STATISTICA, anno LXX, n. 4, 2010

TIME SERIES MODELING AND DECOMPOSITION

Estela Bee Dagum

1. INTRODUCTION

A time series consists of a set of observations ordered in time, on a given phenomenon (target variable). Usually the measurements are equally spaced, *e.g.* by year, quarter, month, week, day. The most important property of a time series is that the ordered observations are dependent through time, and the nature of this dependence is of interest in itself. Examples of time series are the gross national product, the unemployment rate, or the daily closing value of the Dow Jones index. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.

Time series data have a natural temporal ordering. This makes time series analysis distinct from other common data analysis problems, in which there is no natural ordering of the observations (*e.g.* explaining people's income relative to their education level, where the individuals' data could be entered in any order). Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (*e.g.* house prices). A time series model will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values.

Formally, a time series is defined as a set of random variables indexed in time, $\{X_1, ..., X_T\}$. In this regard, an observed time series is denoted by $\{x_1, ..., x_T\}$, where the sub-index indicates the time to which the observation x_t pertains. The first observed value x_1 can be interpreted as the realization of the random variable X_1, which can also be written as $X(t=1, \omega)$ where ω denotes the event belonging to the sample space. Similarly, x_2 is the realization of X_2, and so on. The T-dimensional vector of random variable can be characterized by different probability distribution.

For socio-economic time series the probability space is continuous, and the time measurements are discrete. The frequency of measurements is said to be high when it is daily, weekly or monthly and to be low when the observations are quarterly or yearly.

2. TIME SERIES DECOMPOSITION MODELS

An important goal in time series analysis is the decomposition of a series into a set of non-observable (latent) components that can be associated to different types of temporal variations. The idea of time series decomposition is very old and was used for the calculation of planetary orbits by seventeenth century astronomers. Persons (1919) was the first to state explicitly the assumptions of unobserved components. As Persons saw it, time series were composed of four types of fluctuations:

(1) A long-term tendency or secular trend.
(2) Cyclical movements super-imposed upon the long-term trend. These cycles appear to reach their peaks during periods of industrial prosperity and their troughs during periods of depressions, their rise and fall constituting the business-cycle.
(3) A seasonal movement within each year, the shape of which depends on the nature of the series.
(4) Residual variations due to changes impacting individual variables or other major events such as wars and national catastrophes affecting a number of variables.

Traditionally, the four variations have been assumed to be mutually independent from one another and specified by means of an additive decomposition model:

$$X_t = T_t + C_t + S_t + I_t, \tag{1}$$

where X_t denotes the observed series, T_t the long-term trend, C_t the business-cycle, S_t seasonality and I_t the irregulars.

If there is dependence among the latent components, this relationship is specified through a multiplicative model

$$X_t = T_t \; C_t \; S_t \; I_t, \tag{2}$$

where now S_t and I_t are expressed in proportion to the trend-cycle $T_t \; C_t$. In some cases, mixed additive-multiplicative models are used.

Whether a latent component is present or not in a given time series depends on the nature of the phenomenon and on the frequency of measurement. For example, seasonality is due to the fact that some months or quarters of a year are more important in terms of activity or level. Because this component is specified to cancel out over 12 consecutive months or 4 consecutive quarters, or more generally over 365.25 consecutive days, yearly series cannot contain seasonality.

Flow series can be affected by other variations associated to the composition of the calendar. The most important are the trading-day variations, which are due to the fact that some days of the week are more important than others. Months with five of the more important days register an excess of activity (ceteris paribus) in comparison to months with four such days. Conversely, months with five of the less important days register a short-fall of activity. The length-of-month variation is usually assigned to the seasonal component. The trading-day component is usually considered as negligible in quarterly series and even more so in yearly data. Another important calendar variation is the moving-holiday or moving-festival component. That component is associated to holidays which change date from year to year, e.g. Easter and Labour Day, causing a displacement of activity from one month to the previous or the following month. For example, an early date of Easter in March or early April can cause an important excess of activity in March and a corresponding short-fall in April, in variables associated to imports, exports, hospitality and tourism.

Under models (1) and (2), the trading-day and moving festival components (if present) are implicitly part of the irregular. Young (1965) developed a procedure to estimate trading-day variations which was incorporated in the X-11 seasonal adjustment method by Shiskin et al. in 1967 (for more details see Ladiray and Quenneville, 2001) and its subsequent versions, the X-11-ARIMA (Dagum 1980 and 1988) and X12-ARIMA (Findley et al. 1998) methods. The later two versions also include models to estimate moving-holidays due to Easter.

Considering these new components, the additive decomposition model (1) becomes

$$X_t = T_t + C_t + S_t + D_t + H_t + I_t , \tag{3}$$

where D_t and H_t respectively denote the trading-day and moving-holiday components. Similarly, the multiplicative decomposition model (2) becomes

$$X_t = T_t C_t S_t D_t H_t I_t , \tag{4}$$

where the components S_t, D_t, H_t and I_t are proportional to the trend-cycle $T_t C_t$.

Decomposition models (3) and (4) are traditionally used by seasonal adjustment methods. Seasonal adjustment actually entails the estimation of all the time series components and the removal of seasonality, trading-day and holiday effects from the observed series. The rationale is that these components which are relatively predictable conceal the current stage of the business cycle which is critical for policy and decision making.

Another major objective in time series analysis is the modelling of the observed series mainly for forecasting purposes. In this case, an often used decomposition model for univariate time series is

$$X_t = \eta_t + e_t , \tag{5}$$

where η_t and e_t are referred to as the signal and the noise using electrical engineering terminology. The signal η_t comprises all the systematic components of models (1) to (4), i.e. $T_t C_t$, S_t, D_t and H_t.

Model (5) is classical in signal extraction where the problem is to find the best estimates of the signal $\{\eta_t\}$ given the observations $\{x_t\}$ corrupted by noise $\{e_t\}$. The best estimates are usually defined as minimizing the mean square error.

Signal extraction can be made by means of parametric models or non-parametric procedures. The latter has a long standing and was used by actuaries at the beginning of the 1900's. The main assumption in non-parametric procedures is that η_t is a smooth function of time. Different types of smoothers are used depending on the series under question. The most common smoothers are the cubic splines originally applied by Whittaker (1923) and Whittaker and Robinson (1924) to smooth mortality tables. Other smoother are moving averages and high order kernels used in the context of seasonal adjustment and form the basis of methods such as Census X-11 (Shiskin et al. 1967), X-11-ARIMA (Dagum 1980 1988), X-12-ARIMA (Findley et al. 1998), STL (Cleveland et al. 1990).

Non-parametric signal extraction has also been very much applied to estimate the trend (non-stationary mean) of time series (see among others Henderson 1916; Macaulay 1931; Gray and Thomson 1996 and 2002; Dagum, 1996, Dagum and Luati 2000). Among nonparametric procedures, the 13-term Henderson trend-cycle estimator is the most often applied because of its good property of rapid turning point detection but it has the disadvantages of: (1) producing a large number of unwanted ripples (short cycles of 9 and 10 months) that can be interpreted as false turning points and, (2) large revisions for the most recent values (often larger than those of the corresponding seasonally adjusted data).

The use of longer Henderson filters is not an alternative for the reduction in false turning points is achieved at the expense of increasing the time lag of turning point detection. In 1996, Dagum proposed a new method that enables the use of the 13-term Henderson filter with the advantages of: (1) reducing the number of unwanted ripples, (2) reducing the size of the revisions to most recent trend-cycle estimates and, (3) no increase in time lag of turning point detection.

The Dagum (1996) method basically consists of producing one year of ARIMA extrapolations from a seasonally adjusted series with extreme values replaced by default; extending the series with the extrapolated values and then, applying the Henderson filter to the extended seasonally adjusted series requesting smaller sigma limits (not the default) for the replacement of extreme values. The object is to pass through the 13-term Henderson filter, an input with reduced noise. This procedure was applied to the nine Leading Indicator series of the Canadian Composite Leading Index with excellent results and is currently being used by many statistical agencies. In a recent work, Dagum and Luati (2009) developed a linear approximation to the nonlinear Dagum (1996) method which gave very good results in empirical applications.

