ERASMUS

9217/A

ECONOMETRIC INSTITUTE

THE GOMPERTZ CURVE:
ESTIMATION AND SELECTION

Ph.H. FRANSES

REPORT 9217A

Erasmus

ERASMUS UNIVERSITY ROTTERDAM - P.O. BOX 1738 - 3000 DR ROTTERDAM-THE NETHERLANDS

THE GOMPERTZ CURVE: ESTIMATION AND MODEL SELECTION

by

Philip Hans FRANSES

Econometric Institute, Erasmus University Rotterdam
P.O.Box 1738, *NL* – 3000 *DR Rotterdam, The Netherlands*
phone: +3110 4081273, *fax*: +3110 4527746

This report contains two closely related, but short, papers:

1. Fitting a Gompertz curve

2. A method to select between Gompertz and logistic trend curves

FITTING A GOMPERTZ CURVE

ABSTRACT

In this paper a simple Gompertz curve–fitting procedure is proposed which does not face the problmes of current fitting methods such as the effects of the nonstationary behavior of a time series and the assumptions on the value of the saturation level. The advantages of this method are that the stability of the saturation level over the sample period can now easily be checked, and that any knowledge of its value is not necessary for forecasting. An application to the stock of cars in the Netherlands illustrates its merits.

Keywords: *Time series*, *Forecasting*

ACKNOWLEDGEMENTS

Thanks are due to Erno Kuiper and Nigel Meade for several helpful comments.

1. INTRODUCTION

Forecasting the stock of vehicles or the prospective sales of a new product are examples of practical occasions in which a univariate time series can be usefully, though roughly, described by a Gompertz trend curve. Loosely speaking, this curve has an S–shape which, in contrast to the also often applied logistic curve, is nonsymmetrical. See, e.g., Harrison and Pearce (1972), Mar–Molinero (1980) and Meade (1984) for surveys of the distinct types of trend curves. More precisely, the Gompertz curve assumes that the period of an increasing growth of sales or stock is shorter than the period in which this growth is decreasing and in which the process is adjusting to its saturation level. This level is one of the three unknown parameters in the model, and its value is usually assumed a priori or estimated iteratively. One can imagine that this value can be quite influential for the forecasted future values of a time series process. Another aspect that can establish any fragility of the forecasts from a Gompertz curve is that the variability of the time series processes considered is related to the level of the series. In other words, these processes are not stationary and require differencing. See, e.g., Granger and Newbold (1986) for a survey of definitions and tests for nonstationarity. It is therefore common practice to weigh the observations in the Gompertz curve estimation procedures, or to consider alternative error structures, see Meade (1988). Unfortunately, these weights have to be estimated from the same data, see Harrison and Pearce (1972, p.156), and can therefore also be influential for forecasting.

In section 2 of the present paper I propose a simple Gompertz curve fitting procedure which deals with these issues. In section 3 it is applied to forecasting the stock of cars in the Netherlands. Section 4 concludes.

2. FITTING A GOMPERTZ CURVE

The typical shape of a process x_t that is generated by a Gompertz curve can be observed from figure 1. Three characteristics of this curve are clearly observable. The first is that the point of inflexion, i.e. the point in time in which the rate of growth changes from increasing into decreasing, is occurring before half the saturation level is reached. Second, the rate of growth is always larger than, though decreasing to, zero. Finally, the process is not stationary, in the sense that it does not have a constant mean and constant variance. See Granger and Newbold (1986) for more precise definitions. These characteristics establish the usefulness of fitting a Gompertz growth curve for time series processes like the stock of cars or the sales of a new product. The mathematical representation of the process x_t depicted in figure 1 is given by

$$x_t = \alpha.\exp(-\beta.\exp(-\gamma.t)) \qquad (1)$$

where α, β and γ are unknown positive valued parameters, the first of which is the value of the saturation level. The t is a linear deterministic time trend defined by $t = 0,1,2,..$

To fit model (1) to an empirical series, current estimation procedures often consider three observations of the process x_t, see Meade (1985) and Oliver (1987) for alternative approaches. Substituting these and their corresponding values for t in (1) gives three equations with three unknown parameters. The three observations are usually linear combinations of the first, middle and last observations on x_t. One may also fix the value of α, and estimate the remaining parameters. Furthermore, since the series x_t is nonstationary it is common practice to weigh the observations according to their level. On the other hand, there are also examples in which instead of weighting, the model in (1) is enlarged with a first order autoregressive error process, see, e.g., Mar–Molinero (1980). In that study it is found that the corresponding parameter is quite close to 1, which is an indication that the process x_t is not stationary. In summary, there are several relatively arbitrary choices to be made for the empirical fitting of a Gompertz curve, and it seems that these choices can have a large impact on forecasting performance.

