious that the point forecasts are better with the VR methodology. However, the value of MAPE statistic implies the opposite. This is due to the small value in the denominator of the MAPE formula with the VR methodology. Such cases justify our choice to use more than one statistical criteria for the evaluation of the forecasting performance.

Figure 15 displays the point and interval forecasts for the E&W S.KT3 series. According to Table 2, both the VR methodology and the RWD model use a simple random walk model without drift. Consequently, both methods generate identical point forecasts, which are uninformative, as they are equal to the last observation. However, Table 3 reveals that an additive outlier is identified in the 37th observation using the VR methodology. Despite the use of the same model in both methods, the VR methodology still improved the forecast quality, especially due to the detection of an outlier. This outlier detection resulted in a reduction of the forecast confidence interval. The evaluation of these forecasts can be found in Table 12.

Figure 15. Forecasts and Confidence intervals with both methods for the series E&W S.KT3

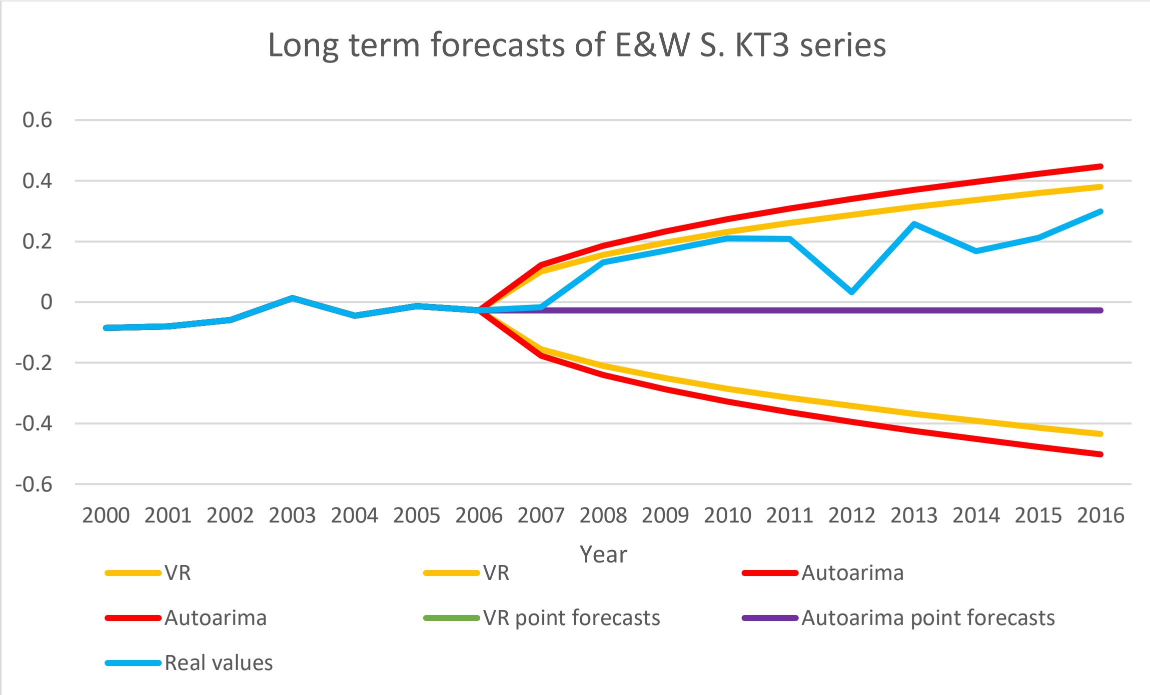


Table 12. Forecast Evaluation Statistics for the series E&W S.KT3

Point Forecasts	**"AUTOARIMA"**	**VR**
MSFE	0.05	0.05
MAPE	114.23%	114.23%
MAE	0.21	0.21
Interval Forecasts	**"AUTOARIMA"**	**VR**
Forecast Standard Error (SE)	0.17	**0.15**

Figure 16 shows the point as well as the interval forecasts for the series E&W L.KT5. The forecast evaluation statistics are presented in the first three columns of Table 9. The aforementioned results indicate that both point, and intervals forecasts are better with the proposed VR methodology of solely outlier adjustment, without any data transformation.

Figure 16. Forecasts and Confidence intervals with both methods for the series E&W L.KT5

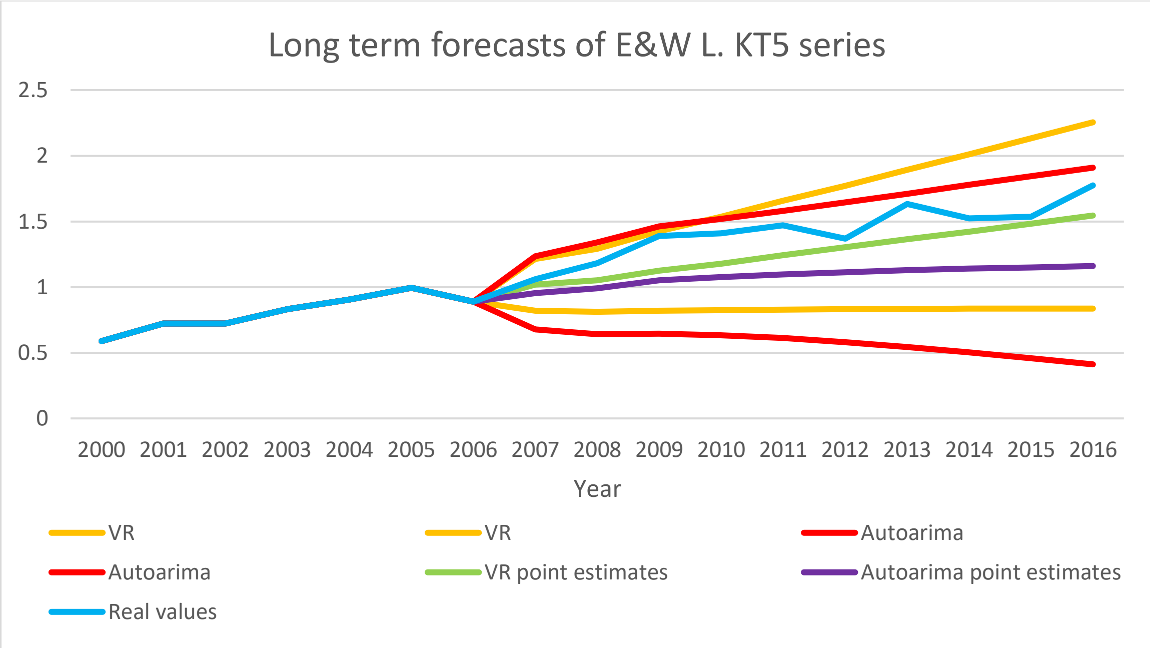
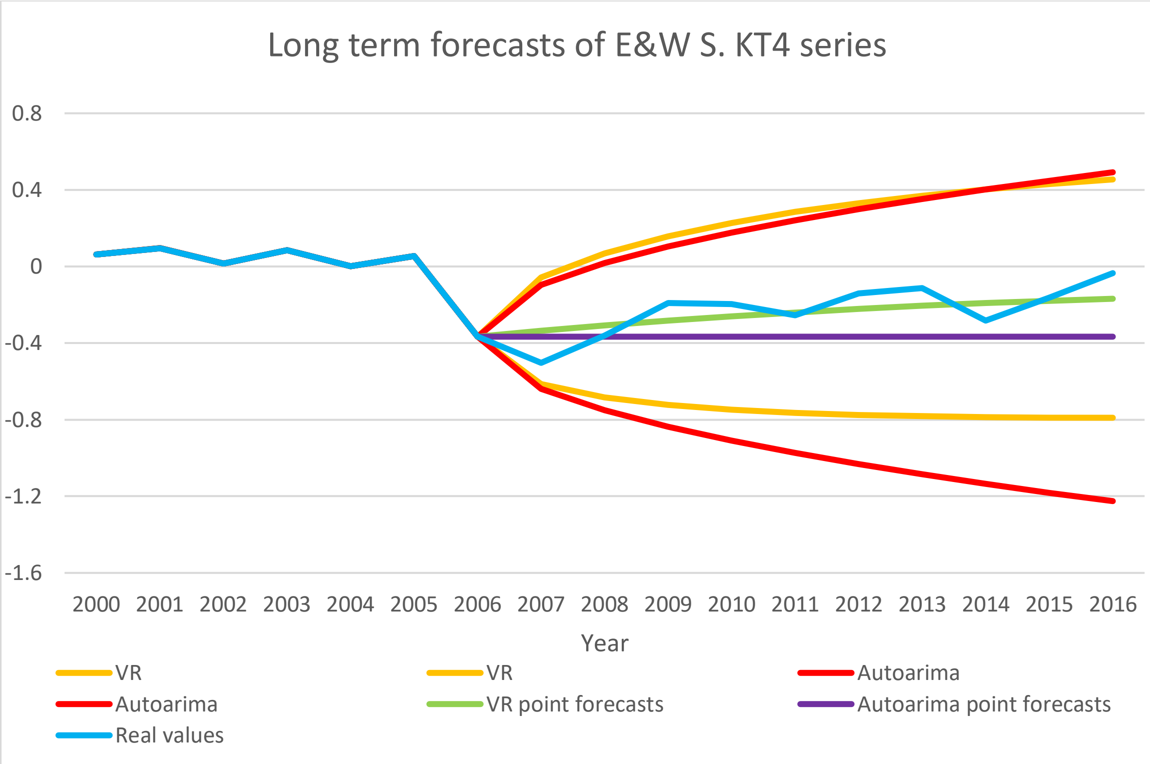


Figure 17 shows the point, as well as the interval forecasts, for the series E&W S.KT4. It is noted that with VR methodology a statistically significant drift (see Table 2) was found. The results indicate that both point, and intervals forecasts are better with the proposed VR methodology. This finding, however, is exclusively due to the ARIMA model identification-forecasting algorithm, as neither any data transformation nor outlier adjustments were used. The relevant forecast evaluation statistics are presented in Table 13.

Table 13. Forecast Evaluation Statistics for the series E&W S.KT4

Point Forecasts	**"AUTOARIMA"**	**VR**
MSFE	0.037	**0.009**
MAPE	177.22%	**71.12%**
MAE	0.170	**0.081**
Interval Forecasts	**"AUTOARIMA"**	**VR**
Forecast Standard Error (SE)	0.31	**0.26**

Figure 17. Forecasts and Confidence intervals with both methods for the series E&W S.KT4

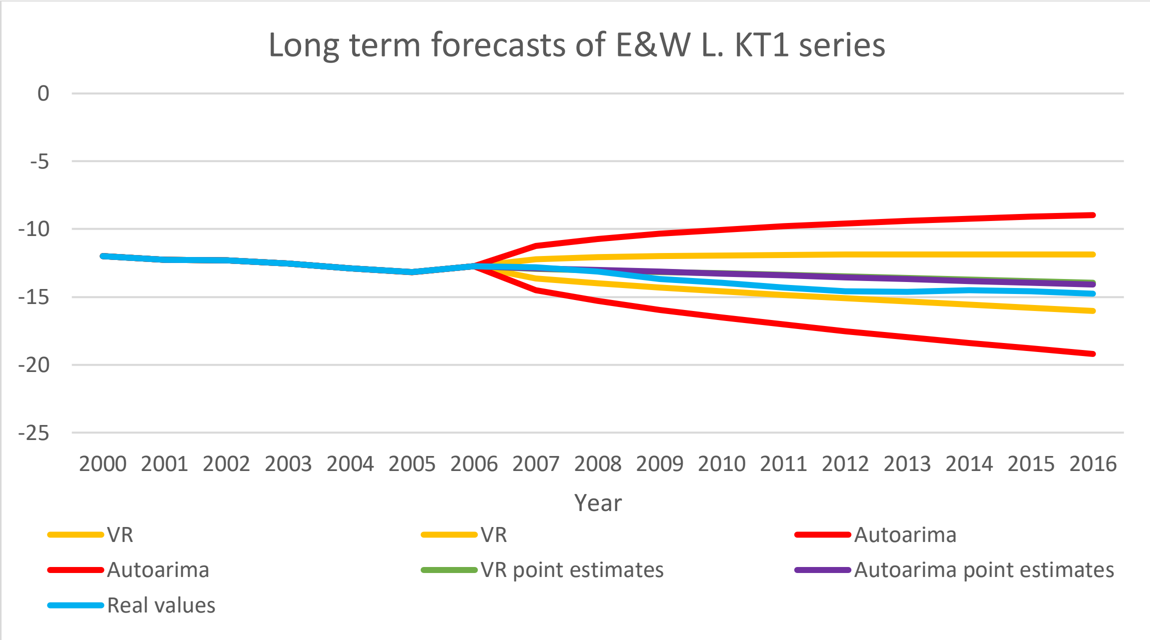


The forecasts and confidence intervals with both methods for the series E&W L.KT1, which is the most important in actuarial sciences among the time series we examined, are presented in Figure 18. From this figure it is obvious that intervals forecasts are better with the proposed VR methodology. This is to be attributed to the outlier adjustment (no data transformation was needed in the particular series). However, point forecasts are slightly better with the "AUTOARIMA" methodology (see Table 14).

Table 14. Forecast Evaluation Statistics for the series E&W L.KT1

Point Forecasts	**"AUTOARIMA"**	**VR**
MSFE	**0.476**	0.479
MAPE	**4.31%**	4.38%
MAE	**0.618**	0.623
Interval Forecasts	**"AUTOARIMA"**	**VR**
Forecast Standard Error (SE)	1.86	**0.76**

Figure 18. Forecasts and Confidence intervals with both methods for the series E&W L.KT1



Figures 19 (series E&W L. KT2) and 20 (series E&W S. KT2) show two cases in which both methods fail, as the real values are outside the confidence interval of the forecasts. The forecast evaluation statistics for these two time series are presented in Tables 15-16.

Table 15. Forecast Evaluation Statistics for the series E&W L.KT2

Point Forecasts	**"AUTOARIMA"**	**VR**
MSFE	1.01	**0.95**
MAPE	21.13%	**20.18%**
MAE	0.95	**0.91**
Interval Forecasts	**"AUTOARIMA"**	**VR**
Forecast Standard Error (SE)	0.54	0.54

Figure 19. Forecasts and Confidence intervals with both methods for the series E&W L.KT2

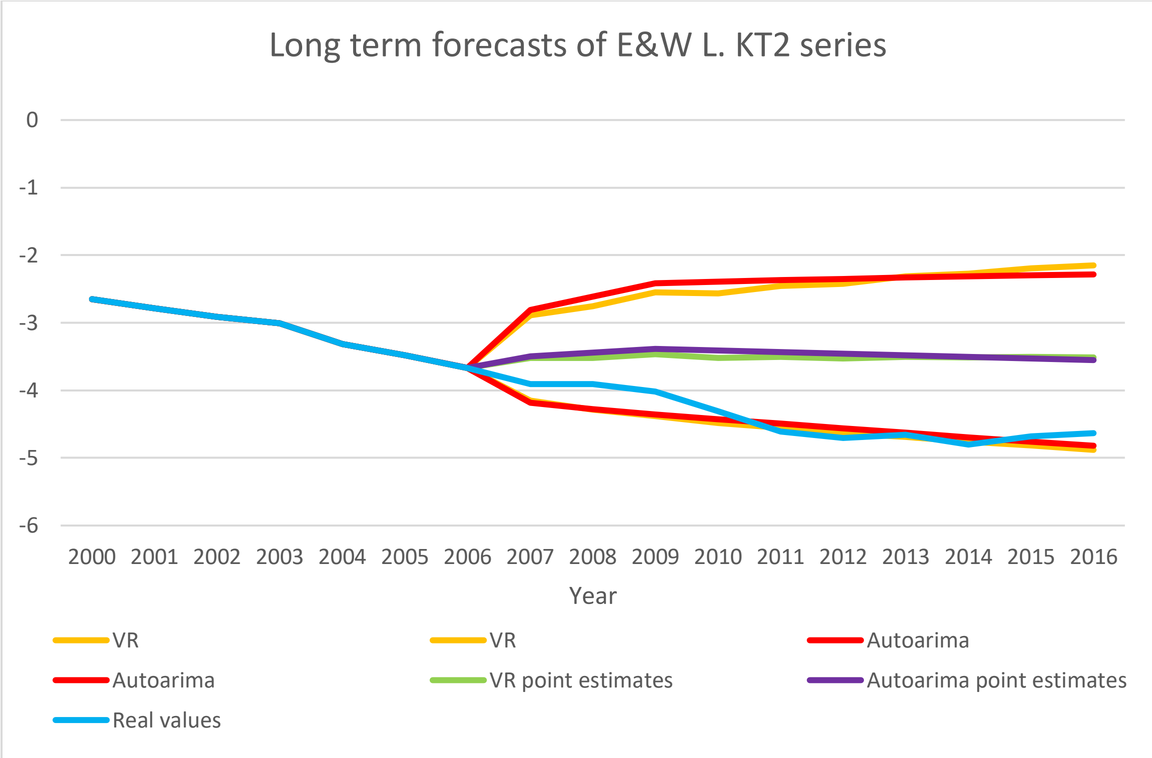


Figure 20. Forecasts and Confidence intervals with both methods for the series E&W S.KT2

