What marketing scholars should know about time series analysis

Horváth, Csilla; Kornelis, Marcel; Leeflang, Peter S.H.

What Marketing Scholars Should Know About Time Series Analysis

Time series applications in marketing

Csilla Horváth   Marcel Kornelis   Peter S.H. Leeflang

SOM-theme F Interactions between consumers and firms

Abstract

In this review, we give a comprehensive summary of time series techniques in marketing, and discuss a variety of *T*ime *S*eries *A*nalysis (TSA) techniques and models. We classify them in the sets (i) univariate *TSA*, (ii) multivariate *TSA*, and (iii) multiple *TSA*. We provide relevant marketing applications of each set, and provide illustrative examples.

Key words: Time series analysis, ARMA models, VAR models, model calibration

1. Introduction

Marketing data often include measures on, for example, sales or marketing mix variables at equally spaced intervals over time. Time series models are uniquely suited to capture the time dependence in these variables.

*T*ime *S*eries *A*nalysis (*TSA*) techniques have been used for (i) forecasting, (ii) the determination of the temporal ordering among some variables through Granger causality tests, and (iii) the determination of the over-time impact of marketing variables or specific discrete events. Notwithstanding these applications, Dekimpe and Hanssens (2000) conclude that marketing scholars have been reluctant to use *TSA* for marketing purposes in the past. The authors identify key factors for this limited acceptance, viz. the scarcity of (i) adequate data sources, (ii) user-friendly software, and (iii) doctoral level training. Other factors are (iv) the reluctance of using data-driven approaches to model specification, and (v) the absence of a substantive marketing area where *TSA* was adopted as primary research tool. In recent years, however, (i) new data sources and longer time spans have become available, especially with the use of scanner data, and (ii) several software packages have been developed for *TSA* applications. Examples are Eviews (2000), Jmulti (2001), RATS (2000), or Microfit (Pesaran and Pesaran 1997). 'General' packages such as Gauss (2000), Matlab (1995), and Ox (Doornik et al. 1998) have developed extensive *TSA* toolboxes. Moreover, more marketing scholars are trained in *TSA* and apply these packages. The application of (iii) Structural *V*ector *A*uto*R*egressive e*X*ogenous variables (*SVARX*) models and co-integration analysis offer more of a confirmatory potential, which gives the data-driven character of TSA more credibility (Dekimpe and Hanssens 2000). Finally, (v) the development of *TSA* techniques gives a natural match between *TSA* and one of marketing's long-lasting interest fields: quantifying the long-run impact of marketing's tactical and strategic decisions. Dekimpe and Hanssens (2000) predict that the most productive use of *TSA* is still ahead. Arguments are (i) the recent growth of marketing data bases, (ii) the fact that models of evolution become critical to shape our understanding of market performance as business cycles shorten and marketing environments change more rapidly, and (iii) the expectation that more *TSA* will be used at the micro-level.

In this review, we discuss a variety of (linear) time series models. We classify these models in the sets (i) univariate *TSA*, (ii) multivariate *TSA*, and (iii) multiple *TSA* and provide relevant marketing applications each of these sets. In the first section, we introduce the basic concepts of *TSA*, and discuss *TSA* models with only one series of observations, such as sales (we call it univariate *TSA* in this review). These models consider a variable, e.g. sales, as a function of its own past, random shocks, seasonal dummies, and time. The *A*uto*R*egressive *I*ntegrated *M*oving *A*verage (*ARIMA*) model is the most general model in this set. We pay special attention to the issue of stationarity, as it is a fundamental concept in distinguishing long- and short- term effects of actions.

We then consider multivariate *TSA*, i.e. models where the dependent variable is explained by another variable, e.g. sales explained by advertising. Finally, we consider models in which more endogenous variables are considered simultaneously (we refer to it as multiple[1] *TSA*, in this review). Vector AutoRegressive models with eXogenous variables (*VARX* models), *S*tructural *VAR* (*SVAR*)models, and *V*ector *E*rror *C*orrection (*VEC*) models belong to this section.

We illustrate the different models by applications of *TSA* on the Lydia Pinkham's vegetable compound data. The Lydia Pinkham vegetable compound was introduced in 1873 as a remedy against menstrual pain and menopausal discomfort. A large court case made the company database public (Palda 1964). The database consists of sales data and data on advertising expenditures. The data set has been attractive to researchers for a long time because (i) advertising was almost the exclusive marketing instrument, (ii) there was no strong competition, and (iii) the data set covers a fairly long time span (see, for example, also Hanssens 1980, or Zanias 1994). Figures 1A and 1B present the available annual (1907-1960) and monthly (1954:01-1960:06) data of Lydia Pinkham's vegetable compound.

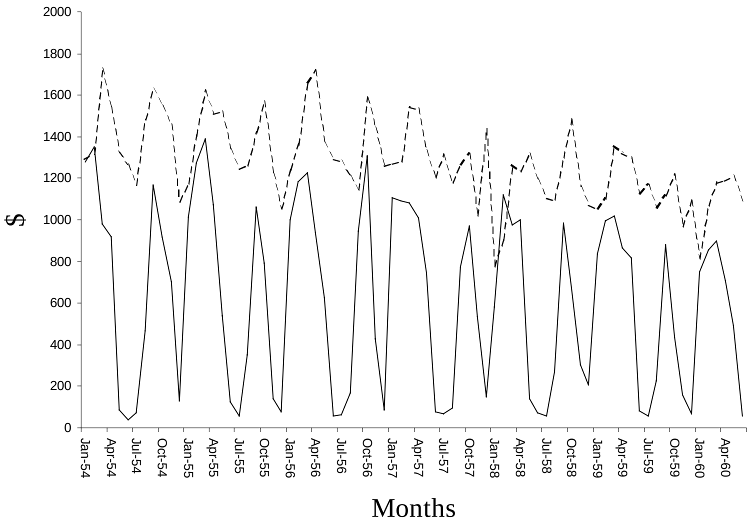
[1] The usage of apellations multivariate time series analysis and multiple time series analysis may be confusing to the reader. Still, we prefer to use these terms aiming at consistecy with the labelling of time series literature.

Figure 1A The Lydia Pinkham dataset of annual observations

--- = Sales, —— = Advertising

Source: Palda (1964, p.23)

Figure 1B The Lydia Pinkham dataset of monthly observations



--- = Sales, ___ = Advertising

Source: Palda (1964, pp. 32, 33)

2. Univariate time series analysis

In this section, we introduce the basic concepts of *TSA* and discuss some techniques that are useful in identifying the dynamic structure of the series. We focus on discrete time series models, and consider equidistant time points, as most marketing data is collected for discrete time periods. In our discussion on univariate time series, we start with *A*uto*R*egressive (*AR*) processes (Section 2.1) in which, say, sales are affected by sales levels in previous periods. In Section 2.2, we introduce *M*oving *A*verage (*MA*) processes. We discuss the *ARMA* processes that are the combination of the first two processes in Section 2. 3. Section 2. 4 introduces the concept of stationarity, which is a basic concept in distinguishing long- and short term effects of marketing actions. Section 2.5 deals with the concepts of trend and difference stationarities. In Section 2.6, we introduce the most general univariate *TS* model, the *ARIMA* model, that captures the behavior of a non-stationary variable. Section 2. 7 describes tests for unit root. Section 2.8. introduces univariate persistence developed by Dekimpe and Hanssens (1995a and 1999), which provides a stong link between *TSA* and marketing applications. Seasonality is discussed in Section 2.9. Table 1 provides a brief verview of the relevant univariate time series applications in marketing.

Table 1 Relevant univariate time series applications in marketing

Study	Country	Entity aggregation	Focus
Univariate $\left(y_t = f(y_{t-j})\right)$			
Geurts and Ibrahim (1975)	USA	Industry sales	Forecasting
Kapoor et al. (1981)	USA	Firm sales; Industry sales	Forecasting
Moriarty (1990)	USA	Sales	Forecasting
Moriarty and Adams (1979)	USA	Firm sales	Forecasting
Moriarty and Adams (1984)	USA	Sales	Forecasting

Univariate time series models seem to have been popular in the 70's, 80's and they were mainly applied for forecasting.

2.1. Autoregressive processes[2]

Let y_t be the sales of a brand in period t. A common and fairly simple way to describe fluctuations in sales over time is with a first-order autoregressive process. In this process, it is assumed that sales at $t-1$ affect sales at period t:

$$y_t = \mu + \varphi y_{t-1} + \varepsilon_t ,\ t = 1,...,T \tag{1}$$

where $\mu =$ a constant, $\varepsilon_t =$ the noise term, $\varphi =$ a parameter, the starting condition is $y_0 = 0$, and T is the sample length. Often, it is assumed that the noise term is 'white noise', which means that it has a mean of zero, a constant variance, and is serially uncorrelated (Enders 1995, p. 65). This model is indicated as *AR*(1), which means 'autoregressive process of order 1'. It can be identified from the data using the *A*uto*C*orrelation *F*unction (*ACF*) and the *P*artial *A*uto*C*orrelation *F*unction (*PACF*) calculated from sample data (see Box and Jenkins 1976 for details).

The *ACF* of a series y_t is defined by:

---

2 We partly based our review on Section 17.3 of Leeflang et al. (2000, p. 458 - 473).

$$\rho_k = \frac{\gamma_k}{\gamma_o} \quad (2)$$

where γ_k is the *k*th order autocovariance of y_t, that is:

$$\gamma_k = E[(y_t - m)(y_{t-k} - m)],\ k = \ldots,-2,-1,0,1,2,\ldots \quad (3)$$

where $m = E(y_t)$ and $E(.)$ is the expected value. The partial autocorrelation at lag i is the regression coefficient on y_{t-i} when y_t is regressed on a constant and on $y_{t-1},\ldots,y_{t-i}$. This is a partial correlation since it measures the correlation of y values that are i periods apart after removing the correlation from the intervening lags[3].

The *ACF* and *PACF* are statistics that can be used to identify the type of time series for a given data set (see also Hanssens et al. 2001). The *ACF* and *PACF* of an *AR*(1) process are shown in Figures 2A and 2B. Specifically, the *ACF* decays exponentially and the *PACF* shows a positive ‘spike’ at lag 1 and equals zero thereafter, if φ is positive in equation (1). The *ACF* shows a damped wavelike pattern and the *PACF* a negative spike at lag 1, if φ is negative.

It is possible that there is a correlation with higher order lags as well. For instance, in monthly data we may not only observe a correlation between two successive months, but also between the same months of different years. In that case, the *ACF* shows a damped pattern and the *PACF* indicates multiple spikes at lag 1 and lag 12. So, we cannot only identify dependence on the immediate past (one lag ago), but also on lags several periods ago.

The order *p* of an *AR*(*p*) process is the highest lag of y_t that appears in the model. The general *p*-order *AR* process is written as:

$$\varphi_p(B)y_t = \mu + \varepsilon_t,\ t = 1,\ldots,T \quad (4)$$

where $\varphi_p(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p)$ and $B =$ the backshift operator defined by $B^k y_t = y_{t-k}$. In this general case, the *ACF* damps down and the *PACF* cuts off after *p* lags.

---

[3] EViews estimates the partial autocorrelation at lag i recursively based on a consistent approximation of the partial autocorrelation. The algorithm is described in Box and Jenkins (1976, Part V, Description of computer programs).

Figure 2 ***ACF*** **and** ***PACF*** **functions of two** ***AR(1)*** **processes**

Figure A. $AR(1)$, $\varphi = 0.8$

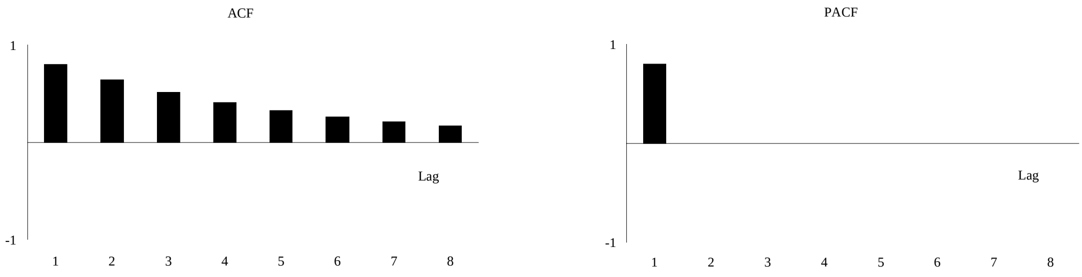
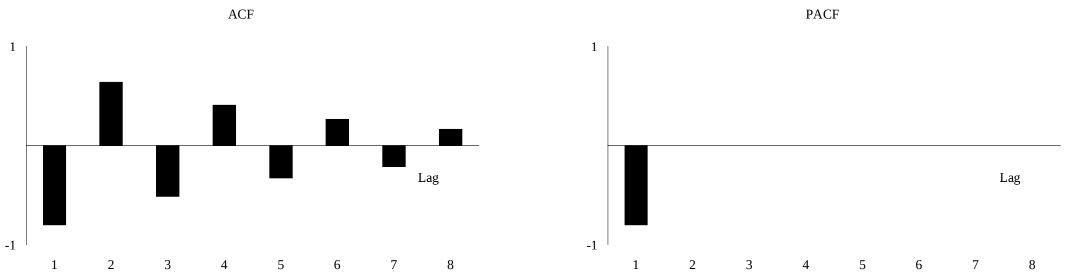


Figure B. $AR(1)$, $\varphi = -0.8$



Source: Pankratz (1991, p.40)

Besides the *ACF* and *PACF* criteria, the order of a time series model, i.e. the number of lagged variables in the model, is often selected on the basis of information criteria, such as the forecasting performance of the model, or *A*kaike's *I*nformation *C*riterion (*AIC*) and *S*chwarz' *B*ayesian *C*riterion (*SBC*) (for non-nested models), or Likelihood Ratio tests (for nested models).

2.2. Moving Average processes

A first-order moving average process assumes that a random shock at $t-1$ affects sales levels at time t:

$$y_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1},\ t = 1,...,T \tag{5}$$

This model is indicated as *MA*(1). Note that the past shock does not come from the past sales (past values of y_t) as in the *AR*(1) model, but it stems from the random component of ε_{t-1}. The *ACF* and *PACF* for the *MA*(1) model are depicted in Figures 3A and 3B. Here, the *ACF* shows a spike at lag 1, which is positive if $\theta < 0$ and negative if $\theta > 0$, while the *PACF* shows exponential decay in the former case, or a damped wavelike pattern in the latter.