Other recent works on nonparametric trend-cycle estimation were done by Dagum and Bianconcini (2006) where these authors derive a Reproducing kernel Hilbert Space (RKHS) representation of the Henderson (1916) and LOESS (due to Cleveland, 1979) smoothers with particular emphasis on the asymmetric ones applied to most recent observations. A RKHS is a Hilbert space characterized by a kernel that reproduces, via an inner product, every function of the space or, equivalently, a Hilbert space of real valued functions with the property that every point evaluation functional is a bounded linear functional. This Henderson kernel representation enables the construction of a hierarchy of kernels with varying smoothing properties. The asymmetric filters are derived coherently with the corresponding symmetric weights or from a lower or higher order kernel within a hierarchy, if more appropriate. In the particular case of the currently applied asymmetric Henderson and LOESS filters, those obtained by means of the RKHS are shown to have superior properties relative to the classical ones from the view point of signal passing, noise suppression and revisions.

In another study, Dagum and Bianconcini (2008) derive two density functions and corresponding orthonormal polynomials to obtain two Reproducing Kernel Hilbert Space representations which give excellent results for filters of short and medium lengths. Theoretical and empirical comparisons of the Henderson third order kernel asymmetric filters were made with the classical ones again showing superior properties of signal passing, noise suppression and revisions. Dagum and Bianconcini (2009, and 2010) provide a common approach for studying several nonparametric estimators used for smoothing functional time series data. Linear filters based on different building assumptions are transformed into kernel functions via reproducing kernel Hilbert spaces. For each estimator, these authors identify a density function or second order kernel, from which a from which a hierarchy of higher order estimators is derived. These are shown to give excellent representations for the currently applied symmetric filters. In particular, they derive equivalent kernels of smoothing splines in Sobolev space and polynomial space. A Sobolev space intuitively, is a Banach space and in some cases a Hilbert space of functions with sufficiently many derivatives for some application domain, and equipped with a norm that measures both the size and smoothness of a function.

Sobolev spaces are named after the Russian mathematician Sergei Sobolev. The asymmetric weights are obtained by adapting the kernel functions to the length of the various filters, and a theoretical and empirical comparison is made with the classical estimators used in real time analysis. The former are shown to be superior in terms of signal passing, noise suppression and speed of convergence to the symmetric filter.

On the other hand, signal extraction by means of explicit models arrived much later. Under the assumption that the entire realization of y_t is observed from $-\infty$ to $+\infty$ and η_t and e_t are both mutually independent and stationary, Kolmogorov (1939, 1941) and Wiener (1949) independently proved that the minimum mean square estimator of the signal η_t is the conditional mean given the observations y_t, that is $\hat{\eta}_t = E(\eta_t | y_t, y_{t-1}, \dots)$. This fundamental result was extended by several authors who provided approximate solutions to the non-stationary signal extraction, particularly Hannan (1967), Sobel (1967) and Cleveland and Tiao (1976). Finally, Bell (1984) provided exact solutions for the conditional mean and conditional variance of vector η when non-stationarity can be removed by applying differences of a finite order. This author used two alternatives regarding the generation of vectors y, η and e.

Model-based signal extraction was also used in the context of seasonal adjustment where the signal η_t is assumed to follow an ARIMA model of the Box and Jenkins (1970) type, plus a regression model for the deterministic variations (see e.g. Burman 1980, Gómez and Maravall 1996, Findley et al. 1998). The latter is applied to estimate deterministic components, such as trading-day variations or moving-holiday effects and outliers. Gersch and Kitagawa (1983) and Koopman et al. (1998) also used signal extraction for seasonal adjustment where the signal η_t is assumed to follow a structural time series component model (Harvey 1989) cast in state-space representation. Signal extraction, parametric and non-parametric, is also widely applied for forecasting purposes.

The feasibility of the decomposition of a time series was proved by Herman Wold in 1938. Wold showed that any second-order stationary stochastic process $\{X_t\}$ can be decomposed in two mutually uncorrelated processes $\{Z_t\}$ and $\{V_t\}$, such that

$$X_t = Z_t + V_t, \tag{6.a}$$

where

$$Z_t = \Sigma_{j=0}^{\infty} \psi_j \; a_{t-j}, \; \psi_0 \equiv 1, \; \Sigma_{j=1}^{\infty} \; \psi_j^2 < \infty, \tag{6.b}$$

with $\{a_t\} \sim WN(0, \sigma_a^2)$.

Model (6.b) is known as an infinite moving average $MA(\infty)$ where the a_t's are the innovations. $\{Z_t\}$ is a convergent infinite linear combination of the a_t's, which are assumed to follow a white noise (WN) process of zero mean, constant variance σ_a^2, and zero autocovariance. The component $\{Z_t\}$ is called the non-deterministic or purely linear component since only one realization of the process is not sufficient to determine future values $Z_{t+\ell}$, $\ell > 0$, without error. Component $\{V_t\}$ can be represented by

$$V_t = \mu + \Sigma_{j=1}^{\infty} \left[\alpha_j \; \sin(\lambda_j t) + \beta_j \; \cos(\lambda_j t)\right], \; -\pi < \lambda_j < \pi \tag{6.c}$$

where μ is the constant mean of process $\{X_t\}$ and $\{\alpha_j\}$, $\{\beta_j\}$ are mutually uncorrelated white noise processes. The series $\{V_t\}$ is called the deterministic part because it can be predicted in the future without error from a single realization of the process by means of an infinite linear combination of past values.

Wold theorem demonstrates that the property of stationarity is strongly related to that of linearity. It provides a justification for autoregressive moving average (ARMA) models (Box and Jenkins 1970) and some extensions, such as the autoregressive integrated moving average (ARIMA) and regression-ARIMA models (RegARIMA).

A stochastic process $\{X_t\}$ is second-order stationary or weakly stationary, if the first two moments are not time dependent, that is, the mean and the variance are constant, and the autocovariance function depends only on the time lag and not on the time origin, that is,

$$E(X_t) = \mu < \infty, \tag{7.a}$$

$$E(X_t - \mu)^2 = \sigma_X^2 < \infty, \; E[(X_t - \mu)\,(X_{t-k} - \mu)] = \gamma(k) < \infty, \tag{7.b}$$

where $k = 0, 1, 2, ...$ denotes the time lag.

3. THE SECULAR OR LONG-TERM TREND

The concept of trend is used in economics and other sciences to represent long-term smooth variations. The causes of these variations are often associated with structural phenomena such as population growth, technological progress, capital accumulation, new practices of business and economic organization. For most economic time series, the trends evolve smoothly and gradually, whether in a deterministic or stochastic manner. When there is sudden change of level and/or slope this is referred to as a structural change. It should be noticed however that series at a higher levels of aggregation are less susceptible to structural changes. For example, a technological change is more likely to produce a structural change for some firms than for the whole industry.

The identification and estimation of the secular or long-term trend have posed serious challenges to statisticians. The problem is not of statistical or mathematical character but originates from the fact that the trend is a latent (non-observable) component and its definition as a long-term smooth movement is statistically vague. The concept of long-period is relative, since a trend estimated for a given series may turn out to be just a long business cycle as more years of data become available. To avoid this problem statisticians have used two simple solutions. One is to estimate the trend and the business cycles jointly, calling it the trend-cycle. The other solution is to estimate the trend over the whole series, and to refer to it as the longest non-periodic variation.

It should be kept in mind that many systems of time series are redefined every fifteen years or so in order to maintain relevance. Hence, the concept of long-term trend loses importance. For example, in Canada, the system of Retail and Wholesale Trade series was redefined in 1989 to adopt the 1980 Standard Industrial Classification (SIC), and again in 2003 to conform to the North American Industrial Classification System (NAICS), following the North American Free Trade Agreement. The following examples illustrate the necessity of such reclassifications. The 1970 Standard Industrial Classification (SIC) considered computers as business machines, e.g. cash registers, desk calculators. The 1980 SIC rectified the situation by creating a class for computers and other goods and services. The last few decades witnessed the birth of new industries involved in photonics (lasers), bio-engineering, nano-technology, electronic commerce. In the process, new professions emerged, and Classification systems had to keep up with these new realities.