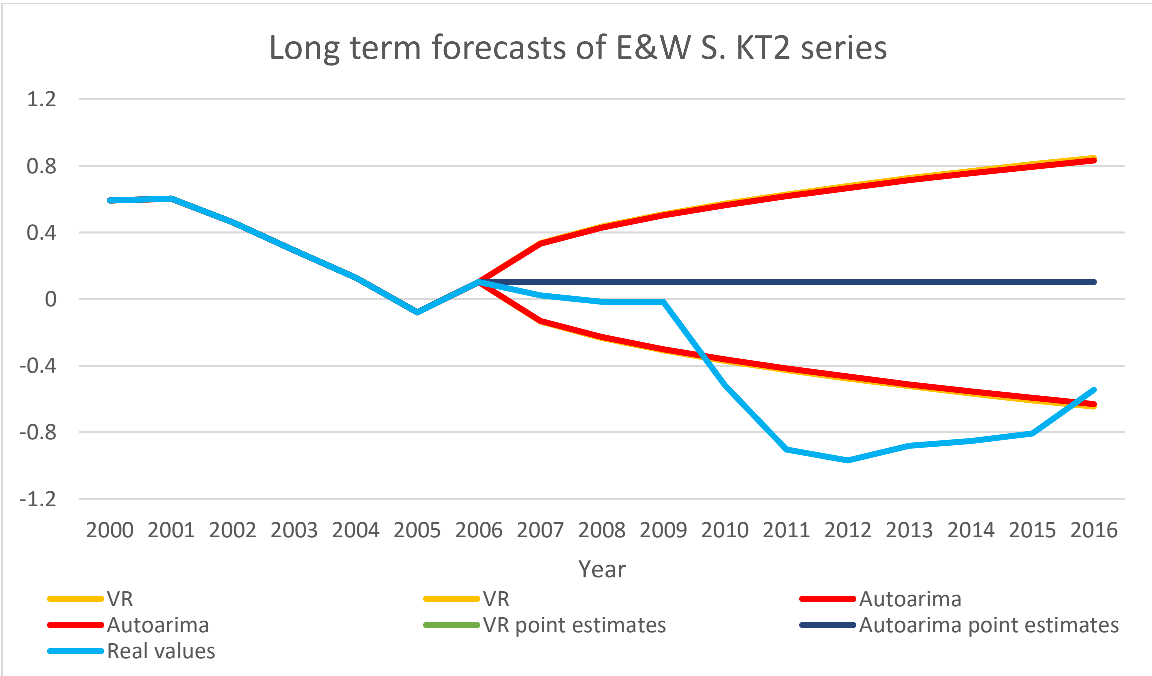


Table 16. Forecast Evaluation Statistics for the series E&W S.KT2

Point Forecasts	"AUTOARIMA"	VR
MSFE	0.57	0.57
MAPE	257.22%	257.22%
MAE	0.65	0.65
Interval Forecasts	**"AUTOARIMA"**	**VR**
Forecast Standard Error (SE)	0.27	0.27

The results for the detailed analysis for the series E&W S. KT1 are shown in Figure 21 and Table 17. The results indicate that both point, and intervals forecasts are better with the "AUTOARIMA" methodology. It is stressed that for this series neither a transformation was necessary, nor any outliers were detected. Hence, any differences in the forecasting performance between the two methods should be attributed solely to differences in the algorithms for the ARIMA model identification and forecasting between the two software products, which in this case are in favour of the "AUTOARIMA" approach (in fact in contrast to what was found in the case of series E&W S.KT4, see Figure 17).

Figure 21. Forecasts and Confidence intervals with both methods for the series E&W S.KT1

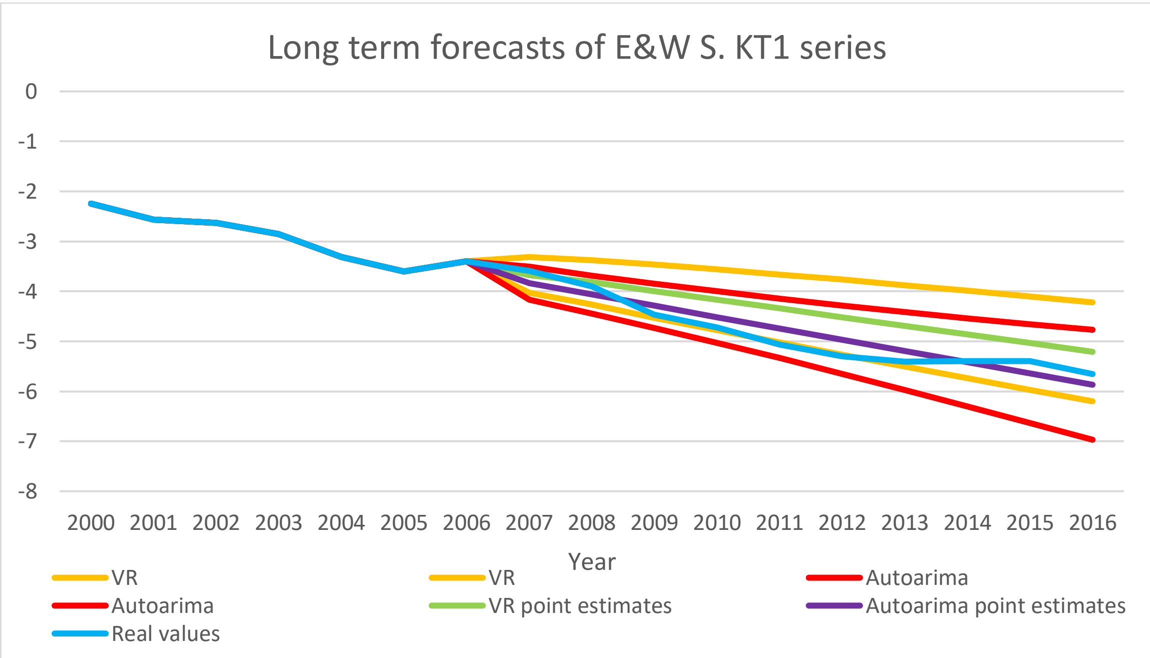


Table 17. Forecast Evaluation Statistics for the series E&W S.KT1

Point Forecasts	"AUTOARIMA"	VR
MSFE	**0.05**	0.28
MAPE	**4.50%**	9.32%
MAE	**0.22**	0.47
Interval Forecasts	**"AUTOARIMA"**	**VR**
Forecast Standard Error (SE)	**0.34**	0.36

From the above analysis it is evident that for those cases where a data transformation and/or outlier adjustments were applied, there is a clear forecasting superiority with the VR method. When no such pre-adjustments were necessary the results are not the same, but they may be either for or against the VR method because the two statistical products obviously use different algorithms for univariate identification and forecasting.

It also interesting to note that averaging across all cases the calculated overall average improvement in interval forecasts is reflected in an approximately 35.4% and 20.4% reduction of the forecast standard error of the VR approach, as compared to the benchmark (RWD), and to the "AUTOARIMA" approach respectively.

4.4.7 The shift towards normality

The non-normal distribution of many time series is another serious issue that needs to be addressed. Nelson and Granger (1979) discovered in their study that data transformations had very little effect on addressing the issue of non-normal distributions in most of the macroeconomic time series they examined.

Table 18 displays the outcomes for the Jarque-Bera statistic for normality, which follows a chi-square distribution with two degrees of freedom. A value in Table 18 marked with an asterisk indicates that the null hypothesis of normality is rejected at a 5% significance level, with the critical value being 5.99. Based on the values in Table 18, it appears that there is a noticeable trend towards normality when moving from the benchmark and "AUTOARIMA" models towards the VR transformation-linearization approach. Milionis and Galanopoulos (2019) obtained similar results in their examination of macroeconomic time series in Greece. Upon closer examination of the results, all the short time series exhibit normality regardless of the approach used.

However, the VR model consistently outperforms the other approaches with better values in terms of the Jarque-Bera statistic. The main discrepancy, however, is noticeable in the long series, where acute non-normality is present in both RWD and "AUTOARIMA" models. This phenomenon is particularly evident is some cases (e.g., in the E&W L.KT1 series). This is a reasonable explanation as outliers that reflect real events, such as world wars, occurred before 1959. Therefore, the use of computational algorithms like maximum likelihood estimation is more justified when using transformed-linearized data.

Table 18. Values of the Jarque –Bera statistic (statistically significant values are indicated with an asterisk)

Time series	RWD	"AUTOARIMA" with further analysis in TSW	VR
E&W L.KT1	4736.*	4736.*	2.840
E&W L.KT2	13.639*	12.23*	7.912 *
E&W L.KT3	0.1057E+05*	4113.*	0.9121
E&W L.KT4	8583.*	2563.*	27.79 *
E&W L.KT5	115.5 *	140.7*	1.747
E&W S.KT1	0.9577	0.9921E-01	0.7816
E&W S.KT2	0.3713E-01	0.3267E-01	0.3267E-01
E&W S.KT3	2.956	3.308	0.3408
E&W S.KT4	2.368	2.584	2.135

4.5 Conclusions – future prospects

In this work we examined the effect of statistical pre-adjustments (data transformation and linearization) on the quality of time series forecasts of mortality rate data. It was found that there is a substantial improvement in interval forecasts which on average are shortened by approximately 35.4% when comparing VR and RWD and 20.4% when comparing VR and "AUTOARIMA". Moreover, there was a less clear improvement in point forecasts. It was also found that for series with unstable, but not functionally dependent on the level, variance the general recommendation of data transformation was not confirmed for the examined case (*E&W L.KT5* series). A case-by-case study for these circumstances seems to be a more sensible approach. Furthermore, it was confirmed that the transformed linearized series satisfy the need for normality to a clearly larger extent as compared to the other alternatives.

The above statistical findings have important implication for the actuarial science. More specifically, the improvement in interval forecasts can significantly affect the Solvency Capital Requirement, and subsequently the Solvency Ratio for a pension fund. Such an improvement might put some pension providers at a competitive advantage as they have less capital locked in their liabilities.

As a further research, we intend to explore more comprehensively the effect of statistical pre-adjustments to the financial impact on Solvency Capital Requirement, under different model structures, actuarial assumptions, and forecast methods. As has been noted previously, the most useful tool for investigating uncertainty over longevity risk is a stochastic mortality projection model. Since, there is a wide choice of such models in the literature, the choice of model can lead to material changes in the best-estimate reserves, while even within a model family there can be major differences (Richards and Currie, 2009). For those models we aim to study the uncertainty over future mortality rates, which is measured as the variance of the mortality forecast values. In particular, we will investigate their respective contributions to the capital requirements for longevity trend risk. Our investigation will be based on the Hatzopoulos and Sagianou (2020) family model structure, which uses time-series methods to project a mortality index. In this respect, we will quantify analytically the respective contributions to capital requirements using VaR calculations. Last but not least, it is apparent that the methodology presented in this work may be used in due course to adjust for the possible effect of the COVID-19 virus on the forecasting of longevity trends.

SUMMARY OF CHAPTER 4

An important risk in the actuarial industry is the longevity risk, therefore the as accurate as possible prediction of mortality rates is very crucial. Such predictions are performed by modelling the mortality rates using mortality models and predicting the future mortality trends. Aiming at possible improvements of such forecasts, we examine the effect of data transformation-"linearization" on the quality of time series forecasts of mortality, using data resulted from mortality models for England-Wales. By time series "linearization" is meant the treatment of causes that disrupt the underlying stochastic process. Results indicate a clear improvement for interval forecasts of mortality as with the transformation of the original time series data and adjustment for outliers the process variance is reduced. However, the result for point forecasts is not as clear. The documented improvement in interval forecasts can significantly affect the Solvency Capital Requirement, rendering some pension providers at a competitive advantage. Furthermore, for series with unstable, but not functionally dependent on the level, variance the general recommendation of data transformation was not confirmed, and a case-by-case treatment seems to be a more sensible approach. It was also confirmed that the transformed-linearized series satisfy better the need for normality as compared to the original series. Moreover, the occurrence of outliers associated with the Covid-19 pandemic would be beneficial to be examined in line with the approach presented in this chapter in future research.

Appendix

A short description of the components of longevity risk

In order to understand and confront the longevity risk, the different potential components of longevity risk can be iterated, and below is a sample list of these components. A risk that is diversifiable may be decreased by expanding the portfolio's size and taking use of the law of big numbers.

In recent decades, the usage of economic models has grown significantly. This has been made possible by technological developments such as improved computing power, new software programs, and novel financial securities. Model Risk in finance refers to the risk associated with utilizing models while making choices. Each model is a condensed representation of reality, but it is never entirely accurate, and failure is always a possibility. It can be challenging to determine whether the forecasting model of choice is accurate. Therefore, it is necessary to set aside some capital in case the model of choice proves to be inaccurate.

Basis Risk comes in a variety of forms. For instance, there is a chance that a change in interest rates will cause the value of a company's or investor's interest-bearing liabilities to alter out of proportion to the value of those assets. This will lead to a loss by raising liabilities and lowering assets. Additionally, in the workplace, models frequently need to be adjusted based on industry or population data rather than the specific portfolio in question. Therefore, it is necessary to set aside some capital in case the mortality trend inferred from a portfolio's data differs from the population used to establish the model.

A negative trend could happen by accident but yet be completely consistent with the selected model, even if the model is correct and there is no basis risk. Some professionals could decide to combine their tolerance for trend risk with one for basis risk.

The risk that occurs by the chance that actuarial calculations are made using estimations that are inaccurate representations of the risk's actual characteristics, is called Parameter Risk.

The risk represented by the difference between actual outcomes and central actuarial estimations based on a random probability would be Process Risk.

Capital must be maintained against the possibility of an uncharacteristically low mortality experience caused by seasonal or environmental fluctuation over the course of a year, such as an exceptionally mild winter and fewer deaths than usual from influenza and other contagious diseases. It should be noted that this Volatility Risk could not be completely diversifiable because a single year of low mortality rates could also signal the beginning of a negative trend.

Market Risk is the possibility of financial loss brought on by shifts in the value of tradable assets. There are a wide range of asset classes (investment rates, bonds, commodities, etc.) and a virtually limitless number of financial products, all of which expose investors to market risk. Diversification cannot completely eliminate market risk, often known as "systematic risk," although it can be hedged (ie offset currency risk). The possibility that a significant natural disaster will cause the market to fall is another illustration of market risk. Political upheaval and changes in interest rates are two more drivers of market risk.

Mis-estimation Risk is the degree of ambiguity surrounding the portfolio's actual mortality rates, which can only be approximated with a degree of confidence corresponding to the size and depth of the data.

The chance of a rapid and temporary increase in the frequency of fatalities is known as Catastrophe Risk. When death benefits are taken into account, it is evident that compensation for catastrophic risk is needed (whereas when life benefits are considered, the gain arises due to higher actual mortality). However, risk transfers can also be taken into consideration as well as the realization of diversity (diversifiable effect).

CHAPTER 5

IMPLICATIONS FOR THE ECONOMETRIC TESTING OF THE HYPOTHESIS OF EFFICIENT MARKETS

5.1 Introduction

The concept of efficient markets was introduced to the academic community in Bachelier's (1900) doctoral thesis. One of his main contributions was the use of the random walk model to describe the prices of financial assets. Bachelier's concepts did not receive immediate recognition within the financial research community, and his contributions remained relatively obscure for a number of decades. Bachelier's ideas started to gain more acknowledgment in the 1950s and 1960s when researchers commenced developing mathematical models to value financial instruments.

The most widely accepted definition of efficient markets has been provided by Fama (1970), stating that a market is efficient when "prices fully reflect all available information". Depending on the available information, there are three forms of market efficiency (Roberts, 1959). More specifically, there is the weak-form efficiency, the semi-strong efficiency, and the strong-form efficiency (for more details, refer to section 1.6.3c). In this chapter, we will focus on the weak-form-market efficiency (WFME), where the available information consists of historical prices of financial assets.

Although the theoretical foundations of market efficiency are laid out in Fama's article (1970), the statistical explanation he presents for market efficiency has faced criticism (LeRoy, 1976; 1989). For this reason, various definitions have been proposed over time (Rubinstein, 1975; Malkiel, 1992; Milionis, 2007) to address these concerns and refine the concept of market efficiency.

Recognizing potential misinterpretations that could arise from the definition Fama proposed in his seminal article in 1970, Fama (1976; 1991) addresses one of the most significant issues, which is the concern of the joint hypothesis of market efficiency with a pricing model (see section 1.6.3c for more details). In 1991, he revised WFME introducing the concept of return predictability (Fama, 1991), acknowledging that in an efficient market, investors cannot achieve excessive returns or profits.

Furthermore, the phenomenon of many researchers frequently using and incorrectly linking the statistical methodology of return predictability testing with the hypothesis of market efficiency has been observed. For this reason, the conditions that must hold in statistical tests (specifically, autocorrelation tests) of return predictability, so that their results are accurately linked to market efficiency are detailed in section 1.6.3c.

More specifically, the time series of financial assets prices (such as stocks, bonds and market indices) are usually non-stationary. However, in the majority of cases, empirical research only examines the first moment (i.e., price levels) and does not test the second moment. This is probably because statistical tests are conducted using asset returns and not asset prices. As assets returns are expressed as differences in the logarithms of the corresponding price relatives it is silently assumed that owing to the logarithmic transformation asset returns are unconditionally variance stationarity. If this is not the case, however, the conditions for testing the weak-form efficiency hypothesis with autocorrelation tests strictly do not hold, as argued in the introductory chapter.

For this reason, in this chapter our primary aim is to test, whether or not, this implicit assumption holds. Additionally, it is also interesting to examine the extent to which the maturity level of a financial market is possibly related to variance non stationarity patterns different that the logarithmic transformation and how this affects market efficiency testing.

The structure of this chapter is the following: The next section provides a literature review of the random walks and the associated tests of market efficiency in developed and emerging financial markets. In section 5.3 the financial markets and the data that will be used in the empirical research and analysis are described. Section 5.4 presents the results of the empirical research, and in section 5.5, the conclusions and proposals for further research are outlined.