Figure 3 *ACF* and *PACF* functions of two *MA(1)* processes

Figure A. *MA*(1), $\theta = -0.8$

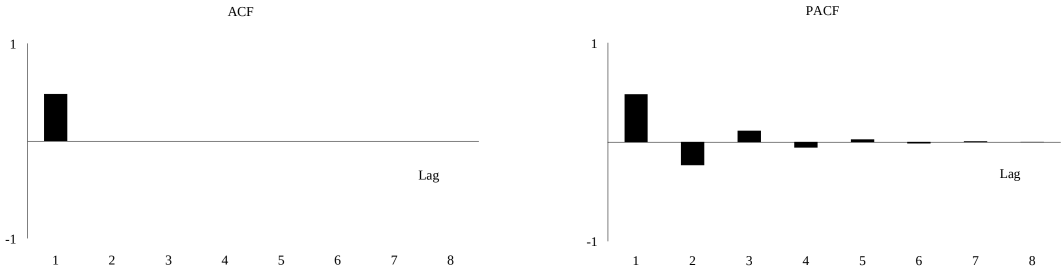
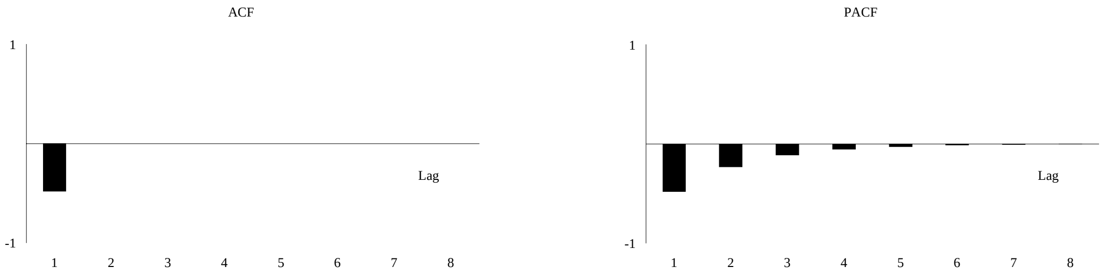


Figure B. *MA*(1), $\theta = 0.8$



Source: Pankratz (1991, p.40)

The order q of an $MA(q)$ process is the highest lag of ε_t that appears in the model. The general q-order MA process is written as:

$$y_t = \mu + \theta_q(B)\varepsilon_t \,,\ t = 1,...,T \tag{6}$$

where $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q)$ and B is the backshift operator defined as before. In this general case, the *PACF* damps down and the *ACF* cuts off after q lags.

2.3. ARMA processes

AR and *MA* processes can be combined into a single model to reflect the idea that both past sales and past random shocks affect y_t. For example, the *ARMA*(1,1) process is:

$$y_t = \mu + \varphi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \,,\ t = 1,...,T \tag{7}$$

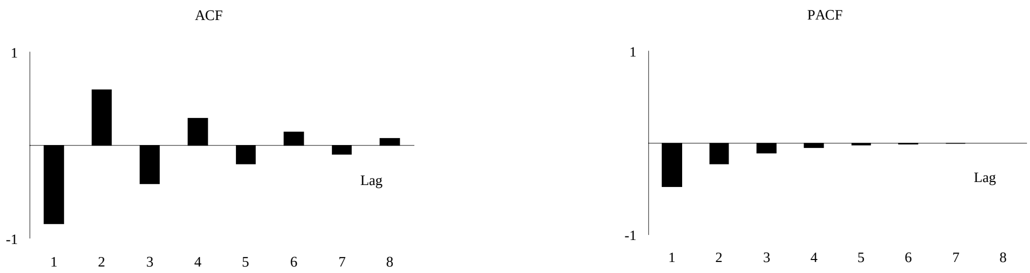
The attractiveness of the *ARMA* model is that it is a parsimonious representation of a stationary stochastic process (Harvey 1990, p. 30). We will discuss the term 'stationary' in Section 2.4.).

The *ACF* and *PACF* for an *ARMA*(1,1) model with $\varphi = -0.7$ and $\theta = -0.7$ are depicted in Figure 4. Here, the *ACF* shows a damped wavelike pattern and the *PACF* shows an exponential decay[4].

---

[4] The selection of the number of lags in *ARMA* models from the *ACF* and *PACF* functions is not always straightforward. In these cases an Extended *ACF* (*EACF*, details can be found in Tsay and Tiao 1984) or information criteria as *AIC* and *SBC* are useful alternatives in identifying the orders of *ARMA* models.

Figure 4 ***ARMA*(1,1),** $\varphi = -0.7, \theta = -0.7$



Source: Enders (1995, p.84).

The orders (*p,q*) of an *ARMA*(*p,q*) process are the highest lags of y_t and ε_t respectively that appear in the model. For example, for an *ARMA*(1,1) process, *p*=1 and *q*=1. The general *ARMA(p,q)* process is formulated as follows:

$$\varphi_p(B)y_t = \mu + \theta_q(B)\varepsilon_t\ ,\ t = 1,...,T \tag{8}$$

with $\varphi_p(B)$ and $\theta_q(B)$ as defined above.

2.4. Stationary and nonstationary processes

In order to use statistical tests for an *ARMA* model selection, we need to estimate the underlying stochastic process. We can do this by deriving the mean, variance, and co-variance of the sample data. However, these quantities are only meaningful (for obtaining the probability distribution and the statistical tests based on it) if they are independent of time[5]. This is the case if the series is stationary. There are different forms of stationarity. The most commonly considered is *covariance* stationarity. A series y_t is said to be *covariance* stationary if the following conditions hold:

[5] Besides, in the case of TS models the assumption that the expected value of the disturbances given the observed information is equal to zero does not hold. This means that unbiasedness of least squares estimator does not necessarily hold. In this case the Gauss-Markov theorem does not apply and we are left only with asymtotic results. The assumption of stationarity ensures consistency of least squares estimator. In sum, the

$$E(y_t) = m \text{, for all } t = 1,\ldots,T \tag{9}$$

$$E\left[(y_t - m)^2\right] = \gamma_0 < \infty \text{, for all } t = 1,\ldots,T \tag{10}$$

$$E[(y_t - m)(y_{t-i} - m)] = \gamma_i \text{, for all } t = 1,\ldots,T \text{ and for all } i = \ldots,-2,-1,0,1,2,\ldots \tag{11}$$

where m, γ_0, and γ_i are all finite-valued numbers (see also Lütkepohl 1993, p. 19). In practice, we often use the requirement that the roots of what are called the characteristic equations, $\varphi_p(.) = 0$, "lie outside the unit circle" [6]. For an *AR*(1) process of (1) this requirement implies that the root is larger than one in absolute value. The characteristic equation is $(1-\varphi B) = 0$. The root equals $\frac{1}{\varphi}$, which is greater than one in absolute sense if $|\varphi| < 1$. The root $\frac{1}{\varphi}$ is called a *unit root* if $|\varphi| = 1$. Thus, the *AR*(1) process is *stationary* if $|\varphi| < 1$ and *non-stationary* (not stationary) if it has a unit root[7]. Figures 2A and 2B present stationary *AR*(1) processes[8].

2.5. Deterministic and stochastic trends

The requirement that the mean level is independent of time implies that, for example, the average sales level at the beginning of the sample period is equal to the average sales level at the end of the sample. If this is the case, the process is called *L*evel *S*tationary (*LS*). Lal and Padmanabhan (1995) investigate the relationships between market share and promotional

---

assumption of stationarity is used preserve the property of consistency of least squares estimator.

[6] It can be shown that condition (10) no longer can be satisfied if the roots are on the unit circle.

[7] The case where $|\varphi| > 1$ is not considered in marketing studies, because it implies that past values of the dependent variable become and more important, which is highly unrealistic (Dekimpe and Hanssens, 1995).

[8] Similar to the stationarity conditions for *AR* processes, *MA* processes need to satisfy conditions for invertibility. An *MA*(1) process is invertible if $|\theta| < 1$ in (2.5). The *ARMA* processes need to satisfy both stationarity and invertibility conditions.

expenditures in the long run. They find that market shares are stationary for a majority of products. However, the *LS* condition is not always realistic in a marketing context. Often, we observe some kind of trending behavior, e.g. the sales in the end of the sample are higher (lower) than the initial sales. Lal and Padmanabhan (1995), for instance, find that a minority of the market shares that they investigated, shows a trend. Moreover, based on data from 400 published models, Dekimpe and Hanssens (1995b) conclude that evolution is the dominant characteristic in marketing series of market performance measures. Hence, we often have to cope with trends when we apply *TSA* techniques on marketing data. Trends in time series can be (i) deterministic or (ii) stochastic (Maddala and Kim 1998, p. 29). A *deterministic* trend imposes that the level is not constant, but can be perfectly predicted if the underlying deterministic function is known. One can approximate the deterministic growth path by a function of time. The linear time trend is the most commonly used function:

$$y_t = \mu + \beta t + \varepsilon_t,\ t = 1,...,T \tag{12}$$

In equation (12), the long-run behavior of y_t is determined by the series' individual (perfectly determined) growth path (βt). Every deviation from this growth path is temporal; in the long run the series always returns to its individual growth path (βt). Therefore, such a series is often called *T*rend *S*tationary (*TS*), because it is stationary around a trend.

If the data exhibit a *stochastic* trend, it implies that the variation is systematic but hardly predictable, because every temporary deviation may change the long-run performance of the series. This phenomenon is called shock *persistence* (see Section 2.8). A simple example of such a series is the *R*andom *W*alk (*RW*) process:

$$y_t = \beta + y_{t-1} + \varepsilon_t,\ t = 1,...,T \tag{13}$$

Repeated substitution reveals the nature of the stochastic trend and the corresponsence with equation (12):

$$y_t = \mu + \beta t + \sum_{i=1}^{t} \varepsilon_{t-i} + \varepsilon_t \tag{14}$$

assuming that $y_0 = \mu$. In equation (14) $\mu + \beta t$ is the deterministic trend, $\sum_{i=1}^{t} \varepsilon_{t-i}$ is the stochastic trend, and ε_t is the noise term.

In the case of a stochastic trend, the inclusion of a time function into the model does not remove the evolution, but the series can be formulated as a (stationary) *ARMA* processes in differences. The differencing operation removes the stochastic trend from the data. Therefore, a time series with a stochastic trend, which becomes stationary after differencing, is called *D*ifference *S*tationary (*DS*). It may be necessary to take the differences of the series more than once before it becomes stationary. Figures 5B and 5C show example of a trend stationary and a difference stationary process respectively. Figure 5A shows a level stationary series to which we could arrive by removing the deterministic trend from the series shown on Figure 5B or taking first differences of the RW process of Figure 5C. The choice between deterministic or stochastic trends depends on the researcher's beliefs or on the outcome of statistical tests (see Section 2.7).

Figure 5 **Examples of different types of time series**

Time Series	Plot

Figure A

LS series

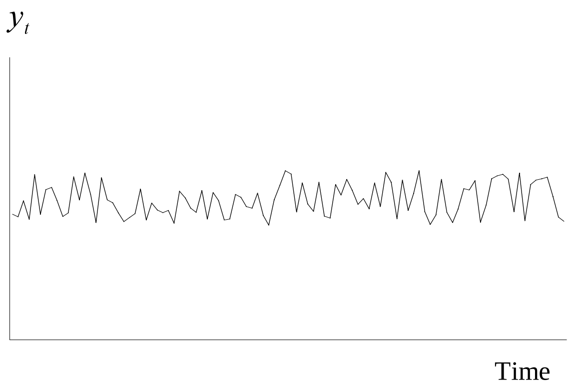


Figure B

TS series

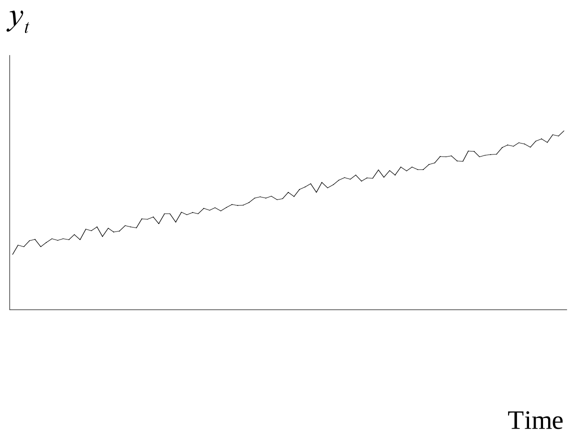
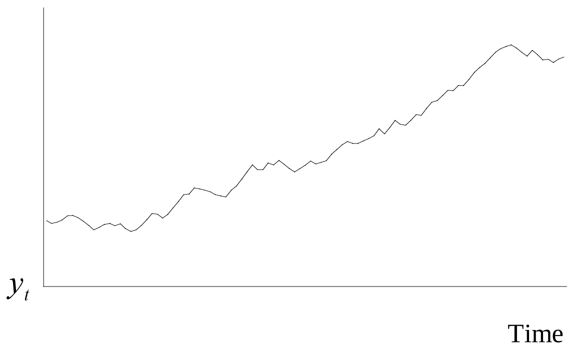


Figure C

DS series



2.6. ARIMA processes

An *ARMA* model for the differences of y_t is called an *ARIMA* (*A*uto *R*egressive *I*ntegrated *M*oving *A*verage) model. See Geurts and Ibrahim 1975 for some nice applications. The terms 'order of integration' and 'order of differencing' are equivalents. As an example, we consider an *ARIMA*(1,1,1) model:

$$\Delta y_t = \mu + \varphi \Delta y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}, \; t = 1,...,T \tag{15}$$

In equation (15), the nonstationary series y_t is differenced once to obtain the stationary series $\Delta y_t (= y_t - y_{t-1})$. Sometimes, it is necessary to differentiate the data more than once to obtain a stationary series. An *ARIMA*(*p,q*) model that is differenced *d* times is denoted by *ARIMA*(*p,d,q*), where *d* is the order of the differencing operator $\Delta^d = (1-B)^d$. For example, if $d = 1$ this amounts to taking $\Delta^1 y_t = (1-B)^1 y_t = y_t - y_{t-1}$. The *ARIMA(p,d,q)* process can now be defined as:

$$\varphi_p(B) \Delta^d y_t = \mu + \theta_q(B) \varepsilon_t, \; t = 1,...,T \tag{16}$$

ARMA and *ARIMA* models are primarily associated with the work of Box and Jenkins (1976). Their methodology for the identification of ARIMA-processes consists of three steps. The first step is the determination of the order of integration of the process. The second step is the analysis of stationary series (such as inspection of ACF, PACF) and finally, is the determination of the orders (p and q) of the ARMA process. In their methodology, they determine the order of integration by analyzing the *ACF* and *PACF* functions for various orders of integration. If these damp down sufficiently rapidly, the integrated series is consistent with the theoretical behavior of the *ACF* and *PACF* functions (Harvey 1990, p.119) for stationary series. More recent studies prefer to determine the order of integration (and the choice between *TS* and *DS* models) statistically on the basis of unit root tests. Section 2.7. discusses such a statistical procedure.

Application

Helmer and Johansson (1977) estimate an *ARIMA* model on the annual advertising expenditures of the Lydia Pinkham vegetable compound for the years 1907-1946. The resulting model is an *ARIMA*(2,1,0) model:

$$\Delta x_t = 0.074\Delta x_{t-1} - 0.407\Delta x_{t-2} + \varepsilon_t ,\ t = 1,...,T \qquad (17)$$

or

$$(1 - 0.074B + 0.407B^2)(1 - B)x_t = \varepsilon_t , t = 1,...,T \qquad (18)$$

where x_t denotes the annual advertising expenditures[9]. Helmer and Johansson use these estimates as a step in their *transfer function* analysis (see Section 3.2. for a discussion of the transfer function analysis). We deduce from equation (17) that x_t is evolving. Evolving implies that a shock in a time series may have a long-run effect on its own future evolution. The current changes in advertising expenditures are positively correlated with the changes in expenditures of one year before and negatively correlated with the changes two years before, indicating a fluctuating pattern.