There is a large number of deterministic and stochastic models which have been proposed for trend estimation (see among many others, Dagum and Dagum, 2006, Alexandrov *et al.*, 2010).

Deterministic models are based on the assumption that the trend can be well approximated by mathematical functions of time such as polynomials of low degree, cubic splines, logistic functions, Gompertz curves, modified exponentials. Stochastic trends models assume that the trend can be better modelled by differences of low order together with autoregressive and moving average terms. Stochastic trend models are appropriate when the trend is assumed to follow a non-stationary stochastic process where the non-stationarity is modelled with finite differences of low order (cf. Harvey 1985, Maravall 1993).

A typical stochastic trend model often used in structural time series modelling, is the so-called random walk with constant drift. In the classical notation this model is

$$\mu_t = \mu_{t-1} + \beta + \xi_t \, , t = 1, 2, ... n \, ; \xi_t \_ N(0, \sigma_\xi^2) \, , \tag{8.a}$$

$\Delta \mu_t = \beta + \xi_t$, where μ_t denotes the trend, β a constant drift and $\{\xi_t\}$ is a normal white noise process. Solving the difference equation (2.15a) and assuming $\xi_0 = 0$, we obtain

$$\mu_t = \beta\, t + \Delta^{-1} \xi_t \;= \beta\, t + \Sigma_{j=0}^{\infty} \;\; \xi_{t-j}, t = 1, ..., n \, , \tag{8.b}$$

which show that a random walk with constant drift consists of a linear deterministic trend plus a non-stationary infinite moving average.

Another type of stochastic trend belongs to the ARIMA (*p,d,q*) class, where *p* is the order of the autoregressive polynomial, *q* is the order of the moving average polynomial and *d* the order of the finite difference operator $\Delta = (1 - B)$. The backshift operator *B* is such that $B^n \; z_t \equiv z_{t-n}$. The ARIMA (*p,d,q*) model is written as

$$\phi_p(B)(1-B)^d \; z_t = \theta_q(B) \;\; a_t \, , \; a_t \sim \; N(0, \sigma_a^2) \, , \tag{9}$$

where z_t now denotes the trend, $\phi_p(B)$ the autoregressive polynomial in *B* of order *p*, $\theta_q(B)$ stands for the moving average polynomial in *B* of order *q*, and $\{a_t\}$ denotes the innovations assumed to follow a normal white noise process. For example, with *p*=1, *d*=2, *q*=0, model (9) becomes

$$(1 - \phi_1 \;\; B)(1-B)^2 \;\; z_t = a_t \, , \tag{10}$$

where z_t now denotes the trend, $\phi_p(B)$ the autoregressive polynomial in *B* of order *p*, $\phi_q(B)$ stands for the moving average polynomial in *B* of order *q*, and

$\{a_t\}$ denotes the innovations assumed to follow a normal white noise process. For example, with p=1, d=2, q=0, model (9) becomes

$$(1-\phi_1 \ B)(1-B)^2 \ z_t = a_t \ , \tag{11}$$

which means that after applying first order differences twice, the transformed series can be modelled by an autoregressive process of order one.

The Hodrick & Prescott (1997) filter follows the cubic smoothing spline approach. The framework used in Hodrick & Prescott is that a given time series X is the sum of a growth component T and a cyclical component C such that $X = T + C$. The measure of the smoothness of the trend T is the sum of the squares of its second order difference. The C are deviations from T and the conceptual framework is that over long time periods, their average is near zero.

The Hodrick-Prescott (HP) filter was not developed to be appropriate, much less optimal, for specific time series generating processes. Rather, apart from the choice of the smoothing parameter λ, the same filter is intended to be applied to all series. Nevertheless, the smoother that results can be viewed in terms of optimal signal extraction literature pioneered by Wiener (1949) and Whittle (1963) and extended by Bell (1984) to incorporate integrated time series generating processes. King & Rebelo (1993) and Ehglen (1998) analyzed the HP filter in this framework, motivating it as a generalization of the exponential smoothing filter. On the other hand, Kaiser& Maravall (1999) showed that under certain restriction the HP filter can be well approximated by an Integrated Moving Average model of order 2, whereas Harvey & Jaeger (1993) interpreted the HP filter in terms of structural time series models.

4. THE BUSINESS CYCLE

The business cycle is a quasi-periodic oscillation characterized by periods of expansion and contraction of the economy, lasting on average from three to five years. Because most time series are too short for the identification of a trend, the cycle and the trend are estimated jointly and referred to as the trend-cycle. As a result the concept of trend loses importance. The trend-cycle is considered a fundamental component, reflecting the underlying socio-economic conditions, as opposed to seasonal, trading-day and transient irregular fluctuations.

The proper identification of cycles in the economy requires a definition of contraction and expansion. The definition used in capitalistic countries to produce the chronology of cycles is based on fluctuations found in the aggregate economic activity. A cycle consists of an expansion phase simultaneously present in many economic activities, followed by a recession phase and by a recovery which develops into the next expansion phase. This sequence is recurrent but not strictly periodic. Business cycles vary in intensity and duration. In Canada for example, the 1981 recession was very acute but of short duration, whereas the 1991 recession was mild and of long duration. Business cycles can be as short as 18 months and as long as 10 years.

A turning point is called a peak or downturn when the next estimate of the trend-cycle indicates a decline in the level of activity; and a trough in the opposite situation. There are many ways to determine when a downturn occurs, but in general, (see e.g. Dagum and Luati 2000, Chhab et al. 1999, Zellner et al. 1991) a downturn is deemed to occur at time t in the trend-cycle of monthly series, if

$$c_{t-l} \leq \ldots \leq c_{t-1} > c_t \geq c_{t+1}; \tag{12.a}$$

and an upturn, if

$$c_{t-l} \geq \ldots \geq c_{t-1} < c_t \leq c_{t+1}. \tag{12.b}$$

Thus, a single change to a lower level c_t, between t+1 and t, qualifies as a downturn, if $c_{t+1} \leq c_t$ and $c_{t-3} \leq c_{t-2} \leq c_{t-1}$; and conversely for an upturn.

The dating of downturns and upturns is based on a set of economic variables related to production, employment, income, trade and so on.

Similarly to the trend, the models for cyclical variations can be deterministic or stochastic. Deterministic models often consist of sine and cosine functions of different amplitude and periodicities. Stochastic models of the ARIMA type, involving autoregressive models of order 2 with complex roots, have been used to model business cycles.

4.1. *Same-Month Comparisons*

In the absence of seasonal adjustment, only the raw series is available. In such cases, it is customary to use same-month comparisons from year to year, $z_t - z_{t-12}$, to assess the stage of the business cycle. The rationale is that the seasonal effect in z_t is approximately the same as in z_{t-12}, under the assumption of slowly evolving seasonality. Same-month year ago comparisons can be expressed as the sum of the changes in the raw series between z_t and z_{t-12},

$$z_t - z_{t\text{-}12} \equiv (z_t - z_{t-1}) + (z_{t-1} - z_{t-2}) + (z_{t-2} - z_{t-3}) + \\ \ldots + (z_{t-11} - z_{t-12}) \equiv \Sigma_{j=1}^{12} (z_{t-j+1} - z_{t-j}). \tag{13}$$

Eq. (13) shows that same-month comparison display an increase, if the increases dominate the decreases over the 13 months involved, and conversely. The timing of $z_t - z_{t-12}$ is $t-6$, the average of t and $t-12$. This points out a limitation of this practise: the diagnosis provided is not timely with respect to t. Furthermore, z_t and z_{t-12} may contain irregular variations affecting one observation positively and the other negatively, hence conveying instability to the comparison. Moreover, for flow data the comparison is systematically distorted by trading-day variations if present.

