UNIVERSITÀ DEGLI STUDI DI PALERMO
FACOLTÀ DI ECONOMIA

---

**Dottorato di Ricerca in:**

**Analisi Economica, Innovazione Tecnologica e Gestione delle Politiche per lo Sviluppo Territoriale**

*XXII CICLO*

# Seasonal Forecasting methods: an empirical analysis at sectoral and territorial levels

**Dottorando:**

*Stefania Rocca*

**Tutor:**

*Ch.mo Prof. Vincenzo Fazio*

**Coordinatore:**

*Ch.mo Prof. Fabio Mazzola*

*Alla mia famiglia e a tutti quelli che in questi anni hanno creduto in me.*
*Un pensiero speciale va alla nonna Franca che in questa occasione sarebbe stata orgogliosa di me.*

# Contents

# Introduction

Many people, institutions, economic agents and companies are involved in producing and using a whole range of different economic forecasts. These may be forecasts of developments in the aggregate economy such as that one of growth in total demand in different markets, of growth in total trade in terms of import-export, of inflation and interest rates or of unemployment. These may be also disaggregated forecasts relating to a specific sector (for example, wearing or leather) or to a specific region. Forecasts may be even more specific, relating to demand for a single product line, or to the reaction of a key competitor. The need for such forecasts arises because people are taking positions and entering into commitments about the future, and therefore need to form expectations, conditional to the current information set, about the possible future consequences of these positions or commitments.

For instance, the government needs to know the likely take-up of a state benefit (for example, the unemployment benefit) in order to plan the relative cost. The company investing in new skills and training processes wants to know that those skills will be relevant in the future. An industry, in order to decide if changing or not the price of a particular product, needs to have an overall view about its competitors and their possible reactions to that change. In all these situations people become the principal users of the forecast procedure, where a view must be taken in order to reach a sensible judgement as to how to act, as to whether the forward commitment is likely to be advantageous.

Hence, based on these findings, it becomes essential to define the object of the analysis and its real applicability. From a technical point of view, this means creating an ad "hoc" structure, usually identified by the choice of the model, of a suitable transformation and of the relevant predictors. For instance, if we want to study the effect of the growth rates of the UK GDP, perhaps it will be more convenient to use a logarithmic transformation instead of analyzing the series in its original scale. This study goes straight on this direction.

The first chapter analyzes the effects of forecasting at regional and national levels. At the beginning we focus the attention on the policy implications in terms of the identification of the key competencies adopting by policy makers to face aggregate and local economic challenges such as national or local shocks. Then we describe the relevant features of a general forecast such as seasonality and business economic fluctuations. Finally, starting from the idea that there exist no unique forecast "winner", we investigate deeper on the limitations supported by forecasting procedure. For instance, at a regional level, the quality of the data used is often less than desirable and there is a clear imbalance between the increased number of regional economies to be forecasted and the time span of the available observations. All these difficulties create the right motivation to discover a new methodology that try not only to avoid these problems but also that is able to capture all the features.

In the second chapter we use transformed time series to produce forecasts. The transformation implemented is that one proposed by Box-Cox (1964) which has the great advantage to include the logarithmic transform as a special case. The model used is the AR(1). A rolling forecast exercise has been done for 125 monthly series of the industrial production index and of tourism demand, related to Italy, France, Germany, UK and Sicily. All the series are not seasonal adjusted, such that we account for seasonality by introducing in the model the calendar effects as exogenous variables. The h-step ahead forecast has been calculated according to different type of transformations whose comparisons have been done by using the standard measure of accuracy and the Diebold Mariano test. All the results seem to suggest that the optimal value of the transformation parameter does not necessary correspond to a gain in forecast accuracy. In addition different transformations are also used to make comparisons for different time horizons. Clearly the results are strictly related to the specific sector analyzed and to the localized economic factors considered.

The last chapter ,instead, presents a non parametric analysis which has been developed in order to see what happens if the optimal value of transformation remains fixed. This involves the computation of the analysis in frequency domain for the prediction error variance. Some Monte Carlo experiments show how the prediction error variance changes according to the sample size chosen and the different bandwidth considered. At the end an application to real data is implemented. This study seems to suggest how the choice of the transformation parameter is essential for better understand the purpose of the analysis both at economic and at methodological level.

# Chapter 1

# The effects of forecasting at national and regional levels

,

## 1.1 Introduction

This chapter describes the role of forecasting both at national and regional levels. Forecasting the direction in which the economy is going is essential for monetary and fiscal policy, but not sufficient since knowledge about the starting position and, possibly, even the past trajectory is required. This is less simple than one would think, because of the key variables are not observable. Indeed, the conduct of economic policy faces challenges which are quite similar to those of economic forecasting, with some additional complications. Thinking about the monetary policy, it affects inflation and output with long and variable lags. These are average lags, measured over a sample period. The actual lags may be longer or shorter, depending on the particular combination of shocks affecting the economy at any point in time. In any case, what the forecasting analysis suggests is that a change in the monetary stance today has an impact on output roughly one year from now and on inflation one to two years from now.

The implication for monetary policy is quite straightforward: policy should be based not only on current output and inflation developments, but (mainly) on expected future developments, one to two years ahead. If one shares this generally recognised principle, one should also agree that a fair assessment about a given monetary policy stance cannot be made referring to the current state of the economy but rather on the basis of the expected state of the economy one or two years ahead. The rationale of a given stance or of a given monetary policy decision can be understood or criticized only on the basis of an economic forecast.

Moreover, any forecast of output and inflation is closely related to the assessment of the current economic conditions, in particular to the cyclical position of the economy, which is called in the literature the "output gap". An incorrect assessment of the underlying position of the economy in the cycle is likely to lead to a forecast error over the relevant forecasting horizon. The converse is also true: forecasts shape our view of current economic conditions.

Also, fiscal and monetary policies have different impacts on different regions [1]. This is because of heterogeneity across regions that creates strong and persistent regional disparities in economic performance. Hence, in order to correctly target their actions and policies, both national and local policy makers need to understand some aspects of such disparities: the determinants of persistence and variation, the region-specificity, the cross-regional dynamics and the aggregate level of the variable considered. Therefore, a forecast of the entire economy may be very important to the planning process of companies, institutions and policy makers operating both at a national and at a regional or local scale.

This demand for both national and regional forecasts challenges the economic researcher to develop models that produce accurate predictions of the major eco-

[1] see Ledyaeva et al. (2008), Longhi et al. (2007), Mayor et al. (2007) and Tena et al. (2010)

nomic variables for regions and countries. In this context it is clear how a comprehensive knowledge of the specific features of the macro-area under observation is essential for choosing the right instruments which allow to react more effectively and to prevent more accurate reactions to aggregate and local shocks.

## 1.2 Why forecasting is so controversial?

Nowadays there is an increasing tendency to describe events that are properly relate to the future. In this sense a forecast should be considered as any statement about the future. Such statements may be well founded or may lack any sound basis; they may be accurate or inaccurate on any given occasion or on average, precise or imprecise, model-based or informal. There are many ways of making forecasts[2], each of them should represent the best choice according to the purpose of the analysis. This means that it is not the method we choose that makes a general forecast successful or not, instead the judgment depends on how much the forecast is "close" to the outcome.

Thus, one must decide on which aspect it is important to be "close" before any choice between methods[3] is possible. Usually accuracy, measured by the notion of unbiasedness, and precision are the two dimensions along which forecasts may be judged. A good compromise between them makes a forecast a possible "winner". Unfortunately, no unique measure of a "winner" is possible in a forecasting competition involving either multi-period or multi-variable forecasts.

In addition to bias and variance considerations, point forecasts are often judged on criteria such as the efficient use of information. Also, forecasts often include forecast intervals and, sometimes the complete density of outcomes, so these are required to be "well calibrated".

[2]see Hendry et al. (2001)
[3]see Granger et al. (2005)

Since it is seldom the case that a single forecast exists for the economic phenomenon of interest, rival forecasts are often available to allow ex-post comparisons of one against the other. Sometimes a combination of one or more forecasts is better than a single forecast, or we may choose a forecast that contains all the useful information which exist in another one.

The properties of forecasting methods can be investigated in both empirical and artificial settings, using mathematical analysis and computer-intensive numerical methods. Forecasting methods should be compared by testing combinations [4] of forecasting models for encompassing or by Monte Carlo simulation. In the last case an investigator generates artificial "data" against which the models are compared in repeated trials, in order to calculate how well such methods perform in a chosen controlled environment. This procedure is most useful when we know the large-sample behaviour of the statistics of interest for the forecasting methods and when we wish to investigate the usefulness of these asymptotic results for samples of the size typically available to the applied researcher.

Empirical comparisons, in the form of forecasting competitions[5], typically look at the performance of different methods for many time series. Because the data generating process is not under the investigator's control and will only be imperfectly known, the results of forecast comparisons for any one series could depend on the idiosyncratic features of the series, thus limiting their general applicability.

For this reason, many series are compared, and often series are selected which share certain characteristics, with the caveat that the results might only be expected to hold for other series with the same characteristics. This highlights a "circularity problem": until we know how empirical economic data are generated, we cannot define the appropriate framework for developing or analyzing methods, so we cannot actually determine how well they should perform.

[4] see Stock et al. (2003)

[5] see Rickman et al. (2009), West et al. (1996) and West (2003)

## 1.3 The evaluation of regional policy through forecasting

Policy makers need to evaluate the total effect of local and regional programs in order to make informed decisions. Development proposals have economic, social, and demographic implications that go well beyond their direct effects. To understand these effects, analysts need to use a comprehensive economic forecasting and simulation approach.

Local and regional policy models show the full effects of policy changes on the local economy, including socioeconomic consequences that may otherwise be unforeseen or unrecognized. Normally, regional and local economic models are used for a variety of policy analysis purposes, including the evaluation of economic development proposals, transportation projects, environmental regulations, and energy programs.

Regional models are also used for planning and programming purposes, particularly when they relate to infrastructure needs including new roads, airports, power plants, water facilities and a broad range of other public and private services. Each of the model used has its strengths and weaknesses. Usually, models are dynamic, with forecasts and simulations generated on an annual basis in relation to behavioural responses to wage, price or other economic variables.

In this sense, in order to evaluate program implications over time, a medium and long term economic forecast is essential to better understand the timing effects strictly related to the model performance over time and to the overall feasibility of a new project. Indeed, once a forecast has been done, it is necessary to control it over time to determine if adjustments should be made in the forecasting model by making the appropriate adjustments to secure the most accurate series of forecasts.

If the model does not simulate future values relatively accurately as time unfolds, the next step is to determine the causes for any deviations. If the review of the causes indicates that forecast values are likely to over- or under-estimating future values of the variable under analysis, then the forecast may be revised. Alternatively, we can find evidence that the trend or causal relationships have changed and that we must determine if model parameters have changed too. That is, whether the coefficients or equations no longer reflect what is actually happening as the future unfolds. Finally, after considering all the evidence, you may decide to generate a new forecast from the existing model or, alternatively, develop a new model.

## 1.4 Which special data features to make predictions?

Starting from the consideration that anything can be forecasted, ranging from next month's rate of consumer price inflation or tomorrow's weather patterns or the value of the Dow Jones index at the start of 2011, there are some data that fit better the forecasting exercise. Indeed, many economic and financial time series possess a number of special features including, in various combinations: seasonality, business-cycle fluctuations, trend growth, successive dependence and changing variability.

More generally, data in economics are often "nonstationary", namely, they change mean and variance over time. These special data features are potentially important for a number of reasons. A failure to allow for such specific characteristics (say, seasonality) may result in inferior forecasts of the aspects of interest (say, turning points, or the underlying trend) especially if, as some recent research suggests, these characteristics are inherently interlinked.

Moreover, some of these characteristics may themselves be the focus of attention in the forecasting exercise. For example, one may wish to forecast a business-cycle characteristic, such as the next recession, and be otherwise uninterested in the level, or rate of growth, of the series.

For both these reasons, models have been developed which attempt to capture special features, and as it will become apparent, many different approaches have been proposed such as the so-called "unit-root" non stationary process which leads the variable to have a stochastic trend.

### 1.4.1 Seasonality

Many economic time series displays seasonality[6]. If seasonality is present, it must be incorporated into the time series model. By seasonality we usually mean periodic and largely repetitive fluctuations that are observed in time series data over the course of a year. As such, it is largely predictable.

A generally agreed definition of seasonality in the context of economics is provided by Hylleberg (1992) as follows: "Seasonality is the systematic, although not necessarily regular, intra-year movement caused by the changes of weather, the calendar, and timing of decisions, directly or indirectly through the production and consumption decisions made by the agents of the economy. These decisions are influenced by endowments, the expectations and preferences of the agents, and the production techniques available in the economy".

This definition implies that seasonality is not necessarily fixed over time, despite the fact that the calendar does not change. Thus, for example, the impact of Christmas on consumption or of the summer holiday period on production may evolve over time, despite Christmas time and summer time remain fixed. Intra-year observations on most economic time series are typically available at

[6]see Chatfield et al. (1973) and Kulendran et al. (2005)

quarterly or monthly frequencies. As a consequence each intra-year observation is related to a "season", by which we mean an individual month or quarter.

Looking at the time series available for the tourism[7] sector, it has been pointed out the tendency of touristic flows to become concentrated into relatively short periods of the year (Allcock, 1989), or that there is a systematic, although not necessarily regular, intra-year movement (Hylleberg, 1992), or a "a temporal imbalance" (Butler, 2001).

Usually, the impact of seasonality on tourist demand has been described as negative, by lookin at effects such as overcrowding, inefficient use of resources, price increases, high social costs for hosting communities, capacity and supply chain issues, underutilization of capital assets, employment concerns, inefficient management of public services at destinations and decrease in quality for tourists. However, other authors have pointed out the positive effects of seasonality, which include opportunities for restoration and recuperation, re-establishing of traditional and socio-cultural patterns in the hosting communities, ecological environment recovery, possibility to complement revenue from tourist activity with other sources and employees involvement in other seasonal activities (for example, agriculture).

The methods for measuring and modelling seasonality are highly developed and well established both in the univariate and the multivariate frameworks. Variables analysed with time-series methods include arrivals and overnight stays, occupancy rates, total tourism receipts and, more recently, per person tourist spending. At regional level, seasonal variation may be combined with structural trends. Tourism demand in Sicily, for instance, has grown up by a 25% in terms of total arrivals between 1998 and 2007, thus becoming a significant factor in the economic development of the region. The seasonal pattern shows differences

[7] see Frechtling (2001) and Zhang (2007)

according to the market of origin (residents versus non residents tourists) and displays a great concentration in the months of July and August.

Thus, seasonal variation is important because it has a great economic impact at destination; it also influences the prices of tourism-related goods and services at destination. When forecasting tourism demand, therefore, seasonal variation cannot be ignored; since it creates volatility in tourism demand. Also if we look at industrial production or at retail sales seasonal variation must be considered. For instance, it is well known the tendency of sales to peak in the Christams or the Easter season and to decline after holidays. Modelling seasonality for forecasting these kind of series allows to adopt a different methodology such as the introduction of seasonal dummies able to capture constant seasonal variations for ARIMA models in which first differences of the series can remove unit roots at zero frequency.

### 1.4.2 Business-cycle

The term business cycle[8] is related to economic fluctuations in production or economic activity over several months or years. These fluctuations occur around a long term growth trend and typically involve shifts over time between periods of relatively rapid economic growth (expansion) and periods of relative stagnation or decline (contraction or recession).

At regional level, one of the stylized facts about regional business cycles is that the fluctuations in different regions of an economy tend to synchronize with each other (Clark, 1992). This synchronization may arise for different reasons. It may come about due to common shocks including national fiscal or monetary policies, common consumer behaviour, world commodity price shocks, or common technology shocks.

[8] see Carlino et al. (2003), Hall et al. (2007) and Hayashida et al. (2009)

Alternatively, the synchronization could come about due to trade and factor flow linkages between different regions of the economy in such a way that regional economies "drive" one another. The synchronization[9] between regions is not predominantly due to a large common exogenous force: evidence has shown that fiscal and monetary policies are not strong enough to enforce synchronization and that they operate on different regions with different strengths and different timing (see Garrison and Chang, (1979) and Kozlowski (1995)).

In addition, we cannot explain the synchronization by referring to a common national business cycle. The national business cycle is the sum of its parts; the national business cycle cannot act like an exogenous force upon the parts. Rather, the regional cycles synchronize with each other through linkages.

## 1.5 Methods of forecasting

The most common methods of forecasting in economics are related to extrapolation, leading indicators, surveys, time-series models and econometric systems. Extrapolation is fine so long as the tendencies persist, but that is itself doubtful: the telling feature is that different extrapolators are used at different points in time. Moreover, forecasts are most useful when they predict changes in tendencies which extrapolative methods are likely to miss.

Forecasting based on leading indicators[10] requires a stable relationship between the variables that "lead" and the variables that are "led". When the reasons for the lead are clear, as with orders preceeding production, then the indicators may be useful, but otherwise are liable to give misleading information. Surveys of consumers and businesses can be informative about future events, but they rely on plans being realized.

[9] see Selaver et al. (2005)
[10] see Shoesmith (2000)

Time-series models which describe the historical patterns of data are popular forecasting methods and have often been found to be competitive relative to econometric systems of equations (particularly in their multivariate forms). They analyze a variable which changes with time and which can be said to depend only upon the current time and the previous values that it took (i.e. not dependent on any other variables or external factors). Econometric systems of equations are the main tool in economic forecasting. They include equations which seek to "model" the behavior of discernible groups of economic agents (consumers, producers, workers, investors, etc.) assuming a considerable degree of rationality, moderated by historical pattern.

The advantages for economists of using formal econometric systems of national economies are to consolidate existing empirical and theoretical knowledge of how economies function, provide a framework for a progressive research strategy leading to increased understanding over time, help to explain their own failures, as well as provide forecasts and policy advice.

## 1.6 Problems in forecasting

One of the main problems with forecasting in economics is that economies evolve over time and are subject to intermittent and sometimes large, unanticipated shocks. Going back to the presence of volatility in tourism demand, visitor volumes fluctuate with the seasons and over annual periods and often produce wide variations. A similar pattern may be considered if we look at the industrial production at any sector. In this situation the more volatile is an activity, the more difficult it is to discern patterns that can help the researcher to make forecast. As a consequence, any kind of mistake could induce poor forecast performance, either from inaccurate (that is biased) or imprecise (that is high variance) forecasts.

However, it transpires that systematic forecast failure is most likely to depend on the behaviour of the deterministic terms, and in particular on unanticipated changes in their values. Such deterministic shifts may reflect changes elsewhere in the economy, interacting with an incomplete or incorrect model specification.

At a national level, there exist ad "hoc" solutions principally related to the introduction of methods that anticipate change by observing it ex-ante, namely before such shocks have occurred. New problems tend to arise if regional forecasts are introduced. First of all the imbalance between the increased number of regional economies to be forecasted for and the time span of the available observations; indeed, given the likely small number of observations, it is impossible to establish which predictors are considered to be relevant for forecasting purpose. If a researcher uses only one of these predictors the forecasts tend to be generally unreliable and unstable.

At the opposite side, if all predictors are included, we may incur in overfitting and poor out-of-sample forecasting accuracy. The forecast may indeed be unfeasible. In this situation a possible strategy is to combine forecasts from many models with alternative subsets of predictors.

From a theoretical perspective, unless one can ex ante identify a particular forecasting model that generates smaller forecast errors than its competitors, forecast combinations offer diversification gains that make attractive to combine individual forecasts rather than relying on forecasts from a single model. Even if a best model could be identified at each point in time, combination may still be an attractive strategy due to diversification gains, although its success will depend on how well the combination weights can be determined.

A second reason for using forecast combinations is that individual forecasts may be very differently affected by structural breaks caused, for example, by institutional change or technological developments. Some models may adapt quickly and will only temporarily be affected by structural breaks, while others have parameters that adjust only very slowly to new post-break data. The more data are available since the most recent break, the better one might expect stable, slowly adapting models to perform relatively to fast adapting ones as the parameters of the former are more precisely estimated. Conversely, if the data window since the most recent break is short, the faster adapting models can be expected to produce the best forecasting performance. Since it is typically difficult to detect structural breaks in "real time", it is plausible that on average, i.e., across periods with varying degrees of stability, combinations of forecasts from models with different degrees of adaptability will outperform forecasts from individual models.