2.7. Testing for a unit root

In Section 2.2, we introduced the concept of a unit root ($|\varphi| = 1$ in equation (1)) where y_t is the dependent variable. The most widely used test for non-stationarity is the Augmented Dickey-Fuller (ADF) unit root test developed by Dickey and Fuller (1979 and 1981). They consider three different regression equations that can be used to test for the presence of a unit root in y_t:

$$\Delta y_t = ay_{t-1} + \sum_{j=1}^{J} \varphi_j \Delta y_{t-j} + \mu + \beta t + \varepsilon_t ,\ t = 1,...,T \qquad (19)$$

$$\Delta y_t = ay_{t-1} + \sum_{j=1}^{J} \varphi_j \Delta y_{t-j} + \mu + \varepsilon_t ,\ t = 1,...,T \qquad (220)$$

$$\Delta y_t = ay_{t-1} + \sum_{j=1}^{J} \varphi_j \Delta y_{t-j} + \varepsilon_t ,\ t = 1,...,T \qquad (21)$$

where Δy_0 is fixed. In equations (19)-(21), the dependent variable is Δy_t. This implies that y_t has a unit root if $a = 0$.

Hence, the null hypothesis of the test equations (19)-(21) states non-stationarity of y_t:

$$H_0: \quad a = 0 \ (y_t \text{ has a unit root}) \tag{22}$$

The three equations differ in the deterministic regressors. The choice between the three equations is an important issue in unit root testing. One problem is that the additional estimated parameters reduce the degrees of freedom and the power of the test. Reduced power means that the researcher may conclude that the process contains a unit root where it is not the case. The second problem is that an appropriate statistic for testing $a = 0$ depends on which regressors are included in the equation. For example, if the data-generating process includes a deterministic trend, omitting the term βt gives an upward bias in the estimated value of a. Additional regressors, however, increase the absolute value of the critical values so that the researcher may fail to reject the null of a unit root.

The test is implemented through the usual t- statistic of $\hat{a}$. The t-statistics of the three models are denoted t_τ, t_μ, and t respectively. Alternatively, Dickey and Fuller (1981) suggest *F*-statistics to test the joint hypotheses $a = \beta = \mu = 0$ (Φ_2) and $a = \beta = 0$ (Φ_3) in equation (19) and the joint hypothesis $a = \mu = 0$ in equation (20), denoted as Φ_1. Under the null hypothesis of non-stationarity the t-statistics and t_τ and t_μ and the *F*-statistic Φ_2 and Φ_3 do not have the standard t- and *F*-distributions, but are functions of Brownian motions. Critical values of the asymptotic distributions of these t-statistics are provided by Fuller (1976) and have recently been improved by MacKinnon (1991) through larger sets of replications. Dickey and Fuller (1981) list critical values for the *F*-statistics of Φ_1, Φ_2, and Φ_3. Dolado et al. (1990) develop a systematic testing strategy between the alternative equations as outlined in Figure 6 (see also Enders 1995, p. 257).
The unit-root testing procedure consists of the following steps:

---

9 Helmer and Johansson (1977) do not report the standard errors of the parameter estimates.

Step 1. In the most unrestricted equation (19) the null hypothesis of stationarity is tested with t_{τ}. If the null hypothesis is rejected, variable y_t is trend stationary and there is no need to proceed any further.

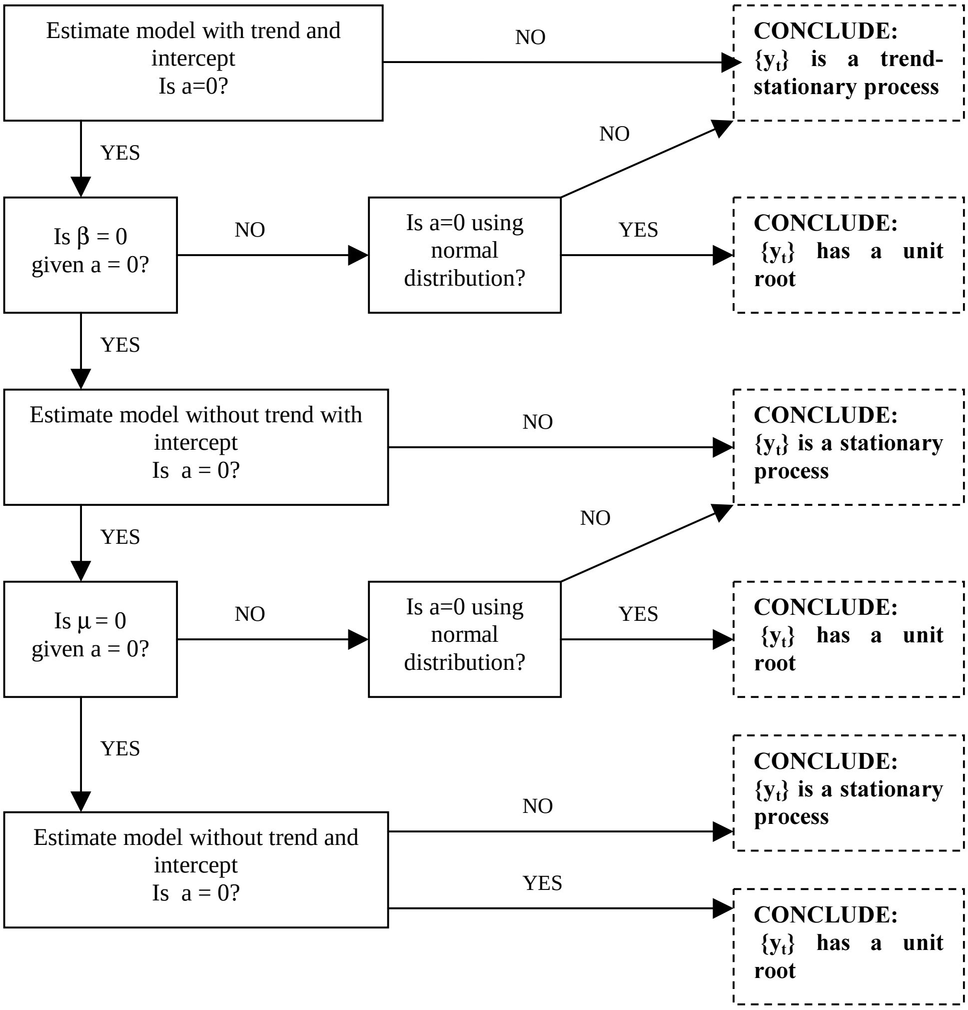
Step 2. If the null hypothesis is not rejected, we test for the significance of the deterministic trend under the null hypothesis $a = \beta = 0$ using the *F*-statistic Φ_3. If it is significant, the presence of the unit root can be tested again, noting that the t-statistic follows now a standard t-distribution.

Step 3. If a and β are jointly insignificant in equation (19), we estimate the equation without the deterministic trend (equation (20)) and test for the unit root using t_{μ} and its critical values. If the null hypothesis is rejected, we may stop again and conclude that variable y_t is stationary.

Step 4. If the null hypothesis is not rejected, we test for the significance of the constant term under the null $a = \mu = 0$ using Φ_1. If the constant term is significant, we test for the unit root using the standard normal distribution.

Step 5. If a and β are jointly insignificant in equation (20), we estimate equation (21) and test for the presence of a unit root. The process ends either with the result that variable y_t is stationary or that y_t contains a unit root.

Figure 6 The unit root testing framework



Sources: Dolado et al. (1990) and Enders (1995) p. 254-258

If we cannot reject the null hypothesis in any of the steps of the strategy, we conclude that y_t is non-stationary and needs to be differenced at least once to become stationary. To detect the order of integration, *d*, of the series y_t we proceed by testing the differenced series until the unit root hypothesis is rejected. So, if y_t is found to be non-stationary and Δy_t is found to be stationary than y_t is called 'integrated of order 1' (denoted as $y_t \sim I(1)$). If we can only reject the null of a unit root after differencing *d* times, we conclude that the series is integrated of order *d*. Stochastic trends in marketing are often linear and sometimes quadratic, so *d* rarely exceeds than 2 (Leeflang et al. 2000, p. 465).

The true order of the autoregressive process is usually unknown to the researcher, so that the problem is to select the appropriate lag length in equations (19)-(21)[10]. The number of lags (J) in equations (19)-(21) is often determined by the *AIC*, *SBC*, or by a selection procedure advocated by Perron (1989).The latter implies that, working backward from a pre-determined number of lags, we choose the first value of J such that that the *t*-statistic on φ_J is greater than 1.6 in absolute value and the *t*-statistic on φ_{J+1} is less than 1.6. To compare the *t*-statistic properly, the sample length is identical for each J and determined as corresponding to the maximum lag length $J = 6$, e.g. if $T = 100$ and the researcher sets the maximum number of lags *a priori* at $J = 6$, than the number of observations for each $J < 6$, will be (100-7=) 93. (Because the dependent variable in the tests is in first differences, we loose an additional observation:6+1=7).

[10] Including too many lags to the unit root test reduces the power of the test to reject the null hypothesis of a unit root since the increased number of lags necessitates the estimation of additional paramaters and hence. This, together with the decreased number of observations lead to a loss in the degrees of freedom. On the other hand, too few lags will not appropriately capture the actual error process, so that a and its standard error will not be well estimated. We note here that including too little lags is more serious than too many.

2.8. Persistence of shocks

In Section 2.4., we introduced the term persistence in relation to a stochastic trend. If a time series exhibits a stochastic trend, shocks may have a long-run impact.[11] Dekimpe and Hanssens (1995a) introduced so-called *persistence modeling* into a marketing context, in which they estimate the long-run impacts of shocks. They distinguish (i) univariate persistence that measures what proportion of any unspecified shock will affect sales permanently and (ii) multivariate persistence that derives the long-run impact of an unexpected change in a control variable. The multivariate persistence is usually considered in a multiple framework (see Section 4). Campbell and Mankiw (1987) developed the *A(1)* measure for univariate persistence which is the ratio of the sum of the moving average coefficients to the sum of the autoregressive coefficients of an *ARIMA*(p,1,q). Thus, the univariate persistence of a shock to y in

$$\varphi_p(B)\Delta y_t = \theta_q(B)\varepsilon_t\,,\ t = 1,\ldots,T \tag{23}$$

is

$$A(1) = \frac{\theta_q(1)}{\varphi_p(1)} = \frac{1-\theta_1-\theta_2-\ldots-\theta_q}{1-\varphi_1-\varphi_2-\ldots-\varphi_p}\ .$$

In stationary time series, the persistence is zero by definition: after a shock, the series returns to its pre-determined level; the shock dies out. The statistical distinction between stationary (stable) and non-stationary (evolving) sales or demand behavior has important implications for marketers (Dekimpe et al. 1999). If the sales are stationary, marketing actions produce at most temporary deviations from the brand's average performance level or around its predetermined deterministic trend, although their effect may die out over a reasonably long (dust-settling) period. If sales are evolving, there is a potential for long-term marketing effectiveness. For example, the advertising expenditures in equation (17) are evolving. The long-run effectiveness of any (i.e. uspecified) shock in the advertising expenditures is given by

---

[11] Early marketing applications focused on the forecasting capabilities of *TSA* (Geurts and Ibrahim 1975 and Kapoor et al. 1981 among others). In more recent studies, *TSA* is used to separate short-term from long-term marketing effects (when marketing variables are included) (Leeflang et al. 2000, p. 458).

$A(1) = \dfrac{\theta_q(1)}{\varphi_p(1)} = \dfrac{1}{1 - 0.074 + 0.407} = 0.75$. So, a portion of 0.75 of each one unit shock in advertising expenditures persists over time and a portion of (1-0.75=) 0.25 dies out.

2.9. Seasonal processes

As we can see in Figure 2.1B, both advertising expenditures and sales figures are much higher in Spring than in Summer. Palda (1960, p.35) reports that the Lydia Pinkham management has the policy of low advertising in late Spring and during two or three months around Christmas, resulting in a highly fluctuating advertising pattern. Many series in marketing display such seasonal patterns caused by managerial decisions, weather conditions, events, holidays, etc. For example, sales of ice cream in Europe are higher in Spring and Summer and lower in Winter . There are a variety of possible approaches to account for seasonal fluctuations in data. Three classes of time series models are commonly used to model seasonality, viz. (i) purely deterministic seasonal processes, (ii) stationary seasonal processes, and (iii) integrated seasonal processes (Maddala and Kim 1998, p. 363). Purely deterministic seasonal processes exhibit a fixed seasonal pattern resulting in systematic fluctuations around the mean level, i.e. some observations are expected to have values above and other observations below the grand mean. A straightforward way to deal with such processes is to include 'seasonal' dummies in the model, such as:

$$y_t = \mu + \sum_{s=1}^{S-1} b_s d_{st} + \varepsilon_t \,,\; t = 1,\ldots,T \qquad (24)$$

where S is the maximum number of seasons (12 for monthly data, 4 for quarterly data, etc.), d_{st} is a dummy variable taking the value one in season s and zero otherwise, and b_s a parameter.

The seasonal effects can also be of the (ii) *ARMA* type if seasonal fluctuations in sales levels or random shocks die out over time in a seasonal way, or (iii) integrated, when nonstationary seasonal patterns exist. A seasonal model may apply, with orders *P,D,* and *Q* respectively for the *AR*, *I*, and *MA* components, denoted by $ARIMA(P,D,Q)_s$, where s is the seasonal lag. To illustrate, suppose there exists a seasonal pattern in monthly data, such that any month's value contains a component that resembles the previous year's value in the same month. Then a purely seasonal $ARIMA(1,0,1)_{12}$ model is written as:

$$y_t = \mu + \varphi y_{t-12} + \varepsilon_t - \theta\varepsilon_{t-12}\,,\ t = 1,...,T \qquad (25)$$

Seasonal processes can be identified from the *ACF* and *PACF* functions, similarly to the nonseasonal *ARIMA* processes discussed above, except that the patterns occur at lags *s*, 2*s*, 3*s*, etc., instead of at lags 1,2,3, etc. Just as in the case of the determining the nonseasonal order of integration, seasonal unit root tests have been developed to detect the order of seasonal integration (see, for example, Franses 1998, Hylleberg 1992).

In practice, seasonal and nonseasonal processes usually occur together. The examination of *ACF* and *PACF* may suggest patterns in these functions at different lags. This general process is indicated as an *ARIMA*(*p,d,q*)(*P,D,Q*)$_s$ process:

$$\varphi_P(B^s)\varphi_p(B)\Delta^D\Delta^d y_t = \mu + \theta_Q(B^s)\theta_q(B)\varepsilon_t\,,\ t = 1,...,T \qquad (26)$$

In equation (26) the seasonal and nonseasonal *AR*, *MA*, and differencing operators are multiplied. In practice, the orders *p,d,q*, and *P,D,Q*, are small, ranging from 0 to 2 in most cases (Leeflang et al. 2000, p. 466).

Application

Hanssens (1980) finds an *ARIMA*(0,0,3)(0,1,1)$_{12}$ model for the monthly sales y_t (see, equation (27)) and an *ARIMA*(0,0,0)(0,1,1)$_{12}$ for the monthly advertising expenditures x_t (see, (28)) of the Lydia Pinkhams vegetable compound for the period 1954:01-1960:06 (Figure 1B):

$$\Delta_{12} y_t = -44.98 + (1 - 0.257B^{12} - 0.621B^{15})\varepsilon_{y,t}\,, \qquad RSS = 2199000\,,$$
$$\chi^2(23) = 10890\,,\ t = 1,...,T \qquad (27)$$

$$\Delta_{12} x_t = (1 - 0.477B^{12})\varepsilon_{x,t}\,,\ RSS = 1083800\,,\ \chi^2(21) = 9.407\,, t = 1,...,T \qquad (28)$$

where RSS is the residual sum of squares and $\chi^2(k)$ is the Box-Pierce Q-square statistic for white noise, which follows a χ^2-distribution. We deduce from equations (27) and (28) that both variables are evolving, because the dependent variable is in annual differences. It is interesting to compare equation (28) with equation (17). Both models indicate a fluctuating pattern, since there is a negative relationship between two successive years.