Seasonal adjustment entails the removal of seasonality, trading-day variations and moving-holiday effects from the raw data, to produce a seasonally adjusted series, which consists of the trend-cycle and the irregular components. The irregular fluctuations in the seasonally adjusted series can be reduced by smoothing, to isolate the trend-cycle and to enable month-to-month comparisons.

5. SEASONALITY

Time series of sub-yearly observations, *e.g.* monthly, quarterly, weekly, are often affected by seasonal variations. The presence of such variations in socio-economic activities has been recognized for a long time. Indeed seasonality usually accounts for most of the total variation within the year.

Seasonality is due to the fact that some months, quarters of the year are more important in terms of activity or level. For example, the level of unemployment is generally higher during the winter and spring months and lower in the other months. Yearly series cannot contain seasonality, because the component is specified to cancel out over 12 consecutive months or 4 consecutive quarters.

5.1. *The Causes and Costs of Seasonality*

Seasonality originates from climate and conventional seasons, like religious, social and civic events, which repeat from year to year.

The climatic seasons influence trade, agriculture, the consumption patterns of energy, fishing, mining and related activities. For example, in North America the consumption of heating oil increases in winter, and the consumption of electricity increases in the summer months because of air conditioning.

Institutional seasons like Christmas, Easter, civic holidays, the school and academic year have a large impact on retail trade and on the consumption of certain goods and services, namely travel by plane, hotel occupancy, consumption of gasoline.

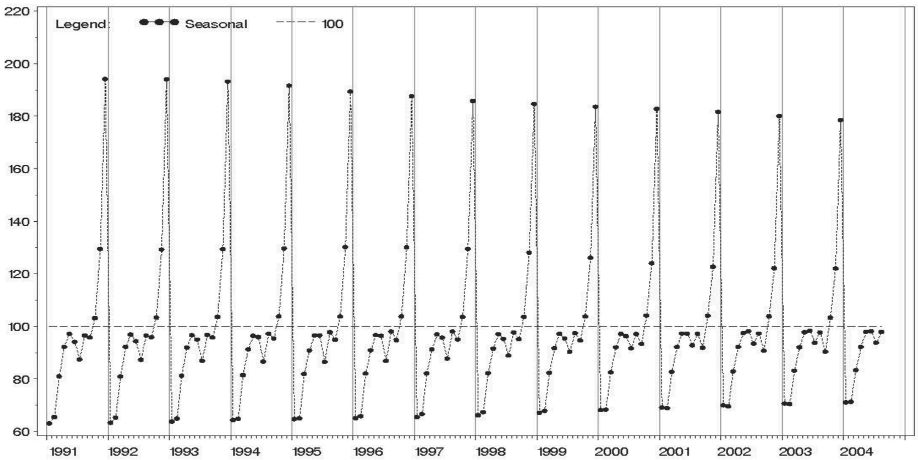

Fig. 1 – Seasonal pattern of Sales by Canadian Department Stores.

In order to determine whether a series contains seasonality, it is sufficient to identify at least one month (or quarter) which tends to be systematically higher or lower than other months. Fig.1 exhibits the seasonal pattern of sales by Canadian Department Stores, where the values are much larger in December and much lower in January and February with respect to other months. The seasonal pattern measures the relative importance of the months of the year. The constant 100% represents an average month or a non-seasonal month. The peak month is December, with sales almost 100% larger than on an average month; the trough months are January and February with sales almost 40% lower than on an average month. The seasonal amplitude, the difference between the peak and trough months of the seasonal pattern, reaches almost 140%.

Seasonality entails large costs to society and businesses. One cost is the necessity to build warehouses to store inventories of goods to be sold as consumers require them, for example grain elevators. Another cost is the under-use and over-use of the factors of production: capital and labour.

Capital in the form of un-used equipment, buildings and land during part of the year has to be financed regardless. For example, this is the case in farming, food processing, tourism, electrical generation, accounting. The cold climate increases the cost of buildings and infrastructure, *e.g.* roads, transportation systems, water and sewage systems, schools, hospitals; not to mention the damage to the same caused by the action of ice.

The labour force is over-used during the peak seasons of agriculture and construction for example; and, under-used in trough seasons sometimes leading to social problems.

A more subtle unwanted effect is that seasonality complicates business decisions by concealing the fundamental trend-cycle movement of the variables of interest.

The four main causes of seasonality are attributed to the weather, composition of the calendar, major institutional deadlines, and expectations. Seasonality is largely exogenous to the economic system but can be partially offset by human intervention. For example, seasonality in money supply can be controlled by central bank decisions on interest rates. In other cases, the effects can be offset by international and inter-regional trade. For example Hydro Québec, a major Canadian electrical supplier, sells much of it excess power during the summer seasonal trough months to the neighbouring Canadian provinces and U.S. states; and imports some of it during the winter seasonal peak months of electrical consumption in Québec. The scarcity of fresh fruits and vegetables in Canada is handled in a similar manner. Some workers and businesses manage their seasonal pattern with complementary occupations: for example landscaping in the summer and snow removal in winter.

To some extent seasonality can evolve through technological and institutional changes. For example the developments of appropriate construction materials and techniques made it possible to continue building in winter. The development of new crops, which better resist cold and dry weather, have influenced the seasonal pattern. The partial or total replacement of some crops by chemical substitutes, e.g. substitute of sugar, vanilla and other flavours, reduces seasonality in the economy.

As for institutional change, the extension of the Canadian academic year to the summer months in the 1970s affected the seasonal pattern of unemployment for the population of 15 to 25 years of age. Similarly the practice of spreading holidays over the whole year impacted on seasonality.

The changing industrial mix of an economy also transforms the seasonal pattern, because some industries are more seasonal than others. In particular, economies which diversify and depend less on "primary" industries (e.g. fishing, agriculture) typically become less seasonal.

In most situations, seasonality evolves slowly and gradually. Indeed the seasonal pattern basically repeats from year to year, as illustrated in Fig. 1. Merely repeating the seasonal pattern of the last twelve months usually provides a reasonable forecast.

5.2. *Models for Seasonality*

The simplest seasonal model for monthly seasonality can be written as

$$S_t = \Sigma_{j=1}^{12} \; \alpha_j \; d_{jt} + u_t \,, \; d_{jt} = \begin{cases} 1, & j = t \pm 12k \,, k = 0,1,2,...,11 \,, \\ 0, & otherwise \,, \end{cases} \tag{14}$$

subject to $\Sigma_{j=1}^{12} \; \alpha_j = 0$, $\{u_t\}$ is assumed white noise. The α_j are the seasonal effects and the $d_{jt}'s$ are dummy variables.

Model (14) can be equivalently written by means of sines and cosines

$$S_t = \Sigma_{j=1}^{6} \; [\alpha_j \; \cos(\lambda_j \; t) \; + \beta_j \; \sin(\lambda_j \; t)] \,, \tag{15}$$

where $\lambda_j = \; 2\pi \; j/12$, $j = 1, 2, ..., 6$ and $\beta_6 = 0$. The λ_j s are known as the seasonal frequencies, with j corresponding to cycles lasting 12, 6, 4, 3, 2.4 and 2 months respectively.

In order to represent stochastic seasonality, the α_j of model (14) are specified as random variables instead of constant coefficients (see Dagum, 2001). Such a model is

$$S_t = S_{t-12} + \omega_t \,, \tag{16.a}$$

$$\text{or} \quad (1 - B^{12}) \, S_t = \omega_t \,, \tag{16.b}$$

subject to constraints $\Sigma_{j=0}^{11} \; S_{t-j} \; = \omega_t$ where ω_t is assumed white noise.