There is however a simple strategy to circumvent these problems of the effects of the nonstationarity of x_t and of the assumptions on α. This is based on transforming the process in (1) via taking first differences and logarithms, see Harvey (1984) for a related approach in the case of a logistic trend curve. Denoting log for the natural logarithm, consider a transformed version of (1),

$$\log x_t = \log\alpha - \beta.\exp(-\gamma.t), \qquad (2)$$

Taking first differences of $\log x_t$, or $\Delta\log x_t = \log x_t - \log x_{t-1}$, gives

$$\begin{aligned} \Delta\log x_t &= -\beta.\exp(-\gamma.t) + \beta.\exp(-\gamma.t + \gamma) \qquad (3) \\ &= \exp(-\gamma.t)(\beta.\exp\gamma - \beta) \end{aligned}$$

where now the value of α is removed from the equation. A linearization of (3) yields the equation

$$\log(\Delta\log x_t) = -\gamma t + \log(\beta.\exp\gamma - \beta) \qquad (4)$$

which can be easily estimated by ordinary least squares (OLS), given certain assumptions on error terms. A regression of $\log(\Delta\log x_t)$ on a contant and a trend yield estimates of μ, say, and γ, respectively. With these $\hat{\mu}$ and $\hat{\gamma}$ one can estimate β via

$$\hat{\beta} = [\exp(\hat{\mu})]/[\exp(\hat{\gamma})-1] \tag{5}$$

and a sequence of saturation levels via

$$\hat{\alpha}_t = \exp(\log x_t + \hat{\beta}\exp(-\hat{\gamma}.t)) \tag{6}$$

which are all simple calculations.

There are several interesting aspects to the equations (4) through (6). First, the problem of the nonstationarity of x_t seems to be overcome since $\log(\Delta\log x_t)$ is a trend stationary variable, see also figure 12 in Harrison and Pearce (1972). Inference on the parameters μ and γ can be carried out using standard procedures. Second, although the level α is assumed to be constant over the sample, there is an opportunity to check its stability by considering the range of α_t values obtained from (6). Structural breaks in the saturation level because of, e.g., technology shocks can now also be detected. Third, though of interest for the understanding of the process under consideration, the values of α and β are not required for forecasting future values of x_t. The forecasts can simply be obtained from a recursion formula implied by (4). The standard errors related to $\hat{\alpha}$ and $\hat{\beta}$ are however difficult to calculate. One suggestion for $\hat{\alpha}$ is to evaluate the mean and variance of $\hat{\alpha}_t$ over the sample. An alternative procedure is to calculate the forecast for x_t from (4) when t goes to infinity. The corresponding value x_∞ should then come close to the saturation level. Finally, the major problem involving the proposed estimation method is that there may be occasions in which $\log(\Delta\log x_t)$ can not be calculated since $\Delta\log x_t$ is below zero. One solution is to treat the corresponding observations as being missing values, and estimate (4) for the remaining data points.

3. AN APPLICATION

To illustrate the proposed Gompertz curve fitting method, I have chosen to consider the stock of cars in the Netherlands, 1965–1989. The observations are given in table 1, and the graph of the series is depicted in figure 2. It can be seen that a Gompertz curve may be a useful description of these data. Furthermore, given that long range forecasts of the stock of cars are quite important for a country as small as the Netherlands, one can imagine that an estimate of the saturation level may also have policy implications.