5.2 Literature review of the random walks and the associated tests of market efficiency in developed and emerging financial markets

As the problem of joint hypothesis relies on adopting a model that captures past security prices, the most prevalent model in the existing literature is the random walk model. If we consider P_{t-1} to be the price of the security at time $t-1$, μ to be the expected change in the security price or the trend, and ε_t to be a stochastic process referred to as increments (see Campbell et al., 1997), then the random walk model is expressed in the following form:

$$\widetilde{P_t} = P_{t-1} + \mu + \widetilde{\varepsilon_t}, \quad (6)$$

where the use of tildes denotes random variables.

Depending on the conditions that hold for the increments, Campbell et al. (1997) distinguish three cases of random walk models. More specifically, if the increments are independent and identically distributed with a mean of 0 and variance of σ^2, then any nonlinear function of the increments is uncorrelated. In this case, the random walk model 1 arises. As the natural logarithm of prices is widely used in empirical literature (Carol, 2009; Bodie et al., 2020; Hull, 2021; Benninga and Mofkadi, 2022), defining $p_{t-1} = lnP_{t-1}$ equation (6) is transformed as follows:

$$\widetilde{p_t} = p_{t-1} + \mu + \widetilde{\varepsilon_t}$$

Since continuously compounded returns are defined as $\tilde{R}_t = \tilde{p}_t - p_{t-1}$, it follows that returns are i.i.d. when the random walk model 1 holds. In other words, continuously compounded returns can be computed by taking the first differences of the natural logarithms of the prices (Zivot, 2023). The first differences of the natural logarithms of the prices is commonly known as the logarithmic return, and it serves as a standard measure in finance for quantifying the percentage change in the price of an asset over a specific period.

The case of the random walk model 2 arises if the increments are independent but not necessarily identically distributed. This case better corresponds to reality, as the prices of securities do not remain identically distributed over extended periods of time.

Additionally, if it is assumed that the increments are uncorrelated but not independent or identically distributed, then the case of the random walk model 3 emerges. This specific case of random walk is the one that is most frequently examined in empirical literature.

It is worth noting that the random walk model 2 encompasses the random walk model 1 as a special case. Furthermore, the random walk model 3 includes both the random walk models 1 and 2 as specific instances. In contrast, if $Cov(\tilde{\varepsilon}_t, \tilde{\varepsilon}_{t-j}) = 0 \ \forall j \neq 0$ and if $Cov(\tilde{\varepsilon}_t^2, \tilde{\varepsilon}_{t-j}^2) \neq 0$ for some $j \neq 0$, then only the conditions for the random walk model 3 are satisfied, as there exists dependence in the squared increments.

The most widely used tests for the assumption of random walk model 1 are the sequences and reversals tests, as well as the runs test. For a comprehensive analysis of the aforementioned tests, as well as the tests for the other 2 cases of random walks, refer to Campbell et al. (1997). The tests applied for the assumption of random walk model 2 are the filter rules and the technical analysis. The most common tests for the assumption of random walk model 3 are the autocorrelation tests using the so-called portmanteau statistics, namely the Box-Pierce statistic (1970) and Ljung-Box statistic (1978) (for more details, refer to section 1.1.6). Also, section 1.6.3c discusses the conditions for autocorrelation tests to be properly applied. Furthermore, many researchers use variance ratio tests to evaluate the assumption of random walk model 3.

Developed markets, characterized by a higher degree of maturity compared to emerging markets, have been shown to exhibit efficiency more frequently than emerging markets, which predominantly demonstrate inefficiency. This could be attributed to the fact that emerging markets have lower capitalization value, thinner volume of daily trade, fewer listed companies, and generally less transparent operations and regulatory framework compared to developed markets.

More specifically, developed markets exhibit mixed results regarding their efficiency, depending on the time period and the specific financial market under examination by each researcher. Until 1990, it was established for developed markets that the hypothesis of WFME was not rejected (Fama, 1970; Dryden, 1970; Brealey and Mayers, 1988; Fama, 1991). However, there were also studies that were not in favor of the random walk theory in developed markets (e.g. Conrad and Juttner, 1973). In contrast, in emerging markets, the results consistently obtained from the literature indicated that the random walk theory is not suitable to describe the behavior of stock prices and indices (Ayadi and Pyun, 1994; Hamid et al., 2010; Nisar and Hanif, 2012; Mehla and Goyal, 2012; Aggarwal, 2018; Malafeyev et al., 2019).

Regarding studies that simultaneously examine emerging and developed markets, Milionis (1998) concluded that for the Athens Stock Exchange (ASE), which belongs to emerging markets (a detailed presentation of how markets are classified as emerging and developed is provided in the next section), the hypothesis of WFME is rejected. On the contrary, for the Standard and Poor's 500 (SPX) and FTSE 100 (UKX) stock indices, used for New York and London respectively, which belong to developed financial markets, it was demonstrated that they exhibit random walks. This suggests that the hypothesis of WFME is not rejected. More specifically, for the ASE, Alexakis and Xanthakis (1995) demonstrated using daily data that the day of the week effect phenomenon exists. The day of the week effect is a financial market anomaly where certain days of the week may be associated with higher or lower returns, and it is not consistent with the hypothesis of WFME.

Borges (2010) reached the same conclusion for the ASE. A similar conclusion, namely the rejection of the hypothesis of WFME, emerged for the stock indices of Portugal, France, and the United Kingdom. On the contrary, efficient markets were found to be the stock markets of Germany and Spain. From Table 1a, it is evident that the stock indices of Portugal, France, the United Kingdom, Germany, and Spain are classified as developed markets. Dias et al. (2020) found that among the 16 stock indices they examined (7 European, 6 Asian, and 3 American), some of which belong to developed markets and others to emerging markets, the hypothesis of WFME is rejected.

Therefore, over time, it becomes evident that conflicting empirical results may arise regarding testing for WFME across different time periods for the same countries. Thus, as market efficiency is observed to evolve it is natural for researchers' interest to remain unabated in examining market efficiency within the same stock markets. This is done using different sets of data and applying diverse statistical tests, thus enriching the existing extensive literature. The concern, as expressed previously in this chapter, is, whether or not, the statistical methodology for efficiency testing has been properly used thus far in the published literature. Indeed, this is to be extensively examined in what follows in sections 5.4 and 5.5.

5.3 Data

The classification of financial markets is extensively utilized by investors for the purpose of evaluating and making investments in various markets. One of the most widespread market classification systems is the Morgan Stanley Capital International (MSCI) market classification framework. According to this framework (MSCI, 2023), a financial market is classified as developed, emerging, frontier, or standalone based on three criteria.

The first criterion is country's economic development, specifically the sustainability of economic development. A market is classified as developed when the country's Gross National Income per capita is 25% higher than the high-income threshold for 3 consecutive years. The high-income threshold is determined by the World Bank using the Atlas method (see https://datahelpdesk.worldbank.org/). The distinction lacks importance between emerging and frontier markets due to the extensive range of developmental stages present within each of these two categories.

The second criterion is related to companies that need to meet certain minimum requirements regarding liquidity and size (company size, security size, security liquidity). The third criterion pertains to market accessibility by international institutional investors and consists of five sub-criteria, which are: i) opennsess to foreign ownership, ii) ease of capital inflows / outflows, iii) efficiency of operational framework, iv) availability of investment instruments, iv) stability of the institutional framework.

Markets that were previously classified as developed, emerging, or frontier and are now categorized as standalone owing to either a significant downgrade in size and liquidity requirements or market accessibility. Additionally, a second reason for classifying a market as standalone is the fulfillment of all three criteria set by MSCI, which were not met by the specific market in previous years or the market was subject to a specific category of investors.

Tables 1a and 1b present the classification of markets, based on how well a country met the three aforementioned criteria (MSCI, 2023). More specifically, Table 1a displays markets that fulfilled the criteria for developed countries, while Table 1b showcases markets that have been categorized as emerging, frontier, or standalone markets. The abbreviations EMEA and APAC, shown in the first row of Table 1, stand for Europe, Middle East, and Africa (EMEA) and Asia Pacific (APAC), respectively.

Table 1a. MSCI Market Classification – Developed Markets

Americas	**EMEA**	**APAC**
Canada	Austria	Australia
USA	Belgium	Hong Kong
	Denmark	Japan
	Finland	New Zealand
	France	Singapore
	Germany	
	Ireland	
	Israel	
	Italy	
	Netherlands	
	Norway	
	Portugal	
	Spain	
	Sweden	
	Switzerland	
	United Kingdom (UK)	

Table 1b. MSCI Market Classification – Emerging, Frontier and Standalone Markets

Emerging Markets			**Frontier Markets**		
Americas	**EMEA**	**APAC**	**Americas**	**EMEA**	**APAC**
Brazil	Czech Republic	China	-	Bahrain	Bangladesh
Chile	Egypt	India		Benin	Pakistan
Colombia	Greece	Indonesia		Burkina Faso	Sri Lanka
Mexico	Hungary	Korea		Croatia	Vietnam
Peru	Kuwait	Malaysia		Estonia	
	Poland	Philippines		Iceland	
	Qatar	Taiwan		Ivory Coast	

	Saudi Arabia	Thailand		Jordan	
	South Africa			Kazakhstan	
	Turkey			Kenya	
	UAE			Lithuania	
				Mauritius	
				Morocco	
				Nigeria	
				Oman	
				Romania	
				Senegal	
				Serbia	
				Slovenia	
				Tunisia	

Standalone Markets		
Americas	**EMEA**	**APAC**
Argentina	Bosnia and Herzegovina	-
Jamaica	Botswana	
Panama	Bulgaria	
Trinidad and Tobago	Lebanon	
	Malta	
	Palestine	
	Ukraine	
	Zimbabwe	

Markets that belong to the MSCI standalone markets index, despite not being included in the categories of the MSCI emerging markets index and the MSCI frontier markets index, utilize the same methodological criteria concerning the size and liquidity of companies (criterion 2) as the markets that fall under either the MSCI emerging markets index or the MSCI frontier markets index.

Based on the interrelation among emerging, frontier, and standalone markets (criterion 2 mentioned earlier, as well as criterion 3 regarding the availability of investment instruments and the stability of the institutional framework — for more details, see MSCI (2023)), along with the distinct separation suggested by criterion 1 between developed markets and the other market categories, for the purpose of this study onwards, emerging markets, frontier markets, and standalone markets will be collectively referred to as emerging markets.

Our dataset will include twenty-five financial market indices, of which fifteen stock indices belong to advanced markets, while ten belong to emerging markets (see Table 2). The dataset of 25 market indices covers the period from 1987 to 2016 except for PCOMP, JCI, SMI (1988-2016) and MERVAL (1989-2016) using daily closing prices (source: Bloomberg database). The year 1987 was selected as the starting point for this study because it marked the beginning of trading for the Athens Stock Exchange. The entire time period will be divided into six five-year intervals (1987-1991, 1992-1996, 1997-2001, 2002-2006, 2007-2011, 2012-2016) in order to examine how the conclusions regarding the weak-form efficiency of the market change in both developed and emerging markets under different conditions prevailing in the global economic context. For the stock indices PCOMP, JCI, and SMI, the initial time period spans from 1988 to 1991, while for the Stock Index MERVAL, the initial time period is from 1989 to 1991. This is because these four indices started trading either in 1988 or 1989. The other five five-year intervals remain the same for all 25 stock indices.

Table 2. Data

Market Indexes	Region	Country
MSCI World Indexes		
AEX	Europe	Netherlands
ATX	Europe	Austria
CAC	Europe	France
CCMP - NASDAQ	Americas	USA
DAX	Europe	Germany
INDU – Dow Jones Industrial Average	Americas	USA
UKX – FTSE 100	Europe	United Kingdom (London)
HEX	Europe	Finland
HSI	Pacific	China (Hong Kong)
IBEX	Europe	Spain
NKY – Nikkei	Pacific	Japan
OMX – Stockholm 30	Europe	Sweden
SPX – S&P 500	Americas	USA
SMI	Europe	Switzerland
SXXP – STOXX Europe 600	Europe	-
MSCI Emerging Markets		
ASE	Europe	Greece
FBMKLCI	Asia	Malaysia
JCI	Asia	Indonesia
KOSPI	Asia	Korea
PCOMP	Asia	Philippines
SET	Asia	Thailand
TWSE	Asia	Taiwan
MSCI Standalone Markets		
JMSMX	Americas	Jamaica
MERVAL	Americas	Argentina
MSCI Frontier Markets		

CSEALL	Asia	Sri Lanka

5.4 Empirical results and comments

5.4.1 Data transformation

It is important to consider that during the typical process of examining a time series of Stock Index prices, the initial step involves testing whether there's a potential requirement to transform the original data for the purpose of stabilizing variance. This precedes any other tasks like identifying outliers. It is noted that owing to the fact that efficiency tests are usually applied in returns rather than prices it is the logarithmic transformation that is used. While this transformation is the proper one in terms of finance, there is no guarantee that this is also the case statistics-wise.

To determine if the time series exhibit stationarity with respect to variance, two specialized algorithms will be employed to assess whether the transformation of the initial data is necessary. One algorithm is embedded within the JDemetra+ software (see section 1.3.1 for more details), while the other algorithm is the one developed in Chapter 2 (refer also to Milionis and Galanopoulos 2018a). The algorithm from Chapter 2 will be denoted as the M-G algorithm from here onwards.

In more detail, JDemetra+ utilizes the test for variance stationarity embedded in TRAMO, which is based on estimating the parameter λ within the framework of the Box-Cox transformation (see section 1.6.1). This estimation is conducted through the maximum likelihood method. If $\lambda = 1$, the logarithmic transformation is recommended, while if $\lambda = 0$, no transformation is suggested.

Therefore, since JDemetra+ only allows the logarithmic transformation, the statistical methodology developed in section 2.2 will be employed. This approach permits the square root transformation and the transformation with negative inverse, as outlined in section 3.3.3. Additionally, the M-G algorithm is used, as TRAMO has been shown to exhibit bias towards the logarithmic transformation (Milionis and Galanopoulos, 2018a; 2018b; Grudkowska, 2016).

The comparison of decisions resulting from the new proposed methodology and the corresponding decisions from the JDemetra+ routine will be compared. Table 6 displays the outcomes regarding the choice of whether to apply the logarithmic transformation to the original time series data or not.

Table 3. Decision about data transformation

Data transformation				
Market indexes	**M-G**		**JDemetra+**	
MSCI World Indexes	Log	No - Log	Log	No - Log
AEX	1	5	2	4
ATX	0	6	3	3
CAC	0	6	3	3
CCMP	1	5	3	3
DAX	1	5	4	2
DOWJONES	0	6	3	3
FTSE 100	1	5	2	4
HEX	1	5	5	1
HSI	0	6	3	3
IBEX	0	6	2	4
Nikkei	0	6	3	3
OMX	2	4	3	3
S&P 500	1	5	3	3
SMI	0	6	3	3
SXXP	0	6	2	4
Total 1	**8 / 90**	**82 / 90**	**44 / 90**	**46 / 90**
MSCI Emerging Markets	**M-G**		**JDemetra+**	
ASE	3	3	6	0
FBMKLCI	1	5	3	3
JCI	0	6	5	1
KOSPI	0	6	3	3
PCOMP	3	3	5	1
SET	4	2	5	1
TWSE	1	5	4	2
Total 2	**12 / 42**	**30 / 42**	**31 / 42**	**11 / 42**
MSCI Standalone Markets	**M-G**		**JDemetra+**	
JMSMX	2	4	6	0
MERVAL	2	4	4	2
Total 3	**4 / 12**	**8 / 12**	**10 / 12**	**2 /12**
MSCI Frontier Markets	**M-G**		**JDemetra+**	
CSEALL	1	5	5	1
Total 4	**1 / 6**	**5 / 6**	**5 / 6**	**1 / 6**
Total 2 + Total 3 + Total 4	**17 / 60**	**43 / 60**	**46 / 60**	**14 / 60**
Total	**25 / 150**	**125 / 150**	**90 / 150**	**60 / 150**

From the results in Table 3, it is obvious that the application of the logarithmic transformation to the stock market prices, which is typically used in the existing financial literature as mentioned above, is erroneous in most of the cases, as far as the purpose of this research is concerned. This arises from the fact that with the M-G methodology, the logarithmic transformation is suggested in only 25 out of 150 cases (or 1/6 cases) that were examined. In all other cases, specifically in 125 out of 150 cases (or 5/6 cases), either some other transformation (such as the square root or negative inverse) should be applied or no transformation at all. Particularly in developed markets, the logarithmic transformation should only be applied in 8.9% of cases (8 out of 90), and in emerging markets, in 26.7% of cases (17 out of 60).

Furthermore, the bias of existing statistical software towards the logarithmic transformation becomes evident once again. Indeed, out of the 90 cases where JDemetra+ recommends the logarithmic transformation, utilizing the M-G algorithm, it was found that in 27 out of these 90 cases, no transformation at all should be applied. It is further noted that in only 25 of the remaining 63 cases for which JDemetra+ recommends the logarithmic transformation the M-G algorithm recommends the logarithmic transformation as well. This difference results from the fact that with the M-G algorithm instead of the logarithmic transformation, the square root and the negative inverse transformations are recommended in 20 and 18 cases respectively.

5.4.2 The effect of "Linearization"

The existence of a benchmark facilitates the process of comparing different approaches. Consequently, proposing a benchmark serves the purpose of establishing a universal foundation for conducting comparisons. As explained earlier (see section 5.2), the first differences of the natural logarithms of prices is widely used in the field of finance and will serve as benchmark for this chapter in comparison to JDemetra+ software and the M-G methodology. Henceforth, the First Differences of the natural Logarithms will be denoted as FDL.