Model (16.a) specifies seasonality as a non-stationary random walk process. Since $(1-B^s) \equiv (1-B)(1+B+...+B^{s-1})$, model-based seasonal adjustment method assigns $(1-B)$ to the trend and $S(B) = \Sigma_{j=0}^{s-1} \; B^j$ to the seasonal component. Hence, the corresponding seasonal model is

$$\Sigma_{j=0}^{s-1} \; S_{t-j} = \omega_t \,, \tag{17}$$

which entails a volatile seasonal behaviour, because the sum is not constrained to 0 but to the value of ω_t. Indeed, the spectrum of $\Sigma_{j=0}^{s-1} \; B^j$ (not shown here) displays broad bands at the high seasonal frequencies, i.e. corresponding to cycles of 4, 3, and 2.4 months.

Model (17) has been used in many structural time series models (see e.g. Harvey 1981, Kitagawa and Gersch 1984). A very important variant to model (17) was introduced by Hillmer and Tiao (1982) and largely discussed in Bell and Hillmer (1984), that is

$$\Sigma_{j=0}^{s-1} \; S_{t-j} = \eta_s(B)\, b_t \,, \tag{18}$$

where $\eta_s(B)$ is a moving average of $s-1$ minimum order and $b_t \sim WN(0, \sigma_b^2)$. The moving average component enables seasonality to evolve gradually. Indeed, the moving average eliminates the afore mentioned bands at the high seasonal frequencies.

Another stochastic seasonality model is based on trigonometric functions (see Harvey 1989) defined as

$$S_t = \Sigma_{j=1}^{[s/2]} \; \gamma_{jt} \,, \tag{19}$$

where γ_{jt} denotes the seasonal effects generated by

$$\begin{bmatrix} \gamma_{jt} \\ \gamma^*_{jt} \end{bmatrix} = \begin{bmatrix} \cos & \sin \\ -\sin \; \lambda_j & \cos \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \gamma^*_{j,t-1} \end{bmatrix} + \begin{bmatrix} \omega_{jt} \\ \omega^*_{jt} \end{bmatrix}, \tag{20}$$

and $\lambda_j = 2\pi\, j/s$, $j = 1,...,[s/2]$ and $t = 1,...,T$. The seasonal innovation ω_{jt} and ω^*_{jt} are mutually uncorrelated with zero means and common variance σ_ω^2.

6. THE MOVING-HOLIDAY COMPONENT

The moving-holiday or moving-festival component is attributed to calendar variations, namely the fact that some holidays change date from year to year.

For example, Easter can fall between March 23 to April 25, and Labour Day on the first Monday of September. The Chinese New Year date depends on the lunar calendar. Ramadan falls eleven days earlier from year to year. In the Moslem world, Israel and in the Far East, there are many such festivals. For example, Malaysia contends with as many as eleven moving festivals, due to its religious and ethnic diversity. These festivals affect flow and stock variables and may cause a displacement of activity from one month to the previous or the following month. For example, an early date of Easter in March or early April can cause an important excess of activity in March and a corresponding short-fall in April, in variables associated to imports, exports, tourism. When the Christian Easter falls late in April (e.g. beyond the 10-th), the effect is captured by the seasonal factor of April. In the long run, Easter falls in April eleven times out of fourteen.

Some of these festivals have a positive impact on certain variables, for examples air traffic, sales of gasoline, hotel occupancy, restaurant activity, sales of flowers and chocolate (in the case of Easter), sales of children clothing (Labour Day).[1] The impact may be negative on other industries or sectors which close or reduce their activity during these festivals.

The festival effect may affect only the day of the festival itself, or a number of days preceding and/or following the festival. In the case of Easter, travellers tend to leave a few days before and return after Easter, which affects air traffic and hotel occupancy, etc., for a number of days. Purchases of flowers and other highly perishable goods, on the other hand, are tightly clustered immediately before the Easter date.

The effect of moving festivals can be seen as a seasonal effect dependent on the date(s) of the festival. Fig. 2 displays the Easter effect on the sales by Canadian Department Stores. In this particular case, the Easter effect is rather mild. In some of the years, the effect is absent because Easter fell too late in April.

In the case exemplified, the effect is felt seven days before Easter and on Easter Sunday but not after Easter. This is evidenced by years 1994, 1996 and 1999 where Easters falls early in April and impacts the month of March. Note that the later Easter falls in April, the smaller the displacement of activity to March; after a certain date the effect is entirely captured by the April seasonal factor. The effect is rather moderate for Department Stores. This may not be the case for other variables. For example, Imports and Exports are substantially affected by Easter, because Customs do not operate from Good Friday to Easter Monday. Easter can also significantly affect quarterly series, by displacing activity from the second to the first quarter.

[1] In Canada and the United States, the school year typically starts the day after Labour Day (the first Monday of September).

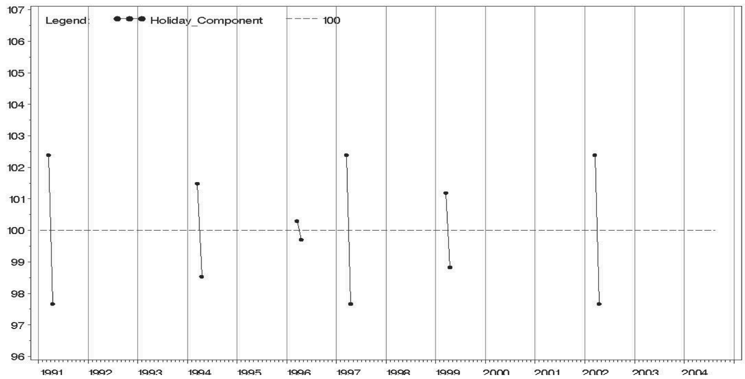

Fig. 2 – Moving Holiday component of the Sales by Canadian Department Stores.

There has been cases of complete reversal on the timing of the Easter effect. For example, Marriages in Canada were performed mainly by the Church during the 1940*s* up to the 1960*s*. The Church did not celebrate marriages during the Lent period, i.e. the 40 days before Easter. Some marriages therefore were celebrated before the Lent period, potentially affecting February and March. However, if Easter fell too early, many of these marriages were postponed after Easter. Generally, festival effects are difficult to estimate, because the nature and the shape of the effect are often not well known. Furthermore, there are few observations, *i.e.* one occurrence per year.

7. THE TRADING-DAY COMPONENT

Flow series may be affected by other variations associated with the composition of the calendar. The most important calendar variations are the trading-day variations, which are due to the fact that some days of the week are more important than others. Trading-day variations imply the existence of a daily pattern analogous to the seasonal pattern. However, these daily factors are usually referred to as daily coefficients.

Depending on the socio-economic variable considered, some days may be 60% more important than an average day and other days, 80% less important. If the more important days of the week appear five times in a month (instead of four), the month registers an excess of activity ceteris paribus. If the less important days appear five times, the month records a short-fall. As a result, the monthly trading-day component can cause increase of +8% or -8% (say) between neighbouring months and also between same-months of neighbouring years. The trading-day component is usually considered as negligible and very difficult to estimate in quarterly series.

For the multiplicative, the log-additive and the additive time series decomposition models, the monthly trading-day component is respectively obtained in the following manner

$$D_t = \Sigma_{\tau \in t}\ d_\tau / n_t \equiv (2800 + \Sigma_{\tau \in t\, 5\ times}\ d_\tau) / n_t\ , \tag{21.a}$$

$$D_t = \exp(\Sigma_{\tau \in t}\ d_\tau / n_t) \equiv \exp((\Sigma_{\tau \in t\, 5\ times}\ d_\tau) / n_t)\ , \tag{21.b}$$

$$D_t = \Sigma_{\tau \in t}\ d_\tau \equiv (\Sigma_{\tau \in t\, 5\ times}\ d_\tau)\ , \tag{21.c}$$

where d_τ are the daily coefficients in the month. The preferred option regarding n_t is to set it equal to the number of days in month t, so that the length-of-month effect is captured by the multiplicative seasonal factors, except for Februaries.[2] The other option is to set n_t equal to 30.4375, so that the multiplicative trading-day component also accounts for the length-of-month effect. The number 2800 in Eq. (21.a) is the sum of the first 28 days of the months expressed in percentage.