A third and related reason for forecast combination is that individual forecasting models may be subject to misspecification bias of unknown form. Hence, combining forecasts across different models can be viewed as a way to robustify the forecast against such misspecification biases and measurement errors in the data sets underlying the individual forecasts. For all these situations described before, one can use the mean, the median or the trimmed mean of these forecasts as final forecasts (Stock and Watson, 2003).

An alternative strategy uses factor augmented forecasts. The methodology allows for factors from all predictors to be estimated by the principal components method, and includes them in a linear forecasting equation for the $h$-step ahead forecast.

Another strategy concerns the application of a testing procedure to decide which predictors to include in the forecasting regression. In this case, a pre-test analysis is conducted for fitting a regression, where it is common to include only those predictors that are significant in predicting the series.

Finally, a bootstrap[11] aggregation should be considered: this involves analyzing a model that includes all potential predictors, generating a large number of bootstrap resamples, applying the pre-test rules to each of resample, and averaging the forecasts from the models selected by the pre-test on each bootstrap sample. By averaging across sample, after variable selection, the prediction mean squared error of the regression model is considerably reduced.

Another problem regards the complex dynamics and economic interdependencies influencing economic performance, which are often difficult to measure and which create difficult specification issues in inferential statistics. Indeed, regions are open, small and highly interconnected economies that show a high degree of interaction with the neighbouring local economies. The economic development of each region is likely to be affected by, and to have a high impact on the economic development of the other regions. In this sense neglecting such spatial correlation and dependence might result in biased coefficients and suboptimal forecasts and create problems of misspecification.

Clearly policy makers who understand the specific characteristics of a region and of interregional dependencies are able to tackle problems more effectively, and to anticipate more accurately necessary reactions to aggregate and local shocks. Moreover a group of (contiguous) regions that share common characteristics has the opportunity to develop common strategies and to react better.

Furthermore the quality of the regional data used in regional forecasts is often less than desirable. Frequent revisions and substantial publication lags mean that the true values of key economic variables at the starting point of the forecast are often uncertain. This can have a significant impact on the quality of the economic forecasts as the true state of the regional economy at the starting point of the forecast is known only very imperfectly. As a consequence, limited data

[11] see Pascual et al. (2005)

availability prevents the completion of a thorough statistical assessment of the errors of the forecasts.

Also, the different frequencies at which the data are available don't help the researcher. In addition the uncertainty increases as one moves down from highly aggregated national data to disaggregated regional and sub-regional data. Moreover, the degree of uncertainty is not eliminated entirely with successive releases of a given year's data.

## 1.7 Conclusions

Based on these initial observations, it is clear how forecasting economic variables is essential for the identification of national and regional key competencies. Governments are important users, and often producers, of macro-economic forecasts. This is because, in setting monetary, fiscal and exchange rate policy, governments must take account of the likely future course of the economy. Macroeconomic models play an important role in the formulation and the analysis of macro-economic policy decisions. For example, in stabilizing whether to raise or lower interest rates, and to raises or lower taxes, governments need to take a view on likely growth and inflation prospects in the economy for the future. Forecasts are essential in this situation.

Instead, at a regional level, forecasting can be used to give an impression of likely effects of policy initiatives on industrial production, income and labour market or to isolate key aspects of the development of competencies in learning clusters. Regarding regional industry policy, in order to see if an industry is over represented in a certain location compared to its representation at the national level, it is important to define the localization coefficients. Hence, an industry which is over represented is likely to contain a cluster of specific competences. At the same time an industry which is over represented will have more capacity to supply the other local industries with inputs.

Therefore when studying inter-industry supply it is natural to use the information from localization coefficients to correct the coefficients in the inter-industry description. Usually, the sectors with an over representation are typically primary and secondary sectors. Further the tertiary sectors which have higher growth are under represented and growth in these sectors is not likely in peripheral regions since they depend on learning capacity and clusters which tend to prefer to locate in the central parts of the economy.

All the analyzed difficulties, push the researcher to discover alternative forecasting techniques that fit better the model under observation and create the right motivation to go deeper inside the problem analyzed. In the next chapter of this dissertation we try to develop a new methodology easily applicable at any level of aggregation by including more than one relevant predictors in transformed time series.

# Chapter 2

# Forecasting transformed time series

## 2.1 Introduction

In empirical time series analysis, it is common to transform[1] the data using power transformation prior to the estimation of the model used for forecasting. There are several reasons for transforming the data before fitting a suitable model. For example, the necessity of stabilizing the increasing variance of trending time series, to reduce the impact of outliers, to make the normal distribution a better approximation to the data distribution. Also the transformed variable may have a convenient economic interpretation; for example, first differenced log-transformed data correspond to growth rates[2]. For any such analysis the usual purpose of the transformation is to change the scale[3] of the measurement in order to make the analysis more valid.

The Box-Cox transformation belongs to the power normal family distributions whose members include normal and lognormal distributions. It has two useful features: first, it includes linear and logarithmic transformations as special cases,

[1] see Hosoya et al. (2009)
[2] see Arino et al. (2000), Bartlett (1947) and Hivkley (1984)
[3] see Kemp (1996)

and, second, it possesses strong scale equivariance properties, including the property that the transformation parameter is unaffected by the rescaling in models with intercepts. Its main disadvantage is that both the domain and the range of the transformation are, in general, bounded.

Unfortunately, the Box-Cox transformation can only be applied to positive $y$, the variable being transformed, and has a range that depends on $\lambda$, the transformation parameter. This is why since the work of Box and Cox(1964), there have been many modifications proposed[4].

Manly (1976), for example, proposed the exponential transformation which allows also for negative values of y. The transformation was reported to be successful in transform unimodal skewed distribution into normal distribution, but is not quite useful for bimodal or U-shaped distribution. John and Draper (1980) proposed the following modification which they called "'Modulus Transformation". Bickel and Doksum (1981) modify the Box-Cox transformation to allow for any $y$ and to have a range that is always $(-\infty, +\infty)$; however, their transformation family obviously does not include the logarithmic transformation as a special case.

The objective of this chapter is to determine the value of the Box-Cox transformation parameter which is optimal for the predictability of a time series. More generally, we want to show how transforming a series should lead an improvement in forecasting accuracy. For this purpose a large recursive forecast exercise was conducted.

This work extends the analysis by Lutkepohl and Xu (2009) and by Bardsen and Lutkepohl (2009), which dealt with logarithmic transformation and evaluate, in a Monte Carlo simulation, the forecasts performance at different time horizons, of the naive and the optimal $h$-step ahead forecasts in a $K$-dimensional VAR(p)

[4]see Burbidge (1988), Yang (2002) and Yeo (2000)

model.

In particular they show how there is no improvement in terms of RMSE reductions from using the optimal relative to the naive forecasts unless the forecasts horizons are very short. In this case, in fact, the RMSE, are very similar. This result is also confirmed by an empirical example based on U.S. quarterly series of real investments and real GNP.

Two considerations arise: how much the transformation considered by the authors is suitable for this analysis and how can we proceed in making forecasts? This work goes straight in this direction. Indeed the idea is to use the Box Cox transformation, which includes the logarithmic transformation as a special case.

## 2.2 The Box-Cox transformation

Let $y_t$ be a time series, I consider the Box-Cox transformation defined as Box and Cox (1964)

$$y(\lambda) = \begin{cases} \frac{y_t^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \\ ln(y_t) & \text{if } \lambda = 0 \end{cases} \tag{2.1}$$

defined for $y_t > 0$ where $\lambda$ is a real constant. The transformation for $\lambda = 0$ follows from the fact that

$$lim_{\lambda \to 0} \frac{y_t^\lambda - 1}{\lambda} = ln(Y_t) \tag{2.2}$$

Subtracting 1 and dividing it by $\lambda$ does not influence the stochastic structure of $Y^\lambda$, and hence, without loss of generality, one often considers the following transformation suggested by Tukey (1957)

$$y(\lambda) = \begin{cases} y_t^{\lambda} & \text{if } \lambda \neq 0 \\ \\ ln(y_t) & \text{if } \lambda = 0 \end{cases} \tag{2.3}$$

instead of the Box-Cox transformation[5].

Hence,different values of $\lambda$ implies different transformations. The value $\lambda = 1$ implies that the series is analyzed in its original scale, $\lambda = 0$ originates the logarithmic transformation. Other important special cases arise for fractional values of $\lambda$, e.g. the square root transform ($\lambda = 1/2$) or the third root transform ($\lambda = 1/3$)

## 2.3 The model

I assume that the series $y_t$ is stationary and follows an AR(p) process

$$y_t = \mu + \sum_{j=1}^{p} \phi_j y_{t-j} + \epsilon_t \tag{2.4}$$

where $\epsilon_t \sim N(0, \sigma^2)$ is a white noise.

In order to obtain the $h$-step ahead forecast $y_{t+h|t}$ we can express the model in the following state space form

$$z_t = F z_{t-1} + v_t \tag{2.5}$$

[5]see Box et al. (1982), Box et al. (1994), Box et al. (1973), Sakia (1992) and Wooldridge (1992)

where

$$z_t = \begin{pmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{pmatrix} \quad F = \begin{pmatrix} \phi_1 & \cdots & \phi_p \\ 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \quad \text{and } v_t = \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2.6}$$

The $h$-step ahead forecast is recursively obtained as

$$y_{t+h|t} = f_{11}^h(y_t - \mu) + f_{12}^h(y_{t-1} - \mu) + \ldots + f_{1p}^h(y_{t-p+1} - \mu) \tag{2.7}$$

where $f_{ij}^h$ denote the $ij$ element of the matrix $F^h$.

Let the forecast error $\epsilon_{t+h|t} = y_{t+h} - y_{t+h|t}$ its variance is given by the last element on the main diagonal of the matrix

$$\sigma^2 \sum_{j=0}^{h-1} F^j F^{jT} \tag{2.8}$$

## 2.4 The simulation procedure: a rolling forecast scheme

An out-of-sample forecasting exercise is conducted in order to compare and evaluate the out-of-sample forecast performance of the different transformations. Following Tashman (2000) the series is split in a pre-forecast period (namely fit period) and in a test period. In practice given a sample of T observations we divide it into R observations to be used for estimation (in-sample) and P observations to be used for out-of-sample evaluation. It follows that timing is $T = R + P$

Starting from a given forecast origin we create a sequence of update origin and produce forecast from each new origin according to the maximum forecast lead chosen. We start considering for a seasonal time series

$$\Delta_{12}(u_t - \beta^{'} x_t) \sim AR(p) \tag{2.9}$$

where $x_t$ represented the calendar effects (trading days + holiday effects) whose coefficients are modelled by a simply OLS regression.

Then we proceed by the following steps: first of all it is useful to select the optimal order of temporal lags (p) by using AIC (or BIC) methods, then we estimate the parameters of the model, we produce forecast up to a given maximum forecast lead $h$ (in my case $h = 24$) making also forecasts comparisons in order to establish which is the forecast that fits better the model chosen.

In particular regarding the latter point we choose the optimal forecast for growth rates where the optimal corresponds to the conditional expectation representing the optimal minimum mean squared error predictor and the naive forecast for levels where the naive represents the minimum mean absolute error which is simply the median of the conditional probability density for the $h$-step ahead forecast.

Finally we run this exercise for 125 monthly time series. Moreover, in order to make comparisons, we define different specification for forecasts according to the transformation chosen. In particular for the logarithmic transformation we can choose between the $h$-step ahead naive forecast (Granger and Newbold, 1976) which is defined as

$$y^{nai}_{t+h|t} = exp(x_{t+h|t}) \tag{2.10}$$

and the $h$-step ahead optimal forecast which is given by

$$y_{t+h|t}^{opt} = exp(x_{t+h|t} + \frac{1}{2}\sigma^2(h)) \tag{2.11}$$

Instead, for the square root (or third root) transformation the choice is between the $h$-step Box-Cox naive forecast, for $\lambda \neq 0$, expressed as

$$y_{t+h|t}^{bcnai} = (\lambda x_{t+h|t}^{bc} + 1)^{\frac{1}{\lambda}} \tag{2.12}$$

and the $h$-step Box-Cox optimal forecast which is obtained as

$$y_{t+h|t}^{bcopt} = y_{t+h|t}^{bcnai} + G \tag{2.13}$$

where $G = 1 + \frac{\sigma^2(h)}{x_{t+h|t}^{bc}}$.

Notice that the optimal forecast differs from the naive one by a multiplicative adjustment factor whose expression for the Box Cox optimal forecast is obtained according to the value of the transformation parameter $\lambda$ which is assumed to be fixed and equal to 0.5 (Pankratz and Dudley, 1987).

## 2.5 Standard measures of accuracy

The diagnostic check typically look at the residuals. Once we have computed the forecast errors, obtaining by subtracting each of the forecast from the know data values of the test period, we proceed by calculating the standard measures of accuracy[6] : ME(Mean Error), MSE(Mean square error), MAE(Mean absolute error), MPE(Mean percentage error), MSPE(Mean square percentage error),

[6]see Clements (2005)

MAPE(Mean absolute percentage error) where

$$ME = \frac{1}{P} \sum_{t=R+h}^{T} \epsilon_{t+h|t} \tag{2.14}$$

$$MSE = \frac{1}{P} \sum_{t=R+h}^{T} \epsilon_{t+h|t}^2 \tag{2.15}$$

$$MAE = \frac{1}{P} \sum_{t=R+h}^{T} \frac{|y_{t+h} - y_{t+h|t}|}{y_{t+h}} \tag{2.16}$$

$$MPE = 100 \times \frac{1}{P} \sum_{t=R+h}^{T} \frac{y_{t+h} - y_{t+h|t}}{y_{t+h}} \tag{2.17}$$

$$MSPE = 100 \times \frac{1}{P} \sum_{t=R+h}^{T} \epsilon_{t+h|t}^2 \tag{2.18}$$

$$MAPE = 100 \times \frac{1}{P} \sum_{t=R+h}^{T} \frac{|y_{t+h} - y_{t+h|t}|}{y_{t+h}} \tag{2.19}$$

Notice that the last three statistic involve percentage errors rather than the rows errors. Such measures are computed for each of the following cases: no transformation, logarithm transformation, square root transformation and third root transformation.

We find also both the exact logarithm and the exact square root (third) transformations simply by taking the relative forecast errors. In this context a Diebold

Mariano test has been implemented in order to permit comparisons in terms of equal predictive accuracy of two competing forecasts.

Let $y_t$ denote the series to be forecast and let $y^1_{t+h|t}$ and $y^2_{t+h|t}$ denote two competing forecasts of $y_{t+h}$ based on the information set $I_t$, given the two associated forecast errors, the idea is to assess the expected loss associated with each of the forecast. In this sense the accuracy of the forecast is measured by a loss function which is usually a direct function of the forecast error. Hence the DM test statistic ,under the null, is given by

$$DM = \frac{\bar{d}}{\left[\frac{\hat{V}}{P}\right]^{\frac{1}{2}}} \tag{2.20}$$

where $\bar{d}$ is the sample mean loss differential given by

$$\bar{d} = \frac{1}{P}\sum_{t=T-P+1}^{T}[g(\epsilon_{it}) - g(\epsilon_{jt})] \tag{2.21}$$

$\hat{V}$ is a consistent estimator of the long run variance computed from the sample autocovariance of the loss differential

$$\hat{V} = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j \tag{2.22}$$

with $\gamma_j = cov(d_t, d_{t-j})$. Notice that the long run variance is used to avoid the problem of serial correlation in the sample of loss differentials for $h > 1$.

Diebold and Mariano (1995) show that under the null of equal predictive accuracy the statistic $S \sim N(0,1)$, so we reject the null of equal predictive accuracy at the 5% level if $|S| > 1.96$.

## 2.6 Empirical application: a database description

The data are monthly series of the Industrial production index (IPI), ordered according to the NACE rev.2 classification and of Tourism demand for the countries: Italy, France, Germany, UK and at regional level for Sicily. All the time series considered are not seasonally adjusted and have been extracted, respectively, from ConIstat (for Italy), from Eurostat (for the other countries) and from the Tourism department of Sicilian Regional Government. In particular the time series for touristic demand involve the number of arrivals and the number of nights spent from resident and non resident tourists respectively in hotels and other complementary structures.

The data employed are available for the following periods: 1990 (1) - 2010 (1) for Italy IPI series, 1991 (1) - 2010 (2) for France, Germany and UK IPI series, 1990 (1) - 2009 (10) for Italy Tourism series, 1994 (1) - 2009 (11) for France, Germany and UK Tourism series and 1998(1) - 2009(12) for Sicily.

Since all the series are not seasonally adjusted we account for seasonality taking the seasonal differencing operator $\Delta_s y_t = y_t - y_{t-s}$ where $s = 12$ is the seasonal period accordingly to the monthly periodicity of the series.
Hence, if we want to focus the attention on the yearly changes of the normalized series, we just calculate $u_t = z_t - z_{t-12}$ Sometimes we should be interested in the yearly growth rates in the original scale, that is

$$\Delta_{12} y_t / y_{t-12}. \tag{2.23}$$

Notice that the prediction of this quantity is not an easy task.

In practice the rolling forecast exercise for monthly time series does the following: fit an AR(p) model to $\Delta_{12}y_t$, with regression effects,

$$\phi_p(L)\Delta_{12}y_t(\lambda) = \beta_0 + \beta^{'}\Delta_{12}x_t + \epsilon_t, \epsilon_t \sim WN(0, \sigma^2) \tag{2.24}$$

where $x_t$ are 6 trading days regressors and an Easter variable, that account for calendar effects and $y_t(\lambda)$ is simply the Box Cox transform of the series.

It is not unusual for the level of a monthly economic time series to be influenced by calendar effects (Bell and Hillmer (1983), Cleveland and Devlin (1980)). For this reason including calendar effects it's important if we aim at modelling the unadjusted series. Such effects arise when changes occur in the level of activity resulting from variations in the composition of the calendar[7] between years.

Two main sources of calendar effects are: trading day corresponding to changes of a particular activity related with the day of the week composition of a certain month and holiday effects representing variations depending on whether a particular month contains a holiday or not (i.e. Easter for Europe). These effects are introduced in our model as exogenous variables.

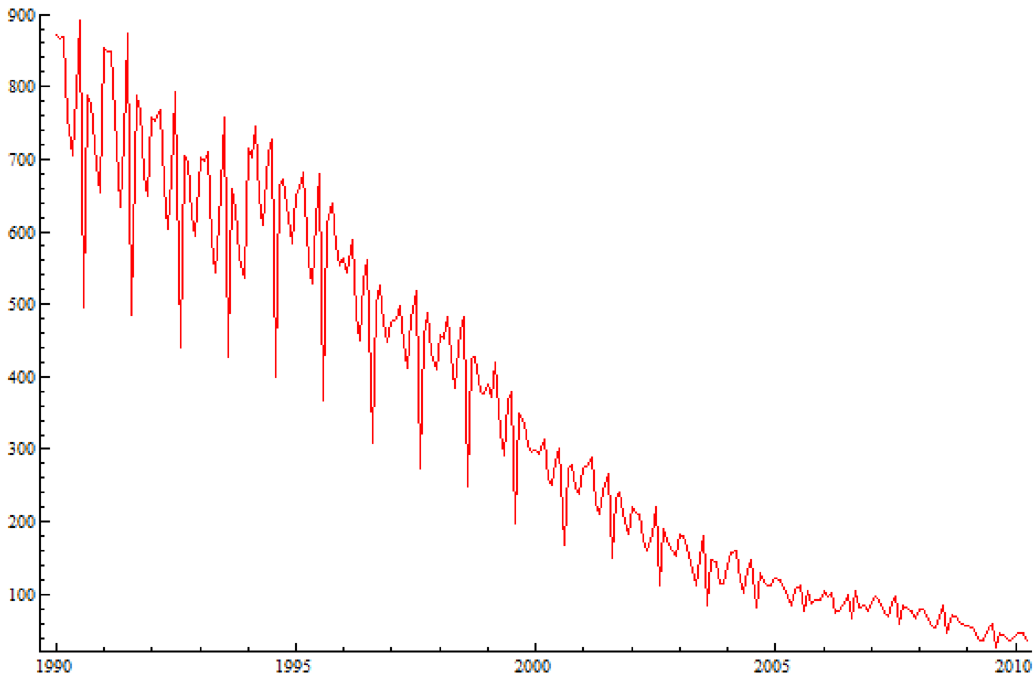
[7] see Cleveland et al. (1982) and Thury et al. (2005)

## 2.7 Simulation results

### 2.7.1 France industrial production index

Looking at the France industrial production index some relevant results are obtained respectively for the wearing, textile and leather sectors. Figure 2.1 plots the monthly index of industrial production for the manufacture of wearing apparel sector (January 1990 - April 2010). The series presents two main related features: a downward trend and a marked seasonal pattern with reduced amplitude. The series is thus a good candidate for analysis the role of the BC transformation.