It is essential to acknowledge that the effects of a transformation are two-fold: direct and indirect. The direct effect is easily noticeable and is tied to the transformation process itself. Meanwhile, the indirect impact centers on how the transformation shapes the identification of outliers. Research has shown that data transformation affects both the quantity and nature of outliers within a time series (Milionis, 2003; 2004; Milionis and Galanopoulos, 2019).

It is remarked that within the JDemetra+ framework, outliers are categorized into three distinct types based on their impact on a time series (refer to section 1.5). Table 4 illustrates the number of detected outliers for each Stock Index across the six examined time periods.

From these results, it is apparent that data transformation significantly influences the pattern of outliers. Particularly in developed markets, the impact of the continuous application of the logarithmic transformation (FDL method) is obvious compared to the M-G method, where 226 fewer outliers are detected. Moreover, applying the logarithmic transformation in nearly half of the cases (see Table 3 for the decision about log-transforming the original price data according to JDemetra+) results in a substantial difference compared to the FDL method (with continuous logarithmic transformation) and a smaller difference compared to the M-G method. In the remaining half of cases, JDemetra+ did not recommend any transformation, a conclusion aligned with the M-G methodology in most of the cases. Therefore, in appropriately transformed data, the observed pattern of detected outliers exhibits distinct variations, a deduction that corresponds to the findings of Milionis (2004) as well.

Conversely, in emerging markets, there are not as pronounced differences among the three methodologies in terms of identifying outliers. The numerous imperfections that emerging markets exhibit make them susceptible to financial events and, consequently, subject to pronounced variability.

Table 4. Detected outliers

Outliers detection			
Market indexes			
MSCI World Indexes	**M-G**	**FDL**	**JDemetra+**
AEX	40	76	41
ATX	44	54	46
CAC	30	46	31
CCMP	37	48	30
DAX	29	47	41
DOWJONES	34	50	41
FTSE 100	33	39	34
HEX	46	43	46
HSI	35	43	34
IBEX	34	55	37
Nikkei	36	38	36
OMX	27	48	33
S&P 500	32	49	32
SMI	37	55	34
SXXP	36	65	45
Total 1	**530**	**756**	**561**
MSCI Emerging Markets	**M-G**	**FDL**	**JDemetra+**
ASE	64	66	66
FBMKLCI	60	69	63
JCI	57	61	58
KOSPI	20	15	13
PCOMP	35	41	39
SET	48	56	54
TWSE	22	15	19
Total 2	**306**	**323**	**312**
MSCI Standalone Markets	**M-G**	**FDL**	**JDemetra+**
JMSMX	98	80	80
MERVAL	31	40	29
Total 3	**129**	**120**	**109**
MSCI Frontier Markets	**M-G**	**FDL**	**JDemetra+**
CSEALL	110	105	98
Total 4	**110**	**105**	**98**
Total 2 + Total 3 + Total 4	**545**	**548**	**519**
Total	**1075**	**1304**	**1080**

5.4.3 Testing of the WFME

Before employing time series for evaluating the WFME, these series generally need certain statistical pre-adjustments, as they are typically unsuitable in their raw form. For example, the original time series data could exhibit non-stationary variance. Consequently, the presence of variance non-stationarity can negatively affect the decision of whether a market is efficient or not, which in turn can have an adverse impact on investors' decisions regarding the investment strategy they will pursue.

To conduct a more in-depth examination regarding the testing of the WFME, three distinct methods will be taken into account. These approaches include: (a) The First Differences of the natural Logarithm (FDL) of the daily prices, a widely used technique in finance as mentioned earlier, and will serve as benchmark, (b) The proposed M-G methodology, which allows the possibility of applying alternative transformations (square root, logarithmic, negative inverse) for stabilizing variance when it is non-stationary. (c) The JDemetra+ software, which includes as an initial step a test concerning the non-stationarity with respect to variance in the original price data. In cases where the logarithmic transformation is suggested, because all time series in our sample are $I(1)$[11], the conclusion regarding market efficiency will be the same as the conclusion of the FDL method. If the logarithmic transformation is not recommended (thus no transformation of the data within the JDemetra+ framework), then if the M-G methodology also does not propose any transformation, the conclusion regarding market efficiency will align with the decision of the M-G method. Finally, if JDemetra+ does not recommend any transformation and the M-G method suggests a transformation (hypothetically, for example, square root transformation), then we separately examine all three methods.

The evaluation and comparison of results regarding market efficiency among the three alternative methods will be based on the possible existence of serial correlation of the transformed and stationary series derived from the original time series of Stock Index prices for all the considered Capital Markets. The LBQ test (see section 1.1.6) which considers autocorrelations for several lags jointly will be used. More specifically, if no statistically significant autocorrelations are detected at any time lag up to the 30th lag,

[11] Space limitations do not allow the presentation of all detailed results, which are available by the author.

then the hypothesis of WFME is not rejected. Conversely, if statistically significant autocorrelations are identified, then the hypothesis of WFME is rejected. The results concerning the decision about WFME are presented in Table 5.

We remind that the assessment of the hypothesis of WFME is examined for each Stock Index in 6 different time periods and employing three different methods. Additionally, the symbol * appearing in the results of Table 5 indicates: i) either that there are no statistically significant autocorrelations in the first 10 (and possibly 20) lags, while there are statistically significant autocorrelations in lags with higher order. In this case, with reservations, the recorded result is that the hypothesis of WFME is not rejected*, ii) or that there are statistically significant autocorrelations in the first 10 (and possibly 20) lags, while there are no statistically significant autocorrelations in lags with higher order. In this case, with reservations, the recorded result is that the hypothesis of WFME is rejected*.

By way of an example, in the sixth row of Table 5, which presents the results for the CAC index, using the M-G method in 4 out of the 6 time periods (second column of Table 5) under examination, the hypothesis of WFME is rejected, in contrast to 2 out of the 6 time periods (third column of Table 5) where it is not rejected. With the FDL method, the notation 6 (2*) (fourth column of Table 5) indicates that the hypothesis of WFME is rejected in all 6 examined time periods, of which 2 times with reservations. Conversely, the hypothesis of weak market efficiency is not rejected in any time period (fifth column of Table 5). Similar interpretations of results with the FDL method arise for the JDemetra+ method.

From the results below, in developed markets we observe that with the M-G method, the hypothesis of WFME is not rejected in 51.1% of the examined cases, while with the established method FDL, the hypothesis is not rejected in 26.7% of the examined cases. Using the JDemetra+ software, where only the logarithmic transformation is allowed, the hypothesis of WFME is not rejected in 44.4% of the examined cases. In the existing literature, the examination of the hypothesis of WFME in developed markets yields mixed results, although in the initial years of studying developed markets, market efficiency had been established (Brealey and Mayers, 1988). This conclusion is primarily supported by the proposed M-G methodology and to some extent by the JDemetra+. In contrast, the benchmark indicates a bias towards rejecting the hypothesis of WFME.

Table 5. Decision about the WFME

WFME						
Market indexes	**M-G**		**FDL**		**JDemetra+**	
MSCI World Indexes	**Rejection**	**Non-Rejection**	**Rejection**	**Non-Rejection**	**Rejection**	**Non-Rejection**
AEX	3 (1*)	3 (1*)	6 (1*)	0	3 (1*)	3 (1*)
ATX	4	2	5	1 (1*)	5	1
CAC	4	2	6 (2*)	0	6 (2*)	0
CCMP	2	4 (2*)	3	3 (2*)	2	4 (2*)
DAX	3 (1*)	3 (1*)	3	3 (3*)	3 (1*)	3 (2*)
DOWJONES	1	5 (2*)	3	3 (2*)	1	5 (3*)
FTSE 100	5	1	6	0	5	1
HEX	4	2 (1*)	4 (1*)	2 (1*)	4 (1*)	2 (1*)
HSI	2 (1*)	4 (2*)	3 (1*)	3 (2*)	2 (1*)	4 (2*)
IBEX	4	2	6	0	4	2
Nikkei	1	5 (1*)	3 (1*)	3 (1*)	3 (1*)	3
OMX	3 (1*)	3 (1*)	4	2 (2*)	3	3 (1*)
S&P 500	1	5 (3*)	3	3 (2*)	2	4 (2*)
SMI	4	2 (1*)	5	1	4	2 (1*)
SXXP	3	3 (2*)	6	0	3	3 (2*)
Total 1	**44 / 90 (4*)**	**46 / 90 (17*)**	**66 / 90 (6*)**	**24 / 90 (16*)**	**50 / 90 (7*)**	**40 / 90 (17*)**
MSCI Emerging Markets	**Rejection**	**Non-Rejection**	**Rejection**	**Non-Rejection**	**Rejection**	**Non-Rejection**
ASE	4	2 (2*)	5	1	5	1
FBMKLCI	5	1 (1*)	5	1 (1*)	5	1 (1*)
JCI	6	0	6	0	6	0
KOSPI	3	3	3	3	3	3
PCOMP	6	0	6	0	6	0
SET	4	2	4	2	4	2
TWSE	2 (1*)	4 (2*)	3	3	2	4 (2*)
Total 2	**30 / 42 (1*)**	**12 / 42 (5*)**	**32 / 42**	**10 / 42 (1*)**	**31 / 42**	**11 /42 (3*)**
MSCI Standalone Markets	**Rejection**	**Non-Rejection**	**Rejection**	**Non-Rejection**	**Rejection**	**Non-Rejection**
JMSMX	5	1	5	1	5	1
MERVAL	3	3 (1*)	4 (1*)	2 (1*)	3	3 (2*)
Total 3	**8 / 12**	**4 / 12 (1*)**	**9 / 12 (1*)**	**3 / 12 (1*)**	**8 / 12**	**4 / 12 (2*)**
MSCI Frontier Markets	**Rejection**	**Non-Rejection**	**Rejection**	**Non-Rejection**	**Rejection**	**Non-Rejection**
CSEALL	6	0	6	0	6	0
Total 4	**6 / 6**	**0 / 6**	**6 / 6**	**0 / 6**	**6 / 6**	**0 / 6**

Total 2 + Total 3 + Total 4	**44 / 60 (1*)**	**16 / 60 (6*)**	**47 / 60 (1*)**	**13 / 60 (2*)**	**45 / 60**	**15 / 60 (5*)**
Total	**88 / 150 (5*)**	**62 / 150 (23*)**	**113 / 150 (7*)**	**37 / 150 (18*)**	**95 / 150 (7*)**	**55 / 150 (22*)**

On the contrary, in emerging markets, the hypothesis of WFME is rejected in 73.3%, 78.3% and 75 % of the examined cases using the M-G method, the benchmark, and JDemetra+ respectively. This finding confirms the existing literature and holds true across all three different methodologies, indicating that the hypothesis of WFME is predominantly rejected in emerging markets.

5.4.4 Cases of different decision about WFME

In the following table (Table 6), all the time periods that yield different conclusions regarding market efficiency are presented for each Stock Index, comparing the three methodologies. More specifically, the 2nd, 3rd, and 4th columns display the time periods with different conclusions between the M-G methodology and the benchmark, the M-G methodology and JDemetra+, and the benchmark and JDemetra+, respectively. A more detailed analysis is provided through Tables 7-11.

Table 6. Time periods with different conclusions about WFME

Market indexes	**Time periods with different decision**		
MSCI World Indexes	**M-G - FDL**	**M-G - JDemetra+**	**FDL - JDemetra+**
AEX	1987-1991, 1997-2001, 2007-2011	-	1987-1991, 1997-2001, 2007-2011
ATX	2002-2006	2002-2006	
CAC	1992-1996, 1997-2001	1992-1996, 1997-2001	-
CCMP	1997-2001, 2007-2011	1997-2001	2007-2011
DAX	-	-	-
DOWJONES	1987-1991, 2002-2006	-	1987-1991, 2002-2006
FTSE 100	1992-1996	-	1992-1996
HEX	1997-2001, 2007-2011	1997-2001, 2007-2011	-
HSI	2007-2011	-	2007-2011

IBEX	2002-2006, 2007-2011	-	2002-2006, 2007-2011
Nikkei	1997-2001, 2012-2016	1997-2001, 2012-2016	-
OMX	2007-2011	-	2007-2011
S&P 500	1987-1991, 1992-1996	1992-1996	1987-1991
SMI	2002-2006	-	2002-2006
SXXP	1992-1996, 2002-2006, 2007-2001	-	1992-1996, 2002-2006, 2007-2011
MSCI Emerging Markets	**M-G - FDL**	**M-G - JDemetra+**	**FDL - JDemetra+**
ASE	2007-2011	2007-2011	-
FBMKLCI	-	-	-
JCI	-	-	-
KOSPI	-	-	-
PCOMP	-	-	-
SET	-	-	-
TWSE	1992-1996	1992-1996	-
MSCI Standalone Markets	**M-G - FDL**	**M-G - JDemetra+**	**FDL - JDemetra+**
JMSMX	-	-	-
MERVAL	2007-2011	-	2007-2011
MSCI Frontier Markets	**M-G - FDL**	**M-G - JDemetra+**	**FDL - JDemetra+**
CSEALL	**-**	**-**	**-**

Emphasis should be placed on cases where, with the logarithmic transformation, the transformed time series of prices are not stationary with respect to variance (see Table 7), and therefore, the analysis using the LBQ test commonly employed by researchers in the existing literature is not valid. The full potential of the proposed M-G methodology becomes evident in the first four cases of Table 7 (rows 3 to 6 of Table 7), where i) with the logarithmic transformation recommended by JDemetra+, non-stationarity with respect to variance persists, and ii) a different conclusion regarding market efficiency is reached. The aforementioned is depicted in detail in Tables 8-9.

Table 7. Cases where the time series are not variance stationary with the log-transformation

Market indexes	Decision about data transformation			
MSCI World Indexes	**Time periods**	**JDemetra+**	**M-G**	**Estimated value of $\widehat{\beta}$ according to M-G method**
ATX	2002-2006	Log	Negative inverse	1.21
CCMP	1997-2001	Log	Negative inverse	1.57
HEX	1997-2001	Log	Negative inverse	1.50
S&P 500	1992-1996	Log	Negative inverse	1.76
ATX	1987-1991	Log	Negative inverse	2.03
MSCI Emerging Markets	**Time periods**	**JDemetra+**	**M-G**	**Estimated value of $\widehat{\beta}$ according to M-G method**
JCI	1988-1991	Log	Negative inverse	1.27
JCI	1992-1996	Log	Negative inverse	1.31
JMSMX	1987-1991	Log	Negative inverse	1.57
JMSMX	2012-2016	Log	Negative inverse	1.46
MERVAL	1987-1991	Log	Negative inverse	1.49
MERVAL	1992-1996	Log	Negative inverse	2.14

More specifically, in Table 8, one case is depicted where, with the M-G methodology, the market is efficient, while with the methodologies of JDemetra+ and the benchmark, the market is not efficient. Similar conclusions, albeit with reservations, are drawn for the Stock Index S&P 500 during the time period 1992-1996.

Table 8. Decision about data transformation and WFME for the series ATX 2002-2006

<table>
<tr><td colspan="7">Data transformation</td></tr>
<tr><td colspan="7">JDemetra+</td></tr>
<tr><td>Time series in</td><td colspan="6">Levels</td></tr>
<tr><td>Decision</td><td colspan="6">Logs</td></tr>
<tr><td colspan="7">M-G</td></tr>
<tr><td>Time series in</td><td colspan="6">Levels</td></tr>
<tr><td>$\hat{\beta}$</td><td colspan="6">1.21</td></tr>
<tr><td>Decision</td><td colspan="6">Negative inverse</td></tr>
<tr><td>Time series in</td><td colspan="6">Negative inverse</td></tr>
<tr><td>$\hat{\beta}$</td><td colspan="6">Not statistical significant</td></tr>
<tr><td colspan="7">Variance stationarity</td></tr>
<tr><td>Time series in</td><td colspan="2">Levels</td><td colspan="2">Logs</td><td colspan="2">Negative inverse</td></tr>
<tr><td>Decision</td><td colspan="2">No</td><td colspan="2">No</td><td colspan="2">Yes</td></tr>
<tr><td>Time series in</td><td colspan="3">First differences of negative inverse (M-G)</td><td colspan="3">First differences of log (FDL and JDemetra+)</td></tr>
<tr><td>Lag</td><td>LBQ</td><td colspan="2">p-value</td><td colspan="2">LBQ</td><td>p-value</td></tr>
<tr><td>10</td><td>6,586</td><td colspan="2">0,764</td><td colspan="2">24,523</td><td>0,006</td></tr>
<tr><td>20</td><td>23,091</td><td colspan="2">0,284</td><td colspan="2">38,170</td><td>0,008</td></tr>
<tr><td>30</td><td>35,716</td><td colspan="2">0,218</td><td colspan="2">47,462</td><td>0,022</td></tr>
<tr><td>WFME</td><td colspan="3">Not Rejected</td><td colspan="3">Rejected</td></tr>
</table>

On the contrary, in Table 9, with the M-G methodology, the hypothesis of WFME is rejected, while the benchmark and JDemetra+ suggest not rejecting the hypothesis of WFME. Similar findings, though with reservations, are reached regarding the CCMP Stock Index between the years 1997 and 2001.