3. Multivariate *TSA*

So far, we have introduced the basic concepts of *TSA,* and discussed some techniques that are useful in identifying the dynamic structure of the series. Univariate time series models can be very useful for out-of-sample forecasting and for descriptive analyses. However, it may be the case that the empirical specification of such univariate models is hampered by fluctuations that can be attributed to one or more variables other than y_t. In marketing, one of the main fields of interest is the determination of the effect of marketing actions of a brand on its sales. In multivariate *TSA*, the dependent variable is explained by past values, random shocks, and explicitly formulated e*X*ogenous variables: the *ARMAX* models. *ARMAX* models are discussed in Section 3.1. We introduce *transfer function models* in Section 3.2. The identification of transfer functions is discussed in Section 3.3. The concept of Granger causality is introduced in Section 3.4. with references to the different causality tests and their applications. Section 3.5. deals with intervention analysis. In this section we discuss (i) pure- and partial change models, (ii) different types of interventions, (iii) possible responses to interventions, (iv) impulses of noises, (v) endogenous change dates, (vi) multiple break dates, and (vii) the effect of interventions on unit root tests. In Section 3.6. we introduce the basic concept of co-integration. In all sections, we provide applications. Table 2 gives a survey of relevant multivariate *TSA* in marketing.

Table 2 Relevant multivariate time series applications in marketing

Study	Country	Entity aggregation	Focus
Multivariate $\left(y_t = f(y_{t-j}, x_{t-k})\right)$			
Aaker et al. (1982)	USA	Sales	Advertising-sales relationship
Ashley et al. (1980)	USA	Macro-variable	Advertising and macroeconomic indicators relationship
Baghestani (1991)	USA	Sales	Advertising-sales
Bass and Pilon (1980)	USA	Market shares	Market-shares and marketing mix relationship
Carpenter et al. (1988)	Australia	Market shares	Market-shares and marketing mix relationship
Dekimpe et al. (1997)	The Netherlands	Sales	Market shake-ups
Didow and Franke (1984)	USA	Macro-variable	Advertising and macroeconomic indicators relationship
Doyle and Saunders (1985)	Europe	Sales	Sales and promotion mix relationship
Doyle and Saunders (1990)	Europe	Industry sales	Sales and marketing mix relationship
Franses (1991)	The Netherlands	Industry sales	Sales, price, advertising, temperature, and consumer expenditures relationships
Hamada (1999)	Japan	Macro-variable	Advertising and macroeconomic indicators relationship
Hanssens (1980)	USA	Sales; Industry sales	Identification of competitive patterns
Hanssens (1998)	USA	Sales	Forecasting

Table 2 continued

Study	Country	Entity aggregation	Focus
Multivariate $\left(y_t = f(y_{t-j}, x_{t-k})\right)$			
Helmer and Johansson (1977)	USA	Sales	Advertising-sales relationship
Jacobson and Nicosia (1981)	USA	Macro-variable	Advertising and macroeconomic indicators relationship
Johnson et al. (1992)	Canada	Industry sales	Relation between industry sales, price, age, personal disposable income, and degree days
Krishnamurthi et al. (1986)	USA	Sales	Advertising-sales relationship
Leeflang and Wittink (1992)	The Netherlands	Sales	Identification of competitive patterns
Leone (1983)	USA	Sales	Advertising-sales relationship
Moriarty (1985b)	USA	Sales	Advertising-sales relationship
Pitelis (1987)	UK	Macro-variable	Advertising and macroeconomic indicators relationship
Pitelis (1991)	UK	Macro-variable	Advertising and macroeconomic indicators relationship
Roy et al. (1994)	USA	Sales	Identification of competitive patterns
Turner (2000)	UK	Macro-variable	Advertising and macroeconomic indicators relationship
Wichern and Jones (1977)	USA	Market shares	Market shake-up analysis

3.1. ARMAX processes

Up till now, we have restricted the discussion to models of one criterion variable such as sales, as a function of past sales, random shocks and time. If we are interested in estimating the effects of marketing variables such as price, advertising, and competitive behavior on sales when the latter variable is also subject to other complex patterns, we can include these variables in an *ARMA* model, and obtain an *ARMA* model with e*X*ogenous variables (*ARMAX*). Time series models with exogenous variables are also known as *transfer function models* (Greene 1997, p. 539). For extensive discussion and numerous examples of transfer functions and *ARMAX* models, see Harvey (1990). Aaker et al. (1982), and Helmer and Johansson (1977) provide marketing applications of transfer function models. Assume that sales (y_t) is explained by one explanatory variable, advertising (x_t). Often, the transfer function takes the form of a linear *D*istributed *L*ag *F*unction (*DLF*). We discuss more general transfer functions in Section 3.2. Thus, these models postulate that sales (y_t) may respond to current (x_t) and previous values of advertising ($x_t, x_{t-1}, \ldots$):

$$y_t = \mu + \nu_0 x_t + \nu_1 x_{t-1} + \varepsilon_t \, , \; t = 1, \ldots, T \tag{29}$$

The general dynamic model formulation for one endogenous variable is:

$$y_t = \mu + \nu_k(B) x_t + \varepsilon_t \, , \; t = 1, \ldots, T \tag{30}$$

where $\nu_k(B) = \nu_0 + \nu_1 B + \nu_2 B^2 + \ldots + \nu_k B^k$ and B is the backshift operator.

Application

Franses (1991) used an *ARMAX* model to analyze the primary demand for beer in the Netherlands. Based on an inspection of 42 bimonthly observations from 1978 to 1984, using *ACF*'s and model tests, Franses obtained the following model:

$$\begin{aligned} \ln y_t = \; & \underset{(3.38)}{0.17} \ln y_{t-6} \underset{(-7.14)}{- 0.06}\delta_1 + \underset{(17.80)}{2.30}\delta_2 + \underset{(16.02)}{2.34}\delta_3 + \underset{(16.11)}{2.51}\delta_4 + \underset{(16.35)}{2.30}\delta_5 + \\ & + \underset{(16.19)}{2.37}\delta_6 \underset{(-12.31)}{- 3.98}\Delta p_t + \underset{(9.96)}{2.27}\Delta p_{t+1} + \varepsilon_t \underset{(-3.90)}{- 0.54}\varepsilon_{t-1} \end{aligned} \tag{31}$$

In this model, ln is the natural logarithm, y_t is the sales variable, δ_1 to δ_6 are bimonthly seasonal dummies, p_t is the price, $\Delta = 1 - B$ is a first-order differencing operator, p_{t+1} is a price expectation variable that assumes perfect foresight, and $t = 1,...,T$. The t-values are given in parentheses. The model contains a lagged endogenous variable, a moving average component, seasonal effects (captured through deterministic dummies rather than through differencing), and accounts for current price and future price effects. Advertising expenditures did not significantly influence the primary demand for beer, so this variable is not introduced in the model. Subsequently, Franses concluded that there are strong price effects in the beer market. Because of this, tax changes may be an effective instrument to change the primary demand for beer. The positive effect of future prices suggests forward buying by consumers.

3.2. Transfer function models

The transfer function is also called the *I*mpulse *R*esponse *F*unction (*IRF*), and the v-coefficients are called the impulse response weights[12]. If sales do not react to advertising in period t, but only to lagged advertising, $v_0 = 0$. In that case, the model is said to have a 'dead time' of one. In general, the dead time is the number of v's equal to zero, starting with v_0. High order transfer functions can be approximated by a ratio of two polynomials of lower order (Helmer and Johansson 1977). Besides, transfer function (and ARMAX) models can also incorporate a noise model. Hence, the general form of the transfer function model can be written as

$$y_t = \mu + \frac{\omega_k(B)B^d}{\alpha_l(B)} x_t + \frac{\Theta_q(B)}{\Phi_p(B)} \varepsilon_t \text{ , } t = 1,...,T \qquad (32)$$

where $\omega_k(B) = \omega_0 + \omega_1 B + \omega_2 B^2 + ... + \omega_k B^k$, which contain the direct effects of changes in x on y over time, $\alpha_l(B) = \alpha_0 + \alpha_1 B + \alpha_2 B^2 + ... + \alpha_l B^l$, which shows the

[12] Note, that the definition of Impulse Responses adopted here, for multivariate TSA, may be conflicting for a researcher who is well versed in VAR methodology. In VAR models (as we will see later) Impulse Responses refer to the responses of endogenous variables to impulses in equation residuals.

gradual adjustment of y to x over time, $B^d =$ the dead time , $\Theta_q(B) = \Theta_0 + \Theta_1 B + \Theta_2 B^2 + ... + \Theta_q B^q$, and $\Phi_p(B) = \Phi_0 + \Phi_1 B + \Phi_2 B^2 + ... + \Phi_p B^p$. A dead time of $d = 0$ corresponds to an immediate effect The *order* of the transfer function is said to be *(l,k,d)* and the noise model is of order *(p.q)*. The transfer function model is straightforward extension of the ARMA model; for $\omega_k(B) = 0$ it is equal to a univariate time series (ARMA) process. The impulse response weights are equal to $v_{k,l}(B) = \frac{\omega_k(B)B^d}{\alpha_l(B)}$.

The simultaneous identification of the transfer function and the *ARIMA* structure of the error in equation like (31) is a very complicated process (we discuss this in Session 1.1.3). An example that frequently occurs in marketing for accounting exogenous variables in the TS model is where the original sales series shows a seasonal pattern, for which an *ARIMA*$(1,0,1)(1,0,0)_{12}$ model is indicated (Leeflang et al. 2000, p. 468). However, if temperature is included as an explanatory variable in the model, an *ARIMA*(1,1,0) may suffice (Leeflang et al. 2000, p. 468). Thus, the identification of the *ARIMA* error structure depends on the exogenous variables included in the model.

Application

As an example, we consider a well-known special case: the Koyck model (see, for instance, Leeflang et al. 2000, Ch. 6, Moriarty 1985a, Winer 1979):

$$y_t = \beta \sum_{j=0}^{\infty} \alpha^j x_{t-j} + u_t \tag{33}$$

where $0 \le \alpha < 1$. It can be easily seen that $v_i = \alpha v_{i-1}$, for $i = 1,...,\infty$. Thus, the response is a constant fraction of the response in the previous time period, it decays exponentially.

Multiplying both sizes of equation (33) at $t-1$ with α, substracting it from equation (33) and rearranging gives:

$$y_t = \alpha y_{t-1} + \beta x_t + u_t - \alpha u_{t-1} \tag{34}$$

So, the Koyck's model is equivalent to a transfer model with $\omega\ (B)=\beta$, $\alpha_l(B)=1-\alpha B$, $\Theta_q(B)=1-\alpha B$, $\Phi_p(B)=1$ and $d=0$.

The formulation in equation (34) is called the (rational) polynomial form of the Koyck model (Leeflang et al. 2000, p. 468). Palda (1964, p. 37) estimates a Koyck type model for the monthly Lydia Pinkham dataset (Figure 1B):

$$y_t = 396 + 0.39 y_{t-1} + 0.27 x_t + 0.21 x_{t-1} + 0.16 x_{t-2} + \varepsilon_t, t = 1,...,T \qquad (35)$$

where y_t (sales) and x_t (advertising expenditures) are both seasonally adjusted. The transfer function equals:

$$\frac{\omega_k(B)B^d}{\alpha_l(B)} = \frac{0.27 + 0.21B + 0.16B^2}{1 - 0.39B} \qquad (36)$$

We conclude from equation (36) that advertising has a positive effect on sales. The numerator represents the direct effect on sales. It is the largest within the same month and has a decreasing effect on sales in the next two months. The denominator in equation (36) represents the indirect effects on sales. It indicates that there is a positive repurchasing mechanism in the Lydia Pinkham sales, so advertising not only directly stimulates sales, but also indirectly, because customers tend to repurchase the product.

3.3. Identification of transfer functions

The explanatory variable (x) can be seen as exogenous if it is independent of the noise in the dependent variable (y) For that reason, the explanatory variable, e.g. advertising, is 'filtered' (through an *ARIMA* model, for example) to remove all its systematic predictable components before it enters the model. To insure that the advertising-sales relationship is not destroyed, the same filter is also applied to the sales variable. The transformed sales variable is considered as potentially predictable from the pre-filtered advertising series (Helmer and Johansson 1977). Procedures that have been proposed for filtering are the *L*inear *T*ransfer *F*unction (*LTF*) method and the *D*ouble *P*re*W*hitening (*DPW*) method. (We refer to Box et al. 1994, Hanssens et al. 2001, or Pankratz 1991, for a description of these methods. Marketing applications can be found in Bass and Pilon 1980, Doyle and Saunders 1985, and Leone 1983, among others). The core of these methods involves fitting univariate time series to the individual series, after which the estimated white noise residuals are used for multivariate analysis. This is called the prewhitening of variables. An important tool in the identification of transfer functions is the *C*ross-*C*orrelation *F*unction (*CCF*), which is the correlation between x and y at lag k: $\rho(y, x_{-k})$. The *CCF* extends the *ACF* for the situation of two series (those of x and of y), with similar interpretation: spikes denote *MA* parameters (in the numerator in equation (32), and decaying patterns indicate *AR* parameters, ending in the denominator in equation (32)).

Application

Box and Jenkins (1976) developed a transfer model selection procedure, which Helmer and Johansson (1977) used to estimate two alternative transfer functions for the advertising-sales relationship of the Lydia Pinkham vegetable compound for the years 1907-1946 (Figure 1A). One of their final models is:

$$\Delta y_t = \underset{(3.27)}{0.4827}\Delta x_t + \underset{(1.39)}{0.1881}\Delta x_{t-1} + \frac{1}{1-\underset{(1.47)}{0.2755}B}\varepsilon_t,\ t = 1,\dots,T \tag{37}$$

where t-values are given in parentheses. We deduce from equation (37) that both variables are evolving and that the change in advertising positively affects the change in sales.

3.4. Granger Causality

It is difficult to establish a feasible definition of causality in a non-experimental setting. Tests based on the stochastic view of time series behavior are based on the assumption that temporal ordering of events can be used to make an empirical distinction between leading and lagging variables. This distinction is the basis of a well-known concept of causality that was introduced by Granger (1969). A variable x is said to 'Granger cause' another variable y with respect to a given information set containing x and y, if future values of y can be predicted better using past values of x and y than using the past of y alone (Leeflang et al. 2000, p. 495). In marketing, causality tests are usually applied (i) to distinguish between causal and noncausal relationships or associations, (ii) to establish direction of causality when variables are related, and (iii) to reduce the large set of potential predictor variables.

Marketing literature uses several types of causality concepts. Most of these are based on the idea of Granger causality[13]. Leeflang and Wittink (1992 and 1996) use the Haugh-Pierce test to investigate competitive reactions between marketing instruments of competing brands. Despite the existence of many alternative bivariate causality tests, this is the test that has been used in marketing applications almost exclusively (Bult et al. 1997). Bult et al. (1997) compare the performance of five different Granger causality tests, and find that the conclusions about causality may depend strongly on the test used. They recommend the Granger-Sargent test for marketing applications, because it is a simple test with substantial amount of power and has a low probability of type-I error.