Figure 2.1: France, industrial production index: manufacture of wearing apparel.



The table below shows the forecast comparisons in terms of MSE for level and growth rates at different time horizon.

Table 2.1: France Wearing: rolling forecast experiment. Mean square forecast error.

| | | Levels $y_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 173.13 | 239.07 | 271.38 | 343.48 | 369.44 | 418.65 | 446.17 | 968.75 |
| 0 | Optimal | 175.23 | 243.63 | 278.44 | 351.98 | 379.67 | 430.52 | 458.04 | 1033.69 |
| 1/3 | Naïve | 194.44 | 266.82 | 280.27 | 349.87 | 372.59 | 412.56 | 419.96 | 863.63 |
| 1/3 | Optimal | 195.84 | 269.68 | 284.46 | 354.77 | 378.41 | 419.23 | 425.95 | 903.36 |
| 1 | Optimal | 331.97 | 410.10 | 417.68 | 479.52 | 531.04 | 549.83 | 618.00 | 1944.25 |
| | | Growth rates $g_t$ | | | | | | | |
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 36.19 | 49.07 | 58.02 | 67.43 | 72.63 | 76.93 | 94.82 | 104.66 |
| 0 | Optimal | 36.37 | 49.43 | 58.73 | 68.41 | 73.73 | 78.18 | 96.29 | 106.85 |
| 1/3 | Naïve | 42.74 | 58.79 | 67.61 | 76.87 | 81.46 | 84.85 | 103.78 | 124.78 |
| 1/3 | Optimal | 42.67 | 58.60 | 67.61 | 76.89 | 81.48 | 84.95 | 103.84 | 124.70 |
| 1 | Optimal | 91.15 | 128.12 | 148.33 | 166.68 | 185.06 | 185.42 | 263.01 | 982.71 |

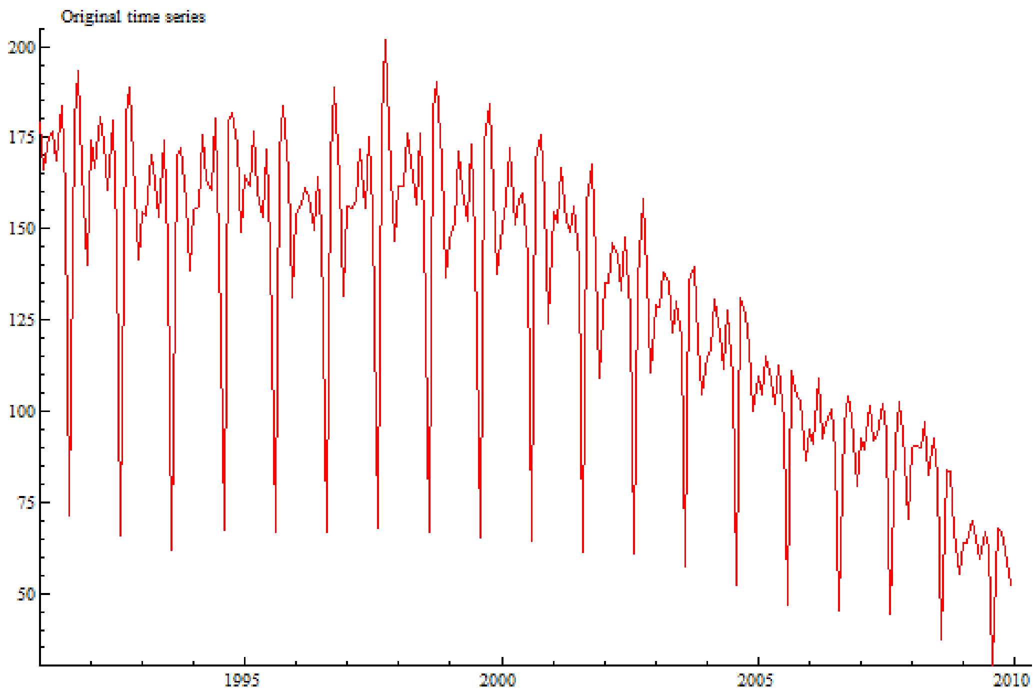
The principal results of the rolling forecast experiment are the following: the worst out of sample performance with an AR model for $\Delta_{12}y_t(\lambda)$ arises in the case the series is analysed in its original scale. For the levels, the logarithmic forecasts fare better at shorter horizons, but are outperformed by the 3rd root forecasts al longer horizons. The differences in performance are statistically significant if one conducts a Diebold Mariano test. Indeed taking for example as time horizon $h = 1$ we have that DM of log versus level is equal to $|5, 58| > 1, 96$ and that DM third root vs logs is equal to $|5, 53| > 1, 96$, so in both cases we reject the null at 5%.

Table 2.2: France Wearing: Diebold-Mariano test of equal forecast accuracy

| | Forecast horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| log vs lev | 5.58** | 4.81** | 3.69** | 3.21** | 3.06** | 2.79** | 2.24** | 2.19** |
| 3rd root vs log | 5.53** | 4.84** | 3.75** | 3.34** | 3.26** | 3.04** | 2.47** | 2.24** |

Figure 2.2 plots the monthly index of industrial production for the manufacture of textile sector (January 1990 - April 2010). The series presents a marked seasonal pattern with high amplitude and it is full of ups and downs movements.

Figure 2.2: France, industrial production index: manufacture of textile.



Moreover the table below shows a similar pattern to the previous one. Indeed the logarithmic transformation prevails for shorter horizons for both levels and growth rates, instead the third root transform seems to fit better for longer horizons.

Table 2.3: France Textile: rolling forecast experiment. Mean square forecast error.

| | | Levels $y_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 27.61 | 30.57 | 36.02 | 48.59 | 48.36 | 53.30 | 72.79 | 154.38 |
| 0 | Optimal | 27.75 | 30.68 | 36.19 | 41.58 | 48.59 | 53.57 | 73.37 | 157.53 |
| 1/3 | Naïve | 28.57 | 32.19 | 37.47 | 42.65 | 49.81 | 54.62 | 72.22 | 152.98 |
| 1/3 | Optimal | 28.66 | 32.25 | 37.57 | 42.76 | 49.93 | 54.77 | 72.59 | 154.99 |
| 1 | Optimal | 35.38 | 40.01 | 44.02 | 48.92 | 56.90 | 60.38 | 75.68 | 167.42 |
| | | Growth rates $g_t$ | | | | | | | |
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 17.31 | 21.30 | 26.78 | 32.38 | 39.28 | 44.94 | 55.35 | 55.89 |
| 0 | Optimal | 17.36 | 21.31 | 26.81 | 32.40 | 39.31 | 44.98 | 55.73 | 56.64 |
| 1/3 | Naïve | 19.42 | 24.24 | 29.72 | 35.48 | 42.51 | 48.21 | 54.23 | 54.60 |
| 1/3 | Optimal | 19.41 | 24.19 | 29.65 | 35.40 | 42.40 | 48.09 | 54.47 | 55.12 |
| 1 | Optimal | 37.91 | 46.25 | 47.99 | 55.24 | 59.72 | 63.32 | 61.06 | 62.52 |

Looking at the DM test results we can see how for the time horizon $h = 1$(logs vs levels) it is equal to $|2,36| > 1.96$ and it is equal to $|2,38| > 1,96$ for third root vs logs, so in both cases we reject the null at 5% level.

Table 2.4: France Textile: Diebold-Mariano test of equal forecast accuracy

| | Forecast horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| log vs lev | 2.36** | 2.22** | 2.59** | 1.94 | 2.80** | 2.67** | 0.90 | 0.70 |
| 3rd root vs log | 2.37** | 4.84** | 3.75** | 3.34** | 3.26** | 3.04** | 1.32 | 0.88 |

Figure 2.3 plots the monthly index of industrial production for the leather sector (January 1990 - April 2010). The series presents a downward trend and a marked seasonal pattern with high amplitude.

Figure 2.3: France, industrial production index: manufacture of leather.

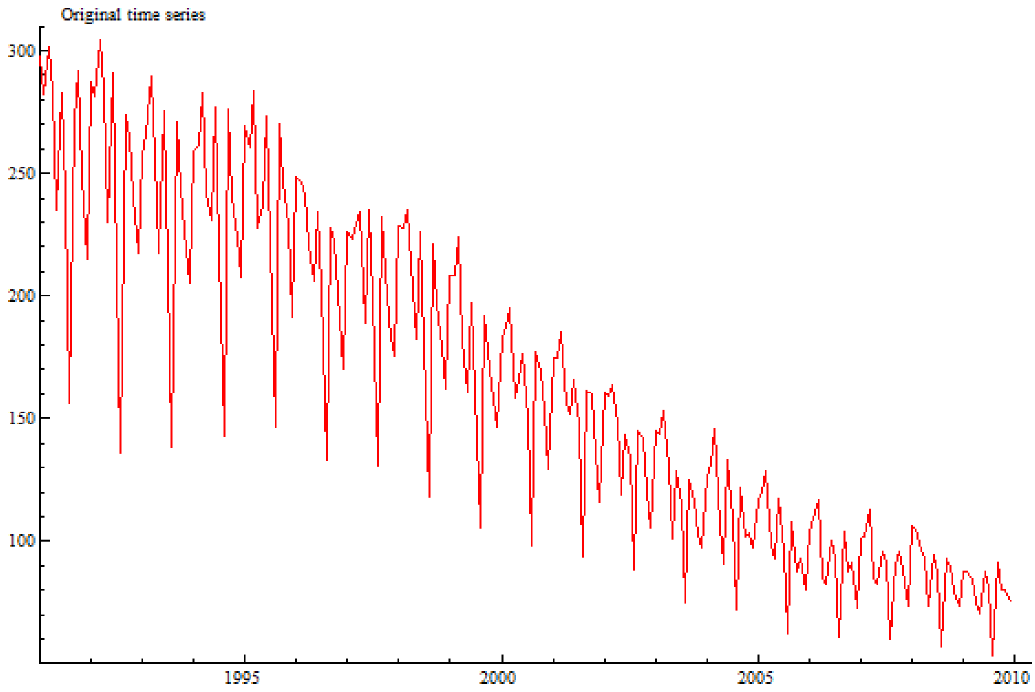


Table 2.5: France leather: rolling forecast experiment. Mean square forecast error.

| $\lambda$ | Type | Levels $y_t$ Forecast horizon 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Naïve | 58.21 | 63.40 | 74.86 | 78.68 | 86.62 | 87.52 | 80.59 | 156.29 |
| 0 | Optimal | 58.58 | 63.79 | 75.39 | 79.22 | 87.18 | 88.10 | 81.20 | 160.79 |
| 1/3 | Naïve | 59.42 | 65.17 | 75.83 | 79.90 | 88.10 | 89.03 | 81.41 | 160.44 |
| 1/3 | Optimal | 59.62 | 62.35 | 76.07 | 80.13 | 88.34 | 89.27 | 81.63 | 162.74 |
| 1 | Optimal | 71.14 | 76.15 | 83.80 | 84.76 | 94.76 | 97.54 | 92.44 | 210.06 |

| $\lambda$ | Type | Growth rates $g_t$ Forecast horizon 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Naïve | 25.24 | 28.37 | 32.08 | 33.28 | 36.48 | 37.25 | 38.02 | 37.91 |
| 0 | Optimal | 25.21 | 28.28 | 32.00 | 33.19 | 36.38 | 37.16 | 37.99 | 38.16 |
| 1/3 | Naïve | 26.07 | 30.10 | 33.78 | 35.82 | 38.85 | 39.91 | 40.64 | 42.46 |
| 1/3 | Optimal | 25.97 | 29.93 | 33.60 | 35.61 | 38.64 | 39.69 | 40.43 | 42.37 |
| 1 | Optimal | 36.61 | 41.76 | 44.55 | 46.07 | 51.18 | 53.20 | 56.01 | 73.57 |

The table above presents some interesting results. Looking at the levels it's possible to see how the logarithmic transformation by using the naive forecast appears to be suitable for all the time horizons except for $h = 2$ where the minimum MSE corresponds to the third root transform by using the optimal forecast. Instead for the growth rates the logarithmic transformation with the optimal forecast is the best option except for $h = 24$ where the same transformation prevails but with a different forecast specification : the naive one. The DM test statistics log versus levels and third root versus logs are respectively equal to $|3.94| > 1.96$ and $|3.81| > 1.96$. Again we reject the null at the 5% level.

Table 2.6: France leather: Diebold-Mariano test of equal forecast accuracy

| Type | Forecast horizon 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
|---|---|---|---|---|---|---|---|---|
| log vs lev | 3.49** | 3.79** | 3.54** | 3.18** | 3.12** | 3.05** | 2.57** | 2.03** |
| 3rd root vs log | 3.81** | 3.96** | 3.78** | 3.40** | 3.23** | 3.20** | 2.78** | 2.12** |

### 2.7.2 Germany industrial production index

Figure 2.4 shows the monthly index of industrial production for the leather sector (January 1990 - April 2010). The series presents a downward trend with some business fluctuations as a sign of a marked seasonal pattern.

Figure 2.4: Germany, industrial production index: manufacture of leather.

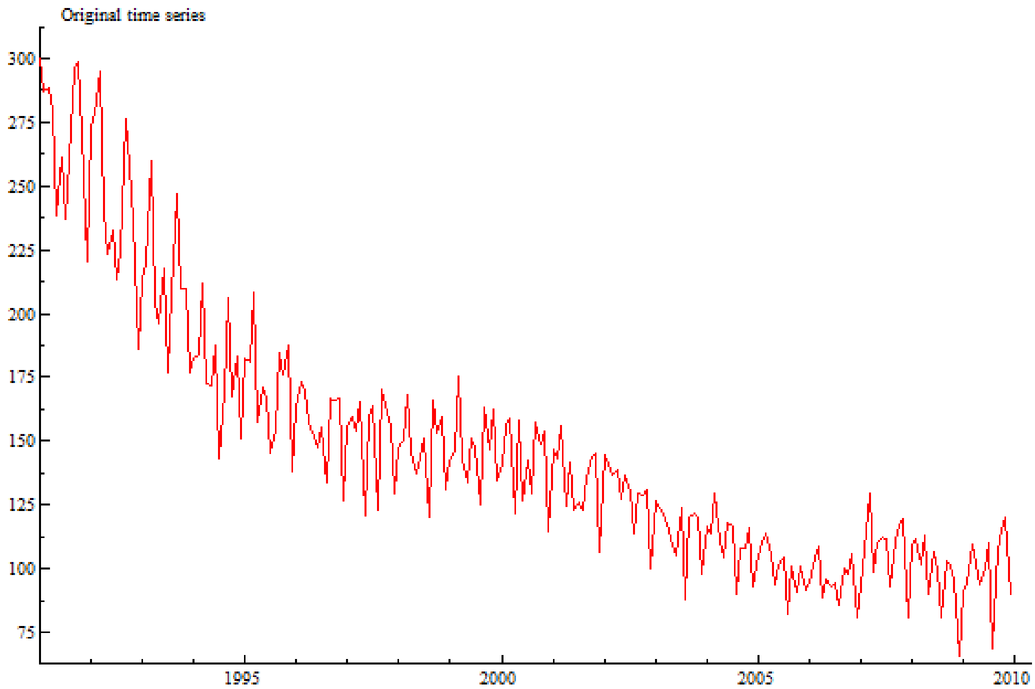


Table 2.7: Germany leather: rolling forecast experiment. Mean square forecast error.

| | | Levels $y_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 94.91 | 92.83 | 109.60 | 118.77 | 127.97 | 134.64 | 159.35 | 269.64 |
| 0 | Optimal | 94.51 | 91.88 | 108.53 | 117.63 | 126.69 | 133.32 | 158.21 | 265.39 |
| 1/2 | Naïve | 94.96 | 95.83 | 114.79 | 124.23 | 135.96 | 145.40 | 174.73 | 335.35 |
| 1/2 | Optimal | 94.60 | 95.02 | 113.80 | 123.11 | 134.69 | 144.05 | 173.29 | 329.57 |
| 1 | Optimal | 100.72 | 100.88 | 122.58 | 134.37 | 148.40 | 159.13 | 200.28 | 451.09 |
| | | Growth rates $g_t$ | | | | | | | |
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 65.20 | 68.99 | 85.83 | 93.55 | 99.07 | 106.63 | 129.85 | 112.33 |
| 0 | Optimal | 64.66 | 68.07 | 84.68 | 92.35 | 97.78 | 105.29 | 128.74 | 110.35 |
| 1/2 | Naïve | 63.86 | 69.62 | 88.55 | 96.11 | 103.65 | 113.20 | 140.35 | 126.06 |
| 1/2 | Optimal | 63.51 | 68.98 | 87.65 | 95.11 | 102.56 | 112.03 | 139.14 | 124.23 |
| 1 | Optimal | 66.83 | 72.10 | 92.97 | 101.90 | 110.83 | 121.65 | 156.97 | 154.93 |

Notice that for this series the square root transformation is the right choice depending from the optimal value of the transformation parameter which is $\lambda = \frac{1}{2}$. From the table is clear how the logarithmic transform prevails at all time horizons for the levels taking as reference the optimal forecast. For the growth rates instead the pattern changes a little bit. Indeed the logarithm one by using the optimal forecast is preferred for the overall time horizons except for $h = 1$ where the square root transform with the optimal forecast appears more suitable and for $h = 5$ where the logarithm ,by adopting the naive forecast, is chosen as the best option. The DM test for log vs level is equal to $|0.59| < 1,96$ meaning that we accept the null at 5% level. Instead the same test for sqrt transform vs logs is equal to $|1,98| > 1.96$, such that we reject the null at 5% level.

Table 2.8: Germany leather: Diebold-Mariano test of equal forecast accuracy

| | Forecast horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| log vs lev | 0.59 | 1.06 | 2.94** | 2.87** | 3.14** | 3.50** | 2.69** | 2.02** |
| sqrt root vs log | 1.98** | 2.21** | 3.63** | 3.52** | 3.31** | 3.20** | 2.53** | 2.07** |

### 2.7.3 UK industrial production index

Figure 2.5 shows the monthly index of industrial production for the manufacture of wearing appeal sector (January 1990 - April 2010). The series presents a not well defined downward trend and many fluctuations with peaks underlining the dominance of the seasonality component in some relevant periods.

Figure 2.5: UK, industrial production index: manufacture of wearing apparel.