Table 9. Decision about data transformation and WFME for the series HEX 1997-2001

Data transformation				
JDemetra+				
Time series in	Levels			
Decision	Logs			
M-G				
Time series in	Levels			
$\hat{\beta}$	1.50			
Decision	Negative inverse			
Time series in	Negative inverse			
$\hat{\beta}$	Not statistical significant			
Variance stationarity				
Time series in	Levels	Logs		Negative inverse
Decision	No	No		Yes
Time series in	First differences of negative inverse (M-G)		First differences of log (FDL and JDemetra+)	
Lag	LBQ	p-value	LBQ	p-value
10	19,357	0,036	7,970	0,632
20	37,519	0,010	19,327	0,501
30	48,249	0,019	38,315	0,142
WFME	Rejected		Not Rejected	

In the other cases listed in Table 7, even though variance doesn't become stationary with the logarithmic transformation, the same conclusion regarding market efficiency arises. More specifically, with all three methodologies, the hypothesis of WFME is rejected, and the markets are not efficient. The result obtained for the specific stock indices (JCI, JMSMX, MERVAL) confirms the existing literature, indicating that in emerging markets there are so many imperfections that regardless of the approach employed, the conclusion about market efficiency remains unchanged (i.e., markets are not efficient). However, the result for the ATX Stock Index (the time period 1987-1991), which now belongs to developed markets, should not surprise us. As noted by Milionis and

Papanagiotou (2008), the ATX Stock Index exhibited many similarities with the ASE Stock Index, which belongs to emerging markets, and in most cases, the hypothesis of WFME is rejected for the ASE index.

Another point that requires particular attention when analyzing the results from Table 7 is that in all cases where the time series are non-stationary in terms of variance, even when the logarithmic transformation is applied, the estimated value of $\hat{\beta}$ using the proposed M-G methodology (see section 2.2) is greater than 1.21 (and can go up to 2.14). This suggests the preference for the negative inverse transformation, which converts the time series to being stationary in terms of variance. This specific finding not only indicates that when the negative inverse transformation needs to be applied, the logarithmic transformation might not make the time series stationary in terms of variance, thus rendering the conclusions about WFME from the benchmark invalid, but also provides a future direction for research into the M-G methodology. More specifically, a future aim is to conduct Monte Carlo simulations in order to accurately determine critical values and intervals for the parameter $\hat{\beta}$ within the framework of the M-G methodology. These simulations will indicate when each specific data transformation should be applied or when no transformation should be applied at all.

Another case that highlights the superiority of the proposed M-G methodology compared to JDemetra+ and the benchmark is presented in Table 10. More specifically, despite the initial time series of prices for the TWSE index during the period 1992-1996 being stationary with respect to variance, JDemetra+ incorrectly suggests applying the logarithmic transformation, once again indicating the bias of the statistical software towards using the logarithmic transformation. However, beyond this, the application of unnecessary and erroneous transformation affects the conclusion regarding market efficiency. In more detail, according to the M-G methodology, it is determined that the markets are not efficient, thus investors have the opportunity to profit from the market by implementing proper investment strategies (similar findings were noted in the cases of the market indices ATX during the time period 2002-2006 (as shown in Table 8) and S&P 500 during the time period 1992-1996). In contrast, using the benchmark and JDemetra+, investors would miss out on this opportunity to gain profits, as they indicate that the markets are efficient.

Table 10. Decision about data transformation and WFME for the series TWSE 1992-1996

Data transformation				
JDemetra+				
Time series in	Levels			
Decision	Logs			
M-G				
Time series in	Levels			
$\hat{\beta}$	Not statistical significant			
Decision	None			
Variance stationarity				
Time series in	Levels		Logs	
Decision	Yes		Yes	
Time series in	First differences of levels (M-G)		First Differences of log (FDL and JDemetra+)	
Lag	LBQ	p-value	LBQ	p-value
10	19,372	0,036	12,352	0,262
20	32,896	0,035	29,010	0,088
30	40,194	0,101	39,260	0,120
WFME	Rejected*		Not Rejected	

However, in all the other cases presented in Table 6, the M-G methodology indicates that the markets are efficient, the benchmark suggests that the markets are not efficient, while JDemetra+ sometimes concludes that the markets are efficient, and other times that the markets are not efficient. All these cases, except for the instances of the market indices HEX during the time period 1997-2001 (presented in Table 9) and CCMP during the same time period, where the further analysis with the LBQ test should not have been pursued at all, as the conditions for stationarity in the second moment are not met, are provided in Table 11.

Table 11. Time periods with different decision about WFME when the time series are variance stationary

Market indexes	Time periods with different decision			
MSCI World Indexes	**Time period**	**M-G**	**FDL**	**JDemetra+**
AEX	1987-1991	Not rejected	Rejected	Not rejected
AEX	1997-2001	Not rejected	Rejected	Not rejected
AEX	2007-2011	Not rejected	Rejected	Not rejected
CAC	1992-1996	Not rejected	Rejected	Rejected
CAC	1997-2001	Not rejected	Rejected	Rejected
CCMP	2007-2011	Not rejected	Rejected	Not rejected
DOWJONES	1987-1991	Not rejected	Rejected	Not rejected
DOWJONES	2002-2006	Not rejected	Rejected	Not rejected
FTSE100	1992-1996	Not rejected	Rejected	Not rejected
HEX	2007-2011	Not rejected	Rejected	Rejected
HSI	2007-2011	Not rejected	Rejected	Not rejected
IBEX	2002-2006	Not rejected	Rejected	Not rejected
IBEX	2007-2011	Not rejected	Rejected	Not rejected
NIKKEI	1997-2001	Not rejected	Rejected	Rejected
NIKKEI	2012-2016	Not rejected	Rejected	Rejected
OMX	2007-2011	Not rejected	Rejected	Not rejected
S&P 500	1987-1991	Not rejected	Rejected	Not rejected
SMI	2002-2006	Not rejected	Rejected	Not rejected
SXXP	1992-1996	Not rejected	Rejected	Not rejected
SXXP	2002-2006	Not rejected	Rejected	Not rejected
SXXP	2007-2011	Not rejected	Rejected	Not rejected
MSCI Emerging Markets				
ASE	2007-2011	Not rejected	Rejected	Rejected
MSCI Standalone Markets		-	-	-
MERVAL	2007-2011	Not rejected	Rejected	Not rejected

From the results above, the bias of the benchmark towards rejecting the hypothesis of WFME is apparent. As 21 out of the 23 cases presented in Table 11 pertain to developed markets, this contradicts the existing literature where mixed results in recent years, or non-rejection of the hypothesis of WFME in the initial years of studying developed markets (Brealey and Mayers, 1988), are often observed. Conversely, this finding further supports the outcomes of the M-G methodology.

5.5 Conclusions-future prospects

Although the concept of market efficiency has been studied for several decades, researchers naturally use extensively asset returns (i.e., the first differences of the natural logarithms of prices) as a standard measure, without checking whether these time series are stationary with respect to variance. According to this research, it becomes evident that utilizing the logarithmic transformation for Stock Index prices is inadequate for stabilizing the variance in most of the cases that were examined. This has the consequence that the use of autocorrelation tests, which are among the most common practices for testing market efficiency, may not be valid. This conclusion was also reached in the present study, where it was indicated that the first differences of the natural logarithms of prices often did not satisfy the condition of variance stationarity. As a result, in some cases the hypothesis of WFME was: i) falsely rejected, and ii) falsely not rejected. These cases emerged in the study of developed markets. In these cases, the same conclusion was reached by JDemetra+, which recommended the logarithmic transformation. In contrast, with the M-G method, which allows for the application of various transformations, it was found that: i) the time series of first differences of prices became stationary with respect to variance, and ii) the proper and valid use of autocorrelation tests led to opposite conclusions regarding market efficiency compared to the first differences of the natural logarithms of prices and JDemetra+.

Furthermore, in emerging markets, it was found that in all cases where the rejection of the hypothesis of WFME emerged as a conclusion by JDemetra+, the same conclusion reached by the M-G method in almost all these cases. This is due to the fact that emerging markets have so many “imperfections” that, regardless of the approach used, in the majority of the existing literature, the hypothesis of WFME is rejected.

Moreover, the divergent conclusions that emerged, primarily in developed markets, using all three methods (first differences of the natural logarithms of prices, JDemetra+, M-G), are caused by the fact that the M-G methodology's findings documented that with the existing methodologies in terms of statistical testing of market efficiency there is a profound bias towards rejecting market efficiency. In that sense, the M-G methodology provides support to the persisting view of Eugene Fama that markets are efficient in the weak sense.

In general, analyzing the entire examined dataset, it was found that the first differences of the natural logarithms of prices exhibits bias in rejecting the weak-form efficiency in developed markets. This contradicts the existing literature, where depending on the time period and the Stock Index under examination, mixed results arise regarding the hypothesis of WFME (although in the initial years of studying developed markets, market efficiency had been established). This finding about mixed results was confirmed by the proposed M-G methodology and partially by JDemetra+. In emerging markets, using all three different methods that were examined, the long-standing conclusion in the existing literature was reaffirmed, i.e., that markets are generally not efficient.

Furthermore, the bias of JDemetra+ regarding the application of the logarithmic transformation was confirmed once again, which consequently affects the detection and number of outliers. The difference in the number of detected outliers (which is of lesser importance as compared to the statistical testing of efficiency) is evident when the logarithmic transformation is continuously applied (using the FDL method), and the results are compared with the M-G method, which suggests the logarithmic transformation only a few times.

A field of future research is the pursuit of determining critical values for the parameter $\hat{\beta}$ within the framework of the M-G methodology, using Monte Carlo simulations. In this way, the value of $\hat{\beta}$ will indicate which transformation (or none) should be applied to satisfy the stationarity criterion in the second moment. Additionally, a future direction is to employ GARCH-type models to capture the conditional heteroskedasticity, which is different from variance non-stationarity, frequently observed in the time series of financial assets prices.

SUMMARY OF CHAPTER 5

In the context of market efficiency, decades of research have very often involved the first differences of natural logarithms of prices (i.e., asset returns), but without adequately verifying the stationarity of these time series in terms of variance. Based on this study, it arises that the application of the logarithmic transformation for stock market prices is in many cases inappropriate for stabilizing the variance of the corresponding price relatives. Consequently, utilizing autocorrelation tests, commonly employed to test market efficiency, may be invalid. Indeed, in this chapter it was documented that autocorrelation tests cannot be legitimately employed as a statistical tool for testing market efficiency in 83.3% of the examined cases. Moreover, it was determined that the usage of the first differences of natural logarithms of prices introduces bias in rejecting the weak-form efficiency in developed markets. This finding was corroborated by the proposed M-G methodology and partially by JDemetra+ software.

It is also remarkable that following the statistical testing based on the M-G methodology the conclusion about the hypothesis of WFME is the opposite of that based on the established methodology in 27.7% of the cases for the developed Capital Markets.

In the case of emerging markets, employing all three methods that were investigated reaffirmed the enduring conclusion in existing literature – namely, that markets are generally not efficient, even though for several cases again the established methodology could not be legitimately employed. For these markets the extent of inefficiency is such that it is consistently detected regardless of the method employed. Thus, there is evidence confirming that the maturity level of the financial market affects market efficiency. Furthermore, the recurring bias of JDemetra+ in applying the logarithmic transformation was confirmed once again, thereby impacting the identification and quantity of outliers. The discrepancy in the number of detected outliers becomes evident when the logarithmic transformation is consistently applied (using the FDL method), and the results are compared to the M-G method, which suggests the logarithmic transformation only on a few cases.

EXTENDED SUMMARY/PROSPECTS

A stochastic time series is a type of time series where the future values can only be determined in terms of a probability distribution. If this probability distribution is constant over time, then the time series is said to be stationary. A less strict condition for stationarity requires that at least the level and variance of the time series be constant over time.

Researchers employ various tests to check for non-stationarity in the level of a time series, but more often than not they neglect to investigate non-stationarity in its variance when conducting applied research. In fact, regarding time series variance, the primary research emphasis is on modeling autoregressive conditional heteroscedasticity, typically using a variety of ARCH-GARCH models.

It is essential to differentiate between two key concepts: variance non-stationarity, often referred to as heteroscedasticity, and conditional heteroscedasticity. Heteroscedasticity implies a functional relation between the variance of a series which is non-stationary in its level and its mean level. This entails non-stationarity in the variance, and the variance is neither conditionally nor unconditionally constant. Consequently, the process is non-homogeneously non-stationary in the sense of Box and Jenkins and cannot be made stationary by simply differencing it. To address variance non-stationarity, one approach is to apply power transformations, such as the well-known Box and Cox transformations. On the other hand, conditional heteroscedasticity, often described using ARCH or GARCH models, signifies that while the conditional variance varies over time, the unconditional variance remains constant. As a result, the series is stationary in the second moment. In the present Ph.D. thesis, the focus is on the series with non-constant variance both conditionally and unconditionally, covering to a certain extent a gap in that area, as the existing research work is relatively scanty.

Indeed, even though it is crucial to deal with non-constant variance in time series modeling, there is a shortage of comprehensive theoretical research on its detection and correction. Moreover, in practical applications, the treatment of non-stationary variance is not only insufficient, as the choice of a specific transformation is often arbitrary, but also, as is documented in Chapter 2, occasionally biased towards over-rejection of the null hypothesis of unconditionally constant variance.

The aim of this Ph.D. thesis is to present a formal econometric approach that not only identifies non-stationary variance and suggests appropriate transformations for correction, but also is robust to the specific partitioning of a time series, which is a necessary step for conducting the test, and the possible presence of outliers. The importance of employing this approach in the fields of macroeconomics, actuarial science and finance is extensively examined and supported in Chapters 3, 4 and 5 respectively.

In Chapter 2, the Ph.D. thesis elaborates on the theoretical foundation of the proposed methodology, which focuses on statistical testing for the existence and the identification of the character of time-varying second moment in its dependence on a non-constant mean level in time series. This approach represents an enhancement over current methods as it combines detection, correction, and robustness.

It is important that during the typical process of analyzing a time series, the initial step is to assess whether the original data requires transformation to make the variance stable. This assessment is carried out before any other actions, including building the univariate ARIMA model, performing seasonal adjustments, etc. Consequently, it is clear that the results of these subsequent actions are influenced by the choice made regarding data transformation.

This is of value in its own right as it leads to the improvement of univariate time series modelling. Furthermore, empirical evidence is presented using real data (Greece's balance of payments and prices of consumer goods and services), as well as simulated data, from which it comes out that an existing test, specifically the widely used algorithm of TSW software, occasionally yields biased results. TSW stands for TRAMO-SEATS for Windows, a Windows version of the DOS programmes TRAMO and SEATS of Gómez and Maravall. TRAMO stands for "Time series Regression with ARIMA noise Missing observations and Outliers" and SEATS stands for "Signal Extraction in ARIMA Time Series". TSW routines are also incorporated in other widely used econometric software. Notably, TSW offers only two alternatives for data transformation: log-transformation or no transformation.

Indeed, by utilizing simulated data, it was feasible to identify one of the possible origins of this bias. More specifically, with simulated homogeneously non-stationary processes, it became evident that the bias of TSW depends on the initial conditions.

Moreover, drawing from the empirical evidence presented, it is argued that the type of data transformation and the entailed correction for variance–non stationarity is also crucial for the detection of outliers and the seasonal adjustment of the original time series. In addition, the empirical results provide evidence of an improved forecasting performance by the proper use of a data transformation, a result that is backed by additional empirical evidence in Chapters 3 and 4.

It was also determined that the consequences of erroneously transforming a time series, which is originally variance stationary, do exist, but are comparatively less severe than the consequences of erroneously not transforming an originally variance non-stationary series. This is a conclusion that holds substantial practical importance.

In Chapter 3, the proposed methodology is applied to macroeconomic time series. As a matter of fact, real-world economic time series are not immediately suitable for forecasting purposes, and they require some statistical preparation and pre-adjustment. This is because raw data time series can often exhibit non-stationary variance. Furthermore, very often there exist causes that disrupt the underlying stochastic process, such as the existence of outliers and calendar effects. Their treatment is referred to as «linearization».

The impact of either data transformation or linearization on the accuracy of forecasts, including both point forecasts and confidence interval forecasts, has not been thoroughly explored until now. This study investigates their impact on univariate forecasting, analyzing each one individually and in combination, employing twenty of the most important time series related to the Greek economy.

For data transformation, two algorithms were utilized, namely those of TSW and Milionis Galanopoulos (M-G henceforth). The M-G algorithm is the statistical methodology developed in Chapter 2, which allows data transformation not only through the logarithmic transformation but also through the square root and negative inverse.

Empirical findings show a significant improvement in forecasts' confidence intervals, but no substantial improvement in point forecasts. Furthermore, there is also evidence that the overall forecasting performance using the M-G data transformation procedure is somewhat better than the one using the data transformation procedure of TSW.

Moreover, the combined transformation-linearization procedure improves substantially the non-normality problem encountered in many macroeconomic time series.

One area where the enhanced forecast confidence intervals documented in Chapter 3 could be particularly advantageous is the field of actuarial science, particularly in dealing with longevity risk. This risk arises from the uncertainty surrounding the future trend of mortality rates of pensioners, as advancements in science and medicine make the prediction of mortality rates a difficult task. To address this issue, one approach is to employ mortality models to forecast the trend of mortality rates and its associated uncertainty in the future. This uncertainty is directly associated with forecast confidence intervals.

In Chapter 4, aiming at possible improvements of such forecasts, it is examined how statistical pre-adjustments (data transformation and linearization) affect the accuracy of time series forecasts of mortality. This analysis was conducted using data derived from mortality models for England-Wales.

To conduct a detailed investigation about statistical forecasting, three distinct methods were considered. These methods were the following: (a) The random walk with drift model, which is widely used in actuarial research due to its simplicity and served as the benchmark. (b) The "AUTOARIMA" command within the programming software "R" for automatic model selection and forecasting, as demonstrated in the published work of Hatzopoulos and Sagianou. (c) ARIMA models implemented after statistical pre-adjustments, which implies Variance Reduction and will be referred to as "VR" forecasts.