3.5. Intervention Analysis

Apart from marketing variables such as price and advertising, we may want to accommodate discrete events in models of sales. Examples include a new government regulation, the introduction of a competitive brand, a one period price discount, and so on. Intervention analysis extends the transfer function approach described above to the estimation of the impact of such events.

Pure and partial change models

The effect of an intervention is represented by changes in the parameters of the model. A *pure* change model is estimated on two (or more) subsamples: the first subsample contains the observations unaffected by the intervention, the second subsample contains the data potentially affected by the intervention. Chow (1960) provides a test for a pure change model:

$$F = \frac{(\varepsilon'_{r,}\,\varepsilon_{r,} - \varepsilon'_{u,}\,\varepsilon_{u,})/k}{\varepsilon'_{u,}\,\varepsilon_{u,}/(T-2k)},\ t = 1,\ldots,T \tag{38}$$

where $\varepsilon_{r,t}$ are residuals of the restricted model that does not allow for an intervention, $\varepsilon_{u,t}$ are the residuals of the unrestricted model that allows for an intervention, T is the number of observations, k is the number of variables, and the residuals are assumed to behave as in any

[13] We do not discuss these here, but rather refer the interested reader to Hanssens et al. (2001, p. 311) and Leeflang et al. (2000, p. 495)

other single equation model (uncorrelated with common variance) (Stewart 1991, p.102). See Leeflang and Naert 1978, p. 300 for a marketing application of this test.

In practice, however, we are often interested in specific changes in the individual parameters. If we want to know which incumbents' sales levels are more influenced, for example, by the introduction of a new brand, *partial* change models, in which only the parameters of interest are allowed to change. Equation (39) is an example of a partial change model in which the level (μ) and the repurchasing parameter (φ) are allowed to change (with the parameters ξ_1 and ξ_2, respectively), while the advertising parameter (ν) remains unaffected.

$$y_t = \mu + \varphi y_{t-1} + \nu x_t + \xi_1 \delta_t + \xi_2 \delta_t y_{t-1} + \varepsilon_t \, , \; t = 1,...,T \qquad (39)$$

where δ_t is a dummy variable that has the value 1 in the time periods of the intervention, and has the value 0 elsewhere (see below for more details). Box and Tiao (1975) developed this approach. Marketing applications can be found in Leone (1987) and Wichern and Jones (1977) among others.

Types of interventions

The intervention dummy variable may take an infinite variety of forms, but there are two main types: (i) a *pulse* dummy variable, which represents a *temporary* event, such as a price discount or a strike, and (ii) a *step* dummy variable that characterizes a *permanent* change, e.g. the introduction of a successful brand or a change in legislation (Pankratz 1991, p. 263).

We can represent a pulse intervention with a dummy variable for which $\delta_{p,t} = 1$ in the time periods of the intervention ($t = T_b$), and $\delta_{p,t} = 0$ in all other periods ($t \neq T_b$), where T_b denotes the timing of the intervention. A step intervention can be represented with a dummy variable defined as $\delta_{s,t} = 1$ in the time periods in which the event occurs and all subsequent time periods ($t \geq T_b$), and $\delta_{s,t} = 0$ at all time periods before the event ($t < T_b$).

Responses to an intervention

The transfer function models defined in equations (2.30) and (32) can be extended to intervention analysis with the δ-variable as defined above. The response of the dependent variable to such an intervention will depend on the dynamics in the model. To illustrate this, we consider three simple sales models in which the intervention (δ) affects sales (y) through a change in the μ -parameter:

*L*evel *S*tationary (*LS*) sales: $y_t = \mu + \varphi y_{t-1} + \xi\delta_t + \varepsilon_t$, $|\varphi| < 1$, $t = 1,...,T$ (40)

*T*rend *S*tationary (*TS*) sales: $y_t = \mu + \varphi y_{t-1} + \beta t + \xi\delta_t + \varepsilon_t$, $|\varphi| < 1$, $t = 1,...,T$ (41)

*D*ifference *S*tationary (*DS*) sales: $\Delta y_t = \mu + \xi\delta_t + \varepsilon_t$, $t = 1,...,T$ (42)

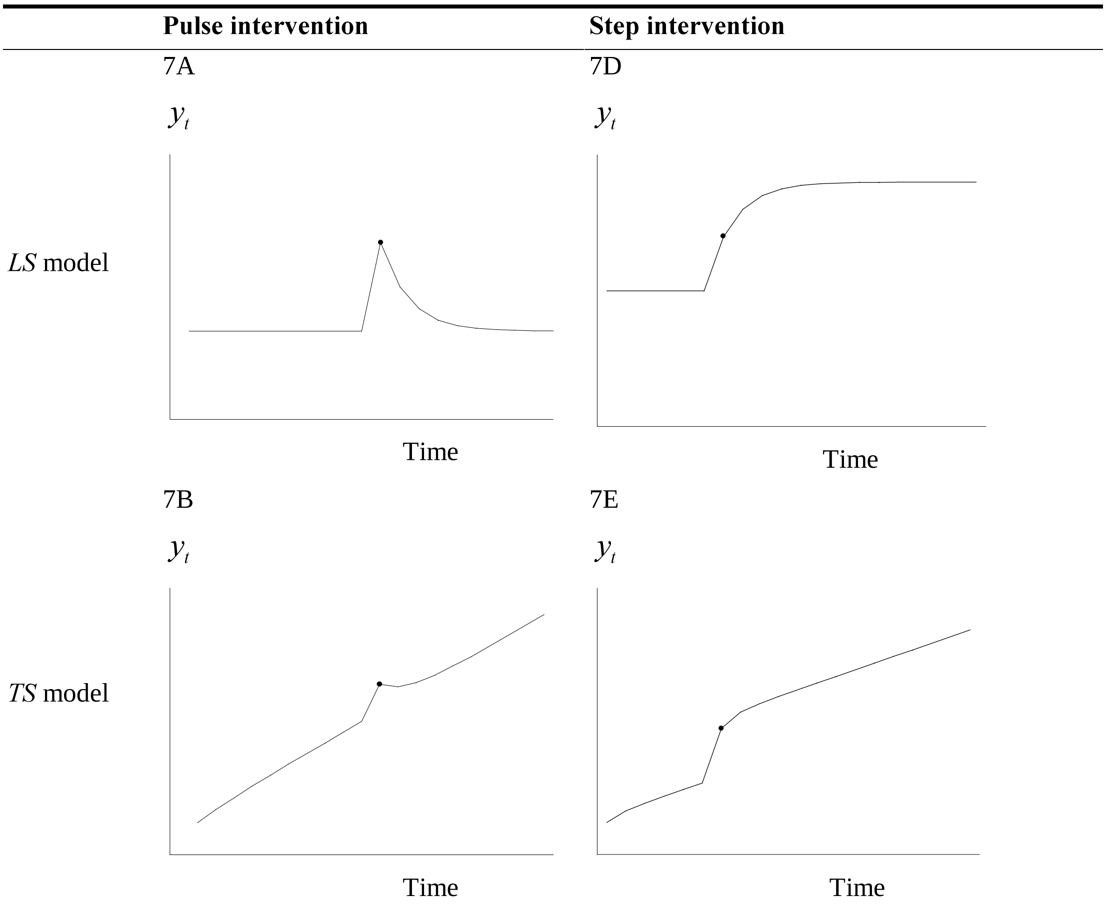
Figures 7A to 7C show the corresponding graphical examples of responses to *pulse* interventions in the three cases, assuming δ_t is a positive *pulse* dummy variable and φ has a positive sign. In Figures 7A and 7B, the effect of the intervention lasts for some periods and dies out in the end. Figure 7C, on the other hand, shows that the long-run behavior of y_t is affected by the temporary event. Hence, the impulse persists over time (see Section 2.8.). Figures 7D to 7F give graphical examples of responses where δ_t is a positive *step* dummy variable and φ has a positive sign. In 7D and 7E, the sales *level* is permanently changed by the intervention. In 7F, the *growth path* of y_t is changed due to the intervention.

Figures 7C to 7F show a permanent (or: *structural*) change in the variable. Therefore, these interventions are often called *structural breaks* in the time series literature (Perron 1989, Zivot and Andrews 1992). Notice that, although both Figures 7C and 7D depict a permanent change in the average sales level, the underlying data generating processes are different. In the former, any impulse may have a long-run impact on y_t , while in the latter, only the step dummy variable has a persistent effect. In Section 2.8., we discussed how to measure the persistence of any impulse in the system.

Impulses of residuals

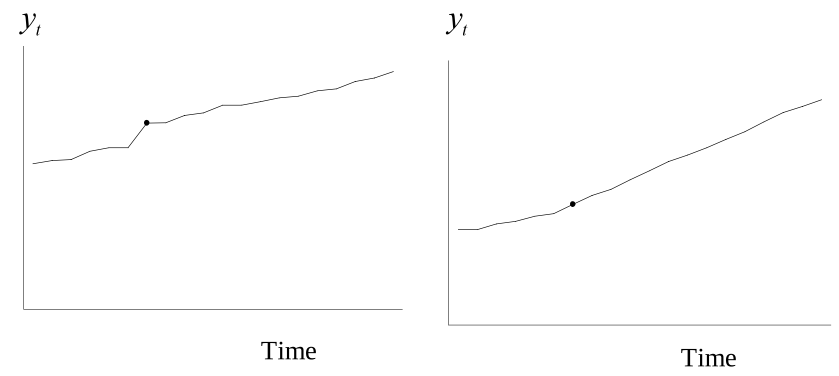
It is important to note that ε_t has a pulse effect on y_t as well. The difference between a series of pulse dummies { $\delta_{p,t}$ } and the residuals { ε_t } stems from the fact we can distinguish { $\delta_{p,t}$ } from { ε_t } by using a priori information about events in the market. In other words, { $\delta_{p,t}$ } are known pulses at irregular points in time, while { ε_t } are unknown pulses at regular points in time (see also the discussions in Balke and Fomby 1991, and Perron 1994).

Figure 7 Examples of responses to pulse and step interventions



7C 7F

DS model



The dot indicates the change date location. Notice that the solid line reflect discrete observations and that the dot indicates the first observation that is subject to a change.

Application

Wichern and Jones (1977) performed an intervention analysis to examine the effects of an advertising campaign for Crest toothpaste based on an endorsement on the market shares of Crest and Colgate in the years 1958-1963. The endorsement of Crest toothpaste as an "important aid in any program of dental hygiene" by the Council on Dental Therapeutics of the American Dental Association led to a permanent change in Crest's market share, and a permanent decrease in Colgate's market share, i.e. a shake up in market share. Wichern and Jones (1977) used weekly data of dentifrice purchasing by members of the Market Research Corporation of America consumer panel during the years 1958-1963. Their final intervention models are:

$$\Delta ms_{1,t} = \underset{(0.047)}{-0.052}\delta_{p_1,t} \underset{(0.048)}{-0.061}\delta_{p_2,t} + \varepsilon_{m_1,t} \underset{(0.037)}{-0.809}\varepsilon_{m_1,t},\ t = 1,\ldots,T \quad (43)$$

and

$$\Delta ms_{2,t} = \underset{(0.045)}{0.065}\delta_{p_1,t} + \underset{(0.045)}{0.112}\delta_{p_2,t} + \varepsilon_{m_2,t} \underset{(0.039)}{-0.779}\varepsilon_{m_2,t},\ t = 1,\ldots,T \quad (44)$$

where $ms_{1,t}$ = the market share of Colgate, $ms_{2,t}$ = the market share of Crest, and $\delta_{p_1,t}$ and $\delta_{p_2,t}$ are pulse dummies associated with the advertising campaign of Crest[14]. The figures in parentheses present the approximate standard errors. Because the models are in differences, the permanent effects on the market shares are the sum of the pulses, (-0.052-0.061=) -0.11 and (0.112+0.065=) 0.18, respectively. Taking the pre-shock market levels into account, the campaign induced a persistent damage for Colgate of 25%, while the shares of Crest roughly doubled in size. This example illustrates the possibility for temporal marketing activities, e.g. Crest's advertising campaign, to have a long-run (or persistent) effect on a performance measure, e.g. market share.

Endogenous change date

The assumption that the structural break coincides exactly with the timing of, for example, the entry of a new player, may be too restrictive. It may take some time before a new entrant affects the behavior of customers/incumbents in the market due to e.g. contract periods. Hence, there are lagged effects (Leeflang et al. 2000, Ch. 6). Consumers may also anticipate on the introduction of a new product or an expected sales promotion, and change their strategies in advance (Van Heerde et al. 2000). Doyle and Saunders (1985) discuss an example of this phenomenon in an industrial market. Not only consumers may adjust their behavior, also incumbents may anticipate the entry of a new player and change their behavior accordingly (Shankar 1999). Hence, the entry timing of, for example, a new player does not necessarily coincide with the date of the shake-up.

Besides conceptual considerations, an a priori determination of the change date has been criticized from a statistical point of view as well (see Christiano 1992, Zivot and Andrews 1992, among others). They suggest to treat the location of the change date as unknown. Different models have been developed to deal with unknown change dates. Quandt (1960) discusses a general likelihood ratio test for intervention models where the intervention takes place at some unknown point in time and the error variance is also allowed to change. Brown et al. (1975) suggest the CUSUM test for this purpose. Andrews (1993) derives asymptotic critical values for the Quandt test as well as the analogous Wald and Langrange Multiplier (LM) tests. He shows that his supF test has better properties than the CUSUM test

[14] Wichern and Jones (1977) decided to include two dummy variables, because their initial examination of the differenced data indicated that the market share adjustment to the

(see also Maddala and Kim 1998, p. 391). The unknown change date tests are usually based on recursive estimation, where the model is estimated for different change date locations and the final model is selected on the basis of a statistical criterion. In practice, it is not very appealing to search for change dates in regions where we already know there are no interventions. It is therefore recommendable to search for these change dates in the neighborhood of the known major events (Maddala and Kim 1998, p. 398). A change date is called endogenous if it is not known *a priori*. Marketing applications that make use of the endogenous change date assumption include Nijs et al. (2001), and Pauwels et al. (2000).

Multiple change dates

Longer time spans of marketing data enable more profound insights in the underlying data generating processes, which helps to build better marketing models. However, the number of interventions that change the market characteristics increases as the time span enlarges, which makes model building more complicated. Dufour (1982), among others, considers tests for multiple regimes in a time series. Kornelis et al. (2001b) consider the possibility of two interventions in a marketing context.

