

Modelling seasonality in innovation diffusion A regressive approach

Mariangela Guidolin Department of Statistical Sciences University of Padua Italy

Renato Guseo Department of Statistical Sciences University of Padua Italy

Abstract: Although the literature on innovation diffusion is very wide, up to now not much attention has been paid on modelling the seasonal component often present in sales data. In this paper we develop two innovation diffusion models that take into account not only the evolutionary trend of sales but also their intra-year oscillations, that are due to seasonal effects. In particular, we treat seasonality as a deterministic component to be estimated with conventional NLS techniques jointly with the trend. The results obtained by applying these models to the life–cycle of a pharmaceutical drug show a clearly better performance in providing short-term forecasts. Moreover, this methodology proves more parsimonious with respect to the autoregressive method, based on SARMA models.

Keywords: Seasonality, Time series decomposition, Bass model, Guseo–Guidolin model, Nonlinear regression, SARMA

Final version (2012-05-09)

Department of Statistical Sciences University of Padua



Contents

1	Introduction	1
2	A diffusion model with seasonal component	3
3	The diffusion of a pharmaceutical drug in the Italian market	5
4	Conclusion	9

Department of Statistical Sciences Via Cesare Battisti, 241 35121 Padova Italy

tel: +39 049 8274168 fax: +39 049 8274170 http://www.stat.unipd.it Corresponding author: Mariangela Guidolin tel:+39 049 827 4168 guidolin@stat.unipd.it http://www.stat.unipd.it/~name

Modelling seasonality in innovation diffusion A regressive approach

Mariangela Guidolin

Department of Statistical Sciences University of Padua Italy

Renato Guseo

Department of Statistical Sciences University of Padua Italy

Abstract: Although the literature on innovation diffusion is very wide, up to now not much attention has been paid on modelling the seasonal component often present in sales data. In this paper we develop two innovation diffusion models that take into account not only the evolutionary trend of sales but also their intra-year oscillations, that are due to seasonal effects. In particular, we treat seasonality as a deterministic component to be estimated with conventional NLS techniques jointly with the trend. The results obtained by applying these models to the life–cycle of a pharmaceutical drug show a clearly better performance in providing short-term forecasts. Moreover, this methodology proves more parsimonious with respect to the autoregressive method, based on SARMA models.

Keywords: Seasonality, Time series decomposition, Bass model, Guseo–Guidolin model, Nonlinear regression, SARMA

1 Introduction

The ability to model and predict the diffusion of innovations is particularly important for firms that develop and launch new products and services in increasingly complex and competitive markets. Diffusion research is aimed at describing the spread of an innovation by modelling its entire life–cycle. There is a quite long tradition in this field: wide research has been produced in order to capture several phenomena visible in sales data. A particular effort has always been devoted to extending the structure of the basic and most known Bass model, [1], by taking into account price dynamics, competition, network externalities, consumer heterogeneity, technological generations (for a review, see for instance [8]), arguing that these may help explain turning points in life-cycle such as take-off, saddle, technological substitution. Indeed, a specific focus of these models is to provide an efficient description and interpretation of the mean trajectory of the life-cycle, and probably for this reason, the presence of a seasonal pattern in sales data has not been much investigated, although many products are clearly characterized by it.