The empirical findings demonstrate a clear improvement in interval forecasts which on average are shortened by approximately 35.4% when comparing VR and RWD and 20.4% when comparing VR and "AUTOARIMA". However, the conclusion for point forecasts is not as clear. The documented improvement in interval forecasts can have a substantial impact on the Solvency Capital Requirement, rendering some pension providers at a competitive advantage. The Solvency Capital Requirement covers all the possible risks that an insurance company may encounter.

Furthermore, for series with unstable but not functionally dependent on the level variance, the conventional recommendation in the literature for transformation of the original data, did not receive confirmation. A case-by-case treatment seems to be a more sensible approach. It was also validated that the series subjected to both transformation and linearization satisfy better the need for normality as compared to the other alternatives.

The above statistical findings have important implication for the actuarial science. More specifically, the improvement in interval forecasts can significantly affect the Solvency Capital Requirement, and subsequently the Solvency Ratio for a pension fund. Such an improvement might put some pension providers at a competitive advantage as they have less capital locked in their liabilities.

As further research, the intention is to explore more comprehensively the effect of statistical pre-adjustments to the financial impact on Solvency Capital Requirement, under different model structures, actuarial assumptions, and forecast methods. As it is noted in Chapter 4, the most useful tool for investigating uncertainty over longevity risk is a stochastic mortality projection model. Since, there is a wide choice of such models in the literature, the choice of model can lead to material changes in the best-estimate reserves, while even within a model family there can be major differences. For those models it is aimed to study the uncertainty over future mortality rates, which is measured as the variance of the mortality forecast values. By this method, it will be quantified analytically the respective contributions to capital requirements using Value at Risk calculations. Last but not least, the overall methodology presented in Chapter 4 may be used also in due course to adjust for the possible effect of the COVID-19 virus on the forecasting of longevity trends.

In the last chapter (Chapter 5), the developed methodology contributes to the improvement of the framework of econometric assumptions and tests in finance, aiming to determine the rejection or non-rejection of the hypothesis of weak-form market efficiency. Weak-form market efficiency (WFME) deals with situations where the available information pertains solely to historical prices of financial assets.

While the concept of market efficiency has been a subject of study for several decades, researchers naturally use extensively asset returns, which are essentially the first differences of the natural logarithms of prices, as a standard measure, without checking whether these time series are stationary with respect to variance. In a risk-unadjusted framework, it is crucial to emphasize that these tests are valid only when the series of logarithmic prices exhibits variance stationarity. If this condition is not satisfied, the significance testing of autocorrelation coefficients in the widely used autocorrelation tests becomes invalid. It is noted that since efficiency tests are typically conducted using returns rather than prices it is the logarithmic transformation that is employed. While this transformation is the proper one in terms of finance, there is no guarantee that this is also the case statistics-wise.

The classification of financial markets is widely employed by investors to assess and make investment decisions across various markets. The examined dataset consists of twenty-five financial market indices, comprising fifteen stock indices from developed markets and ten from emerging markets.

To conduct a more thorough investigation into the testing of the WFME, three different approaches were considered. These methods include: (a) The First Differences of the natural Logarithm (FDL) of the daily prices, a commonly used technique in finance, and served as a benchmark, (b) The proposed M-G methodology, which allows for the application of alternative transformations to stabilize variance when it is non-stationary (c) The JDemetra+ software, which includes as an initial step, a test for non-stationarity with respect to variance in the original price data. In more detail, JDemetra+ utilizes the test for variance stationarity embedded in TRAMO. However, as JDemetra+ only permits the logarithmic transformation, the M-G statistical methodology developed in Chapter 2 was employed. Additionally, the M-G algorithm is chosen because TRAMO has been shown to exhibit bias towards the logarithmic transformation.

According to this study, it is clear that using the logarithmic transformation for Stock Index prices is inadequate for stabilizing the variance in most of the cases that were examined. Consequently, utilizing autocorrelation tests, commonly employed to test market efficiency, may be invalid. Indeed, in this chapter it was documented that autocorrelation tests cannot be legitimately employed as a statistical tool for testing market efficiency in 83.3% of the examined cases.

Furthermore, it was established that employing the first differences of natural logarithms of prices can introduce a bias in rejecting weak-form efficiency in developed markets. This result was supported by the proposed M-G methodology and, to some extent, by the JDemetra+ software. It is noteworthy that, after conducting the statistical testing using the M-G methodology, the conclusion regarding the WFME hypothesis contradicts that obtained through the established methodology in 27.7% of the cases for developed Capital Markets.

In the case of emerging markets, the use of all three methods that were examined reconfirmed the long-standing consensus in existing literature – that is, markets in these regions are typically not efficient. However, it is worth noting that for several cases, the established methodology could not be validly applied. In these markets, the degree of inefficiency is pronounced and consistently detected irrespective of the method used. This confirms the existing research that supports that the maturity level of a financial market affects its efficiency.

Additionally, the previously noted bias of JDemetra+ in its use of the logarithmic transformation was once again verified. Consequently, this bias affects the detection and number of outliers. The difference in the number of detected outliers (which is of lesser importance as compared to the statistical testing of efficiency) becomes apparent when the logarithmic transformation is continuously applied (using the FDL method), and the results are compared with the M-G method, which suggests the application of the logarithmic transformation only a few times.

Regarding future prospects of this research, beyond what has already been mentioned, such a field is the pursuit of determining critical values for the parameter $\hat{\beta}$ within the framework of the M-G methodology, using Monte Carlo simulations. In this way, the value of $\hat{\beta}$ will (more formally) indicate which transformation (or none) should be applied to satisfy the stationarity criterion in the second moment. Additionally, a future direction is to study variance non-stationarity in time series in conjunction with GARCH-type models which capture the autoregressive conditional heteroskedasticity. It is worrisome that researchers often proceed to GARCH models without assurances about stationarity in the second moment.

Finally, it is remarked that the conclusions related to variance non-stationarity of stock index prices and, in sequence, to autocorrelation tests in stock index returns, are unavoidably linked not only to the particular type of assets (stock indices), but also to the particular sampling time interval that was used (five years). Hence, it should not necessarily be taken for granted that these conclusions are identically valid for other types of assets and, more importantly, for the particular asset, but over longer time intervals. There is little doubt that this is yet another field for further future research.

ΕΚΤΕΝΗΣ ΠΕΡΙΛΗΨΗ

Μία στοχαστική χρονοσειρά είναι ένας τύπος χρονοσειράς όπου οι μελλοντικές τιμές μπορούν να καθοριστούν μόνο σε όρους μιας κατανομής πιθανότητας. Αν αυτή η κατανομή πιθανότητας είναι σταθερή με την πάροδο του χρόνου, τότε η χρονοσειρά λέγεται στάσιμη. Μια λιγότερο αυστηρή συνθήκη για τη στασιμότητα απαιτεί τουλάχιστον το επίπεδο και η διακύμανση της χρονοσειράς να είναι σταθερά με την πάροδο του χρόνου.

Οι ερευνητές χρησιμοποιούν διάφορους ελέγχους για να εξετάσουν τη μη-στασιμότητα στο επίπεδο μιας χρονοσειράς, αλλά συχνά παραμελούν να εξετάσουν τη μη-στασιμότητα στη διακύμανσή της, κατά την διεξαγωγή εφαρμοσμένης έρευνας. Πράγματι, όσον αφορά τη διακύμανση της χρονοσειράς, η πρωταρχική ερευνητική έμφαση δίνεται στη μοντελοποίηση της αυτοπαλίνδρομης δεσμευμένης (υπό συνθήκη) ετεροσκεδαστικότητας, συνήθως χρησιμοποιώντας διάφορα ARCH-GARCH τύπου μοντέλα.

Είναι ουσιώδες να διακρίνουμε ανάμεσα σε δύο βασικές έννοιες: την μη-στασιμότητα της διακύμανσης, που συχνά αναφέρεται και ως ετεροσκεδαστικότητα, και της δεσμευμένης ετεροσκεδαστικότητας. Η ετεροσκεδαστικότητα συνεπάγεται μια συναρτησιακή σχέση μεταξύ της διακύμανσης μίας σειράς, που είναι μη-στάσιμη στο επίπεδό της και του μέσου επιπέδου της. Αυτό έχει ως αποτέλεσμα τη μη-στασιμότητα στη διακύμανση, και η διακύμανση είναι μη-σταθερή τόσο υπό συνθήκη, όσο και χωρίς συνθήκη (μη-δεσμευμένη). Συνεπώς, η διαδικασία είναι μη-ομοιογενώς μη-στάσιμη στο πλαίσιο των Box και Jenkins και δεν μπορεί να γίνει στάσιμη απλώς παίρνοντας τις διαφορές. Για να αντιμετωπιστεί η μη-στασιμότητα της διακύμανσης, μια προσέγγιση είναι η εφαρμογή μετασχηματισμών, όπως είναι οι ευρέως γνωστοί μετασχηματισμοί των Box και Cox. Από την άλλη πλευρά, η δεσμευμένη ετεροσκεδαστικότητα, που περιγράφεται συχνά χρησιμοποιώντας μοντέλα του τύπου ARCH-GARCH, υπονοεί ότι ενώ η δεσμευμένη διακύμανση μεταβάλλεται με τον χρόνο, η μη-δεσμευμένη διακύμανση παραμένει σταθερή. Ως αποτέλεσμα, η σειρά είναι στάσιμη στη δεύτερη ροπή. Στην παρούσα διδακτορική διατριβή, επικεντρωνόμαστε σε σειρές με μη-σταθερή διακύμανση τόσο υπό συνθήκη όσο και χωρίς συνθήκη, καλύπτοντας μέχρι ένα βαθμό ένα κενό στην ευρύτερη περιοχή, καθώς η υπάρχουσα ερευνητική βιβλιογραφία είναι σχετικά περιορισμένη.

Πράγματι, αν και είναι ουσιώδες να αντιμετωπιστεί η μη-σταθερή διακύμανση στη μοντελοποίηση χρονοσειρών, υπάρχει έλλειψη συγκροτημένης θεωρητικής έρευνας σχετικά με την ανίχνευση και τη διόρθωσή της. Επιπλέον, στις πρακτικές εφαρμογές, η αντιμετώπιση της μη-στασιμότητας της διακύμανσης δεν είναι μόνο ανεπαρκής, καθώς η επιλογή ενός συγκεκριμένου μετασχηματισμού συχνά είναι αυθαίρετη, αλλά επίσης, όπως τεκμαίρεται στο Κεφάλαιο 2, περιστασιακά είναι μεροληπτική ως προς την υπερβολική απόρριψη της μηδενικής υπόθεσης της μη-δεσμευμένης σταθερής διακύμανσης.

Ο σκοπός της παρούσας διδακτορικής διατριβής είναι να παρουσιάσει μια επίσημη οικονομετρική προσέγγιση που όχι μόνο ανιχνεύει τη μη-στασιμότητα της διακύμανσης και προτείνει κατάλληλους μετασχηματισμούς για τη διόρθωσή της, αλλά επίσης είναι ανθεκτική αφενός ως προς τις διάφορες διαμερίσεις μιας χρονοσειράς, που αποτελεί απαραίτητο βήμα για τη διεξαγωγή του ελέγχου, και αφετέρου την πιθανή ύπαρξη ακραίων τιμών. Η σημαντικότητα της χρήσης αυτής της προσέγγισης στους τομείς της μακροοικονομίας, της αναλογιστικής επιστήμης και της χρηματοοικονομικής εξετάζεται λεπτομερώς και υποστηρίζεται στα Κεφάλαια 3, 4 και 5 αντίστοιχα.

Στο Κεφάλαιο 2 της διδακτορικής διατριβής αναπτύσσονται τα θεωρητικά θεμέλια της προτεινόμενης μεθοδολογίας, η οποία επικεντρώνεται στον στατιστικό έλεγχο για την ύπαρξη και τον προσδιορισμό του χαρακτήρα της χρονικά μεταβαλλόμενης δεύτερης ροπής ως προς την εξάρτησή της από ένα μη-σταθερό μέσο επίπεδο στη χρονοσειρά. Αυτή η προσέγγιση αντιπροσωπεύει μια βελτίωση έναντι των υφιστάμενων μεθόδων καθώς συνδυάζει την ανίχνευση, τη διόρθωση και την ανθεκτικότητα.

Είναι σημαντικό ότι κατά την τυπική διαδικασία ανάλυσης μιας χρονοσειράς, το αρχικό βήμα είναι να αξιολογηθεί εάν τα αρχικά δεδομένα απαιτούν μετασχηματισμό για να γίνει η διακύμανση σταθερή. Αυτή η αξιολόγηση πραγματοποιείται πριν από οποιαδήποτε άλλη ενέργεια, συμπεριλαμβανομένης της κατασκευής του μονομεταβλητού ARIMA υποδείγματος, της πραγματοποίησης εποχικών διορθώσεων, κλπ. Ως αποτέλεσμα, είναι σαφές ότι τα αποτελέσματα αυτών των επόμενων ενεργειών επηρεάζονται από την επιλογή που γίνεται σχετικά με τον μετασχηματισμό των δεδομένων.

Αυτό έχει αξία καθαυτή, καθώς οδηγεί στη βελτίωση της μονομεταβλητής μοντελοποίησης χρονοσειρών. Επιπλέον, παρουσιάζονται εμπειρικά ευρήματα χρησιμοποιώντας πραγματικά δεδομένα της Ελλάδας (ισοζύγιο πληρωμών και τιμές καταναλωτικών αγαθών και υπηρεσιών), καθώς και προσομοιωμένα δεδομένα, από τα οποία προκύπτει ότι ένας υφιστάμενος έλεγχος, ειδικότερα ο ευρέως χρησιμοποιούμενος αλγόριθμος του λογισμικού TSW, μερικές φορές παράγει μεροληπτικά αποτελέσματα. Το TSW αντιπροσωπεύει το TRAMO-SEATS για Windows, μια έκδοση για Windows των DOS προγραμμάτων TRAMO και SEATS των Gómez και Maravall. Το TRAMO αντιπροσωπεύει τη φράση "Time series Regression with ARIMA noise Missing observations and Outliers" και το SEATS αντιπροσωπεύει τη φράση "Signal Extraction in ARIMA Time Series" και ως αλγόριθμος είναι ενσωματωμένος και σε άλλα ευρέως χρησιμοποιούμενα στατιστικά λογισμικά. Επίσης, το TSW προσφέρει μόνο δύο εναλλακτικές για τον μετασχηματισμό των δεδομένων: τον λογαριθμικό μετασχηματισμό ή κανένα μετασχηματισμό.

Πράγματι, μέσω της χρήσης προσομοιωμένων δεδομένων, ήταν δυνατό να εντοπιστεί μία από τις πιθανές αιτίες αυτής της μεροληψίας. Συγκεκριμένα, με προσομοιωμένες ομοιογενώς μη-στάσιμες διαδικασίες, έγινε εμφανές ότι η μεροληψία του TSW εξαρτάται από τις αρχικές συνθήκες.

Επιπλέον, βασιζόμενοι στα παρουσιαζόμενα εμπειρικά ευρήματα, αιτιολογείται ότι ο τύπος του μετασχηματισμού των δεδομένων και η συνοδευόμενη διόρθωση της μη-στασιμότητας της διακύμανσης είναι επίσης ουσιώδης για την ανίχνευση των ακραίων τιμών και την εποχιακή διόρθωση της αρχικής χρονοσειράς. Επιπλέον, τα εμπειρικά αποτελέσματα παρέχουν ενδείξεις βελτιωμένης προβλεπτικής ικανότητας μέσω της σωστής χρήσης του μετασχηματισμού των δεδομένων, ένα αποτέλεσμα που υποστηρίζεται από επιπλέον εμπειρικά ευρήματα στα Κεφάλαια 3 και 4.

Περαιτέρω, τεκμαίρεται ότι οι συνέπειες ενός εσφαλμένου μετασχηματισμού μιας χρονοσειράς, η οποία είναι αρχικά στάσιμη ως προς τη διακύμανση, ναι μεν υφίστανται, αλλά είναι συγκριτικά λιγότερο σοβαρές από τις συνέπειες της μη χρήσης μετασχηματισμού για μία χρονοσειρά που είναι αρχικά μη-στάσιμη ως προς τη διακύμανση. Αυτό είναι ένα συμπέρασμα που έχει ιδιαίτερη πρακτική σημασία.

Στο Κεφάλαιο 3, η προτεινόμενη μεθοδολογία εφαρμόζεται σε μακροοικονομικές χρονοσειρές. Σημειώνεται ότι οι χρονοσειρές οικονομικών δεδομένων από τον πραγματικό κόσμο δεν είναι αμέσως κατάλληλες για προβλεπτικούς σκοπούς και απαιτούν ορισμένη στατιστική προετοιμασία και προ-επεξεργασία. Αυτό συμβαίνει επειδή οι χρονοσειρές των αρχικών δεδομένων συχνά εμφανίζουν μη-στασιμότητα στη διακύμανση, επιπλέον δε πολύ συχνά υπάρχουν αιτίες που διαταράσσουν την υποκείμενη στοχαστική διαδικασία, όπως η ύπαρξη ακραίων τιμών και οι ημερολογιακές επιδράσεις. Ο τρόπος αντιμετώπισής τους αναφέρεται ως «γραμμικοποίηση».

Η επίδραση είτε του μετασχηματισμού των δεδομένων είτε της γραμμικοποίησης στην ακρίβεια των προβλέψεων, συμπεριλαμβανομένων τόσο των σημειακών προβλέψεων όσο και των διαστημάτων εμπιστοσύνης των προβλέψεων, δεν έχει εξεταστεί εκτενώς μέχρι στιγμής. Αυτή η μελέτη εξετάζει την επίδρασή τους στη μονομεταβλητή πρόβλεψη, αναλύοντάς κάθε μία επίδραση τόσο ξεχωριστά όσο και σε συνδυασμό, χρησιμοποιώντας είκοσι από τις πιο σημαντικές χρονοσειρές που σχετίζονται με την ελληνική οικονομία.