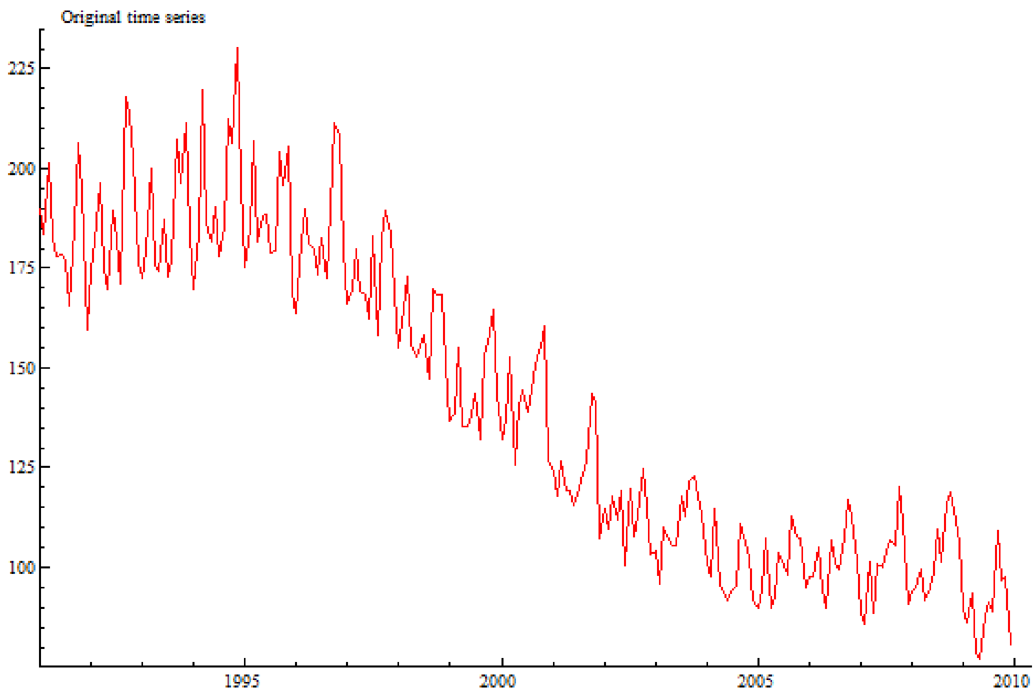


Table 2.9: UK wearing: rolling forecast experiment. Mean square forecast error.

| | | Levels $y_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 55.43 | 69.30 | 78.14 | 89.62 | 96.88 | 102.56 | 115.88 | 226.19 |
| 0 | Optimal | 55.80 | 69.73 | 78.64 | 90.21 | 97.51 | 103.26 | 116.77 | 228.41 |
| 1/2 | Naïve | 56.93 | 72.63 | 82.47 | 93.60 | 100.85 | 105.99 | 120.64 | 240.61 |
| 1/2 | Optimal | 57.05 | 72.72 | 82.57 | 93.73 | 100.97 | 106.13 | 120.82 | 240.55 |
| 1 | Optimal | 59.74 | 76.66 | 87.57 | 98.55 | 106.30 | 111.82 | 128.24 | 271.88 |
| | | Growth rates $g_t$ | | | | | | | |
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 40.75 | 49.39 | 54.75 | 62.86 | 66.78 | 71.10 | 79.01 | 81.11 |
| 0 | Optimal | 40.90 | 49.39 | 54.68 | 62.78 | 60.64 | 70.97 | 78.87 | 80.71 |
| 1/2 | Naïve | 41.81 | 51.87 | 58.01 | 66.09 | 70.59 | 74.50 | 84.12 | 86.35 |
| 1/2 | Optimal | 41.82 | 51.76 | 57.83 | 65.87 | 70.32 | 74.22 | 83.76 | 85.69 |
| 1 | Optimal | 43,92 | 55.09 | 62.00 | 70.64 | 75.73 | 80.00 | 92.00 | 98.49 |

For the levels the logarithmic transformation prevails in all time horizon by taking the naive forecast as better specification. For the growth rates instead the transformation chosen is the same but with the choice of the optimal forecast for all time horizons except for the first one. The DM test for logs versus levels is equal to $|3.16| > 1.96$ and for sqrt vs logs is equal to $|3.28| > 1.96$, such that in both cases we reject the null at 5% level.

Table 2.10: UK wearing: Diebold-Mariano test of equal forecast accuracy

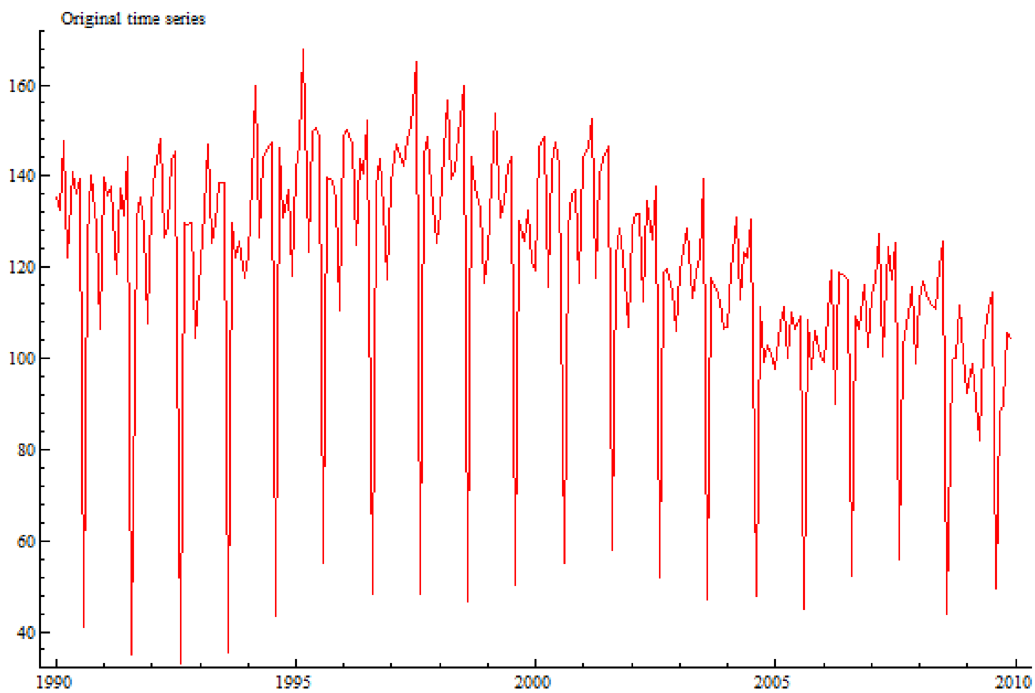
| | Forecast horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| log vs lev | 3.16** | 3.24* | 3.51** | 2.69** | 2.39** | 2.23** | 2.03** | 1.73 |
| sqrt root vs log | 3.28** | 3.34** | 3.17** | 2.56** | 2.34** | 2.22** | 1.98** | 1.78 |

### 2.7.4 Italy industrial production index

Figure 2.6 represents the monthly index of industrial production for the manufacture of wearing, textile and leather sectors (January 1990 - December 2009). The series presents a remarkable pattern of seasonality and sizeable fluctuations with high amplitude.

The square root transform acts better than the others. In particular for the

Figure 2.6: Italy, industrial production index: manufacture of textile, wearing and leather.




levels we choose the square root transformation principally at the short horizons and for $h = 24$ by selecting the naive forecast. For the remaining part, the log-

Table 2.11: Italy wearing, textile and leather: rolling forecast experiment. Mean square forecast error.

| $\lambda$ | Type | Levels $y_t$ Forecast horizon 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Naïve | 44.16 | 52.15 | 59.25 | 61.97 | 63.75 | 63.95 | 84.69 | 176.41 |
| 0 | Optimal | 44.58 | 52.66 | 59.89 | 62.72 | 64.66 | 67.92 | 86.32 | 183.36 |
| 1/2 | Naïve | 39.35 | 48.56 | 55.30 | 61.34 | 63.14 | 68.52 | 84.87 | 175.14 |
| 1/2 | Optimal | 39.41 | 48.61 | 55.39 | 61.46 | 63.27 | 68.68 | 85.35 | 177.17 |
| 1 | Optimal | 40.80 | 50.04 | 56.32 | 65.38 | 68.18 | 72.41 | 87.06 | 180.27 |
| | | Growth rates $g_t$ | | | | | | | |
| $\lambda$ | Type | Forecast horizon 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 38.63 | 44.29 | 49.06 | 56.76 | 55.05 | 55.06 | 64.61 | 64.08 |
| 0 | Optimal | 38.74 | 44.38 | 49.18 | 56.88 | 55.27 | 55.34 | 65.48 | 65.44 |
| 1/2 | Naïve | 35.52 | 41.71 | 46.88 | 62.54 | 59.04 | 57.78 | 65.67 | 64.12 |
| 1/2 | Optimal | 35.41 | 41.54 | 46.71 | 62.23 | 58.73 | 57.60 | 65.86 | 64.47 |
| 1 | Optimal | 42.04 | 46.59 | 51.43 | 81.05 | 74.40 | 66.12 | 69.59 | 68.03 |

arithmic transformation seems to be the right choice. For the growth rates the situation is a little bit different. The square root transform with the optimal forecast prevails for the first three time horizons instead for the others the logarithmic one wins by choosing the naive predictor. The DM test results seem to be significant at 5% for the short time horizons regarding the comparison between the square root transform and the logarithmic one.

Table 2.12: Italy wearing: Diebold-Mariano test of equal forecast accuracy

| Type | Forecast horizon 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
|---|---|---|---|---|---|---|---|---|
| log vs lev | 0.61 | 0.53 | 0.66 | 1.11 | 1.22 | 1.50 | 1.00 | 0.74 |
| sqrt root vs log | 2.57** | 2.62** | 2.04** | 1.19 | 1.36 | 1.88 | 1.10 | 0.95 |

Figure 2.7 exhibits the monthly index of industrial production for the electronic sector (January 1990 - December 2009). The series presents a similar pattern of seasonality to the previous one and suggests the presence of a downward trend.

Figure 2.7: Italy, industrial production index: electronics industry.

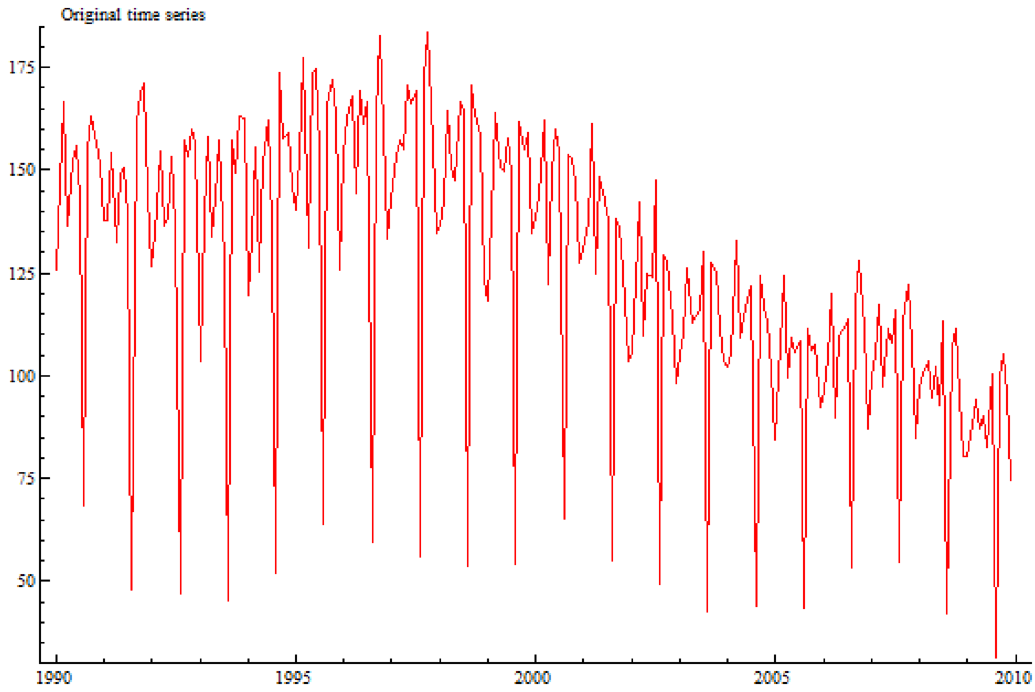


Table 2.13: Italy electronics industry: rolling forecast experiment. Mean square forecast error.

| | | Levels $y_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 64.12 | 77.69 | 77.36 | 75.90 | 76.01 | 77.87 | 88.04 | 209.00 |
| 0 | Optimal | 64.41 | 78.35 | 78.28 | 77.10 | 77.35 | 79.36 | 89.89 | 216.87 |
| 1/2 | Naïve | 59.96 | 74.33 | 73.89 | 73.99 | 74.68 | 76.83 | 86.09 | 204.42 |
| 1/2 | Optimal | 60.00 | 74.49 | 74.14 | 74.33 | 75.06 | 77.26 | 86.62 | 206.52 |
| 1 | Optimal | 60.20 | 75.49 | 75.06 | 75.26 | 76.99 | 78.52 | 87.15 | 211.47 |
| | | Growth rates $g_t$ | | | | | | | |
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 45.45 | 55.05 | 56.49 | 56.33 | 56.35 | 57.88 | 64.69 | 71.28 |
| 0 | Optimal | 45.42 | 55.15 | 56.73 | 56.72 | 56.84 | 58.44 | 65.52 | 72.11 |
| 1/2 | Naïve | 42.92 | 53.23 | 54.70 | 56.12 | 56.34 | 58.55 | 65.52 | 73.03 |
| 1/2 | Optimal | 42.77 | 53.08 | 54.61 | 56.08 | 56.32 | 58.56 | 65.62 | 73.00 |
| 1 | Optimal | 46.90 | 58.20 | 59.51 | 61.06 | 62.02 | 64.43 | 71.64 | 83.59 |

The square root transformation prevails for all the time horizon for the levels by using the optimal predictor. For the growth rates instead notice that for short time the square root transform with the optimal forecast dominates ,for longer time the logarithm one appears to be more suitable with the naive forecast. The DM test confirms the importance of the square root transformation for the series analyzed especially for $h = 4$, $h = 5$ and $h = 6$.

Table 2.14: Italy electronics: Diebold-Mariano test of equal forecast accuracy

| | Forecast horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| log vs lev | 0.25 | 0.68 | 0.70 | 1.19 | 1.42 | 1.54 | 1.11 | 1.21 |
| sqrt root vs log | 1.11 | 1.79 | 1.74 | 1.98** | 2.06** | 2.04** | 1.36 | 1.40 |

### 2.7.5 Italy tourism demand

Figure 2.8 represents the monthly tourism demand for Italy relates to the number of total arrivals in hotels and complementary structures (January 1990 - October 2009). The series presents an upward trend with high seasonal frequencies.

Figure 2.8: Italy, tourism demand: number of total arrivals in hotels plus complementary structures.

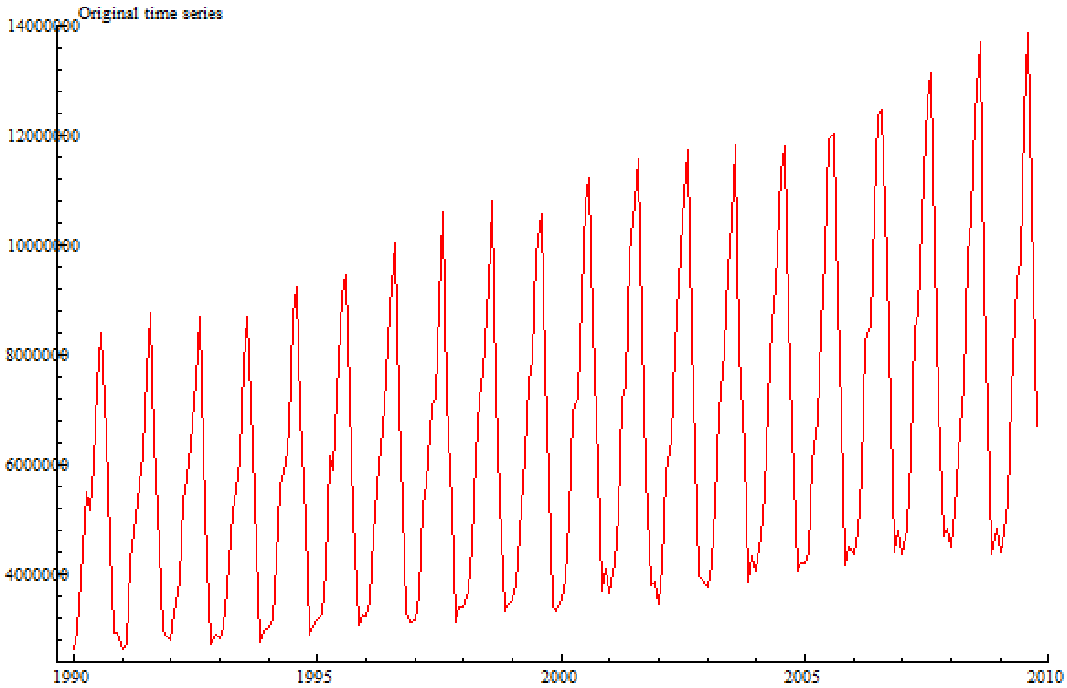


Table 2.15: Italy tourism demand: rolling forecast experiment. Mean square forecast error.

| | | Levels $y_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 98.82 | 91.04 | 114.00 | 128.70 | 108.17 | 110.80 | 117.45 | 222.01 |
| 0 | Optimal | 99.07 | 92.31 | 114.82 | 128.53 | 108.15 | 110.89 | 118.76 | 225.72 |
| 1/2 | Naïve | 90.85 | 86.80 | 96.13 | 104.32 | 97.89 | 102.02 | 115.37 | 212.19 |
| 1/2 | Optimal | 90.61 | 86.92 | 96.35 | 104.74 | 97.85 | 102.47 | 116.21 | 212.40 |
| 1 | Optimal | 91.86 | 88.41 | 95.01 | 95.44 | 97.64 | 104.72 | 118.76 | 226.81 |
| | | Growth rates $g_t$ | | | | | | | |
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 22.62 | 21.06 | 23.35 | 24.50 | 23.69 | 25.52 | 27.30 | 25.03 |
| 0 | Optimal | 22.63 | 21.06 | 23.36 | 24.53 | 23.72 | 25.55 | 27.41 | 25.18 |
| 1/2 | Naïve | 21.53 | 20.63 | 21.76 | 22.92 | 23.04 | 24.95 | 27.97 | 24.94 |
| 1/2 | Optimal | 21.54 | 20.64 | 21.77 | 22.93 | 23.06 | 24.97 | 28.03 | 25.02 |
| 1 | Optimal | 23.38 | 22.20 | 23.56 | 24.04 | 25.81 | 27.38 | 30.22 | 27.02 |

The table above gives a clear evidence of the transformation adopted. In fact, for the levels the square root is the best option especially for longer horizons and for $h = 1$ and $h = 2$. In the other cases no transformation is required. For the growth rates the square root again prevails. The predictor more used is the naive. The DM test confirms the dominance of the square root transformation for longer horizons.

Table 2.16: Italy tourism demand: Diebold-Mariano test of equal forecast accuracy

| | Forecast horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| log vs lev | 0.33 | 0.52 | 0.09 | 0.17 | 0.78 | 0.83 | 1.63 | 1.25 |
| sqrt root vs log | 1.46 | 1.19 | 1.32 | 0.72 | 1.83 | 2.04** | 2.28** | 2.22** |

### 2.7.6 France Tourism demand

Figure 2.9 represents the monthly tourism demand for France (January 1994-November 2009) in terms of total nights spent in hotels and complementary structures by residents and non residents. The series shows an upward trend until 2000 and then it tends to stabilize. The seasonal periodicity is dominant with high fluctuations.

Figure 2.9: France, tourism demand: number of total nights spent in hotels plus complementary structures.