A general definition of seasonality is provided by [6] stating that it is the systematic, although not necessarily regular, intra-year movement caused by the changes of weather, the calendar, and timing of decisions, directly or indirectly through the production and consumption decisions made by the agents of the economy. While seasonality is considered less relevant for long-term evaluations, it is much more important in medium-short periods. A seasonal behaviour is typically visible when data are collected with monthly or quarterly frequency and some works on diffusion concentrated on the issue of temporal aggregation of data: in [9] quarterly or monthly seasonally adjusted data are compared with annual ones, finding that the former perform better in terms of parameter estimates, and in [7] it has been noticed that the smooth development of sales typical of the Bass model matches better with data at yearly frequency than at higher frequency. Indeed, the aggregation of monthly or quarterly data to obtain a smoother shape of sales would result in a loss of information to avoid. Moreover, since product life-cycles are increasingly shortening due to high competition, it becomes more and more necessary to have short term projections, by analyzing accurately not only the trend of sales but also their oscillations within the year. Following [11], typical methods developed in time series analysis for seasonality modelling are: a) the *Regression method*, which assumes that the seasonal component is deterministic and may be described as a linear combination of seasonal dummy variables or as a linear combination of sine-cosine functions of various frequencies; b) the Moving Average method, which estimates the nonseasonal component of a series, N(t), by using a symmetric moving average operator. The seasonal component, S(t), is obtained by subtracting the estimated nonseasonal component N(t) from the original series Y(t). The series with seasonal component removed, Y(t) - S(t) is referred to as the seasonally adjusted series; c) the Autoregressive method, which extends stochastic ARIMA models with the seasonal ARIMA models, SARIMA, developed by [3], assuming a stochastic nature of seasonality. In this work we follow the regression method and consider the seasonal component as deterministic. Such assumption makes it possible to develop a diffusion model where both trend and seasonality may be estimated with NLS techniques (see [10]). In Section 2 we start from the classical additive decomposition of time series and propose a modification of it in order to take into account the influence that the evolution of sales may have on the seasonal pattern. On the basis of this new decomposition we develop two diffusion models with seasonality, that extend the basic Bass and Guseo–Guidolin models. In Section 3 we discuss the application of the new models to the life-cycle of a pharmaceutical drug, highlighting their better performance in terms of global fitting and short-term forecasts. For comparison purposes we also provide the results obtained by applying a SARMA model to the residuals of a Guseo–Guidolin model. By comparing the two methods we see that considering seasonality as a deterministic component produces a more parsimonious model. Section 4 is left for concluding remarks.

2 A diffusion model with seasonal component

Let us consider a classical decomposition of a time series (see [11]) based on an additive structure

$$y(t) = T(t) + S(t) + \varepsilon(t)$$
(1)

where y(t) are instantaneous observed data, T(t) is the trend, S(t) the seasonality, and $\varepsilon(t)$ the accidental component.

In a product life-cycle, the trend of sales T(t), which is not stationary and characterized by phases of introduction, growth, maturity and decline, is represented by a nonlinear growth model, describing the evolution of sales. It is reasonable to imagine that such evolution also has an influence on the effect of seasonality: in particular seasonality may be stronger around the peak of sales and weaker during product launch and decline. To account for this aspect we propose to rewrite equation (1) as

$$y(t) = T(t) + S(t) + \varepsilon(t) = h(t)[M + A(t)] + \varepsilon(t).$$
(2)

In equation (2) T(t) = Mh(t), where h(t) is a probability density function describing the evolution of sales and M is a constant acting as scale parameter of the process, while S(t) = h(t)A(t), with A(t) describing the seasonal effect, modulated by h(t). The error term $\varepsilon(t)$ is commonly assumed as a stochastic stationary process, typically with zero mean, $M(\varepsilon(t)) = 0$, constant variance, $\operatorname{Var}(\varepsilon(t)) = \sigma^2$ and different error terms incorrelated, $\operatorname{Cov}(\varepsilon(t), \varepsilon(t')) = 0, t \neq t'$.

A better specification of T(t) and S(t) needs to define functions h(t) and A(t). A basic option in order to describe the evolution of sales h(t) in a product life-cycle is to start from the simple Bass model,

$$z(t) = m F(t; p, q) \tag{3}$$

where cumulative sales at time t, z(t), are given by the product between the constant market potential m and a cumulative distribution function F(t; p, q),

$$F(t;p,q) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}.$$
(4)

The corresponding instantaneous process, z'(t), is given by

$$z'(t) = m F'(t; p, q) = m f(t; p, q)$$
(5)

where f(t; p, q) is a probability density function. For the purposes of the paper we will work on instantaneous data, so that it is useful to observe that F'(t; p, q) = f(t; p, q) may be efficiently approximated by

$$f(t; p, q) \simeq [F(t+0.5; p, q) - F(t-0.5; p, q)].$$
(6)

In particular, we will assume that

$$h(t) = [F(t+0.5; p, q) - F(t-0.5; p, q)]$$
(7)

so that, by setting m = M, the trend may be described as follows

$$T(t) = Mh(t) = m[F(t+0.5; p, q) - F(t-0.5; p, q)].$$
(8)

An alternative option to the simple Bass model is to consider a diffusion model where the market potential is no longer constant but has a dynamic structure, which depends on the diffusion of information about the new product. This model, presented in [4], has the following cumulative structure,

$$z(t) = K W(t; p_c, q_c, p_s, q_s) = K \sqrt{F(t; p_c, q_c)} F(t; p_s, q_s),$$
(9)

where z(t) are cumulative adoptions, $W(t; p_c, q_c, p_s, q_s)$ is a distribution function, product of two c.d.f, $\sqrt{F(t; p_c, q_c)}$ and $F(t; p_s, q_s)$, respectively describing communication and adoption processes. The constant term K is a scale parameter such that $\lim_{t\to+\infty} KW(t) = K$.