Same-month year-ago comparisons are never valid in the presence of trading-day variations, not even as a rule of thumb. For a given set of daily coefficients, there are only 22 different monthly values for the trading-day component, for a given set of daily coefficients: seven values for 31-day months (depending on which day the month starts), seven for 30-day months, seven for 29-day months and one for 28-day months. In other words, there are at most 22 possible arrangements of days in monthly data.

7.1. *Models for Trading-Day Variations*

A frequently applied deterministic model for trading-day variations was developed by Young (1965),

$$y_t = D_t + u_t\ ,\ t = 1, ,..., n\ , \tag{22.a}$$

$$D_t = \Sigma_{j=1}^{7}\ \alpha_j\ N_{jt}\ , \tag{22.b}$$

where $u_t \sim WN(0,\ \sigma_u^2)$, $\Sigma_{j=1}^{7}\ \alpha_j = 0$, α_j, $j = 1,...,7$ denote the effects of the seven days of the week, Monday to Sunday, and N_{jt} is the number of times day *j* is present in month *t*. Hence, the length of the month is $N_t = \Sigma_{j=1}^{7}\ N_{jt}$, and the cumulative monthly effect is given by (22.b). Adding and subtracting $\overline{\alpha} = (\Sigma_{j=1}^{7}\ \alpha_j) / 7$ to Eq. (22.b) yields

$$D_t = \overline{\alpha}\ N_t + \Sigma_{j=1}^{7}\ (\alpha_j - \overline{\alpha})\ N_{jt}\ . \tag{23}$$

[2] To adjust Februaries for the lengh-of-month, the seasonal factors of that month are multiplied by 29/28.25 and 28/28.25 for the leap and non-leap years respectively.

Hence, the cumulative effect is given by the length of the month plus the net effect due to the days of the week. Since $\Sigma_{j=1}^{7}\ (\alpha_j - \overline{\alpha}) = 0$, model (23) takes into account the effect of the days present five times in the month. Model (23) can then be written as

$$D_t = \overline{\alpha}\ \ N_t + \Sigma_{j=1}^{6}\ (\alpha_j - \overline{\alpha})\ (N_{jt} - N_{7t}), \tag{24}$$

with the effect of Sunday being $\alpha_7 = -\Sigma_{j=1}^{6}\ \alpha_j$.

Deterministic models for trading-day variations assume that the daily activity coefficients are constant over the whole range of the series. Stochastic model for trading-day variations have been rarely proposed. Dagum *et al.* (1992) developed a model where the daily coefficients change over time according to a stochastic difference equation.

8. THE IRREGULAR COMPONENT

The irregular component in any decomposition model represents variations related to unpredictable events of all kinds. Most irregular values have a stable pattern, but some extreme values or outliers may be present. Outliers can often be traced to identifiable causes, for example strikes, droughts, floods, data processing errors. Some outliers are the result of displacement of activity from one month to the other.

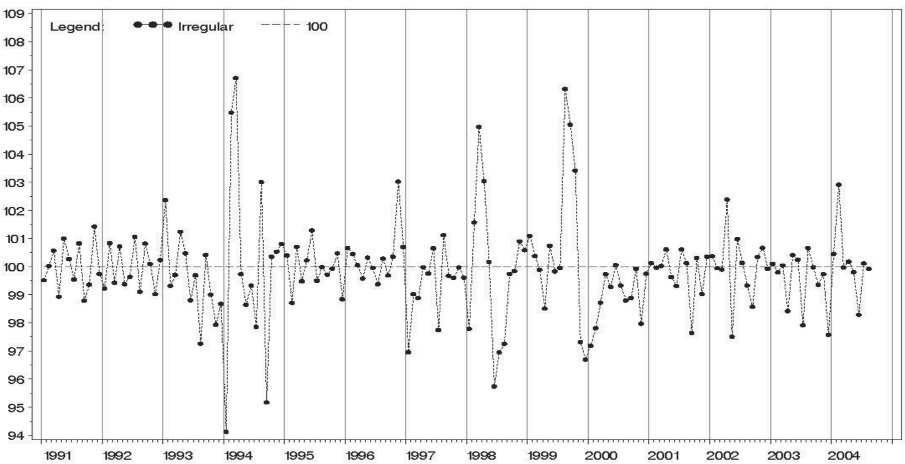

Fig. 3 – Irregular component of Sales by Canadian Department Stores.

Fig. 3 displays the irregular component of Sales by Canadian Department Stores, which comprises extreme values, namely in 1994, 1998, 1999 and Jan 2000. Most of these outliers have to do with the closure of some department stores and the entry of a large department store in the Canadian market.

As illustrated by Fig. 3, the values of the irregular component may be very informative, as they quantify the effect of events known to have happened.

Note that it is much easier to locate outliers in the irregular component than in the raw series because the presence of seasonality hides the irregular fluctuations.

The irregulars are most commonly assumed to follow a white noise process defined by

$$E(u_t)=0\,,\, E(u_t^2)=\sigma_u^2<\infty\,,\, E(u_t\ u_{t-k})=0\ \text{ if } k\neq 0\,.$$

If σ_u^2 is assumed constant (homoscedastic condition), u_t is referred to as white noise in the strict sense. If σ_u^2 is finite but not constant (heteroscedastic condition), u_t is called white noise in the weak sense.

For inferential purposes, the irregular component is often assumed to be normally distributed and not correlated, which implies independence. Hence, $u_t \sim NID(0,\sigma_u^2)$.

There are different models proposed for the presence of outliers depending on how they impact the series under question. If the effect is transitory, the outlier is said to be additive; and if permanent, to be multiplicative.

Box and Tiao (1975) introduced the following intervention model to deal with different types of outliers,

$$y_t=\Sigma_{j=0}^{\infty}\ h_{t-j}x_{t-j}+\eta_t=\Sigma_{j=0}^{\infty}h_j B^j x_j+\eta_t=H(B)x_t+\eta_t \tag{25}$$

where the observed series $\{y_t\}$ consists of an input series $\{x_t\}$ considered a deterministic function of time and a stationary process $\{\eta_t\}$ of zero mean and non-correlated with $\{x_t\}$. In such a case the mean of $\{y_t\}$ is given by the deterministic function $\Sigma_{j=0}^{\infty} h_{t-j}\, x_{t-j}$. The type of function assumed for $\{x_t\}$ and weights $\{h_j\}$ depend on the characteristic of the outlier or unusual event and its impact on the series.

9. LINEAR AND NONLINEAR TIME SERIES MODELS

Models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are the autoregressive (AR) models, the integrated (I) models, and the moving average (MA) models.

These three classes depend linearly on previous data points. Combinations of these three types produce autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models. The autoregressive fractionally integrated moving average (ARFIMA) model generalizes the former two. Extensions of these classes to deal with vector-valued data are available under the heading of multivariate time-series models. An additional set of extensions of these models is available for use where the observed time-series is driven by some "forcing" time-series (which may not have a causal effect on the observed series): the distinction from the multivariate case is that the forcing series may be deterministic or under the experimenter's control. For these models, the acronyms are extended with a final "*X*" for "exogenous".

Non-linear dependence of the level of a series on previous data points is of interest, partly because of the possibility of producing a chaotic time series. However, more importantly, empirical investigations can indicate the advantage of using predictions derived from non-linear models over those from linear models.When dealing with nonlinearities, Campbell, Lo, and MacKinlay (1997) make the distinction between: (1) Linear time series where shocks are assumed to be uncorrelated but not necessarily identically independent distributed (iid), and (2) Nonlinear time series where shocks are assumed to be iid, but there is a nonlinear function relating the observed time series $\{x_t\}$ and the underlying shocks.

Among the most applied non-linear time series models in financial data are those representing changes of variance along time (heteroskedasticity). These models are called autoregressive conditional heteroskedasticity (ARCH) and the collection comprises a wide variety of representation (GARCH, TARCH, EGARCH, FIGARCH, CGARCH, etc). Here changes in variability are related to, or predicted by, recent past values of the observed series. This is in contrast to other possible representations of locally varying variability, where the variability might be modeled as being driven by a separate time-varying process, as in a doubly stochastic model.