Interventions and the unit root test

Interventions may have an important effect on the outcomes of unit root tests (see Section 2.7. for a discussion of a unit root test). Perron (1989) demonstrates that if there is a permanent change in the deterministic trend function, then unit root tests will lead to the misleading conclusion that there is a stochastic trend (a unit root), when in fact there is a deterministic trend with a structural break. Perron provides extensions of the *ADF* test that allow for one structural break at a known date in the trend function. His most general case is the extension of equation (19) with intervention dummies:

$$\Delta y_t = \mu + \beta t + \sum_{j=1}^{J} \varphi_j \Delta y_{t-j} + a y_{t-1} + \xi_1 DU_t(\lambda) + \xi_2 DT_t(\lambda) + \xi_3 DTB_t(\lambda) + \varepsilon_t$$

$, t = 1,\ldots,T$ (45)

intervention was accomplished over two consecutative weeks and that there was no relationships between the two single week adjustments.

where $\lambda = \frac{T_b}{T}$, T = the date of the structural break, $DU_t(\lambda)$ = the structural change in the intercept at date λT, it has the value 1 if $t > \lambda T$ and 0 otherwise, $DT_t(\lambda)$ = the structural change in the trend function at date λT, it has the value $t - \lambda T$ if $t > \lambda T$ and 0 otherwise, and $DTB_t(\lambda) = 1$ if $t = T_b + 1$ and 0 otherwise. The focus of this test is on the unit root hypothesis and not on the intervention itself. Zivot and Andrews (1992) among others, generalize the models of Perron (1989) by endogenizing the change date location. The unit root hypothesis under the allowance of interventions has received much attention in both the economics and statistics literature. We refer to Maddala and Kim (1998) for a more detailed discussion. See Bronnenberg et al. 2000 for a marketing application of a unit root test under the allowance of a structural break.

3.6. Two evolving variables: the case of co-integration

In Section 2.4, we discussed (i) the deterministic trend concept and (ii) the stochastic trend concept. In a multivariate context, we also have (iii) co-integration as a possible long-run relationship between variables with a stochastic trend. We introduce this concept in this section.

If a variable, say sales, only has a deterministic trend, it is not possible to affect sales in the long run through another variable, advertising for instance. Indeed, advertising can only have a temporal effect, since in the end sales will return to their pre-determined growth path. If sales has a stochastic trend, shocks in advertising may affect the future values of sales, but these future values are hard to control. However, a very interesting situation occurs when the two variables under consideration, say sales and advertising, *both* have a stochastic trend. In that case, it is possible that the variables share this stochastic (common) trend so that the growth path of one variable (say sales) can be explained by the growth path of the other (advertising) variable. Such a relationship is called *co-integration*. Co-integration refers to the existence of a stationary linear combination of two or more nonstationary series (Engle and Granger 1987). An example is the linear combination of sales (y_t) and advertising (x_t):

$$e_{y,t} = \mu + y_t - rx_t\,,\; t = 1,...,T \tag{46}$$

where $e_{y,t}$ are the residuals obtained from a regression of y_t on x_t. The co-integrating relationship (43) implies that if $e_{y,t}$ is stationary, the difference between sales (y_t) and a portion of advertising (x_t) is stable over a long period of time. Hence y_t and x_t move together in the long run; there is a long-run equilibrium relationship between them (Maddala and Kim 1998, p.26). Because one is unlikely to observe a perfect equilibrium, a more realistic assumption is that the deviations of (46) are mean reverting around zero (Powers et al. 1991). The co-integrating relationship may also follow a deterministic trend and contain seasonality. To model this, we can include a deterministic trend (t) (Franses 1999), and seasonal dummies ($d_{s,t}$) (For the treatment of the seasonal dummies, see Lee and Siklos 1997, p. 386, Table 2, Note 2) in (46) and obtain:

$$y_t = \mu + \beta t + \sum_{s=1}^{S-1} \gamma_s d_{s,t} + \gamma_S x_t + e_{y,t} \text{ , } t = 1,...,T \tag{47}$$

Engle and Granger (1987) developed a two step procedure (*EG* approach) to test for co-integration (Baghestani 1991). In the first step, the co-integrating relation (47) is estimated. In the second step, the *ADF* test is performed on the residuals, $e_{y,t}$, but, with other critical values than in the *ADF* test. The *EG* critical values depend on the inclusion of trend and intercept and on the number of regressors in (46). (These critical values can be found in Engle and Yoo 1987, pp. 157-158, and in Philips and Ouliaris 1990, pp. 189-190).

If cointegration exists between two or more variables, these long-run connections should be accounted for in the models that capture short-run relationships between them. Engle and Granger (1987) showed that this can be achieved through an error-correction mechanism. This is a model in differences, which is augmented by the lagged equilibrium error:

$$\Delta y_t = \mu + \alpha \Delta y_{t-1} + \nu \Delta x_t + \upsilon e_{t-1} + \varepsilon_t \text{ , } t = 1,...,T \tag{48}$$

The lagged equilibrium error parameter (ν) is a measure of the speed of adjustment towards the long-run equilibrium.

If the model contains, say N evolving variables, there may exist $N-1$ co-integrating relationship between those variables. We will return to this issue in Section 4.9., where we discuss the co-integration test developed by Johansen (1988).

Application

Baghestani (1991) applied the *EG* methodology to the Lydia Pinkham annual data set. He found that the vegetable compound sales share a common trend with advertising expenditures, indicating a long-run equilibrium for the advertising-sales relationship. The long-run sales are related to the long-run advertising expenditures as follows:

$$\underset{}{y_t} = \underset{(3.84)}{488.83} + \underset{(11)}{1.43}x_t + e_{y,t}\,,\ t = 1,\ldots,T \tag{49}$$

where the t-ratios are in parentheses. We can consider the present sales as a deviation from the long-run equilibrium of equation (49). To model the adjustment of present sales towards this long-run equilibrium, we can incorporate the lagged (stationary) residuals of equation (49), $e_{y,t-1}$, into a *ARIMAX* model and obtain

$$\varphi_{y,p}(B)\Delta y_t = \mu + \rho e_{y,t-1} + \nu_{x,q}(B)\Delta x_t + \varepsilon_t\,,\ t = 1,\ldots,T \tag{50}$$

Because, we can consider a deviation as an 'error', equation (50) is called an *E*rror *C*orrection (*EC*) model and it links short run instability to long-run stability. The variables y_t and x_t are in differences (Δ) as they both have a stochastic trend. Baghestani's (1991) final EC model for the Lydia Pinkham vegetable compound sales is

$$\Delta y_t = \underset{(0.30)}{8.298} - \underset{(-2.05)}{0.226}e_{y,t-1} + \underset{(1.58)}{0.228}\Delta y_{t-1} + \underset{(1.41)}{0.221}\Delta y_{t-2} + \underset{(4.46)}{0.832}\Delta x_t - \underset{(-1.49)}{0.183}\Delta x_{t-5} + \varepsilon_t\,,\ t = 1,\ldots,T \tag{51}$$

where the t-ratios are in parentheses. We deduce from equation (51) that the current change in sales are explained by the intercept, the adjustment towards the long-run relation with advertising, the change in sales of two years ago, the change in advertising expenditures of the same year, and the change in advertising expenditures of five years ago. To assure that the variable returns to the long-run co-integrating relationship, the sign of the adjustment parameter (-0.226) should be negative. long-run

Regressing sales on advertising refers to the same co-integrating relation as regressing advertising on sales. Theoretically, the residuals of both operations are perfect linear transformations of each other. In practice, however, these are approximations (Baghestani 1991). Therefore, if we consider the EC model of sales, we usually normalize on sales, e.g. regress sales on advertising to obtain the co-integrating relationship. Baghestani (1991) gives the following co-integrating relation (46) and EC model (50) for the annual advertising expenditures of Lydia Pinkhams vegetable compound:

$$\underset{}{x_t} = \underset{(0.33)}{27.93} + \underset{(11)}{0.49} y_t + e_{x,t}, \; t = 1,\ldots,T \tag{52}$$

and

$$\Delta x_t = \underset{(-0.13)}{-2.787} - \underset{(-3.18)}{0.382} e_{x,t-1} + \underset{(2.31)}{0.239} \Delta x_{t-4} + \underset{(3.27)}{0.356} \Delta y_t + \underset{(1.66)}{0.186} \Delta y_{t-1} - \underset{(-2.92)}{0.295} y_{t-2}$$

$$, \; t = 1,\ldots,T \tag{53}$$

where $e_{x,t}$ is the co-integrating relation normalized on advertising (x_t). In his study, Baghestani (1991) finally concluded that less than two fifths of the adjustment towards the long-run equilibrium condition occurred within a year through changes in the company's advertising expenditures. Other marketing applications that consider co-integrated relationships include, among others, Dekimpe and Hanssens 1999, Franses et al. 2000, Kornelis et al. 2001a.

4. Multiple Time Series Analysis

In marketing practice, it is not always known *a priori* whether the time path of the "dependent" variable has affected the "independent" variable. An example is performance feedback. Bass (1969) warned that advertising may be influenced by current and past sales, and should not automatically be treated as exogenous. This means that not only marketing activities may influence sales, but (changes in) sales may also induce marketing activities. Marketing managers may track, for example, own-brand market share or sales, and if they observe a drop in either performance measure, they may tend to compensate it with changes in marketing activities. In its most basic form, multiple *TSA* treats all variables symmetrically without making reference to the issue of dependence versus independence and permits causality testing of all variables simultaneously. This is a major advantage of multiple *TSA* models compared to the multivariate time series models (Enders 1995, Franses 1998, and Hanssens 1980).

Moriarty and Salamon (1980) concluded that their multiple *TSA* model provides substantial improvement in parameter estimation efficiency and forecasting performance in comparison with multivariate models. Takada and Bass (1998) find that their multiple *TSA* models outperform multivariate *TS* models in goodness-of-fit measures as well as in forecasting performance.

Another advantage of multiple *TSA* models results from the fundamental philosophy of *TSA* in general, i.e. to let the data rather than the researcher specify the model. Multiple *TSA* methodology applies iterative processes that identify basic models, the lag structure, and relationships between variables, estimate the parameters, and check the estimated model. Because of this, multiple *TSA* modeling has been labeled a-theoretical (Jacobs 1998, p. 31). In our view, this is an exaggeration. Economic/marketing theory is essential in the selection of sets of variables, in the identification, in the interpretation of impulse responses and *F*orecast *E*rror *V*ariance *D*ecomposition results (see Section 4.8.), as well as in co-integration analysis. Dekimpe and Hanssens (2000) make a similar claim, especially with the CI analysis, MTSA also has *confirmatory* rather than only *explanatory* value. They also emphasize the importance of Structural VAR models to supplement sample-based information with marketing theory. All in all, multiple *TSA* models have been shown (i) to be extremely flexible in capturing the dynamic inter-relationships between a set of variables, (ii) to be able to treat several variables endogenously, (iii) not to require firm prior knowledge on the nature of the different relationships, (iv) to be able to capture both short- and long-run inter-relationships, and (v) to outperform multivariate *TSA* models in parameter efficiency, goodness-of-fit measures as well as in forecasting performance. We summarize the marketing applications of MTS models in Table 4.

The remainder of this section is organized as follows. First, in Section 4.1 we introduce *V*ector *A*uto*R*egressive *M*oving *A*verage (*VARMA*) models. Next, the estimation of such models is discussed in Section 4.2. In Section 4.3, we shortly discuss the issue of exogeneity. We consider the system with exogenous variables (*VARMAX* model) in Session 4.4 Section 4.5 introduces *I*mpulse *R*esponse (*IR*) analysis, which involves some identification issues that are discussed in Section 4.6. Section 4.7 discusses *S*tructural *VAR* (*SVAR*) models. We discuss multivariate persistence in Section 4.8. Dynamic multipliers are introduced in Section 4.9 that measure marginal the impact of changes in the exogenous variables. *F*orecast *E*rror *V*ariance *D*ecomposition (*FEVD*), a useful tool in detecting the interrelationships between the endogenous variables of the model, is introduced in Section 4.10. Finally, we discuss co-integration in Section 4.11.

Table 3 Relevant multiple time series applications in marketing

Study	Country	Entity aggregation	Focus
Multiple $\left(\begin{bmatrix} y_t \\ x_t \end{bmatrix} = f\left(\begin{bmatrix} y_{t-j} \\ x_{t-k} \end{bmatrix}\right)\right)$			
Bronnenberg et al. (2000)	USA	Market share	Distribution and market share relation
Bronnenberg et al. (2001)	USA	Sales	Competition
Chowdhury (1994)	UK	Macro-variable	Advertising and macroeconomic indicators relationship
Dekimpe and Hanssens (1995a)	USA	Industry sales; Sales	Advertising sales relationship
Dekimpe and Hanssens (1995b)	-	-	Empirical generalizations
Dekimpe and Hanssens (1999)	USA	Sales	Sales and marketing mix instruments
Dekimpe et al. (1999)	USA	Industry Sales; Sales	Sales and price promotions
Dekimpe et al. (2001)	The Netherlands	Sales	Competition
Franses (1994)	The Netherlands	Industry sales	Forecasting
Franses et al. (1999)	-	Market share	Market share and marketing mix
Franses et al. (2000)	USA	Market share	Identification of dynamic patterns
Hanssens and Ouyang (2000)	USA	Sales	Modeling market hysteresis
Horvath et al. (2001)	USA	Sales	Competition
Grewal et al. (2001)	USA	Sales	Analysing marketing interactions in dynamic environment

Table 3 continued

Study	Country	Entity aggregation	Focus
Multiple $\left(\begin{bmatrix} y_t \\ x_t \end{bmatrix} = f\left(\begin{bmatrix} y_{t-j} \\ x_{t-k} \end{bmatrix}\right)\right)$			
Jung and Seldon (1995)	USA	Macro-variable	Advertising and macroeconomic indicators relationship
Kornelis et al. (2001a)	The Netherlands	Industry sales	Market shake-ups
Kornelis et al. (2001b)	The Netherlands	Macro-variable	Advertising and macroeconomic indicators relationship
McCullough and Waldon (1998)	USA	Macro-variable	Substitutability
Moriarty and Salamon (1980)	USA	Industry sales	Forecasting
Nijs et al. (2000)	The Netherlands	Industry sales	Sales and price promotions
O'Donovan (2000)	New Zealand	Macro-variable	Advertising and macroeconomic indicators relationship
Parsons et al. (1979)	USA	Macro-variable	Advertising and macroeconomic indicators relationship
Pauwels et al. (2000)	USA	Sales; Industry sales	Sales and Price promotions
Seldon and Jung (1995)	USA	Macro-variable	Advertising and macroeconomic indicators relationship
Srinivasan and Bass (2000)	USA	Market share; Sales; Industry sales	Competition
Srinivasan et al. (2000)	USA	Market share	Market share and price
Umashankar and Ledolter (1983)	USA	Sales	Forecasting
Zanias (1994)	USA	Sales	Advertising-sales relationship

4.1. VARMA processes

Multiple time series models are a natural extension of the univariate *ARMA* models in the sense that a vector of dependent variables replaces the dependent variable y_t. Such models are called *V*ector *A*uto*R*egressive *M*oving *A*verage (*VARMA*) processes. Takada and Bass (1998) build *VARMA* models to analyze competitive marketing behavior and detect causality of marketing mix variables and sales. In marketing, the most commonly applied multiple time series model is the *V*ector *A*uto*R*egressive (*VAR*) model[15] (Dekimpe and Hanssens 1999, Srinivasan et al. 2000, Nijs et al. 2001, Horváth et al. 2001, among others). *VAR* models have mainly become popular in marketing in the analysis of competitive marketing systems, where the identification of competitive structures combined with the need to capture various dynamic relationships between marketing variables is a rather complex and difficult task (Hanssens 1980, Bass and Pilon 1980, Aaker et al. 1982, and Takada and Bass 1998). A *VAR* model has the following structure:

$$\mathrm{P}_0 y_t = \sum_{j=1}^{J} \mathrm{P}_j y_{t-j} + \varepsilon_t ,\ t = 1,\ldots,T \tag{54}$$

where y_{t-j} are k-dimensional vectors of endogenous variables at time $t-j$, P_0, P_j are the $k \times k$ parameter matrices, ε_t is a vector of disturbances with $\varepsilon \sim N(0, \sigma^2 I)$, and J is the order of the model.