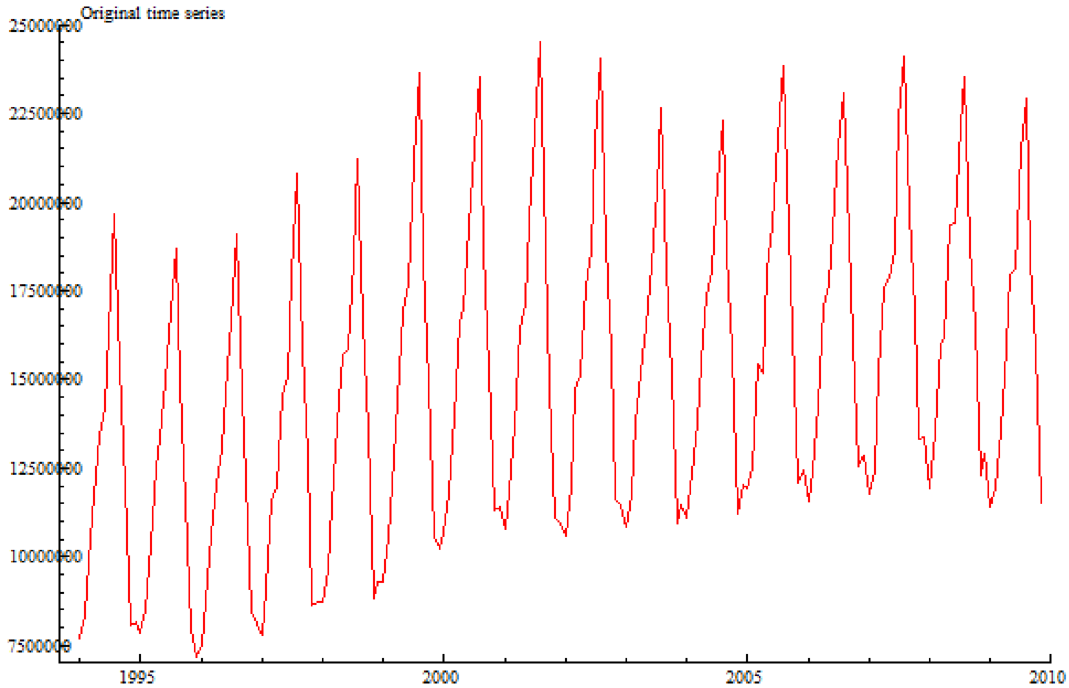


Table 2.17: France tourism demand: rolling forecast experiment. Mean square forecast error.

| | | Levels $y_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 38.88 | 49.47 | 62.14 | 75.77 | 84.38 | 93.92 | 98.47 | 277.35 |
| 0 | Optimal | 39.10 | 49.37 | 62.97 | 75.58 | 85.11 | 94.42 | 99.80 | 285.35 |
| 1/2 | Naïve | 38.57 | 48.93 | 59.88 | 69.53 | 78.28 | 86.96 | 92.52 | 242.38 |
| 1/2 | Optimal | 39.75 | 48.47 | 59.59 | 70.95 | 78.90 | 86.95 | 92.51 | 245.67 |
| 1 | Optimal | 39.33 | 50.39 | 62.90 | 73.52 | 78.45 | 83.94 | 88.99 | 209.02 |
| | | Growth rates $g_t$ | | | | | | | |
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 16.19 | 20.14 | 24.59 | 28.45 | 31.63 | 34.58 | 37.40 | 36.40 |
| 0 | Optimal | 16.27 | 20.23 | 24.73 | 28.60 | 31.81 | 34.80 | 37.80 | 37.06 |
| 1/2 | Naïve | 16.09 | 19.45 | 23.27 | 26.71 | 29.66 | 32.88 | 36.10 | 33.86 |
| 1/2 | Optimal | 16.14 | 19.51 | 23.34 | 26.78 | 29.75 | 32.98 | 36.30 | 34.19 |
| 1 | Optimal | 16.26 | 20.13 | 25.18 | 29.96 | 32.17 | 33.76 | 35.73 | 31.46 |

Notice that for levels there is no unique transformation that dominates. For $h = 1$ and $h = 2, 3$ the logarithmic transformation is preferred respectively by employing the naive and the optimal predictor. For $h = 4$ and $h = 5$ the square root is used with respectively the optimal and the naive forecast. For longer horizon no transformation seem to be required. For the growth rates it's easy to see how the square root transformation by including the naive forecast is chosen. The DM test confirms the dominance of the square root transform for $h = 3$, $h = 4$ and $h = 5$.

Table 2.18: France tourism demand: Diebold-Mariano test of equal forecast accuracy

| | Forecast horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| log vs lev | 0.08 | 0.01 | 0.38 | 0.71 | 0.23 | 0.34 | 1.31 | 1.54 |
| sqrt root vs log | 1.56 | 1.36 | 2.14** | 1.97** | 2.92** | 1.19 | 0.39 | 1.47 |

### 2.7.7 Germany Tourism demand

Figure 2.10 represents the monthly tourism demand for Germany (January 1994- November 2009) in terms of total nights spent in hotels and complementary structures by residents and non residents. The series shows an upward trend with clear seasonal components.

Figure 2.10: Germany, tourism demand: number of total nights spent in hotels plus complementary structures.

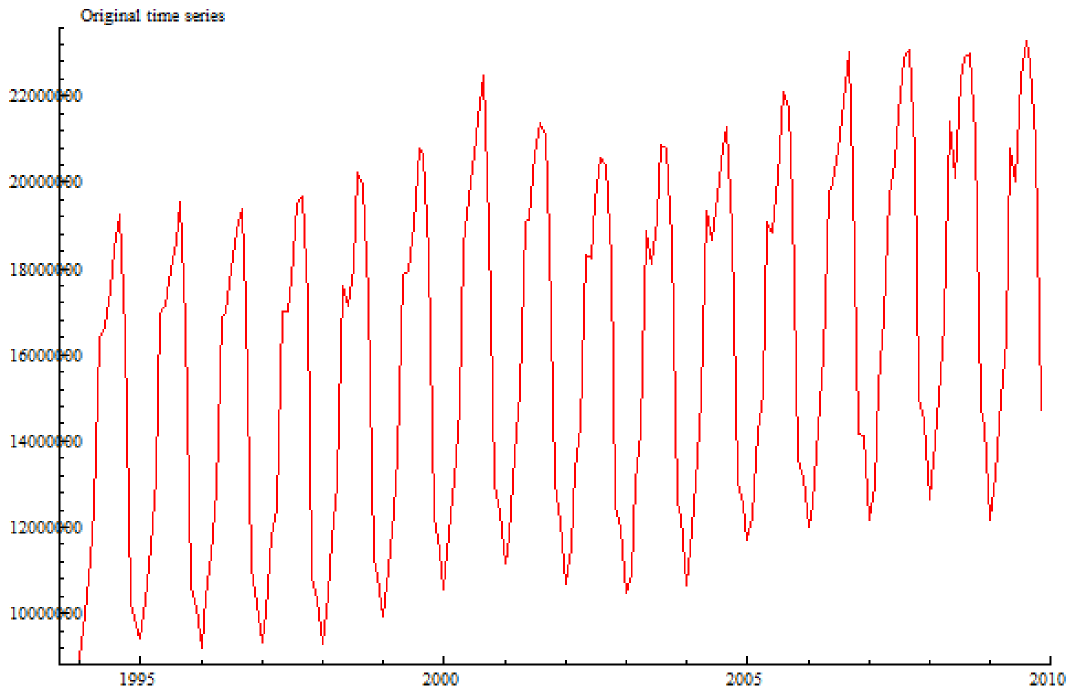


Table 2.19: Germany tourism demand: rolling forecast experiment. Mean square forecast error.

| | | Levels $y_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 24.90 | 27.98 | 30.85 | 34.63 | 37.09 | 40.17 | 61.41 | 169.96 |
| 0 | Optimal | 24.46 | 27.22 | 30.28 | 34.44 | 37.58 | 40.42 | 61.80 | 168.20 |
| 1/3 | Naïve | 23.98 | 26.71 | 29.52 | 32.40 | 35.50 | 38.46 | 64.31 | 198.62 |
| 1/3 | Optimal | 23.92 | 26.29 | 29.44 | 32.89 | 35.76 | 38.76 | 64.75 | 196.12 |
| 1 | Optimal | 24.16 | 27.88 | 30.80 | 34.65 | 37.80 | 40.11 | 65.58 | 210.70 |
| | | Growth rates $g_t$ | | | | | | | |
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 8.99 | 10.13 | 11.28 | 12.41 | 13.47 | 14.59 | 22.22 | 23.11 |
| 0 | Optimal | 9.00 | 10.12 | 11.26 | 12.37 | 13.43 | 14.54 | 22.14 | 22.85 |
| 1/3 | Naïve | 8.70 | 9.78 | 10.93 | 12.06 | 13.08 | 14.29 | 22.99 | 25.95 |
| 1/3 | Optimal | 8.70 | 9.77 | 10.90 | 12.03 | 13.04 | 14.24 | 22.92 | 25.72 |
| 1 | Optimal | 9.37 | 10.40 | 11.79 | 13.59 | 15.27 | 16.26 | 22.89 | 26.32 |

Looking at the table above we can say that for levels the third root transformation, choosing as best predictor the naive, is predominant at short time horizon ($h = 4, 5, 6$) and for $h = 12$. For $h = 1$ no transformation is necessary. For $h = 24$ the best choice corresponds to the logarithmic transform by selecting the optimal forecast. Regarding the growth rates the third root transform is preferred for short horizon by using the optimal predictor. For longer periods the logarithmic wins. The DM test partially confirms the results obtained with the rolling forecast procedure.

Table 2.20: Germany tourism demand: Diebold-Mariano test of equal forecast accuracy

| | Forecast horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| log vs lev | 0.60 | 0.38 | 0.49 | 0.67 | 0.80 | 0.77 | 0.34 | 2.97** |
| 3rd root vs log | 1.28 | 1.08 | 1.03 | 1.07 | 2.02** | 2.16** | 0.02 | 2.31** |

### 2.7.8 UK Tourism demand

The figure below describes the monthly tourism demand in terms of number of arrivals for residents in hotels and complementary structures (January 1994 - November 2009). The series seems to be characterized by some relevant fluctuations probably due to the different flow of tourists according to the month chosen. Notice that at the beginning we have a downward trend with sizeable peak and immediately after a recovery. The seasonality component is still dominant.

Figure 2.11: UK, tourism demand: number of resident arrivals in hotels and complementary structures.

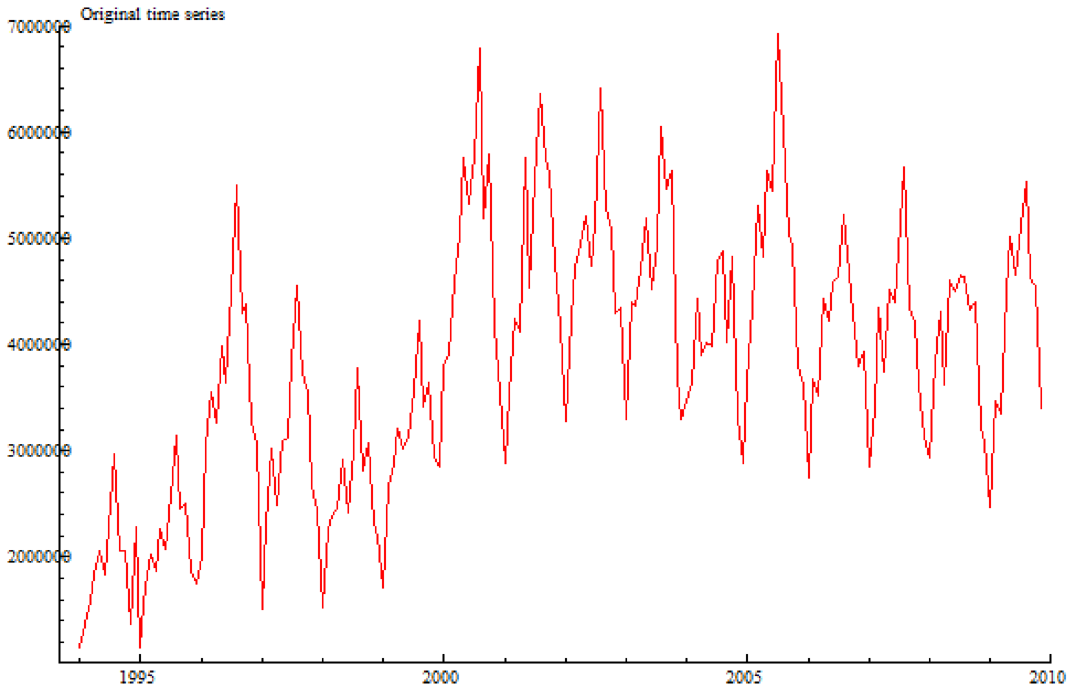


Table 2.21: UK tourism demand: rolling forecast experiment. Mean square forecast error.

| | | Levels $y_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 410.64 | 474.04 | 641.26 | 704.07 | 818.64 | 934.53 | 1058.32 | 1893.92 |
| 0 | Optimal | 419.05 | 486.59 | 666.47 | 738.45 | 871.10 | 999.61 | 1181.00 | 2937.18 |
| 1/2 | Naïve | 397.02 | 468.64 | 615.51 | 699.48 | 796.42 | 898.08 | 1025.22 | 1632.30 |
| 1/2 | Optimal | 398.95 | 469.74 | 618.20 | 704.42 | 804.51 | 906.66 | 1041.61 | 1817.68 |
| 1 | Optimal | 421.56 | 495.07 | 636.45 | 749.97 | 853.35 | 989.34 | 1036.52 | 1630.47 |
| | | Growth rates $g_t$ | | | | | | | |
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 318.31 | 340.82 | 435.76 | 467.31 | 526.67 | 591.71 | 675.77 | 326.71 |
| 0 | Optimal | 317.22 | 334.93 | 427.61 | 456.47 | 519.57 | 582.34 | 682.44 | 443.49 |
| 1/2 | Naïve | 296.89 | 328.22 | 416.44 | 463.43 | 523.11 | 584.47 | 662.77 | 300.96 |
| 1/2 | Optimal | 294.13 | 323.24 | 408.45 | 453.93 | 512.71 | 571.10 | 647.74 | 326.36 |
| 1 | Optimal | 307.59 | 342.34 | 427.89 | 503.43 | 562.25 | 644.02 | 666.14 | 310.49 |

The square root transform seems to be more suitable. For the levels the minimum MSE for each horizon, except for the longer one where no transformation is required, corresponds to the square root transformation by using the naive as predictor. For the growth rates the situation is similar. The square root prevails for all the time horizon. The performance obtained by implementing DM test show how the square root transform is required only for $h = 5$ and $h = 6$.

Table 2.22: UK tourism demand: Diebold-Mariano test of equal forecast accuracy

| | Forecast horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| log vs lev | 0.30 | 0.06 | 0.21 | 0.87 | 0.84 | 1.07 | 0.69 | 0.61 |
| sqrt root vs log | 0.81 | 1.27 | 1.03 | 1.85 | 2.92** | 2.25** | 0.75 | 0.71 |

### 2.7.9 Sicily Tourism demand

A regional dimension has also been analyzed. This allows a better comparison with national data from the European framework. The most interesting results concerns the number of night spent in hotels by residents and the number of total arrivals in hotels. Figure 2.12 describes the monthly tourism demand in terms of number of night spent in hotels by residents (January 1998 - December 2009). It underlines the presence of seasonality for each year analyzed. The series tends to fluctuate up and down according to the given periodicity.

Figure 2.12: Sicily, tourism demand: number of night spent in hotels by residents.

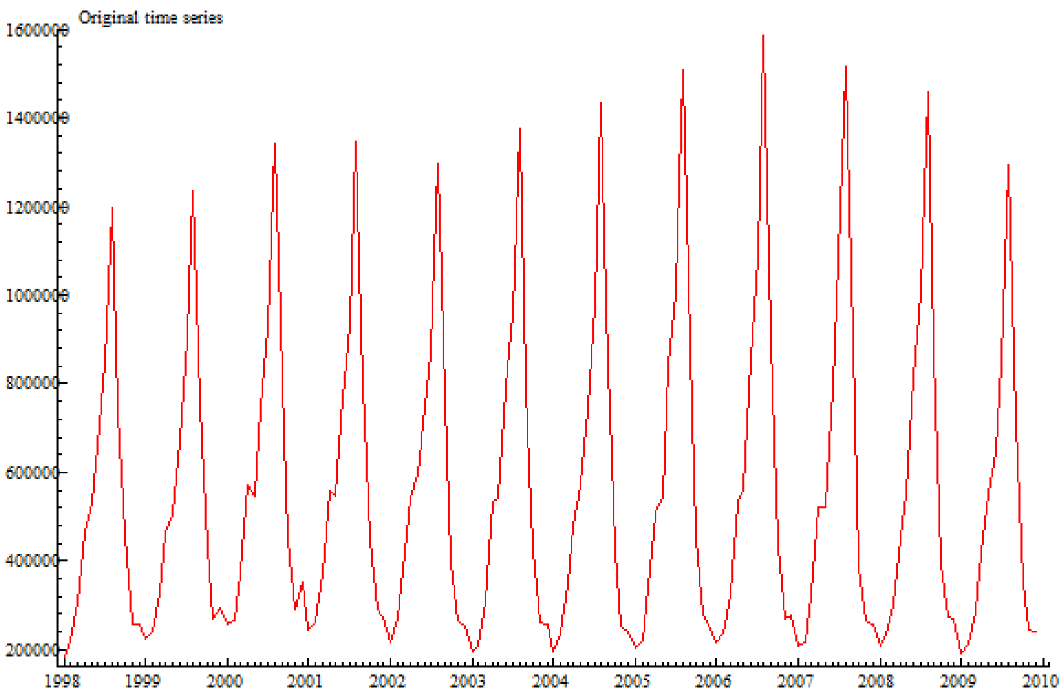


Table 2.23: Sicily tourism demand: rolling forecast experiment. Mean square forecast error.

| | | Levels $y_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 219.67 | 261.86 | 307.79 | 321.60 | 317.93 | 299.68 | 304.95 | 700.65 |
| 0 | Optimal | 220.31 | 261.48 | 305.50 | 319.50 | 315.94 | 297.99 | 304.89 | 696.88 |
| 1/3 | Naïve | 202.23 | 246.52 | 287.85 | 297.79 | 297.71 | 287.79 | 296.14 | 656.89 |
| 1/3 | Optimal | 398.95 | 469.74 | 618.20 | 704.42 | 804.51 | 906.66 | 1041.61 | 1817.68 |
| 1 | Optimal | 203.64 | 247.24 | 287.95 | 297.83 | 297.47 | 287.52 | 297.28 | 657.70 |

| | | Growth rates $g_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 57.06 | 61.09 | 61.15 | 62.12 | 61.23 | 60.03 | 63.30 | 61.32 |
| 0 | Optimal | 57.44 | 61.32 | 61.09 | 61.97 | 61.05 | 59.74 | 63.42 | 61.76 |
| 1/3 | Naïve | 55.54 | 59.69 | 59.41 | 59.46 | 60.32 | 60.19 | 63.87 | 62.46 |
| 1/3 | Optimal | 55.92 | 60.05 | 59.63 | 59.63 | 60.44 | 60.26 | 64.17 | 63.01 |
| 1 | Optimal | 61.52 | 70.42 | 69.67 | 63.06 | 61.01 | 62.46 | 70.26 | 70.44 |

The table above gives the following results: for the levels the third root transform, given tha naive as predictor, dominates for all time horizon. Instead for the growth rates the logarithmic transformation, given the naive as forecast, prevails for $h = 1, 12, 24$. The remaining part is characterized by the third root transform by using as best predictor the naive. The optimal for the logarithmic one is only considered for $h = 6$. The Dm test behaves in a similar way. Indeed, for longer horizons the third root transform seems to prevail, on the contrary for short horizons the logarithmic one dominates.

Table 2.24: Sicily tourism demand: Diebold-Mariano test of equal forecast accuracy

| | Forecast horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| log vs lev | 0.96 | 2.13** | 2.17** | 0.20 | 0.05 | 0.49 | 1.51 | 1.41 |
| 3rd root vs log | 1.34 | 1.24 | 1.39 | 1.01 | 0.22 | 0.70 | 2.23** | 2.70** |

Figure 2.13 shows the monthly tourism demand in term of number of total arrivals by residents (January 1998-December 2009). It seems to be characterized by an high degree of seasonality and a double peak in the interval between one year and the other.

Figure 2.13: Sicily, tourism demand: number of total arrivals in hotels.