In explicit form

$$W(t; p_c, q_c, p_s, q_s) = \sqrt{\frac{1 - e^{-(p_c + q_c)t}}{1 + \frac{q_c}{p_c}e^{-(p_c + q_c)t}}} \frac{1 - e^{-(p_s + q_s)t}}{1 + \frac{q_s}{p_s}e^{-(p_s + q_s)t}}.$$
(10)

Instantaneous adoptions, z'(t), are given by

$$z'(t) = K W'(t; p_c, q_c, p_s, q_s) = K w(t; p_c, q_c, p_s, q_s)$$
(11)

Also in this case $w(t; p_c, q_c, p_s, q_s)$ may be approximated by

$$w(t; p_c, q_c, p_s, q_s) \simeq [W(t+0.5; p_c, q_c, p_s, q_s) - W(t-0.5; p_c, q_c, p_s, q_s)]$$
(12)

and we may assume

$$h(t) = [W(t+0.5; p_c, q_c, p_s, q_s) - W(t-0.5; p_c, q_c, p_s, q_s)].$$
(13)

By setting M = K, the instantaneous trend T(t) will be

$$T(t) = Mh(t) = K \left[W(t+0.5; p_c, q_c, p_s, q_s) - W(t-0.5; p_c, q_c, p_s, q_s) \right].$$
(14)

Depending on the shape of h(t) the seasonal behaviour S(t) = h(t)A(t) will be described accordingly,

$$S(t) = h(t)A(t) = [F(t+0.5; p,q) - F(t-0.5; p,q)] A(t)$$
(15)

following the Bass framework, or by

$$S(t) = h(t)A(t) = [W(t+0.5; p_c, q_c, p_s, q_s) - W(t-0.5; p_c, q_c, p_s, q_s)] A(t).$$
(16)

if the dynamic model presented in [4] determines the basic trend component. By combining T(t) and S(t) we will obtain

$$T(t) + S(t) = [m + A(t)][F(t + 0.5; p, q) - F(t - 0.5; p, q)].$$
(17)

or, alternatively

$$T(t) + S(t) = [K + A(t)][W(t + 0.5; p_c, q_c, p_s, q_s) - W(t - 0.5; p_c, q_c, p_s, q_s)].$$
(18)

According to the regression method, the seasonal effect may be modelled as a linear combination of trigonometric functions (see for instance [2] and [11])

$$A(t) = \sum_{j=1}^{\left[\frac{s}{2}\right]} \left[a_j \cos\left(\frac{2\pi jt}{s}\right) + b_j \sin\left(\frac{2\pi jt}{s}\right) \right]$$
(19)

where a_j and b_j are parameters to be estimated with standard regressive techniques, s is the period, and [s/2] is the integer part of s/2. A simplified version of equation (19) just considers one harmonic function,

$$A(t) = \left[a\cos\left(\frac{2\pi t}{s}\right) + b\sin\left(\frac{2\pi t}{s}\right)\right].$$
 (20)

Moreover, a translation parameter c may be included in order to add flexibility to the harmonic function,

$$A(t) = \left[a\cos\left(\frac{2\pi(t-c)}{s}\right) + b\sin\left(\frac{2\pi(t-c)}{s}\right)\right].$$
 (21)

3 The diffusion of a pharmaceutical drug in the Italian market

We discuss here the application of the models developed in section 2 to the life-cycle of a pharmaceutical drug launched in the Italian market in September 2004. The product, that we label "Itm", is a patch containing a non-steroidal anti-inflammatory drug for the topical treatment of acute pain due to minor strains, sprains and contusions. Figure 1 displays monthly non cumulative sales data of the two versions available, the package with 5 patches and the more recent with 10, although in our analysis we will just focus on sales of the first, that we label "Itm5". We observe that the earlier version with 5 patches experienced a quite fast growth in the first three months of its life-cycle and, after the peak, reached in March 2007 (t = 31), the decline has begun, probably accelerated by the competing effect of the newer version with 10 patches launched in October 2006 (t = 26). We may also record the presence of a quite clear seasonal pattern, with a period of about 12 months, which appears stronger during the growing phase of the life-cycle. Such seasonal behaviour is understandable in the case of this drug, which is mostly used in winter time, when joint inflammations are more frequent.