Because multiplication of (54) with any nonsingular $k \times k$ matrix results in an equivalent representation of the process generating y_t, we can estimate the so-called *reduced* form of the model (Lütkepohl 1993, p. 325). The reduced form of the system is obtained by pre-multiplying (50) with P_0^{-1}, which gives:

$$\Pi(L) y_t = u_t ,\ t = 1,\ldots,T \tag{55}$$

where $\Pi(L)$ is a matrix polynomial with lag operator L:

$$\Pi(L) = I - \Pi_1 L - \Pi_2 L^2 - \ldots - \Pi_J L^J ,\ t = 1,\ldots,T \tag{56}$$

and $\Pi_j = \mathrm{P}_0^{-1}\mathrm{P}_j$, $j = 1,\ldots,J$, $u = \mathrm{P}_0^{-1}\varepsilon$, and $u_t \sim N(0, \Omega)$.

Since only lagged values of the endogenous variables appear on the right-hand side of each equation, simultaneous effects are no longer directly visible. Their presence can be obtained from Ω through an identification procedure (we elaborate below on the directional

[15] Sims (1980) states that a general VARMA model can be approximated by low order VAR models.

issues involved). The optimal lag length (J) selection is usually based on some kind of information criteion, such as, the Akaice Information Criterion, the Schwartz' criterion, the Hannan-Quinn criterion, or the Final Prediction Error criterion. Alternative method is to use the likelihood-ratio test. Lütkepohl (1993, p. 128-135) provides extensive description of the lag selection procedure.

For a two-variable case, such as the relation between the evolving advertising (x_t) and sales (y_t) variables, the reduced form *VAR(J)* model is:

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \sum_{j=1}^{J} \begin{bmatrix} \pi_{11}^j & \pi_{12}^j \\ \pi_{21}^j & \pi_{22}^j \end{bmatrix} \begin{bmatrix} \Delta x_{t-j} \\ \Delta y_{t-j} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}, \; t = 1,..., T \tag{57}$$

This model includes brand- or firm- specific decision rules (π_{11}), performance feedback effects (π_{12}), lagged effect of advertising activity ($\pi_{21}{}^j$), and repurchase effect ($\pi_{22}{}^j$), $j = 1,..., J$.

4.2. Estimation of VAR models[16]

Because the disturbances of the reduced *VAR* system are, in general, contemporaneously correlated ($\Omega \neq \sigma^2 I$ in equation 56), a system estimator is applicable. In this case, Zellner's (1962) *S*eemingly *U*nrelated *R*egressions (*SUR*) estimator is required to gain efficiency from the cross-equation correlations of the disturbances. If the equations have identical right-hand-side variables (unrestricted *VAR* model), and the order J is known, each equation in the system can be estimated by *O*rdinary *L*east *S*quares (*OLS*). In that case, the *OLS* estimates are consistent and asymptotically efficient even if the errors are correlated across equations (see Srivastava and Giles 1987, Ch. 2).

[16] In this session we mainly devote our attention to VAR models. The issues we discuss can easily be extended and implemented for VARX models, too. See Lütkepohl (1993), Ch. 10.

4.3. Exogeneity and causality

So far, we have assumed (in Session 4.1) that all variables of a system are determined within the system. In other words, the VAR model describes a joint generation process of all the observable variable of interest. In practice the generation process of a variable may be affected by other observable variables which are determined outside the system of interest. Such variables are called *exogenous*. In contrast, the variables determined within the system are called *endogenous*. The distinction between "exogenous" and "endogenous" variables in a model is subtle and it is a subject of a long literature[17]. As a result, several exogeneity concepts, such as predeterminedness, strict, weak, strong, predictive, and super exogeneity, have been distinguished, which extend the categorization of variables of Hanssens et al. (2001, p. 184) and Leeflang et al. (2000, p. 60)[18]. The relevant concept of exogeneity in marketing is so far strong exogeneity since the main interest of marketing applications is in conditional forecasting involving no changes in the conditional distribution of the data. This concept involves Granger causality that has been introduced in Session 3.4. The concept of Granger causality bears similarities with the concept of (strong) exogeneity in the sense that it allows us to draw inference on dynamic impact of one variable on another.

4.4. VARMAX processes

Ideally all considered variables are treated as and exogeneity should be tested during the model-building process[19]. However, this requires to start from the most general VAR setting, which is often not feasible. The common practice in marketing is to allow the most relevant variables to be endogenous and to control for the effects of other variables by considering them exogenously (Dekimpe and Hanssens 1999, Dekimpe et al. 1999 and 2001, Horváth et al. 2001, Nijs et al. 2001, Srinivasan and Bass 2001, and Srinivasan et al. 2000). This, i.e. the imposition of exogeneity, can imply a reduction of the number of parameters and also an improved precision of forecasting. These models are called *V*ector *A*uto*R*egressive models with e*X*ogenous variables (*VARX* models) and they can be expressed the following way:

---

[17] See, for example, Engle et al. (1983), Osiewalski and Steel (1996). Gourieroux and Monfort (1997, Chapter 10) provides a clear distinction between the different exogeneity concepts.

[18] We do not discuss the different exogeneity concepts in detail here since they have not yet been applied in marketing, but refer the reader to the references listed in Footnote 16.

[19] Gourieroux and Monfort (1997, p. 391) provide nested hypothesis tests involving the different exogeneity concepts.

$$\Pi(L)y_t = \Gamma(L)x_t + u_t \text{ , } t = 1,...,T \tag{57}$$

where x_t is an *h*-dimensional vector of exogenous variables and $\Gamma(L)$ is is a matrix polynomial with lag operator L: $\Gamma(L) = \Gamma_0 - \Gamma_1 L - \Gamma_2 L^2 - ... - \Gamma_S L^S$ and Γ_i are $h \times k$ coefficient matrices, $i = 0,..,S$.

The model is referred to as a VARX(J,S) process. If u_t is an MA(Q) process the model becomes a VARMAX(J,S,Q) process.

4.5. Impulse Response Analysis

Traditionally, *VAR* studies do not report estimated parameters or standard test statistics. Coefficients of estimated *VAR* systems are considered of little use in themselves and also the high (i.e. $J \times (k \times k)$ autoregressive coefficients) number of them does not invite for individual reporting. Instead, the approach of Sims (1980) is often used to summarize the estimated *VAR* systems by *I*mpulse *R*esponse *F*unctions (*IRF*s). *IRF*s trace out the effect over time of an exogenous shock or an innovation in an endogenous variable on all the endogenous variables in the system, to provide an answer to the following question: “What is the effect of a shock of size δ in the system at time t on the state of the system at time $t + \tau$, in the absence of other shocks?” In marketing *IRF*s are often used to estimate the effects of a marketing action on brand performance over time when indirect effects (through for example feedbacks, competitive reactions, and firm-specific decisions) are considered (Dekimpe and Hanssens 1999, Dekimpe et al. 1999, Horváth et al. 2001, Srinivasan and Bass 2000, and Takada and Bass 1998) Promotions, for instance, then are operationalized as one-time, hence temporary, deviations from the expected price level.

Assuming stationarity, equation (55) can be transformed into a *V*ector *M*oving *A*verage (*VMA*) representation (Lütkepohl 1993, p. 13):

$$y_t = C(L)u_t = \sum_{j=0}^{\infty} C_j u_{t-j} \text{ , } t = 1,...,T \tag{58}$$

where $C(L) = \Pi(L)^{-1}$ and $C_0 = I$. For this transformation, the model needs to be stable[20]. A sufficient condition is that the variables in the system are stationary (Lütkepohl 1993, p. 12). From this representation, the response of $y_{i,t+\tau}$ to a one-time impulse in $y_{j,t}$ can easily be obtained:

$$\frac{\partial y_{i,t+\tau}}{\partial u_{j,t}} = c_{ij,\tau},\ t = 1,...,T \tag{59}$$

where $c_{ij,\tau}$ is the row i, column j element of the $k \times k$ matrix of coefficients C_τ, the coefficient matrix of the τ-th lag of the *VMA* representation, $\tau = 0,...,\infty$. A plot of these values as a function of τ is called the graphical representation of the Impulse Response Function (*IRF*).

4.6. Identification of IRFs

The innovations of the reduced *VAR* model (equation (55)) may be contemporaneously correlated, i.e. $u_t \sim N(0,\Omega)$ where $\Omega \neq I$. As a result, shocks may affect multiple variables in the current period, which makes it impossible to isolate the effect of a given shock. The usual treatment of this identification problem is to impose some structure on the system of equations, based on a priori information. The most widely used approach suggested by Sims (1980) is assuming a causal ordering based on Cholesky decomposition of the covariance matrix, which was adopted by Dekimpe and Hanssens (1995a). This approach requires the ranking of variables from the most pervasive (a shock to this variable affects all other variables in the current period) to the least pervasive (a shock to this variable does not affect any other variable in the current period). We note that it is extremely rare for managers to appropriately order variables in a competitive marketing environment. Especially in instances in which leader-follower roles are not obvious (Dekimpe and Hanssens 1999) this ranking is almost impossible. Furthermore, Cholesky decomposition a rather arbitrary method of attributing common effects, because a change in the order of the equations can dramatically change the impulse responses. *G*eneralized *I*mpulse *R*esponse (*GIR*) analysis proposed by Pesaran and Shin 1998 and Pesaran and Smith 1998 overcomes the problem of ordering. This

[20] For explanation about stability of a VAR system we refer the reader to Lütkepohl (1993), p. 9-13.

approach, unlike the traditional *IRF* analysis, does not require a priori information or orthogonalization of shocks, and is invariant to the ordering of variables in the model. It measures the effect on the endogenous variables of a typical shock to the system, based on the estimated covariances between the reduced form shocks in the estimation period. The generalized impulse response function, denoted $\psi_j(m)$, to one unit standard error chock to the j th equation on expected values of x at time $t+m$ is:

$$\psi_j(m) = \frac{C_m \Omega e_j}{\sqrt{\sigma_{jj}}} \qquad (60)$$

where e_j is a k times 1vector with unity in its j th element and zeros elsewhere σ_{jj} the variance of the disturbances in equation j, C_m is the $k \times k$ coefficient matrices of the VMA representation in equation (2.54) and Ω is the variance-covariance matrix of the disturbances. This method overcomes the problem of ordering, although it is not based on economic or marketing theory. (Marketing applications in the spirit of the *GIR* approach are in Dekimpe and Hanssens 1999, and Nijs et al. 2001).

4.7. Structural VAR processes

Dekimpe and Hanssens (2000) recognize the importance of combining the data-driven *VAR* technique with marketing theory, and they suggest the application of *S*tructural *VAR* (*SVAR*) models. The *SVAR* approach, (proposed by Bernanke 1986, and Sims 1986), involves employing additional information based on economic theory or “ conventional wisdom”.

The idea behind *SVAR* modeling is the following. The residuals, u_t, obtained from the reduced form, are related to the structural disturbances, ε_t, according to $u_t = P_0^{-1}\varepsilon_t$. The identification of P_0^{-1} requires the imposition of restrictions on P_0^{-1}. So, if k equations are included in the model, full identification of this matrix requires $\frac{k^2-k}{2}$ restrictions[21] (Hamilton 1994, p. 332). As long as the parameters in P_0^{-1} are not identified, it is impossible to identify structural shocks (ε_t) from reduced-form estimation. The effects of the *j*th structural disturbance on subsequent values of the variables of the system are (Giannini 1992, p. 44-51):

$$\frac{\partial y_{t+\tau}}{\partial \varepsilon_{jt}} = \frac{\partial y_{t+\tau}}{\partial u_t'} \cdot \frac{\partial u_t}{\partial \varepsilon_{jt}} = C_\tau p_j \,,\ t = 1,...,T \qquad (61)$$

where C_τ is the $k \times k$ matrix of coefficients for the τ-th lag of the *VMA* representation (equation (58) and p_j is the *j*th column of P_0^{-1}.

Cholesky decomposition is a special case of *SVAR* modeling, as it sets the parameters of P_0^{-1} above the diagonal equal to zero, which imposes a triangular pattern on P_0^{-1}. It is also possible that *SVAR* models contain e*X*ogenous variables. These models are referred to as *SVARX* models and have been applied in a marketing context by, for instance, Horváth et al. (2001).

4.8. Multivariate persistence

Dekimpe and Hanssens (1995a) introduce *multivariate persistence* modeling in a marketing context. Multivariate persistence derives the total long-run impact in a dependent variable (say, sales) of an unexpected change in an control variable (say advertising). Their approach strongly relies on the tools of *VAR* modeling, especially on *IRF* analysis. In multivariate

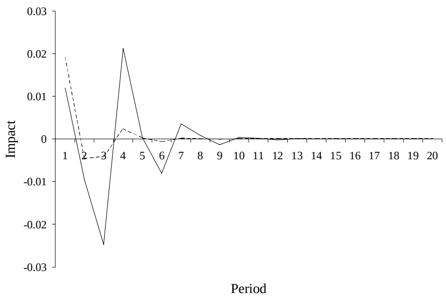
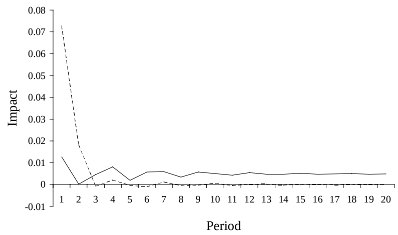
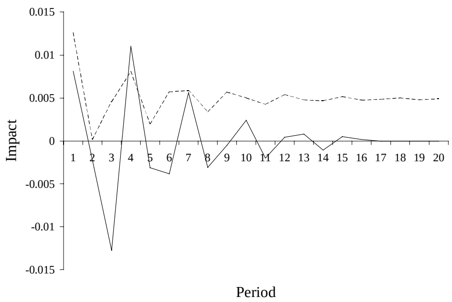
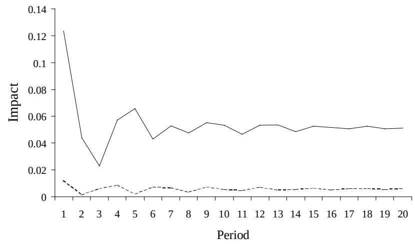
---

[21] The variance-covariance matrix of the reduced VAR model (Ω)contains $\frac{k^2+k}{2}$ distinct elements since $E(u_t u_t') = E(P_0 u_t u_t' P_0') = E(P_0 P_0')$. Given that the diagonal elements of P_0 are all unity, it contains $k^2 - k$ unknown variables. Therefore, in order to identify the structural VAR model it is necessary (but not sufficient) to impose (at least) $\frac{k^2-k}{2}$ restrictions on the model. We demonstrate this in the empirical part of this paper.

persistence modeling, the (non)stationarity property of the performance *and/or* marketing instruments variables has important implications for marketers. The combination of stationary (nonstationary) sales responses to stationary (nonstationary) marketing efforts leads to four possible situations. Dekimpe and Hanssens (1999) label and discuss these four situations that may benefit strategic marketing decision making. Figure 8 displays the four strategic scenarios. According to Figure 8, we classify the Crest advertising case as market hysteresis, i.e., for Crest temporary advertising actions cause sustained sales change.

Dekimpe and Hanssens (1999) and Hanssens et al. (2001) discuss the four situations extensively, provide illustrations from practice for each scenario, and describe its positive and negative consequences for long-term profitability.