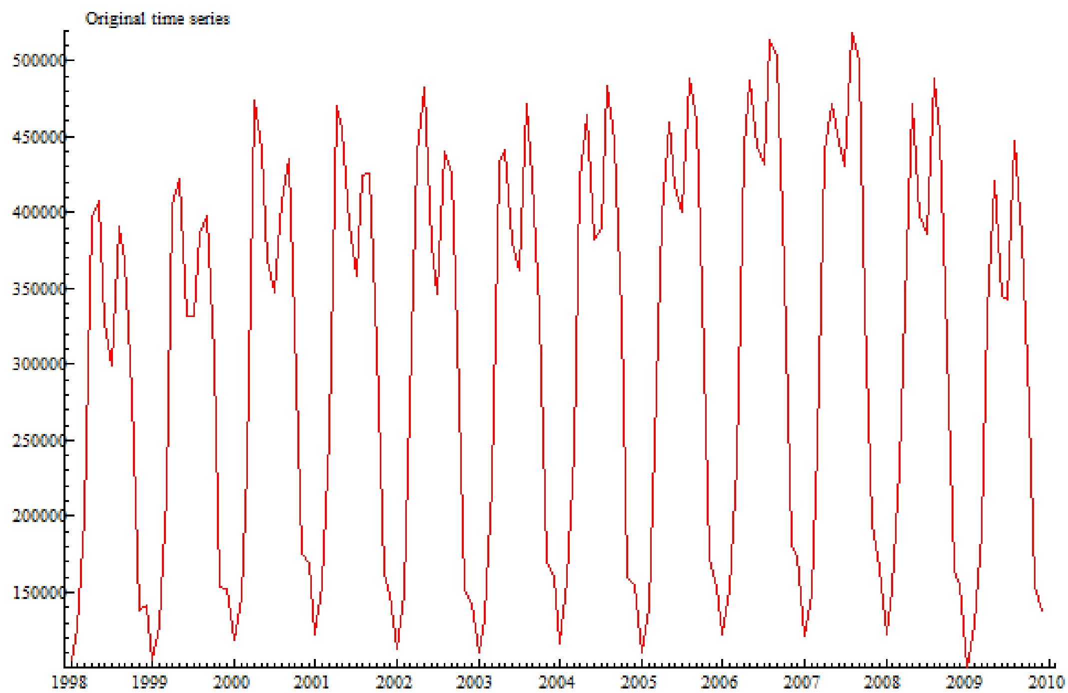


Table 2.25: Sicily tourism demand: rolling forecast experiment. Mean square forecast error.

| | | Levels $y_t$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 268.43 | 330.91 | 395.92 | 536.78 | 555.35 | 550.71 | 711.78 | 244.75 |
| 0 | Optimal | 268.60 | 330.80 | 395.86 | 536.16 | 554.23 | 550.15 | 719.43 | 248.38 |
| 1/3 | Naïve | 284.83 | 312.18 | 378.39 | 466.60 | 481.13 | 493.28 | 712.93 | 237.72 |
| 1/3 | Optimal | 284.17 | 312.39 | 378.17 | 467.82 | 482.36 | 494.67 | 716.36 | 239.35 |
| 1 | Optimal | 354.51 | 412.74 | 493.34 | 548.45 | 585.88 | 592.61 | 719.89 | 226.26 |
| | | Growth rates $g_t$ | | | | | | | |
| $\lambda$ | Type | Forecast horizon | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| 0 | Naïve | 30.63 | 39.31 | 41.70 | 54.10 | 55.94 | 59.95 | 89.96 | 92.20 |
| 0 | Optimal | 30.65 | 39.38 | 41.65 | 54.07 | 55.90 | 59.91 | 90.47 | 93.96 |
| 1/3 | Naïve | 34.00 | 35.71 | 39.31 | 47.37 | 52.41 | 59.81 | 86.64 | 93.03 |
| 1/3 | Optimal | 34.05 | 35.81 | 39.37 | 47.44 | 52.48 | 59.82 | 87.02 | 94.16 |
| 1 | Optimal | 52.88 | 59.61 | 62.22 | 66.60 | 76.98 | 80.88 | 88.30 | 104.40 |

From the table we can see that for levels the third root transform is preferred except for $h = 1$ where the logarithmic transformation wins by using the naive as predictor and $h = 24$ where no transformation is required. For growth rates again the third root one appears more suitable choosing the naive as principal forecast except for $h = 1$ and $h = 24$ where prevails the logarithmic transform given the naive as predictor. The DM test supports these results. Indeed the statistic for $h = 1$ of logs versus levels and of 3rd root versus logs is respectively equal to $|2.01| > 1.96$ and $|1.98| > 1.96$. Hence, we reject the null in both cases at 5% level.

Table 2.26: Sicily tourism demand: Diebold-Mariano test of equal forecast accuracy

| | Forecast horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 |
| log vs lev | 2.01** | 1.28 | 1.48 | 1.01 | 1.49 | 1.56 | 0.15 | 1.45 |
| 3rd root vs log | 1.98** | 1.75 | 1.86 | 1.85 | 1.93 | 1.74 | 0.17 | 1.48 |

## 2.8 Conclusions

From an economic point of view, the analysis shows how the sectors analyzed are particularly useful in making accurate forecasts. In fact, the industrial production measures changes in output for the industrial sector of the economy which includes manufacturing, mining, and utilities. Although these sectors contribute only a small portion of GDP (Gross Domestic Product), they are highly sensitive to interest rates and consumer demand. This makes the industrial production index an important tool for forecasting the short or medium run evolution of the GDP and the economic performance. Moreover it could be considered as the most important and widely analyzed high-frequency indicator, given the relevance of the manufacturing activity as a driver of the whole business cycle.

With regard to the tourist sector, all the transformations adopted could be particularly useful for measuring the growth of the tourism demand for different countries. In addition, it is well known how these forecasts help marketers and other managers in reducing the risk of decisions regarding the future.

For example, tourism marketers use demand forecasts to set marketing goals, either strategic or for the annual marketing plan, to explore potential markets as to the feasibility of persuading them to buy their product and the anticipated volume of these purchases and to simulate the impact of future events on demand, including alternative marketing programmes as well as uncontrollable developments such as the course of the economy and the actions of competitors.

On the other side policy makers use tourism demand forecasts to predict the economic, social-cultural, environmental consequences of visitors, to assess the potential impact of regulatory policies, such as price regulation and environmental quality controls, to project public revenues from tourism for the budgeting process and to ensure adequate capacity and infrastructure, including airports and airways, bridges and highways, and energy and water treatment utilities.

From all these illustrations, we learn that choosing the optimal value of the transformation parameter, on the basis of the fit within the sample, does not necessarily correspond to a gain in forecasting accuracy.

Another important piece of evidence is that the optimal predictor does not lead to an improvement with respect to the naïve predictor using the same $\lambda$. Moreover the logarithmic forecasts are better suited to shorter time interval; on the contrary more time means a sizeable gain in terms of predictive accuracy by choosing other transformations such as square root or third root.

These considerations underline the importance of working not only with series at its original scale. Also, they stress the relevance of the choice of the transformation parameter $\lambda$. The last point introduces the analysis of the next chapter where a non parametric methodology has been developed in order to produce the best choice for $\lambda$, whose value so far has been fixed exogenously.

# Chapter 3

# Understanding forecasting through a non parametric analysis

## 3.1 Introduction

The objective of this chapter is to provide a non parametric[1] complementary analysis to the one displayed in the previous chapter. Indeed, we may notice that the optimal value of the transformation parameter $\lambda$, previously described, corresponds exactly, in a non parametric framework, to the minimizer of the prediction error variance. For this purpose, starting from the consideration that the Box-Cox transformation belongs to the class of power normal family distributions and calling $z_t$ the power transformed variable, a fast non parametric method, based on the estimation of the prediction error variance for a normal stationary process, is proposed.

The knowledge of the spectrum seems to yield greater immediate insight into the structure of the process under observation. Hence, an analysis in the frequency domain is carried out. This involves the computation of the sample spectrum, sample frequencies , autocovariance generating functions and the log

[1]see Buckinsky (1995)

periodogram at sample frequencies. The latter is particularly useful for the estimation of the prediction error variance. In fact, it is well known from the literature that the prediction error variance can be estimated non parametrically by a bias-corrected geometric average of the periodogram.

Many authors have proposed different versions for the estimation of the prediction error variance. For example Hannan and Nicholls (1977) proposed an estimator which is strongly consistent and asymptotically normal, replacing the raw periodogramm ordinates (Davide and Jones, 1968) by their non overlapping averages of $m$ consecutive ordinates. The multi-step case is also considered compared to the one step forecasting procedure. Also a number of Monte Carlo simulation experiments are obtained in order to illustrate better the properties of the HN estimator of the prediction error variance and its reliability.

For the first experiment the model chosen is a simple AR(1) model. Each observations is generated from the model and the procedure is replicated 10,000 times. Moreover, in order to control the smoothness of the HN prediction error variance, we use different bandwidth. The results show the differences between the asymptotic variance and the MC estimate of the variance related to che choice of the bandwidth and the sample size.

A second MC experiment was carried out to assess the sampling distribution of the grid search estimator of $\lambda$ which is the minimiser of the prediction error variance of the normalised Box Cox transform. In this case we generate 1000 replicates of a seasonal random walk process making comparisons between the case of transformation and no transformation. The results show that the choice of the bandwidth is relevant in order to find a good compromise between the variance and the bias.

At the end it is interesting to see how this method behaves once applied to real data. Relevant time series are considered: monthly industrial production index for France relating to the manufacture of wearing apparel and monthly tourism demand both for Germany and Sicily ,respectively, in terms of total nights spent in hotels and complementary structures and number of total arrivals in hotels.

## 3.2 The normalized power transformation

Let $y_t$ be a time series, we consider the Box-Cox transformation defined as Box and Cox (1964)

$$y(\lambda) = \begin{cases} \frac{y_t^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ \\ ln(y_t) & \text{if } \lambda = 0 \end{cases} \tag{3.1}$$

defined for $y_t > 0$. Hence, different values of $\lambda$ implies different transformations. The value $\lambda = 1$ implies that the series is analyzed in its original scale, $\lambda = 0$ originates the logarithmic transformation. Other important special cases arise for fractional values of $\lambda$, e.g. the square root transform ($\lambda = 1/2$).

There are many advantages in using this kind of power transformation including the scale equivariance property in models with intercepts meaning that the transformation parameter is unaffected by rescaling and the fact that linear and logarithm transformations are included. The main drawback is that both the domain and the range of the transformation are bounded depending on the parameter $\lambda$. Usually, to compare different value of $\lambda$ it is necessary to evaluate the log- likelihood[2] in relation to that of the original observations $y$ which is given by:

$$L(\beta, \sigma^2, \lambda, y) = (2\pi\sigma^2)^{-\frac{n}{2}} exp \left\{ -\frac{(Y(\lambda) - X\beta)'(y(\lambda) - X\beta)}{2\sigma^2} \right\} J \tag{3.2}$$

[2]see Amemiya et al. (1981)

where $J$ is the Jacobian defined as

$$J = \Pi \left| \frac{\partial y_i(\lambda)}{\partial y_i} \right| \tag{3.3}$$

The Jacobian allows for the change of scale of the response due to the operation of the power transformation $y(\lambda)$.

The maximum likelihood estimates of the parameters $\beta$ and $\sigma^2$ are easily found as the solution for a least square problem with response $y(\lambda)$. By replacing the estimate obtained in the expression for the log likelihood, it's common to derive the expression for the profile log-likelihood defined as:

$$L_{max}(\lambda) = -\frac{n}{2} log 2\pi - \frac{n}{2} log + \hat{\sigma^2(\lambda)} + log J \tag{3.4}$$

An equivalent form for the partially maximized likelihood is found by working with the normalized transformation. The normalized power transformation[3] is obtained by dividing the power transformation $y(\lambda)$, by $\sqrt[n]{J}$ such that

$$z(\lambda) = \frac{y(\lambda)}{\sqrt[n]{J}} \tag{3.5}$$

where $J = \prod_{t=1}^{n} \left| \frac{\partial y_t(\lambda)}{\partial y_t} \right|$ is the Jacobian of the transformation, which is equal to $g_y^{\lambda-1}$, where $g_y = [\prod_{t=1}^{n} y_t]^{1/n}$ is the geometric average of the original observations (Atkinson, 1985).

Hence,

$$z_t(\lambda) = g_y^{1-\lambda} y_t(\lambda) \tag{3.6}$$

The greater advantage is that working with $z(\lambda)$ rather than $y(\lambda)$ neutralizes the effect of the change of scale.

[3] see Atkinson, (1973), Caroll, (1980), Caroll et al. (1981), De Bruin et al. (1999), Freeman et al. (2006), Guerrero (1993) and Schlesselman (1971)

## 3.3 Estimating the prediction error variance : a non parametric approach

Given $z_t$, the normalized power transformed variable, the purpose is to determine the optimal value of $\lambda$, i.e. the value of the transformation variable for which the process $z_t$ is more predictable. Assuming that there exists a stationary[4] representation for $z_t$, $u_t = \Delta(L)z_t, \quad t = 1, \ldots, n$; the one step ahead prediction error variance for $z_t$ is the same as that of $u_t$, since $u_t - \mathrm{E}(u_t|\mathcal{F}_{t-1}) = z_t - \mathrm{E}(z_t|\mathcal{F}_{t-1})$.

Notice that the above measure can be interpreted as a coefficient of determination, i.e. as the proportion of the variance of $u_t$ that can be predicted from knowledge of its past realization. Moreover letting $f(\omega)$ denote the spectral density of $u_t$, the one-step-ahead prediction error variance is defined as the geometric average of the spectral density, (Szegö-Kolmogorov formula),

$$\sigma^2 = 2\pi \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\omega) d\omega\right] \tag{3.7}$$

This formula gives an explicit relationship between the spectral density $f(\omega)$ and the variance of the noise sequence $u_t$.

Notice that if log $f(\omega)$ is not integrable this can be only because the integral diverges to $-\infty$, since $log f(\omega) \leq f(\omega)$. When this is so, it's possible to interpret the right side of the formula above as zero and the theorem continues to hold. Given that the spectrum is always positive and that the geometric average is no larger than the the arithmetic average, predictability is always in the range (0,1).

The prediction error variance[5] can be estimated non parametrically by a bias-corrected geometric average of the periodogram. Letting $\omega_j = \frac{2\pi j}{n}$, $j = 1, \ldots, [n/2]$, denote the Fourier frequencies, where $[\cdot]$ is the integral part of the argument, the

[4] see Grenander et al. (1952)

[5] see Grenander (1958) and Jones (1976)

periodogram is defined as

$$I(\omega_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} (u_t - \bar{u}) e^{-\imath \omega_j t} \right|^2 \tag{3.8}$$

with $\bar{u} = \frac{1}{n} \sum_t u_t$ denotes the sample spectrum.

Non parametric estimation of the prediction error variance is based on a fundamental large sample result given by Brockwell and Davies (1991), establishing that the periodogram is asymptotically distributed as a scaled chisquare random variable:

$$I(\omega_j) \sim \begin{cases} \frac{1}{2} f(\omega_j)\chi_2^2, & 0 < \omega_j < \pi \\ f(\omega_j)\chi_1^2, & \omega_j = \pm\pi \end{cases} \tag{3.9}$$

where $\chi_m^2$ denotes a chisquare random variable with $m$ degrees of freedom and $f(\omega)$ is the spectral density of $u_t$ at a given frequency $\omega$.

Davis and Jones (1968) instead proposed an estimator for $\sigma^2$

$$\hat{\sigma}^2 = \exp \left[ \frac{1}{n^*} \sum_{j=1}^{n^*} \ln I(\omega_j) + \gamma \right]. \tag{3.10}$$

where $n^*$ denote $n/2-1$, if $n$ is even, and $(n-1)/2$, if $n$ is odd and $\gamma \cong 0.577221$ is Euler's constant (minus the expectation of a log chi-squared variable). The need for the factor $\exp\gamma$ originates from the fact that, in large samples, $\mathrm{E}[\ln I(\omega_j)] = \ln f(\omega_j) - \gamma$, so that the log-periodogram is a biased estimator of the log-spectrum. Also the variance of $lnI(\omega_j) = \frac{\pi^2}{6}$.

They show also that $ln\hat{\sigma}^2$ is asymptotically normal

$$ln\hat{\sigma}^2 \sim N\left(\ln \sigma^2, \frac{\pi^2}{6[n/2]}\right) \tag{3.11}$$

Hannan and Nicholls (1977) proposed a more generalized version of the formula (10) by replacing the raw periodogram ordinates by their non overlapping averages of $m$ consecutive ordinates.

$$\hat{\sigma}^2(m) = m \exp\left[\frac{1}{M}\sum_{j=0}^{M-1}\sum_{k=1}^{m} \ln I(\omega_{jm+k}) - \psi(m)\right]. \tag{3.12}$$

where $M = [(n-1)/(2m)]$ and $\psi(m)$ is the digamma function. Notice that the estimator (10) is obtained in the case $m = 1$ and the bias correction increases as m increases. The large sample distributions of (12) and $\ln \hat{\sigma}^2(m)$ are, respectively

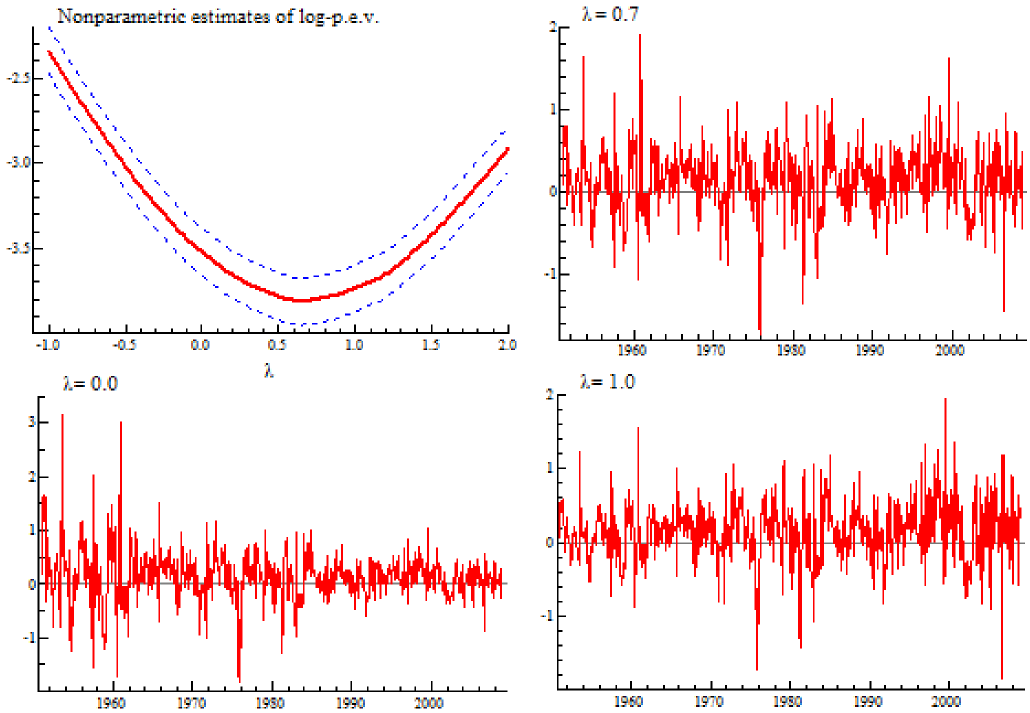
$$\hat{\sigma}^2(m) \sim N\left(\sigma^2, \frac{2\sigma^4 m\psi'(m)}{n}\right), \quad \ln \hat{\sigma}^2(m) \sim N\left(\sigma^2, \frac{2m\psi'(m)}{n}\right) \tag{3.13}$$

For a better explanation the figure below underlines the importance of the prediction error variance and its relative estimates. In particular Figure 3.1 displays the estimates of the logarithmic p.e.v. of $u_t = \Delta z_t$, where $z_t$ is the normalized power transformation of Industrial Production series for the US (considered here for the sample period 1950.1-2007.12 so as to exclude the latest recession from the analysis). The estimator used is the one in(3.10).

The plot seems to suggest that the optimal value of $\lambda$ is around 0.7 which corresponds in practice to the minimizer of the prediction error variance. Moreover if an approximate confidence interval for $\lambda$ is obtained by inverting the confidence bands for $\ln \sigma^2$, it can be seen that the logarithmic transformation is not included. The left bottom panel confirms that the series $u_t$ corresponding to log transform is heteroscedastic (the volatility declines with time). On the contrary, the predictability on the original scale is not significantly different.

As it can be seen from the bottom right panel, there is no manifest heteroscedasticity, and $u_t$ closely resembles the one corresponding to the optimal transformation.