We begin our analysis by applying a simple Bass model on instantaneous data as described in equation (8), in order to confirm the hypothesis of a finite life-cycle for this drug. From the results of this first application, obtained under a standard nonlinear least squares approach (Levemberg-Marquardt, see [10]), and presented in Table 1 and Figure 2a, we see that the Bass model identifies quite well the mean trajectory of the series, as confirmed by the acceptable R^2 . The Durbin-Watson

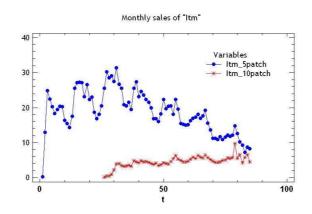


Figure 1: "Itm": number of monthly sold packages in the two available versions.

Table 1: Parameter estimates of a Bass model on instantaneous data. (): marginal linearized asymptotic 95% confidence limits

m	p	q	R^2
1816.61	0.00886	0.03299	0.619892
(1703.62)	(0.00792)	(0.02704)	SSE: 1137.4
(1929.6)	(0.00981)	(0.03894)	(D-W: 0.703545)

statistic highlights the presence of autocorrelated residuals, although we believe that part of the strong variability still visible in data is due to seasonal effects.

We therefore extend the analysis following the structure proposed in equation (16), where A(t) has the form described in equation (20), so that in this case c = 0, and apply to our series a Bass model with a seasonal component of period 12, s = 12, namely

$$y(t) = \left\{ m + \left[a \cos\left(\frac{2\pi t}{12}\right) + b \sin\left(\frac{2\pi t}{12}\right) \right] \right\} \left[F(t+0.5) - F(t-0.5) \right] + \varepsilon(t).$$
(22)

The results obtained with this second methodology are significantly better and highlight the role of function A(t) in improving forecasts. From Table 2 we see that parameter estimates of m, p and q are essentially stable, but the model global fitting has notably increased, $R^2 = 0.802401$, and the *SSE* has been halved with respect to the standard Bass model. Moreover, the modulation of the seasonal effect according to the phase of the life–cycle appears very accurate, since this tends to fade as long as sales decline. A clear advantage of using the model with seasonal component is the ability to provide better forecasts in the short period.

Although the analysis performed so far seems enough satisfactory, we also decide to apply a Guseo–Guidolin model with seasonal component to our series. By

Table 2: Parameter estimates of a Bass model with seasonal component. (): marginal linearized asymptotic 95% confidence limits

\overline{m}	p	q	a	b	R^2
1815.61	0.00887	0.03289	-308.174	140.060	0.802401
(1732.57)	(0.00819)	(0.02858)	(-387.168)	(61.849)	SSE: 591.277
(1898.65)	(0.00954)	(0.03720)	(-229.180)	(218.272)	(D-W: 1.05691)

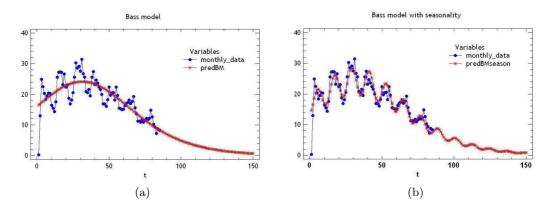


Figure 2: Itm5: (a) standard Bass model and monthly non cumulative data, (b) Bass model with seasonal component and monthly non cumulative sales data

assuming a dynamic market potential, this model has shown a greater efficiency in describing the first part of the life-cycle (see [4], which is typically over-estimated by the Bass model (as one may distinctly observe in Figure 2a). The applied model is

$$y(t) = \left\{ K + \left[a \cos\left(\frac{2\pi t}{12}\right) + b \sin\left(\frac{2\pi t}{12}\right) \right] \right\} \left[W(t+0.5) - W(t-0.5) \right] + \varepsilon(t).$$
(23)

From Table 3 we may observe that parameter estimates, obtained with standard NLS techniques, are quite unstable, but we may see in Figure 3 that model (23) indeed produces a better description of the first part of the series. In order to test the global significance of this model with respect to the simpler Bass model with seasonal component, we may compute a squared multiple partial correlation coefficient, $\tilde{R}^2 = (R_{GuGuS}^2 - R_{BMS}^2)/(1 - R_{BMS}^2)$, normalized in the interval [0, 1], and the corresponding F-ratio, $F = \tilde{R}^2(N - k)/[(1 - \tilde{R}^2)s]$ where N is the number of observations, k the parameter cardinality of the extended model, and k - s is the number of parameters of the nested model (a typical critical value for F is 4). In this case $\tilde{R}^2 = 0.229980$ and F = 11.648, which indicate a not minor significance of the extended model.