Figure 8 Strategic scenarios resulting from temporary versus permanent change/ response

<table>
<tr><th></th><th colspan="2">Change in variable I</th></tr>
<tr><th>Response of variable II</th><th>Temporal</th><th>Sustained</th></tr>
<tr><td>Temporal</td><td>Business as usual

</td><td>Escalation

</td></tr>
<tr><td>Persistent</td><td>Hysteresis

</td><td>Evolving business practice

</td></tr>
</table>

where: —— = variable I, ---- = variable II

Source: Dekimpe and Hanssens (1999)

4.9. Dynamic Multipliers

In a VARMAX model the marginal impact of changes in the exogenous variables can be investigated. For example, if the exogenous variables are marketing instruments, such as prices the consequences (of changes) in these instruments can be investigated (if they are endogenous we apply impulse response analysis). A brand manager may, for instance, desire to know about the effects of a price change. In this case policy simulation is of interest. In other cases the consequences of changes in exogenous variables that are not under control of any decision maker may be of interest. It may be, for example, desirable to study the future consequences of present weather conditions on the production (supply) of an agricultural product. The dynamic effects of exogenous variables on the endogenous variables is captured by the *dynamic multipliers* (Lütkepohl 1993, p. 338):

$$D(L) = \sum_{i=0}^{\infty} D_i L^i = \Pi(L)^{-1}\Gamma(L), \qquad (62)$$

where $\Pi(L)$ and $\Gamma(L)$ are defined in equations (55) and (57). From this representation, the response of $y_{i,t+\tau}$ to a unit change x_{jt} can easily be obtained:

$$\frac{\partial y_{i,t+\tau}}{\partial x_{j,t}} = d_{ij,\tau}\,,\ t = 1,...,T \qquad (63)$$

where $d_{ij,\tau}$ is the row i, column j element of the $h \times k$ matrix of coefficients D_τ, the coefficient matrix of the τ-th lag equation (62), $\tau = 0,...,\infty$.

4.10. Forecast Error Variance Decomposition

Another way of characterizing the dynamic behavior of a system is through *F*orecast *E*rror *V*ariance *D*ecomposition (*FEVD*) (see e.g. Hamilton 1994, Franses 1998, Chapter 9, and Lütkepohl 1993). While *IRF*s trace the effects of a shock in one variable on other variables in the *VAR*-system, the *FEVD* separates the variation in an endogenous variable into component shocks to the system. If, for example, shocks to one variable fail to explain the forecast error variances of another variable (at all horizons), the second variable is said to be exogenous with respect to the first one. The other extreme case is if the shocks to one variable explain all forecast variances of the second variable at all horizons, so that the second variable is entirely endogenous with respect to the first.

The *FEVD* can be derived from the *VMA* representation of the model described in equation (58) (Lütkepohl 1993, p. 56). The τ -period forecast error is equal to:

$$y_{t+\tau} - E_t y_{t+\tau} = \sum_{i=0}^{\tau-1} C_i \varepsilon_{t+\tau-i} \text{ , } t = 1,..., T \tag{64}$$

where C_i are the $k \times k$ parameter matrices of the *VMA* representation in (58), and E_t denotes expectations formulated at time t , based on the estimated *VAR* model. Focusing, for example, on $y_{1,t}$, the first element of vector y_t , the forecast error can be written as:

$$y_{1,t+\tau} - E_t y_{1,t+\tau} = \sum_{r=1}^{k} \sum_{i=0}^{\tau-1} c_{1r,i} \varepsilon_{r,t+\tau-i} \text{ , } t = 1,..., T \tag{65}$$

where $c_{1r,i}$ is the element of the C_i matrix in the 1st row and *r*th column and $\varepsilon_{r,t+\tau-i}$ is the *r*th element of the $\varepsilon_{t+\tau-i}$ vector. Since the variances of the disturbance terms are all equal to one, the τ -step ahead forecast error variance of $y_{1,t}$ can be derived from the following expression:

$$\sigma_{y_1}^2(\tau) = \sum_{r=1}^{k} \sum_{i=0}^{\tau-1} c_{1r,i}^2 \text{ , } t = 1,..., T \tag{66}$$

where $\sigma_{y_1}^2(\tau)$ denotes the forecast error variance of variable y_1 at step τ . We note that this expression is a summation of nonnegative terms, so that the forecast error variance is nondecreasing with the forecast horizon τ . The forecast error variance can be decomposed into contributions of each of the variables in the system. The proportions of $\sigma_{y_1}^2(\tau)$ that can be attributed to shocks in each variable y_r , $r = 1,..., k$ at step τ are:

$$\frac{\sum_{i=0}^{\tau-1} c_{1r,t}^2}{\sigma_{y_1}^2(\tau)} \text{ , } t = 1,..., T \tag{67}$$

The variance decomposition is subject to the same identification problem inherent to the impulse response analysis. To overcome the identification problem we can apply the same approaches, such as Cholesky decomposition, the method of *GIF* analysis, and *SVAR* modeling.

4.11. Co-integration between more than two variables

In Section 3.6., we introduced and discussed the concept of co-integration. The number of possible co-integration relations increases with the number of time series considered in the model, which implies an increasing ambiguity in determining the empirical validity of equation (2.43). The Johansen (1988) *FIML* estimator is often applied in marketing to test for the presence of multiple co-integrating vectors (Dekimpe and Hanssens 1999, Nijs et al. 2001). Consider the *VAR* model in levels equation (56) where w_t is assumed to be a $k \times 1$ vector of stochastic I(1) variables. This equation can be reformulated in a *V*ector *E*rror-*C*orrection form (Engle and Granger 1987):

$$\Delta y_t = D_1 \Delta y_{t-1} + ... + D_{J-1} \Delta y_{t-J+1} + H y_{t-J} + e_t,\ t = 1,...,T \qquad (68)$$

with parameter matrices $D_1,..., D_{J-1}$ and H, where $D_i = -(I_k - \Pi_1 - ... - \Pi_i),\ i = 1,..., J-1,\ , H = -(I_k - \Pi_1 - ... - \Pi\ \ )$, and Δy_0 is fixed. Matrix H is the *E*rror *C*orrection term and contains information about long-run relationships between the variables in the data vector. The Johansen procedure relies heavily on the relationship between the rank of H and its characteristic roots. The rank of H, denoted *r*, is called the co-integration rank. Johansen and Juselius (1990) distinguish three cases:

Case 1. $\text{Rank}(H) = k$, i.e., the matrix H has a full rank, indicating that the vector process w is stationary and that the *VAR* can be estimated in levels;

Case 2. $\text{Rank}(H) = 0$, equation (68) reduces to a *VAR* model in first differences;

Case 3. $0 < \text{Rank}(H) = r < k$, implying that there are r cointegrating vectors. In this case, H can be written as $\alpha\beta'$ with α and β $k \times r$ full rank matrices. The columns of βt give an estimate of the CI relations, while the α parameters describe the speed of adjustment towards the long-run equilibrium.

The third case is the most interesting one. Accordingly, the hypothesis of co-integration can be formulated as $H(r): H = \alpha\beta'$ for a positive rank ($0 < r < k$). The term $\beta' y_t$ is a vector of r cointegrating residuals, which can be given a long-run equilibrium interpretation. The elements in α called the factor loadings, are interpreted as the average speed of adjustment of each variable in the direction of each of the long-run equilibrium relationships (Johansen 1991). Estimates of equation (68) can be found by maximum likelihood. The Johansen *FIML* co-integration testing method aims to test the rank of matrix H using reduced rank regression technique based on canonical correlations[22]. The interpretation in terms of long-run equilibribria is not straightforward as any linear combination of the co-integration relationships will reserve the stationarity property[23]. Hence, the long-run relationships must be identified. This can easily be seen, since if $H = \alpha\beta'$, then $H = (\alpha K)K^{-1}\beta'$ also holds for any K. So, if there exists a cointegrating vector, β needs to be normalized. In order to ensure the uniqueness of α and β, conditions based on marketing theory need to be imposed. Marketing researchers are just beginning to use cointegration to study marketing interactions. Some, quite receint, applications of cointegration in marketing are Bagestani (1991) Dekimpe and Hanssens (1999), Dekimpe et al. (1999), Grewal et al. (2001), Kornelis et al. (2001b), Nijs (2001), and Zanias (1994)

5. A testing framework for *TSA*

In the previous sections, we have discussed the development in *TSA* from univariate to multiple models. The question is which model to choose and which test to use. The *VAR* or *VEC* models can capture all previous concepts and are general representations of dynamic markets. So, *VAR/VEC* models constitute the central part in many recent studies (see for instance, Dekimpe and Hanssens 1999, Horváth et al. 2001, Kornelis et al. 2001a, and Nijs et al. 2001). To build the appropriate *VAR* or *VEC* models, univariate and multivariate pre-tests are needed, e.g. the unit root test and co-integration tests. Figure 9 displays a testing scheme that is commonly used in marketing (See, for example, Dekimpe and Hanssens 1999, Horváth

---

[22] For details about the determination of the co-integrating rank, r, we refer the interested reader to Lütkepohl (1993), p. 384-387, Enders (1995) p. 385-386, Franses (1998), p. 218-233, Dekimpe and Hanssens (1999), Bronnenberg et al. (2000), Srinivasan et al. (2000), and Srinivasan and Bass (2001).

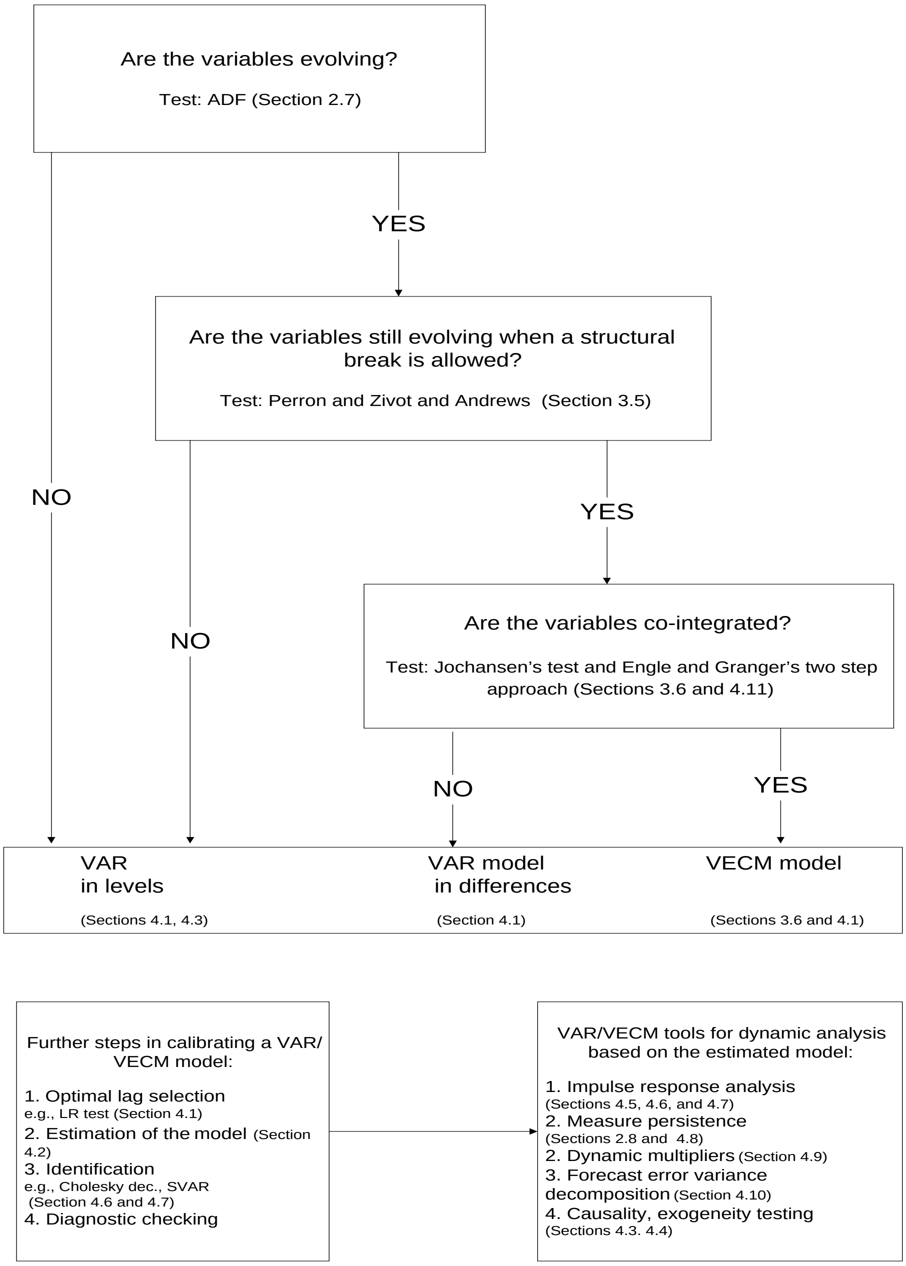
[23] The software packages EViews and RATS (Holden and Juselius (1995)) have implemented the Johansen-procedure for analysing multivariate cointegration models. RATS offers a procedure of five basic steps: (i) model checking, (ii) determination of the cointegration rank, (iii) estimation of the cointegration space, (iv) graphical analysis, and (v) tests of structural hypotheses in the parameter space.

et al. 2001, Kornelis et al. 2001a, and Nijs et al. 2001)[24]. We indicate the sessions in which the different steps in the testing scheme are discussed.

The first step in the model building procedure is the investigation of stationarity issues. As interventions may affect the outcomes of unit-root tests one should also apply tests that incorporate structural breaks. If a series contains a unit root, we have to consider the persistence for that series. If not, only temporary changes can be induced in that series by shocks in the system. If all variables are (level or trend) stationary, a VAR model should be built in levels. If we find more than one variable with a unit-root, we have to test for the existence of long-run equilibrium relationship(s) among the non-stationary variables. In case the series contain unit-root but are not cointegrated, a VAR model in differences needs to be built. If we find cointegrating series, we build a VEC model. If stationary as well as non-stationary variables are considered in the system, a mixed model need to be built. After deciding about which type of model to consider we still need to decide about the lag structure of the model, we have to consider identification of the model and finally, we have to apply some diagnostic checking to test whether the assumptions of the model apply. As a last step, which should be the first and most important thing to consider for the researcher, we point out what kind of questions can be investigated using these models.

[24] We indicate the main steps in the model building process. One may need alter this process subject to the aim of the study.

Figure 9 Model building framework



6. Concluding remarks

In this article, we give a brief overview of the most important time series methods and provide illustrations from marketing applications. Our aim is to introduce the basic concepts of TSA and their interpretations to marketing researchers and to provide a reference from which they can further develop their TSA skills. We point out references from marketing applications, econometrics, and methodological sources, which may serve as a bibliography for the interested reader. We believe that such an article can be very useful for marketing researchers who plan to apply time series models.

As we have pointed out, time series modeling requires combining data-driven techniques and marketing knowledge. With the growing availability of data that consints of repeated observations over a(n) (increasing) time-span (e.g., scanner data), development of user-friendly softwares, and with the increasing interest in dynamic mechanisms of markets, time series modeling is becoming more important for scholars and practitioners. We hope that the increasing tendency towards TSA will further develop the relationship between dynamic marketing concepts and time series econometrics.