Figure 3.1: U.S. monthly industrial production, 1950:1-2007.12. Nonparametric estimates of $\ln \sigma^2$ (with asymptotic 95% confidence limits), and plot of $u_t = \Delta z_t$ for different values of $\lambda$.




## 3.4 A multi-step forecasting procedure

When dealing with the multi-step case, a non parametric approach for the mean square prediction error variance can be based on Bloomfield's (1973) exponential model for the spectral density of a stationary process. Since the logarithm of the spectral density function is a smooth function, it can be approximated in some ways.

Bloomfield's exponential model is based on the truncated Fourier series approximation of the logarithm of the spectral density function of $u_t$, assuming that $u_t = z_t - z_{t-12}$ represents the yearly change in $z_t$.

Wahba (1980) instead use as an approximation for the log spectral density function a smoothing spline. In particular we can write the spectral density as

$$f(\omega) = \frac{1}{2\pi} \exp \left[ \alpha_0 + 2 \sum_{k=1}^{K} \alpha_k \cos k\omega \right] \tag{3.14}$$

where the coefficients $\alpha_k, k = 0, 1, \ldots, K$, can be estimated by a least squares regression of the log-periodogram on a constant and $K$ trigonometric terms. Hence, taking logarithms of the periodogram we have

$$\ln I(\omega_j) \approx \ln f(\omega_j) - \gamma + \epsilon_j, \quad 0 < \omega_j < \pi \tag{3.15}$$

where $\gamma = 0.57722$ is Euler's constant and $\epsilon_j \sim \log(\frac{1}{2}\chi_2^2) + \gamma$, i.e. has a centred log-chisquare distribution with two degrees of freedom, so that $\mathrm{E}(\epsilon_j) = 0$ and $\mathrm{Var}(\epsilon_j) = \pi^2/6$ (Davis and Jones, (1968)). Notice that we exclude the frequencies $\omega = 0, \pi$ from the above representation.

The Exponential model is thus formulated as follows as the linear regression model

$$\log(2\pi I(\omega_j)) + \gamma = \alpha_0 + 2 \sum_{k=1}^{K} \alpha_k \cos k\omega_j + \epsilon_j, \qquad \omega_j = \frac{2\pi j}{n}, j = 1, \ldots, J \tag{3.16}$$

where $J$ is such that $J > K$, and $\omega_J < \pi$. The one-step ahead prediction error variance is obtained as

$$\sigma^2 = \exp \alpha_0. \tag{3.17}$$

This follows from Kolmogorov's formula, since

$$\sigma^2 = 2\pi \exp \left[ \frac{1}{\pi} \int_0^{\pi} \log f(\omega) d\omega \right] \tag{3.18}$$

The function $lnf\omega$ has the absolutely and uniformly convergent Fourier series expansion

$$lnf(\omega) = \alpha_0 + \sum \alpha_j e^{ij\omega} + \sum \alpha_{-j} e^{-ij\omega} \tag{3.19}$$

Moreover the long run variance, saying the spectral density at the origin, is obtained by setting $\omega = 0$ in the expression (14). Thus, it is equal to

$$g(0) = \exp\left(\alpha_0 + 2\sum_k \alpha_k\right) \tag{3.20}$$

For forecasting purpose it's essential to derive a one-sided moving average rapresentation of the process. For this reason notice that the coefficients $\alpha_k$'s are related to those of the Wold decomposition. $u_t = \psi(L)\xi_t$, are obtained recursively, see (Hurvich (2002)), from the following formula

$$\psi_j = j^{-1} \sum_{r=1}^{j} r\alpha_r \psi_{j-r}, \quad j = 1, 2, \ldots, \tag{3.21}$$

with $\psi_0 = 1$ and $\alpha_r = 0, ifr > K$. Once the $\alpha_j$ are estimated by the log-periodogram regression, the h-step ahead prediction error variance can be computed as

$$\sigma_h^2 = \hat{\sigma}^2(1 + \hat{c}_1^2 +, \ldots, \hat{c}_{h-1}^2) \tag{3.22}$$

This represents an alternative way of estimating the multi step prediction error variance.

## 3.5 A Monte Carlo simulation experiment

A number of simulations were carried out to illustrate better the properties of the HN estimator of the prediction error variance and its reliability. The model chosen is an AR(1) of the form:

$$(1 - \beta(L))x(t) = \epsilon_t \tag{3.23}$$

where L is the usual lag operator and $\epsilon_t \sim NID(0, 1)$. Each observation is generated from the model and the procedure is replicated 10,000 times. Figure 3.2 shows a grid of non parametric density estimates for HN prediction error variance using bandwidth m , for $m = 1, \ldots, 11$.

Notice that $m$ has the usual function to control the smoothness of the estimates proposed. Figure 3.3 instead displays the histogram and a non parametric density estimate for a positive AR coefficient equal to 0.8. It's important to underline that the AR parameter affects only the bias, not the variance of the distribution. The true value of the prediction error variance is $\sigma^2 = 1$.

Notice that there are some differences according to the sampling size. Indeed, in large sample, i.e. $n = 1000$, the asymptotic variance is a good approximation of the true variance. For values of $m = 1$ and $m = 3$, the Monte Carlo[6] estimates of the variance are 0.0034 and 0.0025, respectively, whereas the asymptotic variances $2m\psi'(m)/n$ are 0.0033 and 0.0024. In small sample the asymptotic variance understates the true variance. For $n = 200$ and $m = 1$, $2\psi'(1)/200 = 0.0118$, but the MC estimate of the variance is 0.0168.

A second MC experiment was carried out to assess the sampling distribution of the grid search estimator of $\lambda$ which is the minimiser of the prediction error variance of the normalised BC transform, computed according to 12. For $\lambda =$

[6] see Showalter (1994), Spitzer (1978) and Spitzer (1984)

$0,5$, square root transformation, we simulated a seasonal random walk of length $n = 240$ (20 years of monthly data) on the transformed scale

$$yt(\lambda) = y_{t-12}(\lambda) + \mu + \sigma\epsilon_t \tag{3.24}$$

with $\epsilon \sim NID(0,1)$ and generated yt as the inverse transformation

$$yt = (1 + 0.5yt(\lambda))^2 \tag{3.25}$$

Then we estimate the optimal transformation parameter for yt by a grid search. Figure 3.4 displays the estimated densities for m = 1 (upper panel) and m = 3 (left panel). The main outcome of the experiment is that m = 3 provides the most reasonable compromise between bias and variance. Higher values of $m$ yield a small reduction in the variance and may yield large biases in small samples.

Figure 3.2: Non parametric density estimates corresponding to different bandwidth.

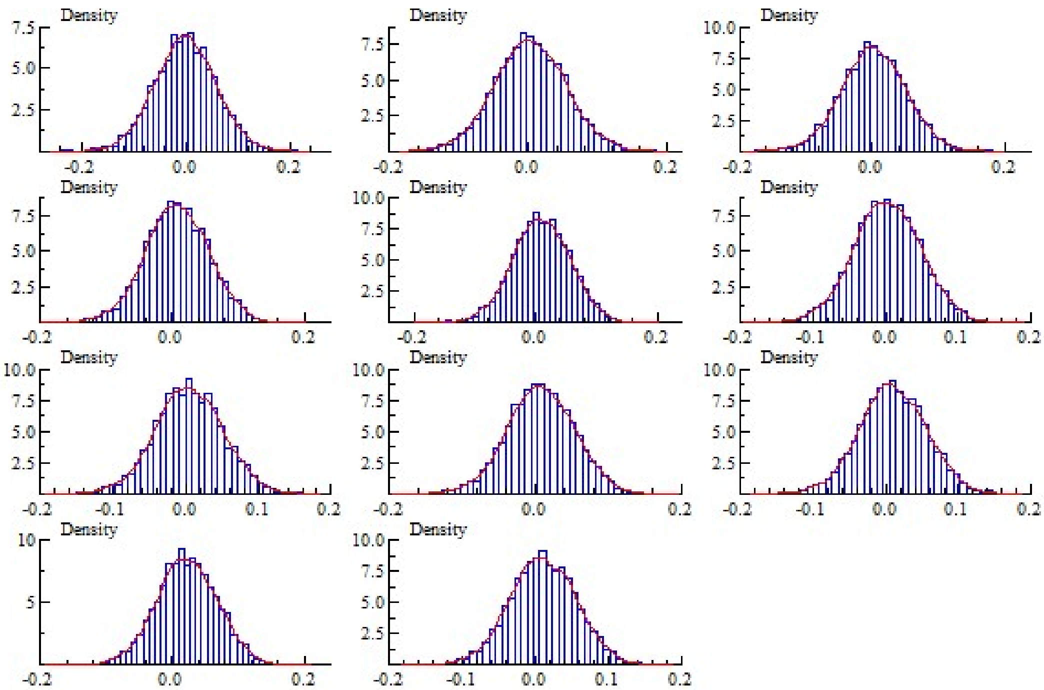


Figure 3.3: Non parametric density estimates corresponding to different bandwidth.

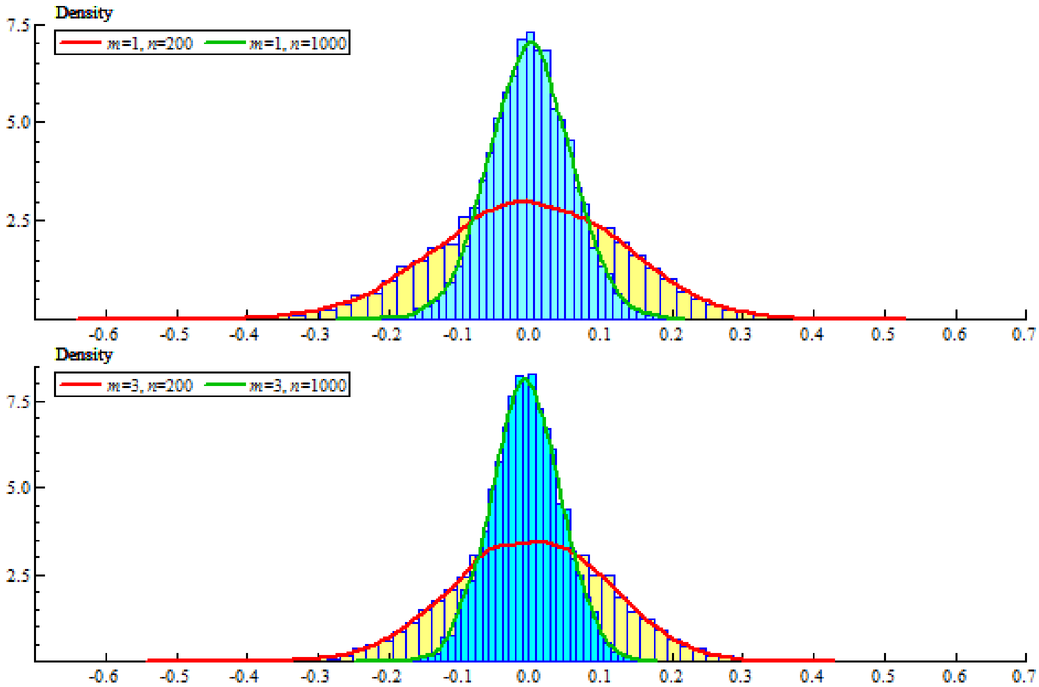
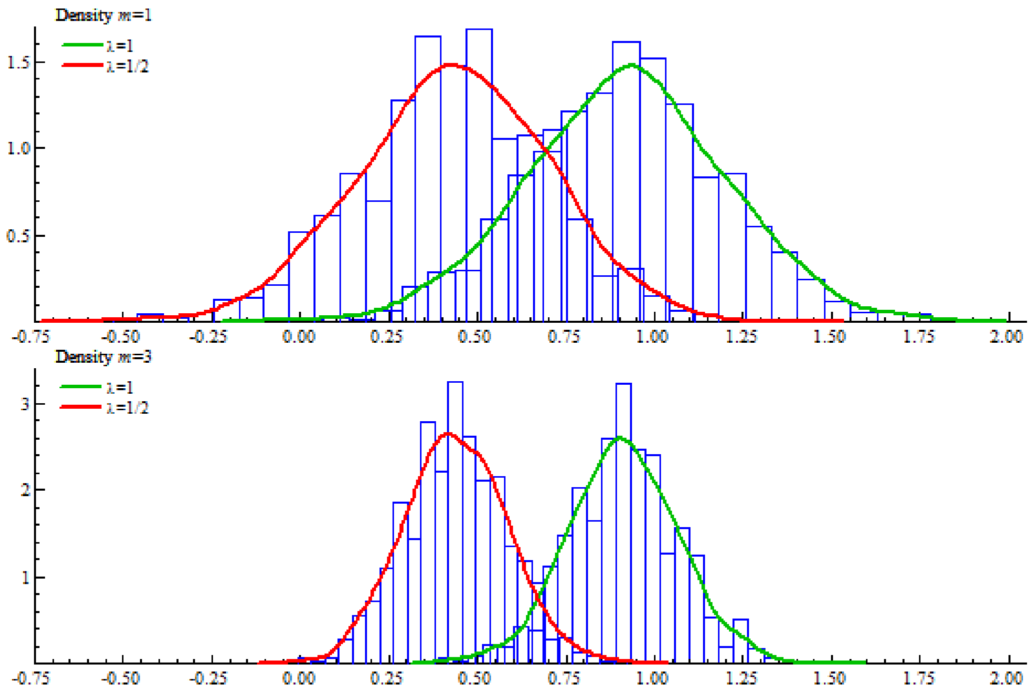


Figure 3.4: Non parametric density estimator of the transformation parameter $\lambda$, generated from $1,000$ replicates of a seasonal random walk process



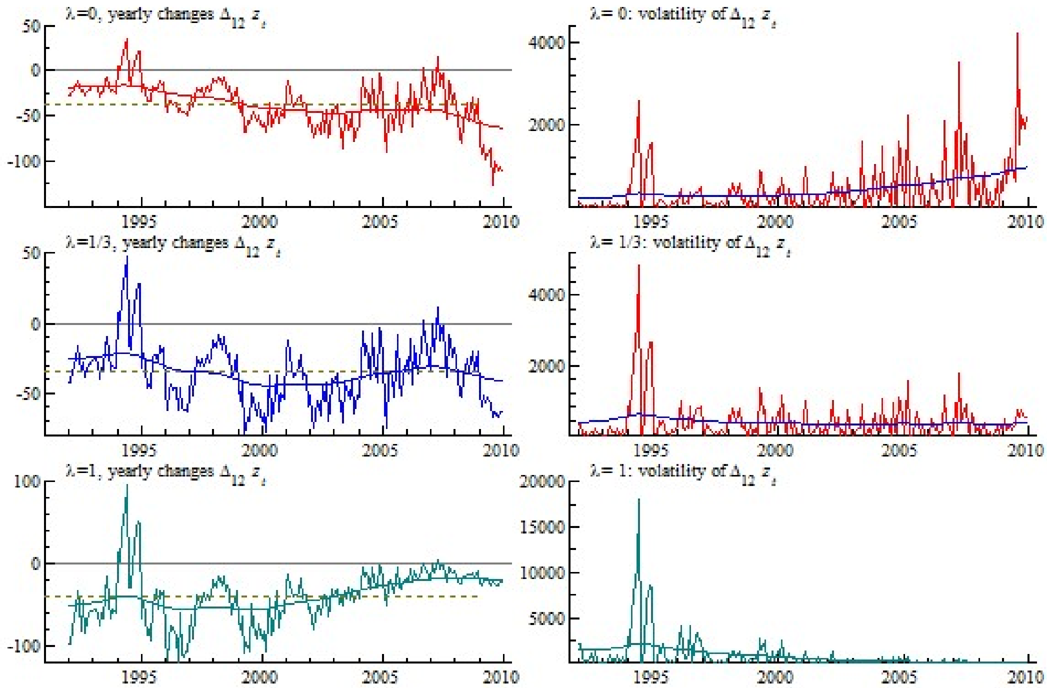
## 3.6 Exploring the optimal choice of $\lambda$ by using real data

The first relevant result concerns the monthly industrial production index for France in terms of wearing industry. The value of $\lambda$ which minimises the prediction error variance of $\Delta_{12} z_t(\lambda)$ is equal to 0.32, when the standard HN estimator of the prediction error variance with $m = 3$ is used. If we delete the seasonal and trading day frequencies, then the estimate is 0.38 (the 95% c.i. is (0.29, 0.44)). This is coincident with the estimated $\lambda$ arising from the cepstrum approach, using $K = 8$ and pulse dummies for the trading day frequencies. The value of $\lambda$ minimising the multi-step prediction error variance is either 0.32 or 0.38 depending on the horizon.

In sum, the third root transformation appears to be suitable for this series. Figure 3.5 displays the patterns of $u_t = \Delta_{12} z_t(\lambda)$ and $(u_t - \bar{u})^2$, for three values of $\lambda$, corresponding to the logarithmic transformation (top panels), the third root, and the original scale.

The local estimates of the underlying levels and the volatilities are obtained by a two-sided exponential smoothing filter with smoothing parameter providing a cut-off at 12 years. The plot illustrates quite clearly why the optimal transformation parameter is 0.33: both the level and the variance are most likely to be time invariant, whereas for the other $\lambda$ values they are trending either upwards or downwards.

Figure 3.5: France, industrial production index: manufacture of wearing apparel. Plot of $\Delta_{12} z_t(\lambda)$ and its volatility for $\lambda = 0, 1/3, 1$.

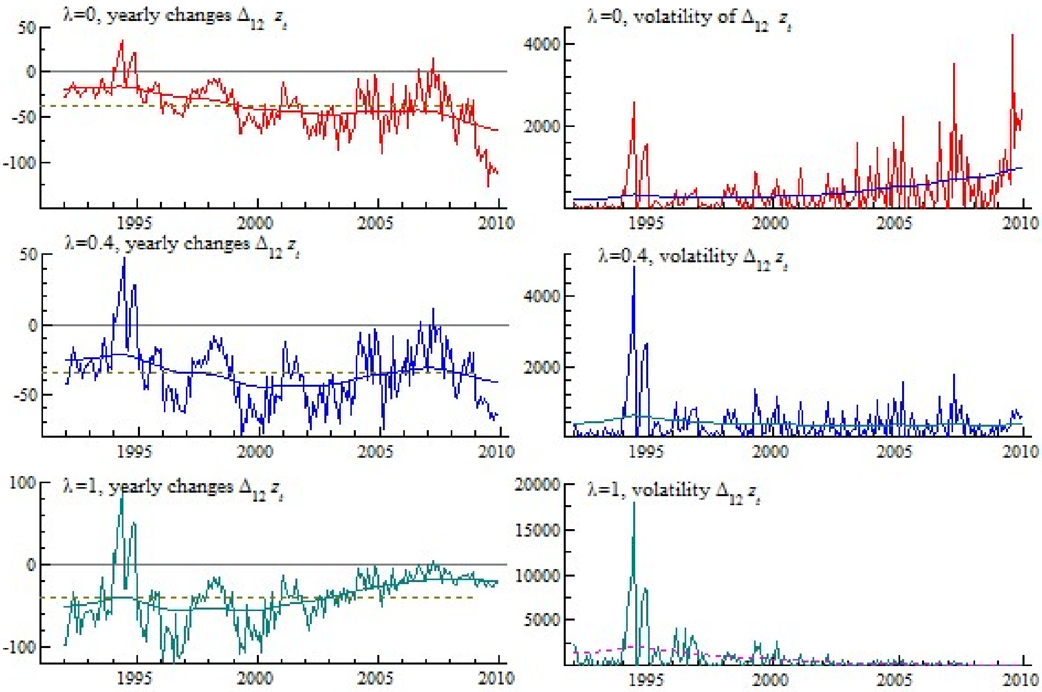



Another relevant result concerns the tourism demand for Germany in terms of total nights spent in hotels and complementary structures by residents and no residents. The value of $\lambda$ which minimises the p.e.v. of $\Delta_{12} z_t(\lambda)$ is equal to 0.44, when the standard HN estimator of the p.e.v. with $m = 3$ is used. If we delete the seasonal and trading day frequencies, then the estimate is 0.41 (the 95% confidence interval is (0.23, 0.56)). This is coincident with the estimated $\lambda$ arising from the cepstrum approach, using $K = 8$ and pulse dummies for the trading day frequencies.