9.1. *Autoregressive Conditional Heteroskedasticity (ARCH) Model*

Autoregressive Conditional Heteroskedasticity (ARCH) models are used to characterize and model observed time series. They are used whenever there is reason to believe that, at any point in a series, the terms will have a characteristic size, or variance. In particular ARCH models assume the variance of the current error term or innovation to be a function of the actual sizes of the previous time periods' error terms: often the variance is related to the squares of the previous innovations. Such models are often called ARCH models (Engle, 1982), although a variety of other acronyms is applied to particular structures of model which have a similar basis. ARCH models are employed commonly in modeling financial time series that exhibit time-varying volatility clustering, *i.e.* periods of swings followed by periods of relative calm.

Suppose one wishes to model a time series using an ARCH process of order q. Let ε_t denote the error terms (return residuals, with respect to a mean process) i.e. the series terms. These ε_t are split into a stochastic part z_t and a time-dependent standard deviation σ_t characterizing the typical size of the terms so that

$$\varepsilon_t = \sigma_t z_t \tag{26}$$

where z_t is a random variable drawn from a Normal distribution centered at 0 with standard deviation equal to 1. (*i.e.* $z_t \xrightarrow{iid} N(0,1)$) and where the series $\sigma_t{}^2$ are modeled by

$$\sigma_t{}^2 = -\alpha_0 + \alpha_1 \varepsilon_t{}^2 + ... + \alpha_q \varepsilon_{t-q}^2 = -\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \tag{27}$$

and where $\alpha_0 \geq 0,\ \alpha_i \geq 0, i > 0$. An ARCH($q$) model can be estimated using ordinary least squares. A methodology to test for the lag length of ARCH errors using the Lagrange multiplier test was proposed by Engle (1982).

9.2. *Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Model*

If an ARMA model is assumed for the error variance, the model is a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model introduced by Bollerslev in 1996. In that case, the GARCH(p, q) model (where p is the order of the GARCH terms $\sigma_t{}^2$ and q is the order of the ARCH terms ε_t) is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_t^2 + ... + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + ... + \beta_p \sigma_{t-p}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{28}$$

The Nonlinear GARCH (NGARCH) also known as Nonlinear Asymmetric GARCH(1,1) (NAGARCH) was introduced by Engle and Ng in 1993

$$\sigma_t{}^2 = \omega + \alpha(\varepsilon_{t-1} - \theta\sigma_{t-1})^2 + \beta\sigma_{t-1}^2 . \tag{29}$$

$\alpha, \beta \geq 0; \omega > 0$. For stock returns, parameter θ is usually estimated to be positive; in this case, it reflects the leverage effect, signifying that negative returns increase future volatility by a larger amount than positive returns of the same magnitude.

9.3. *Self-Exciting Threshold AutoRegressive (SETAR) Model*

Another type of nonlinear time series models are the Self-Exciting Threshold AutoRegressive (SETAR) models introduced in a seminal paper by Tong and Lim (1980). They are typically applied as an extension of autoregressive models, in order to allow for higher degree of flexibility in model parameters through a regime switching behavior. Given a time series of data x_t, the SETAR model is a tool for understanding and, perhaps, predicting future values in this series, assuming that the behavior of the series changes once the series enters a different regime. The switch from one regime to another depends on the past values of the x series (hence the Self-Exciting portion of the name). The model consists of k autoregressive (AR) parts, each for a different regime. The model is usually referred to as the SETAR(k, p) model where k is the number of regimes and p is the order of the autoregressive part (since those can differ between regimes, the p portion is sometimes dropped and models are denoted simply as SETAR(k).They allow for changes in the model parameters according to the value of weakly exogenous threshold variable z_t, assumed to be past values of y, e.g. y_{t-d}, where d is the delay parameter, triggering the changes. Defined in this way, SETAR model can be presented as follows:

$$y_t = X_t \gamma^{(j)} + \sigma^{(j)} \varepsilon_t \quad \text{if} \quad r_{j-1} < z_t < r_j$$

where $X_t = (1, y_{t-1}, y_{t-2}, ..., y_{t-p})$ is a column vector of variables; $-\infty = r_0 < r_1 < ... < r_k = +\infty$ are k-1 non-trivial thresholds dividing the domain of z_t into k different regimes. In each of the k regimes, the AR(p) process is governed by a different set of p variables: $\gamma^{(j)}$. In such setting, a change of the regime (because the past values of the series y_{t-d} surpassed the threshold) causes a different set of coefficients: $\gamma^{(j)}$. to govern the process y. The SETAR model is a special case of Tong's general threshold autoregressive models (Tong 1990). The latter allows the threshold variable to be very flexible, such as an exogenous time series in the open-loop threshold autoregressive system, a Markov chain in the Markov-chain driven threshold autoregressive model which is now also known as the Markov switching model.

Department of Statistics ESTELA BEE DAGUM
University of Bologna

REFERENCES

ALEXANDROV, T.A.; BIANCONCINI, S.; DAGUM, E.B.; MAAS, P. AND MCELROY, S.T. (2010): "Review of some modern approaches to the problem of trend extraction", *Econometric Reviews*, forthcoming.

BELL, W.R. (1984): "Signal Extraction for Non-stationary Time Series", *Annals of Statistics*, 646-664.

BELL W.H., HILLMER, S.C., (1984): "Issues Involved with the Seasonal Adjustment of Economic Time Series", *Journal of Business and Economic Statistics*, 2, 3, pp. 291-320.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31:307-327.

BOX, G.E.P., JENKINS, G.M. (1970): *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, U.S.A (seconda edizione, 1976).

BOX, G.E.P., TIAO, G.C. (1975): "Intervention Analysis with Applications to Economic and Environmental Problems", *Journal of the American Statistical Association,* 70, pp. 70-79.

BURMAN, J.P. (1980): "Seasonal adjustment by signal extraction", *Journal of the Royal Statistical Society, Series A*, 143, pp. 321-337.

CAMPBEL, J.Y., LO A.W., AND MACKINLAY, A.C. (1997): *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.

CHAAB, N., MORRY, M., DAGUM, E.B. (1999): "Further Results on Alternative Trend-Cycle Estimators for Current Economic Analysis", *Estadistica*, 6, pp. 3-73.

CLEVELAND, R.B., CLEVELAND, W.S., MCRAE, J.E., TERPENNING, I.J. (1990): "A Seasonal Trend Decomposition Procedure Based on Loess", *Journal of Official Statistics*, 6, pp. 3-73.

CLEVELAND, W.P., TIAO, G.C. (1976): "Decomposition of Seasonal Time Series: a Model for Census X11 Program", *Journal of the American Statistical Association*, 71, pp. 581-587.

DAGUM, C AND DAGUM, E.B. (2006): "Stochastic and Deterministic trend Models", *Statistica*, 3, 269-280.

DAGUM, E.B. (1980): *The X-11-ARIMA Seasonal Adjustment Method*, Statistics Canada, Ottawa, Canada.

DAGUM, E.B. (1982a): "Revisions of Time Varying Seasonal Filters", *Journal of Forecasting*, 1, pp. 173-187.

DAGUM, E.B. (1982b): "The Effects of Asymmetric Filters of Seasonal Factor Revisions", *Journal of the American Statistical Association*, 77, pp. 732-738.

DAGUM, E.B. (1982c): "Revisions of Seasonally Adjusted Data Due to Filter Changes", Proceedings of the Business and Economic Section, American Statistical Association, pp. 39-45.

DAGUM, E.B. (1983): "Spectral Properties of the Concurrent and Forecasting Seasonal Linear Filters of the X11ARIMA Method", *The Canadian Journal of Statistics*, 11, pp. 73-90.

DAGUM, E.B. (1988): *The X11ARIMA/88 Seasonal Adjustment Method – Foundations and User's Manual*, Time-series research and analysis centre, Statistics Canada, Ottawa, Canada.