The Durbin-Watson statistic still diagnoses the presence of autocorrelated residuals, so that we complete the analysis by applying an ARMA(p,q) to the series of residuals, $\hat{\varepsilon}(t) = y(t) - \hat{y}(t)$, (as proposed for instance in [5]). The results obtained after the ARMA refinement, and presented in Table 4, appear enough satisfactory: $R^2 = 0.899546$. Figure 3b displays the final forecasts, obtained by summing up a

D-W: 1.24932

(19755.300)

K	p_c	q_c	p_s	q_s	a	b	R^2
2262.58	0.00634	0.04065	0.03505	-0.03490	-366.242	207.083	0.847845
(-211454)	(0.00273)	(0.01851)	(-3.23216)	(-3.23216)	(-34953.600)	(-19341,200)	SSE: 455.295

(3.18670)

(34221.100)

Table 3: Parameter estimates of a Guseo–Guidolin model with seasonal component. () marginal linearized asymptotic 95% confidence limits

Table 4: ARMA (3,2) on autocorrelated residuals; () t-statistic, [] p value

(3.30227)

AR(1)	AR(2)	AR(3)	MA(1)	MA(2)	R^2
1.44332	-1.41704	0.49632	1.04727	-0.80375	0.899546
(11.2463)	(-12.5819)	(4.37409)	(10.24270)	(-9.16953)	SSE: 3.75736
[0.00000]	[0.00000]	[0.00003]	[0.00000]	[0.00000]	d.f. 80

Guseo–Guidolin model with seasonal component and an ARMA (3,2) on autocorrelated residuals.

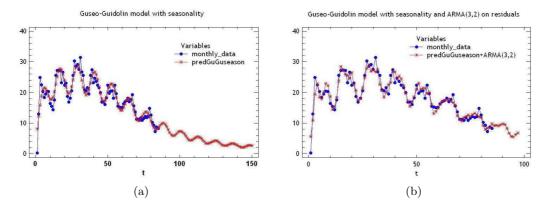


Figure 3: Itm5: (a) Guseo–Guidolin model with seasonal component and monthly non cumulative sales data, (b) Guseo–Guidolin model with seasonal component + ARMA (3,2) on autocorrelated residuals, and monthly non cumulative sales data

For comparative purposes we propose here the results of an alternative analysis which assumes that seasonality is a stochastic component correlated with nonseasonal ones. In this case, its effect may be studied under a seasonal autoregressive approach within residual analysis, once the trend has been estimated. ARMA processes have been extended by Box and Jenkins (see [3]) in order to include a periodic behaviour due to seasonality, giving rise to seasonal ARMA models, SARMA, that have the following structure

$$\phi(B)\Phi(B^S)y(t) = \vartheta(B)\Theta(B^S)a_t.$$
(24)

The underlying feature of SARMA models is the ability to describe simultaneously two types of correlation: the first one, among successive observations called *within period correlation* and the second, among observations delayed of a multiple of the period, called *between period correlation*.

(215979)

(0.00994)

(0.06279)

In particular, we estimated the trend with a simple Guseo–Guidolin model on instantaneous data (see equation (14)) and subsequently we applied a SARMA model to the series of residuals. Results of this alternative procedure are presented in Table 5, Table 6 and Figure 4. It is not surprising that this methodology produces a higher fitting, $R^2 = 0.945809$, since the model used has a larger set of parameters. Indeed, the choice to model the seasonal component according to a deterministic view results in a more parsimonious structure.