Για τον μετασχηματισμό των δεδομένων, χρησιμοποιήθηκαν δύο αλγόριθμοι. Πιο συγκεκριμένα, χρησιμοποιήθηκαν ο αλγόριθμος του TSW και ο αλγόριθμος των Milionis-Galanopoulos (M-G εφεξής). Ο αλγόριθμος M-G είναι η στατιστική μεθοδολογία που αναπτύχθηκε στο Κεφάλαιο 2, η οποία επιτρέπει τον μετασχηματισμό των δεδομένων όχι μόνο μέσω του λογαριθμικού μετασχηματισμού, αλλά και μέσω της τετραγωνικής ρίζας και του αρνητικού αντίστροφου.

Τα εμπειρικά ευρήματα δείχνουν σημαντική βελτίωση στα διαστήματα εμπιστοσύνης των προβλέψεων, αλλά καμία ουσιώδη βελτίωση στις σημειακές προβλέψεις. Επιπλέον, υπάρχουν ενδείξεις ότι η συνολική προβλεπτική ικανότητα χρησιμοποιώντας την διαδικασία M-G για τον μετασχηματισμό των δεδομένων είναι κάπως καλύτερη από εκείνη που χρησιμοποιεί τη διαδικασία μετασχηματισμού δεδομένων του TSW. Περαιτέρω, η συνδυαστική διαδικασία μετασχηματισμού-γραμμικοποίησης βελτιώνει σημαντικά το πρόβλημα της μη-κανονικής κατανομής που εντοπίζεται σε πολλές μακροοικονομικές χρονοσειρές.

Μία ερευνητική περιοχή, όπου τα βελτιωμένα διαστήματα εμπιστοσύνης των προβλέψεων που καταγράφονται στο Κεφάλαιο 3, θα μπορούσαν να είναι ιδιαίτερα χρήσιμα είναι ο τομέας της αναλογιστικής επιστήμης, ιδιαίτερα όσον αφορά τον κίνδυνο μακροζωίας. Αυτός ο κίνδυνος προκύπτει από την αβεβαιότητα που περιβάλλει τη μελλοντική τάση των ποσοστών θνησιμότητας των συνταξιούχων, καθώς οι βελτιώσεις στην επιστήμη και την ιατρική καθιστούν δύσκολη την πρόβλεψη των ποσοστών θνησιμότητας. Για να αντιμετωπιστεί αυτό το πρόβλημα, μια προσέγγιση είναι η χρησιμοποίηση μοντέλων θνησιμότητας για να προβλέπουν την τάση των ποσοστών θνησιμότητας και τη συνοδευόμενη αβεβαιότητα στο μέλλον. Αυτή η αβεβαιότητα συνδέεται άμεσα με τα διαστήματα εμπιστοσύνης των προβλέψεων.

Στο Κεφάλαιο 4, με στόχο τις πιθανές βελτιώσεις σε τέτοιες προβλέψεις, εξετάζεται πώς η στατιστική προ-επεξεργασία (μετασχηματισμός δεδομένων και γραμμικοποίηση) επηρεάζουν την ακρίβεια των προβλέψεων χρονοσειρών θνησιμότητας. Αυτή η ανάλυση διεξήχθη χρησιμοποιώντας δεδομένα που προέρχονται από μοντέλα θνησιμότητας για την Αγγλία-Ουαλία.

Για να διεξαχθεί μια λεπτομερής έρευνα σχετικά με τις στατιστικές προβλέψεις, λήφθηκαν υπόψη τρεις διαφορετικοί μέθοδοι. Αυτές οι μέθοδοι ήταν οι εξής: (α) Το υπόδειγμα του τυχαίου περιπάτου με τάση, το οποίο χρησιμοποιείται ευρέως στην αναλογιστική έρευνα λόγω της απλότητάς του, και χρησίμευσε ως σημείο αναφοράς. (β) Η εντολή "AUTOARIMA" του λογισμικού "R" για την αυτόματη επιλογή μοντέλου και πρόβλεψη, όπως παρουσιάστηκε σε δημοσιευμένη εργασία των Hatzopoulos και Sagianou. (γ) Τα υποδείγματα ARIMA που εφαρμόζονται μετά από στατιστική προ-επεξεργασία, που συνεπάγεται μείωση της διακύμανσης (Variance Reduction) και θα αναφέρονται ως "VR" προβλέψεις.

Τα εμπειρικά ευρήματα υποδεικνύουν μια σαφή βελτίωση στις προβλέψεις διαστήματος, τα οποία κατά μέσο όρο μειώνονται περίπου 35.4% όταν συγκρίνεται το VR με το RWD και περίπου 20.4% όταν συγκρίνεται το VR με το "AUTOARIMA". Ωστόσο, το συμπέρασμα για τις σημειακές προβλέψεις δεν είναι τόσο σαφές. Η τεκμηριωμένη βελτίωση στις προβλέψεις διαστήματος μπορεί να έχει σημαντική επίδραση στην Κεφαλαιακή Απαίτηση Φερεγγυότητας (Solvency Capital Requirement), καθιστώντας ορισμένους παρόχους συντάξεων σε συγκριτικό πλεονέκτημα. Η Κεφαλαιακή Απαίτηση Φερεγγυότητας καλύπτει όλους τους πιθανούς κινδύνους που μπορεί να αντιμετωπίσει μια ασφαλιστική εταιρεία.

Επιπλέον, για χρονοσειρές με ασταθή, αλλά όχι συναρτησιακά εξαρτώμενη από το επίπεδο διακύμανση, η συμβατική στη βιβλιογραφία σύσταση του μετασχηματισμού δεδομένων δεν επιβεβαιώθηκε, και μια πρόταση μελέτης κατά περίπτωση φαίνεται να είναι μια πιο λογική προσέγγιση. Επίσης, επιβεβαιώθηκε ότι οι σειρές που υποβλήθηκαν τόσο σε μετασχηματισμό δεδομένων όσο και σε γραμμικοποίηση ικανοποιούν καλύτερα την ανάγκη για τη κανονική κατανομή σε σύγκριση με τις άλλες εναλλακτικές.

Τα παραπάνω στατιστικά ευρήματα έχουν σημαντικές συνέπειες για την αναλογιστική επιστήμη. Συγκεκριμένα, η βελτίωση στις προβλέψεις διαστήματος μπορεί να επηρεάσει σημαντικά την Κεφαλαιακή Απαίτηση Φερεγγυότητας, και συνεπώς, τον Δείκτη Φερεγγυότητας (Solvency Ratio) για μία εταιρεία διαχείρισης συντάξεων (pension fund). Μια τέτοια βελτίωση μπορεί να θέσει ορισμένους παρόχους συντάξεων σε ανταγωνιστικό πλεονέκτημα, καθώς έχουν αποθηκεύσει λιγότερο κεφάλαιο για τις υποχρεώσεις τους.

Ως περαιτέρω έρευνα, θα μπορούσε να μελετηθεί πιο διεξοδικά ο αντίκτυπος της στατιστικής προ-επεξεργασίας στην χρηματοοικονομική επίδραση της Κεφαλαιακής Απαίτησης Φερεγγυότητας, υπό διάφορες δομές υποδειγμάτων, αναλογιστικές υποθέσεις και μεθόδους πρόβλεψης. Όπως αναφέρεται στο Κεφάλαιο 4, το πιο χρήσιμο εργαλείο για την εξέταση της αβεβαιότητας ως προς τον κίνδυνο μακροζωίας είναι ένα στοχαστικό μοντέλο προβολής θνησιμότητας. Δεδομένου ότι υπάρχει ευρεία επιλογή τέτοιων μοντέλων στη βιβλιογραφία, η επιλογή του μοντέλου μπορεί να οδηγήσει σε σημαντικές αλλαγές για τις καλύτερες-εκτιμήσεις αποθεμάτων, καθώς ακόμα και εντός μιας οικογένειας μοντέλων, μπορεί να υπάρχουν σημαντικές διαφορές. Για αυτά τα μοντέλα, ο σκοπός είναι η μελέτη της αβεβαιότητας σχετικά με τα μελλοντικά ποσοστά θνησιμότητας, που μετριούνται ως η διακύμανση των τιμών πρόβλεψης της θνησιμότητας. Με τον τρόπο αυτό θα ήταν δυνατή η ποσοτική εκτίμηση στις αντίστοιχες συνεισφορές στις κεφαλαιακές απαιτήσεις χρησιμοποιώντας υπολογισμούς της Αξίας σε Κίνδυνο (Value at Risk). Αξίζει ακόμα να σημειωθεί ότι η όλη μεθοδολογία που αναπτύσσεται στο Κεφάλαιο 4 μπορεί να χρησιμοποιηθεί στο μέλλον και για την προσαρμογή της δυνητικής επίδρασης της πανδημίας COVID-19 στην πρόβλεψη των τάσεων μακροζωίας.

Στο τελευταίο κεφάλαιο (Κεφάλαιο 5), η προτεινόμενη μεθοδολογία συμβάλει στη βελτίωση του πλαισίου των οικονομετρικών υποθέσεων και ελέγχων στη χρηματοοικονομική, με στόχο τον προσδιορισμό της απόρριψης ή μη-απόρριψης της υπόθεσης της αποτελεσματικότητας της αγοράς υπό την μορφή ασθενούς ισχύος. Αυτή η μορφή αποτελεσματικότητας της αγοράς (Weak-form market efficiency - WFME) συνδέεται με καταστάσεις όπου οι διαθέσιμες πληροφορίες αφορούν αποκλειστικά τις ιστορικές τιμές των χρηματοοικονομικών περιουσιακών στοιχείων.

Ενώ η έννοια της αποτελεσματικότητας των αγορών έχει αποτελέσει αντικείμενο μελέτης για αρκετές δεκαετίες, οι ερευνητές χρησιμοποιούν εκτενώς τις αποδόσεις των περιουσιακών στοιχείων, οι οποίες ουσιαστικά είναι οι πρώτες διαφορές των φυσικών λογαρίθμων των τιμών, ως ένα πρότυπο μέτρο, χωρίς να ελέγχουν εάν αυτές οι χρονοσειρές είναι στάσιμες ως προς τη διακύμανση. Σε ένα πλαίσιο όπου δεν λαμβάνεται υπόψη ο κίνδυνος, είναι σημαντικό να σημειωθεί ότι αυτοί οι έλεγχοι είναι έγκυροι μόνο όταν οι σειρές των λογαρίθμων των τιμών εμφανίζουν στάσιμη διακύμανση. Εάν αυτή η συνθήκη δεν πληρείται, ο έλεγχος σημαντικότητας των συντελεστών αυτοσυσχέτισης στους ευρέως χρησιμοποιούμενους ελέγχους αυτοσυσχέτισης δεν είναι έγκυρος. Σημειώνεται ότι καθώς οι έλεγχοι αποτελεσματικότητας πραγματοποιούνται συνήθως χρησιμοποιώντας αποδόσεις και όχι τιμές, χρησιμοποιείται ο λογαριθμικός μετασχηματισμός. Ενώ αυτός ο μετασχηματισμός είναι ο κατάλληλος από πλευράς χρηματοοικονομικής, δεν υπάρχει καμία εγγύηση ότι αυτό είναι επίσης ορθό από πλευράς στατιστικής.

Η κατηγοριοποίηση των χρηματοοικονομικών αγορών χρησιμοποιείται ευρέως από τους επενδυτές για να αξιολογήσουν και να λάβουν επενδυτικές αποφάσεις σε διάφορες αγορές. Το εξεταζόμενο σύνολο δεδομένων αποτελείται από είκοσι πέντε δείκτες χρηματοοικονομικών αγορών, συμπεριλαμβανομένων δεκαπέντε μετοχικών δεικτών από ανεπτυγμένες αγορές και δέκα μετοχικών δεικτών από αναδυόμενες αγορές.

Για να διεξαγάγουμε μια πιο λεπτομερή έρευνα σχετικά με τον έλεγχο της WFME, τρεις διαφορετικές προσεγγίσεις εξετάστηκαν. Αυτές οι μέθοδοι περιλαμβάνουν: (α) Τις Πρώτες Διαφορές του φυσικού Λογαρίθμου (First Difference of the natural Logarithm-FDL) των ημερήσιων τιμών, μια κοινή χρησιμοποιούμενη τεχνική στη χρηματοοικονομική, που χρησίμευσαν ως σημείο αναφοράς, (β) Την προτεινόμενη μεθοδολογία M-G, η οποία επιτρέπει την εφαρμογή εναλλακτικών μετασχηματισμών για τη σταθεροποίηση της διακύμανσης όταν αυτή είναι μη-στάσιμη, (γ) Το λογισμικό JDemetra+, το οποίο περιλαμβάνει ως αρχικό βήμα, έναν έλεγχο για την μη-στασιμότητα ως προς τη διακύμανση στα αρχικά δεδομένα τιμών. Πιο συγκεκριμένα, το JDemetra+ χρησιμοποιεί τον έλεγχο για τη στασιμότητα της διακύμανσης που περιλαμβάνεται στο TRAMO. Ωστόσο, καθώς το JDemetra+ επιτρέπει μόνο τον λογαριθμικό μετασχηματισμό, χρησιμοποιήθηκε η στατιστική μεθοδολογία M-G που αναπτύχθηκε στο Κεφάλαιο 2. Επιπλέον, ο αλγόριθμος M-G επιλέγεται επειδή έχει αποδειχθεί ότι το TRAMO εμφανίζει μεροληψία προς τον λογαριθμικό μετασχηματισμό.

Σύμφωνα με την παρούσα μελέτη, είναι σαφές ότι η χρήση του λογαριθμικού μετασχηματισμού για τιμές Χρηματιστηριακών Δεικτών είναι ανεπαρκής για τη σταθεροποίηση της διακύμανσης στις περισσότερες από τις περιπτώσεις που εξετάστηκαν. Ως αποτέλεσμα, η χρήση των ελέγχων αυτοσυσχέτισης, που συνήθως χρησιμοποιούνται για τον έλεγχο της αποτελεσματικότητας της αγοράς, μπορεί να μην είναι έγκυρη. Πράγματι, σε αυτό το κεφάλαιο καταγράφηκε ότι οι έλεγχοι αυτοσυσχέτισης δεν μπορούν να χρησιμοποιηθούν έγκυρα ως στατιστικό εργαλείο για τον έλεγχο της αποτελεσματικότητας της αγοράς στο 83.3% των εξεταζόμενων περιπτώσεων.

Επιπλέον, διαπιστώθηκε ότι η χρήση των πρώτων διαφορών των φυσικών λογαρίθμων των τιμών μπορεί να εισάγει μεροληψία ως προς την απόρριψη της αποτελεσματικότητας ασθενούς ισχύος στις ανεπτυγμένες αγορές. Αυτό το αποτέλεσμα υποστηρίχθηκε από την προτεινόμενη μεθοδολογία M-G και, έως ένα βαθμό, από το λογισμικό JDemetra+. Είναι σημαντικό να σημειωθεί ότι μετά τη διενέργεια των στατιστικών ελέγχων χρησιμοποιώντας τη μεθοδολογία M-G, τα συμπεράσματα σχετικά με την WFME υπόθεση ήρθαν σε αντίθεση με αυτά που προέκυψαν μέσω της εδραιωμένης μεθοδολογίας στο 27.7% των περιπτώσεων για τις ανεπτυγμένες Αγορές Κεφαλαίου.

Στην περίπτωση των αναδυόμενων αγορών, η χρήση και των τριών μεθόδων που εξετάστηκαν, επανεπιβεβαίωσε την πολυετή ομοφωνία στην υπάρχουσα βιβλιογραφία, δηλαδή, ότι οι αναδυόμενες αγορές δεν είναι συνήθως αποτελεσματικές. Ωστόσο, αξίζει να σημειώσουμε ότι σε αρκετές περιπτώσεις, η εδραιωμένη μεθοδολογία δεν μπορούσε να εφαρμοστεί έγκυρα. Σε αυτές τις αγορές, ο βαθμός της μη-αποτελεσματικότητας είναι τόσο έντονος ώστε να ανιχνεύεται ανεξαρτήτως της μεθόδου που χρησιμοποιείται. Αυτό επιβεβαιώνει τις υπάρχουσες ερευνητικές εργασίες που υποστηρίζουν ότι το επίπεδο ωριμότητας μιας χρηματοοικονομικής αγοράς επηρεάζει την αποτελεσματικότητα της.

Επιπλέον, η προηγουμένως αναφερθείσα μεροληψία του JDemetra+ στη χρήση του λογαριθμικού μετασχηματισμού επαληθεύτηκε για μια ακόμη φορά. Ως εκ τούτου, αυτή η μεροληψία επηρεάζει την ανίχνευση και τον αριθμό των ακραίων τιμών. Η διαφορά στον αριθμό των ανιχνεύσιμων ακραίων τιμών (που είναι λιγότερο σημαντική σε σχέση με το στατιστικό έλεγχο της αποτελεσματικότητας) γίνεται εμφανής όταν ο λογαριθμικός μετασχηματισμός εφαρμόζεται συνεχώς (με τη χρήση της μεθόδου FDL), και τα αποτελέσματα συγκρίνονται με τη μέθοδο M-G, η οποία υποδεικνύει την εφαρμογή του λογαριθμικού μετασχηματισμού μόνο σε λίγες περιπτώσεις.