DAGUM, E.B. (1996): "A New Method to Reduce Unwanted Ripples and Revisions in Trend-Cycle Estimates from X11ARIMA", *Survey Methodology*, 22, 1, pp. 77-83.

DAGUM, E.B. (2001): "Time Series: Seasonal Adjustment", in *International Encyclopedia of the Social and Behavioral Sciences*, vol. II - Methodology, Fienberg, S. e Kadane, J. Editors, Elsevier Sciences Pub.

DAGUM, E.B. AND BIANCONCINI, S. (2006): "Local Polynomial Trend-cycle Predictors in Reproducing Kernel Hilbert Spaces for Current Economic Analysis. *Anales de Economía Aplicada*, pp. 1-22.

DAGUM, E.B AND BIANCONCINI, S. (2008): "The Henderson Smoother in Reproducing Kernel Hilbert Space", *Journal of Business and Economic Statistics*, 26 (4), pp. 536-545.

DAGUM, E.B. AND BIANCONCINI, S. (2009): "Recent Developments in Short-term Trend Prediction for Real Time Analysis", Proceedings of the American Statistical Association, Business and Economic Statistics Section, Washington D.C., August (invited paper).

DAGUM, E.B. AND BIANCONCINI, S. (2010): "A Unified Probabilistic View of Nonparametric Predictors via Reproducing Kernel Hilbert Spaces", under review.

DAGUM, E.B. AND LUATI, A. (2000): "Predictive Performance of some Nonparametric Linear and Nonlinear Smoothers for Noisy data", *Statistica*, Anno LX, vol. 4, pp. 635-654.

DAGUM, E.B. AND LUATI, A. (2009): "A Cascade Linear Filter to Reduce Revisions and False Turning Points in Real Time Trend- cycle Estimation", *Econometric Reviews*, vol. 28, 1-3, pp. 40-59.

DAGUM, E.B., QUENNEVILLE, B., SUTRADHAR, B. (1992): "Trading-Day Variations Multiple Regression Models with Random Parameters", *International Statistical Review*, 60, 1, pp. 57-73.

EHGLEN, J. (1998), 'Distortionary effects of the optimal Hodrick–Prescott filter', *Economics Letters* 61, 345-349.

ENGLE, R.F. (1982): "Heteroscedasticity with Estimates of Variance of United Kingdom Inflation", *Econometrica* 50: 987-1008.

FINDLEY, D.F., MONSELL, B.C., BELL, W.R., OTTO, M.C., CHEN B. (1998): "New Capabilities and Methods of the X12-ARIMA Seasonal Adjustment Program", *Journal of Business and Economic Statistics*, 16, 2.

GERSCH, W., KITAGAWA, G. (1983): "The Prediction of Time Series with Trends and Seasonalities", *Journal of Business and Economic Statistics*, 1, pp. 253-263.

GOMEZ, V., MARAVALL, A. (1996): *TRAMO-SEATS*, Bank of Spain, Madrid.

HANNAN, E.J. (1967): "Measuring of Wandering Signal Amid Noise", *J. of Applied Probability*,4, 90-102.

HARVEY, A.C. (1981, 1993): *Time Series Models*, Harvester- Wheatsheaf, Great Britain.

HARVEY, A.C. (1985): "Trends and Cycles in Macroeconomic Time Series", *Journal of Business and Economic Statistics*, 3, pp. 216-227.

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

HARVEY, A.C. AND JAEGER, A. (1993), Detrending, Stylized Facts and the Business Cycle, *Journal of Applied Econometrics*, 8, 231-247.

HENDERSON, R. (1916): "Note on Graduation by Adjusted Average", *Transaction of the Actuarial Society of America*, 17, pp. 43-48.

HILLMER, S.C., TIAO, G.C. (1982): "An ARIMA Model Based Approach to Seasonal Adjustment", *Journal of the American Statistical Association*, 77, pp. 63-70.

KAISER, R. AND MARAVALL A. (1999), "Estimation of the Business-Cycle: A Modified Hodrick-Prescott filter", *Spanish Economic Review*, 1, 175-206.

KING, R.G. AND REBELO, S.T. (1993), "Low Frequency Filtering and Real Business Cycles", *Journal of Economic Dynamics and Control*, 17, 207-233.

KITAGAWA, G., GERSCH, W. (1984): "A Smoothness Priors-State Space Modeling of Time Series with Trend and Seasonality", *Journal of the American Statistical Association*, 79, pp. 378-389.

KOLMOGOROV, A.N., (1939), Sur l'interpolation et l'extrapolation des suites stationnairres: *C.R. Acad.Sci.*, 208, 2043-2045.

KOLMOGOROV A.N., (1941). "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers, "Proceedings of the USSR Academy of Sciences 30: 299-303. (Russian), translated into English in Proceedings of the Royal Society of London, Series A: Mathematical and Physical Sciences 434 (1991): 9-13.

KOOPMAN, S.J., HARVEY, A.C., DOORNIK, J.A., SHEPHARD, N. (1998): *Structural Time Series Analyser, STAMP (5.0)*, International Thomson Business Press, London, England.

HODRICK, R.J. AND PRESCOTT, E.C. (1997): "Post war U.S. Business Cycles: An Empirical Investigation", *Journal of Money, Credit and Banking*,. 29, nº 1, 1-16.

HILLMER, S.C., TIAO, G.C. (1982): "An ARIMA Model Based Approach to Seasonal Adjustment", *Journal of the American Statistical Association*, 77, pp. 63-70.

LADIRAY, D., QUENNEVILLE, B. (2001): "*Seasonal Adjustment with the X11 Method*", Lecture Notes in Statistics no. 158, Springer-Verlag, New York.

MACAULAY, F.R. (1931): *The Smoothing of Time Series*, National Bureau of Economic Research, Washington, D.C.

MARAVALL, A. (1993). "Stochastic and Linear Trends: Models and Estimators", *Journal of Econometrics*, 56, pp. 5-37.

PERSONS, W.M. (1919): "Indices of Business Conditions", *Review of Economic Statistics*, 1, pp. 5-107.

SHISKIN, J., YOUNG, A.H., MUSGRAVE, J.C. (1967): "*The X-11 Variant of Census Method II Seasonal Adjustment*", Technical Paper n. 15, Washington, D.C.: U.S. Bureau of Census.

SOBEL, E.L. (1967): "Prediction of Noise-distorted Multivariate Non-stationary Signal" *J. of Applied Probability*, 4, 330-342.

TONG, H. AND LIM, K. S. (1980) "Threshold Autoregression, Limit Cycles and Cyclical Data (with discussion)", *Journal of the Royal Statistical Society, Series B*, 42, 245-292.

TONG H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press.

WHITTAKER, E.T. (1923), "On a New Method of Graduation", Proceedings of the Edinburgh Mathematical Society 41, 63-75.

WHITTAKER, E.T. AND ROBINSON, G. (1924), *Calculus of Observations: a Treasure on Numerical Calculations*, Blackie and Son, London.

WIENER, N. (1949): *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, John Wiley and Sons, New York.

WHITTLE, P. (1963): *Prediction and regulation by linear least-square methods*, English Universities Press, London.

WOLD, H.O. (1938): *A Study in the Analysis of Stationary Time Series*, Almquist and Wicksell, Uppsala, (second edition 1954).

YOUNG, A.H. (1965): "*Estimating Trading-Day Variation in Monthly Economic Time Series*", Technical Paper 12, Washington, D.C.: U.S. Bureau of Census.

ZELLNER, A. HONG,C. AND MIN, C. (1991): "Forecasting Turning Points in International Output Growth Rates using Bayesian Exponentially Weighted Autoregression, Time Varying Parameters and Pooling Techniques", *Journal of Econometrics*, 48, 275-304.

SUMMARY

Time series modeling and decomposition

The paper provides an overview of techniques and methods in time series modeling and decomposition with focus on the business cycle, models for seasonality, the moving holiday component, the trading-day component and the irregular component.