Table 5: Parameter estimates of a Guseo–Guidolin model. () marginal linearized asymptotic 95% confidence limits

K	p_c	q_c	p_s	q_s	R^2
2152.29	0.00526	0.04409	0.04361	-0.04314	0.66568
(-63364)	(0.00166)	(0.01637)	(-1.28522)	(-1.44758)	SSE: 1000.39
(67668)	(0.00887)	(0.07180)	(1.37245)	(1.36128)	D-W: 0.71311

Table 6: SARMA (3,2)x(3,1) on autocorrelated residuals; () t-statistic, [] p value

AR(1)	AR(2)	AR(3)	MA(1)	MA(2)
-0.60577	-0.23148	0.62984	-1.22715	-0.98628
(-19.73200)	(-5.60227)	(33.17000)	(-18.99980)	(-21.05570)
[0.00000]	[0.00000]	[0.00003]	[0.00000]	[0.00000]
SAR(1)	SAR(2)	SAR(3)	SMA(1)	R^2
-0.48916	0.50621	0.82765	-0.83824	0.945809
(-7.56970)	(7.46820)	(19.89580)	(-13.58180)	SSE: 2.1336
[0.00000]	[0.00000]	[0.00003]	[0.00000]	d.f. 76

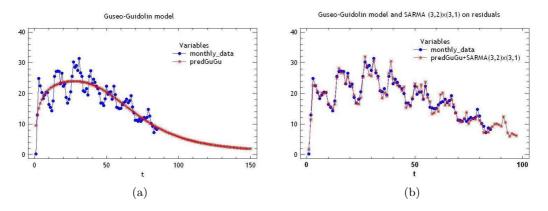


Figure 4: Itm5: (a) Guseo–Guidolin model and monthly non cumulative sales data, (b) Guseo–Guidolin model + SARMA (3,2)x(3,1) on autocorrelated residuals, and monthly non cumulative sales data

4 Conclusion

In this paper we have proposed two diffusion models that extend the basic Bass and Guseo–Guidolin structures by taking into account the presence of a seasonal pattern in sales. In particular, both trend and seasonality are assumed as deterministic components to be estimated with standard nonlinear regression techniques, and the seasonal effect is made proportional to the trend, in order to account for the effect that life–cycle evolution may exert. An application to the monthly sales of a pharmaceutical drug has confirmed the better performance of such models in describing the intra-year oscillations, suggesting their possible use for short-term forecasts. Indeed, this is just one possible option to deal with seasonality: an alternative path may be applying SARMA models to the series of residuals, having assumed a stochastic nature of seasonality.

Our choice for a deterministic structure is inspired by the aim of creating new extensions of basic diffusion models and those proposed in this paper are just a first step for further developments. In particular, future research should focus on a model where the seasonal pattern contributes to the formation of demand and is therefore included in the structure describing the latent market potential. This model, which would represent an extension of the Guseo–Guidolin one, should ideally be able to take into account the presence of different regimes in the level of potential demand during a product life–cycle.

References

- F.M. Bass, A new product growth model for consumer durables, *Manage. Sci.* 15 (1969) 215 227.
- [2] P. Bloomfield, Fourier Analysis of Time Series, An Introduction, 2nd ed., Wiley Series in Probability and Statistics, Wiley and Sons, 2000.
- [3] G.E.P. Box, G.M. Jenkins, Time Series Analysis, Forecasting and Control, revised edition, Holden–Day, Oakland, California, 1976.
- [4] R. Guseo, M. Guidolin, Modelling a dynamic market potential: A class of automata networks for diffusion of innovations, *Technol. Forecast. Soc. Change*, **76** (6) (2009) 806 820.
- [5] R. Guseo, M. Guidolin, Market potential dynamics in innovation diffusion: modelling the synergy between two driving forces, *Technol. Forecast. Soc. Change*, 78 (1) (2011) 13 24.
- [6] S. Hylleberg, Modelling Seasonality, Oxford University Press, 1992.
- [7] Y. Peers, D. Fok, P.H. Franses, Modelling seasonality in new product diffusion, *ERIM Report Series Research in Management*, July 2010, available at: www.erim.eur.nl
- [8] R. Peres, E. Muller, V. Mahajan, Innovation diffusion and new product growth models: A critical review and research directions, *Intern. J. of Research in Marketing* 27 (2010) 91 106.

- [9] W.P. Putsis, Temporal aggregation in diffusion models of first-time purchase: does choice of frequency matter? *Technol. Forecast. Soc. Change*, **51** (1996) 265 279.
- [10] G.A.F. Seber, C.J. Wild, Nonlinear regression, Wiley, New York, 1989.
- [11] W.W.S. Wei, Time series analysis, Univariate and Multivariate Methods, 2nd ed. Addison Wesley, 2006.

Working Paper Series Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it Most of the working papers can also be found at the following url: http://wp.stat.unipd.it