Όσον αφορά μελλοντικές προοπτικές της παρούσας έρευνας, πέραν των ήδη αναφερθέντων, ένα τέτοιο πεδίο είναι η επιδίωξη καθορισμού κρίσιμων τιμών για την παράμετρο $\hat{\beta}$ στο πλαίσιο της μεθοδολογίας M-G, χρησιμοποιώντας προσομοιώσεις Monte Carlo. Με αυτό τον τρόπο, η τιμή του $\hat{\beta}$ θα υποδεικνύει (και τυπικά) ποιος μετασχηματισμός (ή κανένας) πρέπει να εφαρμοστεί για να ικανοποιείται το κριτήριο της στασιμότητας στη δεύτερη ροπή. Επιπλέον, μια μελλοντική κατεύθυνση είναι η μελέτη της μη-στασιμότητας ως προς τη διακύμανση σε χρονοσειρές σε συνδυασμό με μοντέλα τύπου GARCH που περιγράφουν την αυτοπαλίνδρομη δεσμευμένη ετεροσκεδαστικότητα. Είναι ανησυχητικό το γεγονός ότι οι ερευνητές προκρίνουν την χρησιμοποίηση GARCH υποδειγμάτων χωρίς να υπάρχει επιβεβαίωση σχετικά με τη στασιμότητα στη δεύτερη ροπή.

Τέλος, σημειώνεται ότι τα συμπεράσματα που σχετίζονται με τη μη-στασιμότητα της διακύμανσης των τιμών των μετοχικών δεικτών και, κατά συνέπεια, με τους ελέγχους αυτοσυσχέτισης στις αποδόσεις των μετοχικών δεικτών, συνδέονται αναπόφευκτα όχι μόνο με το συγκεκριμένο είδος περιουσιακών στοιχείων (δείκτες μετοχών), αλλά και με το συγκεκριμένο χρονικό διάστημα δειγματοληψίας που χρησιμοποιήθηκε (πέντε έτη). Επομένως, δεν πρέπει απαραίτητα να θεωρείται δεδομένο ότι αυτά τα συμπεράσματα ισχύουν πανομοιότυπα για άλλα είδη περιουσιακών στοιχείων, και ακόμη σημαντικότερα, για τα συγκεκριμένα μεν περιουσιακά στοιχεία, αλλά σε μεγαλύτερα χρονικά διαστήματα. Δεν υπάρχει αμφιβολία ότι αυτό αποτελεί ένα ακόμη πεδίο για μελλοντική έρευνα.

REFERENCES

Aggarwal, D. (2018). Random walk model and asymmetric effect in Korean composite stock price index. *Afro-Asian Journal of Finance and Accounting*, *8*, 85-104.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Alexakis, P., and Xanthakis, M. (1995). Day of the week effect on the Greek stock market. *Applied Financial Economics*, *5*, 43–50.

Ayadi, O. F., and Pyun, C. S. (1994). An application of variance ratio test to the Korean securities market. *Journal of Banking and Finance*, *18*, 643–658.

Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques De L Ecole Normale Superieure*, *17*, 21–86.

Bell, W. R., and Monsell, B. C. (1991). Using principal components in time series modeling and forecasting of age-specific mortality rates. *Proceedings of the American Statistical Association*, Social Statistics Section, 154-159.

Benninga, S. and Mofkadi, T. (2022). Financial Modeling, 5th edition, The MIT Press.

Bodie, Z., Kane, A. and Marcus, A. (2020). Investments, 12th edition, McGraw Hill.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*, 307–327.

Borges, M. R. (2010). Efficient market hypothesis in European stock markets. *European Journal of Finance*, *16*, 711–726.

Box, G. E. P., and Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *26*, 211–243.

Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control, Revised Edition, Holden-Day, San Francisco.

Box, G. E. P., and Pierce, D. A. (1970). Distribution of residual autocorrelations in Autoregressive-Integrated Moving Average Time Series models. *Journal of the American Statistical Association*, *65*, 1509-1526.

Box, G. E. P., and Tiao, G. C. (1975). Intervention Analysis with Applications to Economic and Environmental Problems. *Journal of the American Statistical Association*, *70*, 70–79.

Brealey, R. A. and Myers, S. C. (1988). Principles of Corporate Finance, 3rd edition, McGraw Hill.

Brillinger, D. R. (1986). The Natural Variability of Vital Rates and Associated Statistics. *Biometrics*, *42*, 693-734.

Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, *31*, 373–393.

Burman, J. P. (1980). Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society*, *143*, 321-337.

Cairns, A. J. G., Blake, D., and Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, *73*, 687–718.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G., Epstein, D., and Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, *48*, 355–367.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). The econometrics of financial markets, Princeton University Press.

Caporello, G., and Maravall, A. (2004). Program TSW: Revised reference manual. *Bank of Spain Working Paper 0408*.

Carol, A. (2009). Market Risk Analysis, Wiley.

Chen, C., and Liu, L. (1993). Forecasting time series with outliers. *Journal of Forecasting*, *12*, 13–35.

Conrad, K., and Jüttner, D. J. (1973). Recent behaviour of stock market prices in Germany and the random walk hypothesis. *Kyklos*, *26*, 576–599.

Currie, I. D., Durbán, M., and Eilers, P. H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, *4*, 279–298.

Currie, I. (2006) Smoothing and forecasting mortality rates with P-splines. Talk given at the Institute of Actuaries, June 2006

De Jong, P., and Tickle, L. (2006). Extending Lee–Carter mortality forecasting. *Mathematical Population Studies*, *13*, 1–18.

Deaton, A and Paxson, C. (2004). Mortality, Income, and Income Inequality Over Time in the Britain and the United States. *Technical Report 8534, National Bureau of Economic Research Cambridge*, MA:. http://www.nber.org/papers/w8534.

Department of Health and Mental Hygiene (2023). Summary of Vital Statistics 2020 [online]. *Bureau of Vital Statistics, New York.* https://www.nyc.gov/assets/doh/downloads/pdf/vs/2020sum.pdf (Accessed 7 April 2023).

Dias, R., Heliodoro, P., Teixeira, N., and Godinho, T. (2020). Testing the Weak Form of Efficient Market Hypothesis: Empirical Evidence from Equity Markets. *International Journal of Accounting, Finance and Risk Management*, *5*, 40-51.

Dickey, D. A., and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, *74*, 427–431.

Dickey, D. A., and Fuller, W. A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, *49*, 1057-1072.

Dowd, K., Blake, D., and Cairns, A. J. G. (2010). Facing up to uncertain life expectancy: The longevity fan charts. *Demography*, *47*, 67–78.

Dryden, M. M. (1970). A statistical study of U.K. share prices. *Scottish Journal of Political Economy*, *17*, 369–389.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, *64*, 813-836.

Elton, E.J., Gruber, M.J., Brown, S. J., and Goetzmann, W. N. (2014). Modern Portfolio Theory and Investment Analysis, 9th edition, Wiley.

Enders, W. (1995). Applied Econometric Time Series, 1st edition, Wiley.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*, 987-1007.

Fama, E. F. (1970). Efficient Capital Markets: A review of theory and Empirical work. *The Journal of Finance*, *25*, 383-417.

Fama, E. F. (1976). Efficient Capital Markets: Reply. *The Journal of Finance*, *31*, 143-145.

Fama, E. F. (1991). Efficient Capital Markets: II. *The Journal of Finance*, *46*, 1575–1617.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., and Chen, B. (1998). New capabilities and methods of the X-12-ARIMA Seasonal-Adjustment Program. *Journal of Business & Economic Statistics*, *16*, 127–152.

Fischer, B., and Planas, C. (2000). Large Scale fitting of Regression Models with ARIMA Errors. *Journal of Official Statistics*, *16*, 173–184.

Gao, Q. and Hu, C. (2009). Dynamic factor model with conditional heteroskedasticity. *Insurance: Mathematics & Economics*, *45*, 410–423.

Gómez, V. and Maravall, A. (1996). Programmes SEATS and TRAMO: instructions for the User. *Bank of Spain Working Paper 9628*.

Granger, C. W. J., and Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society*, *38*, 189–203.

Grudkowska, S. (2016). JDemetra+ Reference Manual Version 2.1. *Narodowy Bank Polski, Department of Statistics*.

Gujarati, D. N. (2003). Basic Econometrics, 4th edition, McGraw-Hill.

Haberman, S., and Renshaw, A. E. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, *48*, 35–55.

Hamid, K., Suleman, M. H., Shah, S. Z. A. and Akash, R. S. I. (2010). Testing the Weak form of Efficient Market Hypothesis: Empirical Evidence from Asia-Pacific Markets. *International Research Journal of Finance and Economics, 58*, 121-133.

Hamilton, J. D. (1994). Time Series analysis, 1st edition, *Princeton University Press*.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, *23*, 365–380.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011). The model Confidence set. *Econometrica*, *79*, 453–497.

Hatzopoulos, P., and Haberman, S. (2011). A dynamic parameterization modeling for the age–period–cohort mortality. *Insurance: Mathematics and Economics*, *49*, 155–174.

Hatzopoulos, P., and Sagianou, A. (2020). Introducing and Evaluating a new Multiple-Component Stochastic Mortality Model. *The North American Actuarial Journal*, *24*, 393–445.

Hay, A. R., and McCleary, R. (1979). Box-Tiao Time Series models for impact assessment. *Evaluation Review* , *3*, 277–314.

Hilmer, S. C., Bell, W. R. and Tiao, G. C. (1983). Modelling considerations in the seasonal adjustment of economic time series. In A. Zellner (edt) *Applied Time Series Analysis of Economic Data*, 74-100 Washington DC, Bureau of the Census.

Hull, J. (2021). Options, Futures, and Other Derivatives, 11th edition, Pearson.

Hyndman, R., and Ullah, M. S. (2005). Robust forecasting of mortality and fertility rates: a functional data approach. *Department of Economics and Business Statistics, Monash University, Working Paper.*

Hyndman, R. J., and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, *51*, 4942–4956.

International Monetary Fund (2009). Balance of Payments and International Investment Position, 6th edition, Washington D.C.

Jarque, C. M., and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*, 255–259.

Kaiser, R., and Maravall, A. (2001). Measuring business cycles in economic time series, Springer.

Kalligeris, E., Karagrigoriou, A., Ladopoulos, K., and Parpoula, C. (2019). Adjusted Transformation Methods for reproduction Quality control. *Journal of Mathematics and Statistics*, *15*, 273–279.

Kleinow, T., and Richards, S. J. (2017). Parameter risk in time-series mortality forecasts. *Scandinavian Actuarial Journal*, *2017*, 804–828.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*, 159-178.

Lazar, D. T., and Denuit, M. (2009). A multivariate time series approach to projected life tables. *Applied Stochastic Models in Business and Industry*, *25*, 806–823.

Ledermann, S., and Breas, J. (1959). Les dimensions de la mortalite. *Population*, *14*, 637-682.

Lee, R. D., and Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, *87*, 659–671.

Lee, R. D. and Miller, T. (2001). Evaluating the performance of the Lee–Carter approach to modelling and forecasting. *Demography*, *38*, 537–549.

LeRoy, S. F. (1976). Efficient Capital Markets: Comment. *The Journal of Finance*, *31*, 139-141.

LeRoy, S. F. (1989). Efficient capital markets and martingales. *Journal of Economic Literature*, *27*, 1583–1621.

Lovász, E. (2011). Analysis of Finnish and Swedish mortality data with stochastic mortality models. *European Actuarial Journal*, 1, 259–289.

Ljung, G. M., and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*, 297–303.

Makridakis, S., and Hibon, M. (1979). Accuracy of Forecasting: An Empirical investigation. *Journal of the Royal Statistical Society*, *142*, 97-145.

Makridakis, S., Wheelwright, S. C., and Hyndman, R. J. (1998). Forecasting Methods and Applications, 3rd edition, Wiley.

Malafeyev, O., Awasthi, A., Kambekar, K. S., and Kupinskaya, A. (2019). Random walks and market efficiency in Chinese and Indian equity markets. *Statistics, Optimization and Information Computing*, *7*, 1-25.

Malkiel, B. (1992). Efficient market hypothesis, In: Newman, P., Milgate, M. and Eatwell, J., Eds., New Palgrave Dictionary of Money and Finance, McMillan.

Maravall, A. (1995). Unobserved components in economic time series, In Handbook of Applied Econometrics (Eds) M. H. Pesaran and M. Wickens, Blackwell.

Mehla, S., and Goyal, S. (2012). Empirical evidence on weak form of efficiency in Indian stock market. *Asia-Pacific Journal of Management Research and Innovation*, *8*, 59–68.

Meese, R., and Geweke, J. (1984). A comparison of autoregressive univariate forecasting procedures for macroeconomic time series. *Journal of Business & Economic Statistics*, *2*, 191-200.

Milionis, A.E. (2003). Modeling economic time series in the presence of variance non-stationarity. *Bank of Greece Working Paper No 7*.

Milionis, A. E. (2004). The importance of variance stationarity in economic time series modelling. A practical approach. *Applied Financial Economics*, *14*, 265–278.

Milionis, A. E. (2007). Efficient capital markets: A statistical definition and comments. *Statistics & Probability Letters*, *77*, 607–613.

Milionis, A. E. (2016). Lecture Notes in Financial Econometrics, Master Program in Applied Economics and Finance, Department of Economics of the National and Kapodistrian University of Athens in collaboration with the Department of Statistics and Actuarial Financial Mathematics of the University of the Aegean.

Milionis, A. E., and Davies, T. D. (1994). Regression and Stochastic models for Air Pollution, Part II. *Atmospheric Environment*, *28*, 2801-2810.

Milionis, A. E., and Galanopoulos, N. G. (2017). Time series with interdependent level and second moment: testing, consequences for modelling; preliminary results. *Proceedings of the 30th Panhellenic Statistics conference*, 109-123.

Milionis, A. E., and Galanopoulos, N. G. (2018a). Time series with interdependent level and second moment: statistical testing and applications with Greek external trade and simulated data. *Bank of Greece Working Paper 246*.

Milionis, A. E., and Galanopoulos, N. G. (2018b). Time series with interdependent level and second moment: testing, consequences for modelling; applications (part II). *Proceedings of the 31th Panhellenic Statistics conference,* 302-313.

Milionis, A. E., and Galanopoulos, N. (2019). Forecasting economic time series in the presence of variance instability and outliers. *Theoretical Economics Letters*, *9*, 2940–2964.

Milionis, A. E., and Moschos, D. (2000). On the validity of the weak–form efficient markets hypothesis applied to the London stock exchange: comment. *Applied Economics Letters*, *7*, 419–421.

Milionis, A. E., Moschos, D, and Xanthakis, M. (1998). The influence of foreign markets on the Athens Stock Exchange. *«ΣΠΟΥΔΑΙ», 48*, 140–156.

Milionis, A. E., and Papanagiotou, E. (2008). A Note on the Use of Moving Average Trading Rules to Test for Weak Form Efficiency in Capital Markets. *Bank of Greece Working Paper No. 91*.

Mills, T. C. (1990). Time Series Techniques for Economists, Cambridge University Press.

MSCI (2023). MSCI market classification framework, June 2023.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, *6*, 318–334.

Nelson, H. L., and Granger, C. W. J. (1979). Experience with using the Box-Cox transformation when forecasting economic time series. *Journal of Econometrics*, *10*, 57–69.

Nisar, S., and Hanif, M. (2012). Testing Weak Form of Efficient Market Hypothesis: Empirical Evidence from South-Asia. *World Applied Sciences Journal*, *17*, 414-427.

Oeppen, J., and Vaupel, J. W. (2002). Demography. Broken limits to life expectancy. *Science*, *296*, 1029–1031.

Oikonomou, G., Sampethai, I., and Symigiannis, G. eds, (2010). Greece's balance of payments: reasons for imbalances and policy proposals (in Greek).

Pankratz, A., & Dudley, U. (1987). Forecasts of power-transformed series. *Journal of Forecasting*, *6*, 239–248.

Phillips, P. C. B., and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*, 335–346.

Pitacco, E., Denuit, M., Haberman, S. and Olivieri, A. (2009). Modelling Longevity Dynamics for Pensions and Annuity Business, Oxford University Press.

Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, *45*, 393–404.

Renshaw, A. E., and Haberman, S. (2003). Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, *33*, 255–272.

Renshaw, A. E., and Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, *38*, 556–570.

Richards, S. J., and Currie, I. D. (2009). Longevity Risk and Annuity Pricing with the Lee-Carter Model. *British Actuarial Journal*, *15*, 317–343.

Roberts, H. (1959). Stock-Market "Patterns" and Financial Analysis: Methodological suggestions. *The Journal of Finance*, *14*, 1-10.

Romano, J. P., and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, *73*, 1237–1282.

Rubinstein, M. (1975). Securities market efficiency in an Arrow-Debreu economy. *The American Economic Review*, *65*, 812–824.

Said, S. E., and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, *71*, 599–607.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.

System of National Accounts – 2008 (2009).

Solvency II (Directive 2009/138/EC). Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II) (recast) (Text with EEA relevance)

Tsay, R. (1984). Regression Models with Time Series Errors. *Journal of the American Statistical Association*, *79*, 118–124.

Tsay, R. (1986). Time series model specification in the presence of outliers. *Journal of the American Statistical Association*, *81*, 132–141.

Tukey, J. W. (1961). Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics*, *3*, 191-219.

Villegas, A. M., Kaishev, V. K., and Millossovich, P. (2018). StMoMo: An R Package for Stochastic Mortality Modeling. *Journal of Statistical Software*, *84*.

White, H. (2000). A reality check for data snooping. *Econometrica*, *68*, 1097–1126.

Wold, H. (1938). A Study in the Analysis of Stationary Time Series, 2nd Edition, Almqristand Witsett, Stockholm.

World Bank Atlas method. datahelpdesk.worldbank.org/knowledgebase/articles/378832

Zivot, E. (2023). Introduction to Computational Finance and Financial Econometrics, 1st edition, Chapman and Hall/CRC.