The value of $\lambda$ minimising the multi-step prediction error variance is either 0.44 or 0.41 depending on the horizon. The figure below displays the patterns of $u_t = \Delta_{12} z_t(\lambda)$ and $(u_t - \bar{u})^2$, for three values of $\lambda$, corresponding to the logarithmic transformation (top panels), $\lambda = 0.4$, and the original scale.

The local estimates of the underlying levels and the volatilities are obtained by a two-sided exponential smoothing filter with smoothing parameter providing a cut-off at 12 years. The plot illustrates quite clearly why the optimal transformation parameter is 0.4: both the level and the variance are most likely to be time invariant, whereas for the other $\lambda$ values they are trending either upwards or downwards.

Figure 3.6: Germany, Tourism Demand: Total nights spent in hotels and complementary structures. Plot of $\Delta_{12} z_t(\lambda)$ and its volatility for $\lambda = 0, 0.4, 1$.



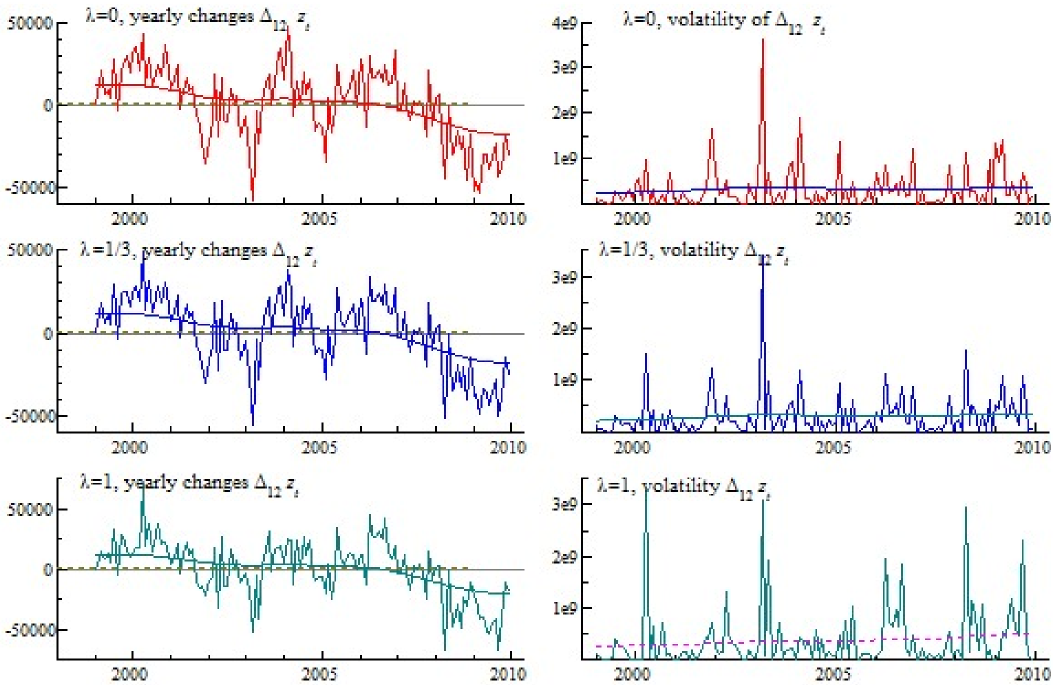

At a regional level, we consider the Sicilian Tourism demand in terms of total arrivals in hotels we see how the third root transformation appears to be more suitable for this series. The value of $\lambda$ which minimises the p.e.v. of $\Delta_{12} z_t(\lambda)$ is equal to 0.35, when the standard HN estimator of the p.e.v. with $m = 3$ is used.

If we delete the seasonal and trading day frequencies, then the estimate is 0.29 (the 95% confidence interval is (0.11, 0.47)).

This is coincident with the estimated $\lambda$ arising from the cepstrum approach, using $K = 8$ and pulse dummies for the trading day frequencies. The value of $\lambda$ minimising the multi-step p.e.v. is either 0.35 or 0.29 depending on the horizon. The figure below displays the patterns of $u_t = \Delta_{12} z_t(\lambda)$ and $(u_t - \bar{u})^2$, for three values of $\lambda$, corresponding to the logarithmic transformation (top panels), the third root transform, and the original scale.

The local estimates of the underlying levels and the volatilities are obtained by a two-sided exponential smoothing filter with smoothing parameter providing a cut-off at 12 years. The plot illustrates quite clearly why the optimal transformation parameter is $\frac{1}{3}$: both the level and the variance are most likely to be time invariant, whereas for the other $\lambda$ values they are trending either upwards or downwards.

Figure 3.7: Sicily, Tourism Demand: Total arrivals in hotels. Plot of $\Delta_{12} z_t(\lambda)$ and its volatility for $\lambda = 0, 1/3, 1$.




## 3.7 Conclusions

The results just obtained confirm how important is the choice of a ad "hoc" transformation. The spectral analysis suggests an alternative method to go deeper into the structure of the process analyzed by the researcher. The method should be applied to any relevant time series by selecting a referee model whether in an univariate or in a multivariate framework.

Indeed, the spectral decomposition of any time series represents a very useful tool which can be used for solving some problems such as finding trends of different resolution, smoothing, extracting of seasonality components or of cycles with small and large periods or with varying amplitudes. In addition, since many time series exhibit nonlinear behaviour, a method that works well for both linear and nonlinear, stationary and non stationary time series is the best choice for modelling and forecasting the real time series data. Moreover the use of smoother filters (i.e, the Kalman filter) allows the detection and subsequent treatment of influential observations which are considered as noises. This procedure cleans the series and gives more accurate estimates. Hence, the study of the optimal value for $\lambda$ becomes easier especially if it is possible to evaluate the extent to which the parameter can vary.

# Chapter 4

# Conclusions

From this study we learn how there is no "unique" method in forecasting economic time series. As a consequence, the methodology adopted is strictly related to the specific object of the analysis, to the availability of the data and to the choice either of the sector or of the level of aggregation.

In this sense a possible future research deals with the topic of temporal aggregation of a time series under transformation. In particular the idea is that sometimes the choice of the frequency clearly influences the estimation results although estimated models for different frequencies should be related.

For instance, a model for quarterly data should be related to a model for annual data, as the latter is a temporal aggregation of the former along the year.

Therefore, not only are the annual data a function of the quarterly data, but the annual model is also a function of the quarterly model.

Moreover, the quarterly estimated model is richer, information wise, as the number of observations used for estimation is four times larger than for the annual model.

Based on these considerations and evaluating other economic variables, the idea is to aggregate monthly series to create annual observations in order to see the consequences in terms of model structure and the effectiveness of the transformation adopted.

Also, it could be interested to see how different combination of forecasts behave when we include in our model the interaction of some variables representing different economic specifications of the analysis.

# References

Allcock, J. B. (1989). Seasonality, *Tourism Marketing and Management Handbook, Prentice Hall, London*, 387–392.

Amemiya, T. and J. L. Powell (1981). A comparison of the Box-Cox maximum likelihood estimator and the non-linear two stage least square estimator, *Journal of Econometrics*, 17, 351–381.

Ariño, M. A. and P. H. Franses (2000). Forecasting the levels of vector autoregressive log-transformed time series, *International Journal of Forecasting*, 16, 111–116.

Atkinson, A. C. (1973). Testing transformations to normality, *Journal of the Royal Statistical Society B*, 35, 473–479.

Atkinson, A. C. (1985). Plots, transformations and Regression. An introduction to graphical methods of diagnostic regression analysis, *Oxford Universiy Press.*

Bårdsen, G. and Lütkepohl, H. (2009). Forecasting levels of log variables in vector autoregressions, *Working Paper EUI ECO 2009/24, European University Institute.*

Bartlett, M. S. (1947). The use of transformations, *Biometrics*, 3, 1, 39–52.

Bell, W. R. and S. C. Hillmer (1983). Modelling time series with calendar variation, *Journal of the American Statistical Association*, 78, 383, 526–534.

Bhansali, R. J. (1977). Asymptotic Properties of the Wiener-Kolmogorov predictor, *Journal of the Royal Statistical Society*, 39, 1, 66–72.

Bickel, P. J. and K. A. Doksum (1981). An Analysis of transformations revisited, *Journal of the American Statistical Association*, 76, 374, 296–311.

Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series, *Biometrika*, 60, 2, 217–226.

Box, G. E. P. and D.R. Cox (1964). An analysis of transformations (with discussion), *Journal of the Royal Statistical Society, B*, 26,211–246.

Box, G. E. P. and D.R. Cox (1982). An analysis of transformation revisited, rebutted, *Journal of the American Statistical Association*, 77, 377, 209–210.

Box, G. E. P., G. M. Jenkins and G. C. Reinsel (1994). Time series analysis: forecasting and control, *Prentice–Hall International*

Box, G. E. P. and G.M. Jenkins (1973). Some Comments on a Paper by Chatfield and Prothero and on A Review by Kendall, *Journal of the Royal Statistical Society, A*, 136, 337–352.

Brockwell, P. J. and R. A. Davis (1987). Time series: theory and methods, *Springer–Verlag, New York.*

Buchinsky, M. (1995). Quantile regression, Box-Cox transformation model and the U. S. wage structure, 1963–1987, *Journal of Econometrics*, 65, 109–154.

Burbidge, J. B. , L. Magee and A. L. Robb (1988). Alternative transformations to handle extreme values of the dependent variable, *Journal of the American Statistical Association*, 83, 401, 123–127.

Butler, R. W. (2001). Seasonality in tourism: issues and implications, *Seasonality in Tourism, Pergamon, Oxford*, 5–21.

Carlino, G. A. and R. H. De Fina (2003). How strong is co-movement in employment over the business cycle? Evidence from state/ industry data, *Working paper n 03-5, Federal Reserve Bank of Philadelphia.*

Caroll, R. J. (1980). A robust method for testing transformations to achieve approximate normality, *Journal of the Royal Statistical Society, B*, 42, 1, 71–78.

Caroll, R. J. and D. Ruppert (1981). On prediction and the power transformation family, *Biometrika*, 68, 3, 609–615.

Chatfield C. and D. L. Prothero (1973). Box-Jenkins seasonal forecasting: problems in a case-study, *Journal of the Royal Statistical Society, A*, 136, 3, 295–336.

Clark, T. (1992). Business Cycle Fluctuations in U. S. regions and industries: the roles of national, region-specific and industry-specific shocks, *Research Working paper n 05, Federal Reserve Bank of Kansas City.*

Clements, M. P. and D. F. Hendry (1998). Forecasting Economic Time Series, *Cambridge, MA: MIT Press.*

Clements, M. P. and D. F. Hendry (2005). Evaluating a model by forecast performance, *Oxford Bulletin of Economics and Statistics* 67, 931–956.

Cleveland, W. S. and S. J. Devlin (1980). Calendar effects in monthly time series: detection by spectrum analysis and graphical methods, *Journal of the American Statistical Association*, 75, 371, 487–496.

Cleveland, W. S. and S. J. Devlin (1982). Calendar Effects in Monthly Time Series: Modelling and Adjustment, *Journal of the American Statistical Association* 77, 520–528.

Cutler, H. , S. England and S. Weiler (2007). Urban and regional distinctions for aggregating time series data, *Papers in Regional Science*, 86, 4, 575–595.

Davis, H. T. and R. H. Jones (1968). Estimation of the Innovation Variance of a Stationary Time Series, *Journal of the American Statistical Association*, 63, 321, 141–149.

De Bruin, P. and P. H. Franses (1999). Forecasting power-transformed time series data, *Journal of Applied Statistics*, 26, 7, 807–815.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy, *Journal of Business and Economic Statistics*, 13, 253–263.

Frechtling, D. C. (2001). Forecasting Tourism Demand: methods and startegies, *Butterworth-Heinemann, Oxford*.

Freeman, J. and R. Modarres (2006). Inverse Box-Cox: the power-normal distribution, *Statistics and Probability Letters*, 76, 764–772.

Garrison, C. B. and H. S. Chang (1979). The effect of monetary and fiscal policies on regional business cycles, *International Regional Science Review*, 4, 2, 167–180.

Granger, C. W. J. and P. Newbold (1976). Forecasting transformed time series, *Journal of the Royal Statistical Society, B*, 38, 2, 189-203.

Granger, C. W. J. and M. J. Machinov (2005). Forecasting and decision theory, *Handbook of Econometrics chapter: Department of Economics, University of California, San Diego.*

Grenander, U. and M. Rosemblatt (1952). On spectral analysis of stationary time series, *Proceeding of the National Academy of Sciences of the United States of America*, 38, 6, 519–521.

Grenander, U. (1958). Bandwidth and variance in estimation of the spectrum, *Journal of the Royal Statistical Society*, 20, 1, 152–157.

Guerrero, V. M. (1993). Time-series analysis supported by power transformations, *Journal of Forecasting*, 12, 37–48.

Guerrero, V. M. and R. Perera (2004). Variance stabilizing power transformation for time series, *Journal of Modern Applied Statistical Methods*, 3, 2, 357–369.

Hall, V. B. and C. J. McDermott (2007). Regional business cycle in New Zealand: Do they exist? What might drive them?, *Papers in Regional Science*, 86, 2, 167–191.

Hamilton, J. D. (1994). Time Series analysis, *Princeton University Press*.

Hannan, E. J. and D. F. Nicholls (1977). Estimation of the prediction error variance, *Journal of the American Statistical Association*, 72, 360, 834–840.

Hayashida, M. and G. J. D. Hewings (2009). Regional Business Cycle in Japan *International Regional Science Review*, 32, 2, 119–147.

Hendry, D. F. and N. R. Ericsson (2001). Understanding Economic Forecasts, *Cambridge University Press*.

Hinkley, D. V. and G. Runger (1984). The Analysis of Transformed Data, *Journal of the American Statistical Association*, 79, 386, 302–309.

Hosoya, Y. and T. Terasaka (2009). Inference on transformed stationary time series, *Journal of Econometrics*, 151, 129–139.

Hurvich, C. M. (2002). Multistep forecasting of long memory series using fractional exponential models, *International Journal of Forecasting*, 18, 167–179.

Hylleberg, S. (1992). General introduction, *Modelling Seasonality, Oxford University Press*

John, J. A. and N. R. Draper (1980). An alternative family of transformations, *Applied Statistics*, 29, 2, 190–197.

Jones, R. H. (1976). Estimation of the innovation generalized variance of a multivariate stationary time series, *Journal of the Americal Statistical Association*, 71, 354, 386–388.

Kemp, G. C. R. (1996). Scale equivariance and the Box- Cox transformation, *EConomics Letters*, 51, 1–6.

Kim, M. and R. C. Hill (1995). Shrinkage estimation in non linear regression. The Box-Cox transformation, *Journal of Econometrics*, 66, 1–33.

Kozlowski, P. J. (1995). Money and Interest rates as predictors of regional economic activity, *The Review of Regional Studies*, 25, 2, 143–157.

Kulendran, N. and K. K. F. Wong (2005). Modelling seasonality in Tourism Forecasting, *Journal of Travel Research*, 44, 163–170.

Ledyaeva, S. and M. Sirkjarvi (2008). Forecasting regional economy: the case study for the Russian city of Saint-Petersburg, *Center for markets in transition, Helsinki School of Economics.*

Longhi, S. and P. Nijkamp (2007). Forecasting Regional Labour Market Developments under Spatial Autocorrelation, *International Regional Science Review*, 30, 2, 110–119.

Lutkepohl, H. (2005). New Introduction to Multiple Time Series Analysis, *Sprimger-Verlag, Berlin.*

Lutkepohl, H. and F. Xu (2009). The Role of the log transformation in forecasting economic variables, *Working paper n 391, European University Institute.*

Lutkepohl, H. (2009). Forecasting aggregated time series variables: a survey, *Working paper, European University Institute.*

Manly, B. F. (1976). Exponential data transformation, *The Statistician*, 25, 37–42.

Mayor, M. , A. J. Lopez and R. Pérez (2007). Forecasting Regional Employment with Shift-Share and ARIMA Modeling, *Regional Studies*, 41, 4, 543–551.

Neyman, J.N. and E.L. Scott (1960). Correction for bias introduced by a transformation of variables, *The Annals of Mathematical Statistics*, 31, 3, 643-655.

Pankratz, A. and U. Dudley (1987). Forecasts of power-transformed series, *Journal of Forecasting*, 6, 239–248.

Pascual, L. , J. Romo and E. Ruiz (2005). Bootstrap prediction intervals for power-transformed time series, *International Journal of forecasting*, 21, 219–235.

Poirier, D. J. (1978). The use of the Box-Cox transformation in limited dependent variable models, *Journal of the American Statistical Association*, 73, 362, 284–287.

Proietti, T. and M. Riani (2009). Transformations and Seasonal Adjustment: analytic solutions and case studies, *Journal of Time Series Analysis*, 30, 1, 47–69.

Rickman, D. S. , S. R. Miller and R. Mckenzie (2009). Spatial and sectoral linkages in region models: a Bayesian vector autoregression forecast evaluation, *Papers in Regional Science*, 88, 1, 29–40.

Sakia, R. M. (1992). The Box-Cox transformation technique: a review, *The Statistician*, 41, 2, 169–178.

Schlesselman, J. (1971). Power Families: a note on the Box-Cox transformation, *Journal of the Royal Statistical Society*, 33, 2, 307-311.

Selaver, D. D. , V. J. Roderick and J. Kroll (2005). Mode-Locking and regional business cycle synchronization, *Journal of Regional Science*, 45, 4, 703–745.

Shoesmith, G. L. (2000). The time series relatedness of State and national indexes of leading indicators and implications for regional forecasting, *International Regional Science Review*, 23, 3, 281–299.

Showalter, M. H. (1994). A Monte Carlo investigation of the Box-Cox model and a non lnear least squares alternative, *The Review of Economics and Statistics*, 76, 3, 560–570.

Spitzer, J. J. (1978). A Monte Carlo investigation of the Box-Cox transformation in small samples, *Journal of the American Statistical Association*, 73, 363, 488–495.

Spitzer, J. J. (1984). Variance estimates in models with the Box-Cox transformation: implication for estimation and hypothesis testing, *The Review of Economics and Statistics*, 66, 4, 645–652.

Stock, J. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association*, 97, 1167–1179.

Stock, J. and M. W. Watson (2003). Forecasting output and inflation: the role of asset prices, *Journal of Economic Literature*, 41, 788–829.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review, *International Journal of Forecasting*, 16, 437–450.

Tena, J. D. D. , A. Espasa and G. Pino (2010). Forecasting Spanish inflation using the maximum disaggregation level by sectors and geographical areas, *International Regional Science Review*, 33, 2, 181–204.

Taylor, J. M. G. (1986). The retransformed mean after a fitled power transformation, *Journal of the American Statistical Association*, 81, 393, 114–118.

Tukey, J. W. (1957). On the comparative anatomy of transformations, *Annals of Mathematical Statistics*, 28, 602-632.

Thury, G. and M. Zhou (2005). Calendar effects in monthly time series models, *Journal of Systems Science and Systems Engineering*, 14, 2, 218–230.

Wahba, G. (1980). Automatic smoothing of the log periodogram, *Journal of the American Statistical Association*, 75, 369, 122–132.

West, C. T. and T. M. Fullerton (1996). Assessing the historical accuracy of regional economic forecasts, *Journal of Forecasting*, 15, 1, 19–36.

West, C. T. (2003). The status of evaluating accuracy of regional forecasts, *The Review of Regional Studies*, 33, 1, 85–103.

Wooldridge, J. M. (1992). Some alternatives to the Box-Cox regression model, *International Economic Review*, 33, 4, 935–955.

Yang, Z. (2002). A modified family of power transformations, *Working paper series n 12, Singapore Management University.*

Yeo, I. K. and R. A. Johnson (2000). A new family of power transformations to improve normality or simmetry, *Biometrika*, 87, 4, 954–959.

Zhang, J. , B. Madsen and C. J. Butler (2007). Regional economic impacts of tourism: the case of Denmark, *Regional Studies*, 41, 6, 839–853.