Denoting s_t as the stock of cars, the estimation results of the model in (4) are (with standard errors in parentheses)

$$\log(\Delta \log s_t) = \underset{(0.1314)}{-1.8161} - \underset{(0.0095)}{0.0988t} \qquad (7)$$

This model is estimated with 25 observations, and it includes a dummy for 1982 to capture the effects of an outlying observation. The values of the R^2 and of an F test for first order autocorrelation are 0.863 and 1.474, respectively. Hence, together with the large values of the t ratios, the model seems to fit the data reasonably well.

An estimate for the β in (1) is obtained via (5). It yields a value of 1.567. Furthermore, 25 estimates of the saturation level α can be found from (6), and they are displayed in table 2 and figure 3. It appears that these estimates are quite constant over the sample, and that the mean of the $\hat{\alpha}_t$ is about 6200. This implies a saturation level of 6.2 million cars in the Netherlands. Moreover, the relative values of this level correspond to several economic events. For example, the decrease in 1986 is related to the stringent application of a quality control measure for cars.

To forecast the stock of cars for the period 1990–2010, I have chosen to use equation (7). The results are displayed in the third column of table 3. Furthermore, I have generated forecasts when $\hat{\beta}$ and $\hat{\gamma}$ and the maximum and minimum values of the estimated saturation level in table 3 are substituted in (1). The corresponding forecasts are reported in the columns 2 and 4 of table 3. In figure 4, the graphs of these forecasts are given.

4. CONCLUDING REMARKS

In this paper I propose a simple Gompertz curve fitting procedure which does not face the problems of current fitting methods such as the effects of the nonstationary behavior of a time series and the assumptions on the value of the saturation level. The advantages are that estimates of this level can yield insights in the stability of the saturation level over the sample period, and that knowledge of its value is not necessary for forecasting. Furthermore, the other two parameters can be estimated via standard methods as ordinary least squares. An application to the stock of cars illustrates its merits. An unreported application to the tractors in Spain series in Mar–Molinero (1980) also yields successful outcomes.

A possible drawback of the procedure is that it assumes that the growth of the process is always positive. For practical series this may however not always be the case. One way to circumvent this problem is to delete the observations that invalidate this assumption, and to estimate the model for the remaining observations. Unreported experience with the data sets given in Harrison and Pearce (1972) indicates the usefulness of this approach.

TABLES

Table 1. The stock of cars in the Netherlands, 1965-1989

Year	Stock (×1000)	Year	Stock (×1000)
1965	1273	1978	4056
1966	1502	1979	4312
1967	1696	1980	4515
1968	1952	1981	4594
1969	2212	1982	4630
1970	2465	1983	4728
1971	2702	1984	4818
1972	2903	1985	4901
1973	3080	1986	4950
1974	3214	1987	5118
1975	3399	1988	5251
1976	3629	1989	5371
1977	3851		

Note: The observation for 1964 is 1059 and it is used as a starting-value.

Table 2. Estimated saturation level when a Gompertz curve is fitted

Year		Year	
1965	6097	1978	6258
1966	6208	1979	6387
1967	6134	1980	6445
1968	6255	1981	6342
1969	6353	1982	6200
1970	6411	1983	6160
1971	6423	1984	6123
1972	6361	1985	6090
1973	6268	1986	6026
1974	6118	1987	6116
1975	6090	1988	6171
1976	6155	1989	6217
1977	6215		

mean = 6225 standard deviation = 120

$\hat{\alpha}_{max} = 6445$ $\hat{\alpha}_{min} = 6026$

Table 3. Forecasting the stock of cars, 1990-2010

Year	with $\hat{\alpha}_{max}$	without using $\hat{\alpha}$	with $\hat{\alpha}_{min}$
1990	5645	5445	5278
1991	5716	5514	5344
1992	5781	5576	5405
1993	5840	5633	5461
1994	5895	5686	5512
1995	5944	5734	5558
1996	5990	5777	5600
1997	6031	5817	5639
1998	6069	5854	5674
1999	6103	5887	5707
2000	6135	5917	5736
2001	6163	5944	5763
2002	6189	5969	5787
2003	6213	5992	5809
2004	6234	6013	5829
2005	6254	6032	5847
2006	6272	6049	5864
2007	6288	6064	5879
2008	6302	6078	5893
2009	6316	6091	5905
2010	6328	6103	5916

FIGURES

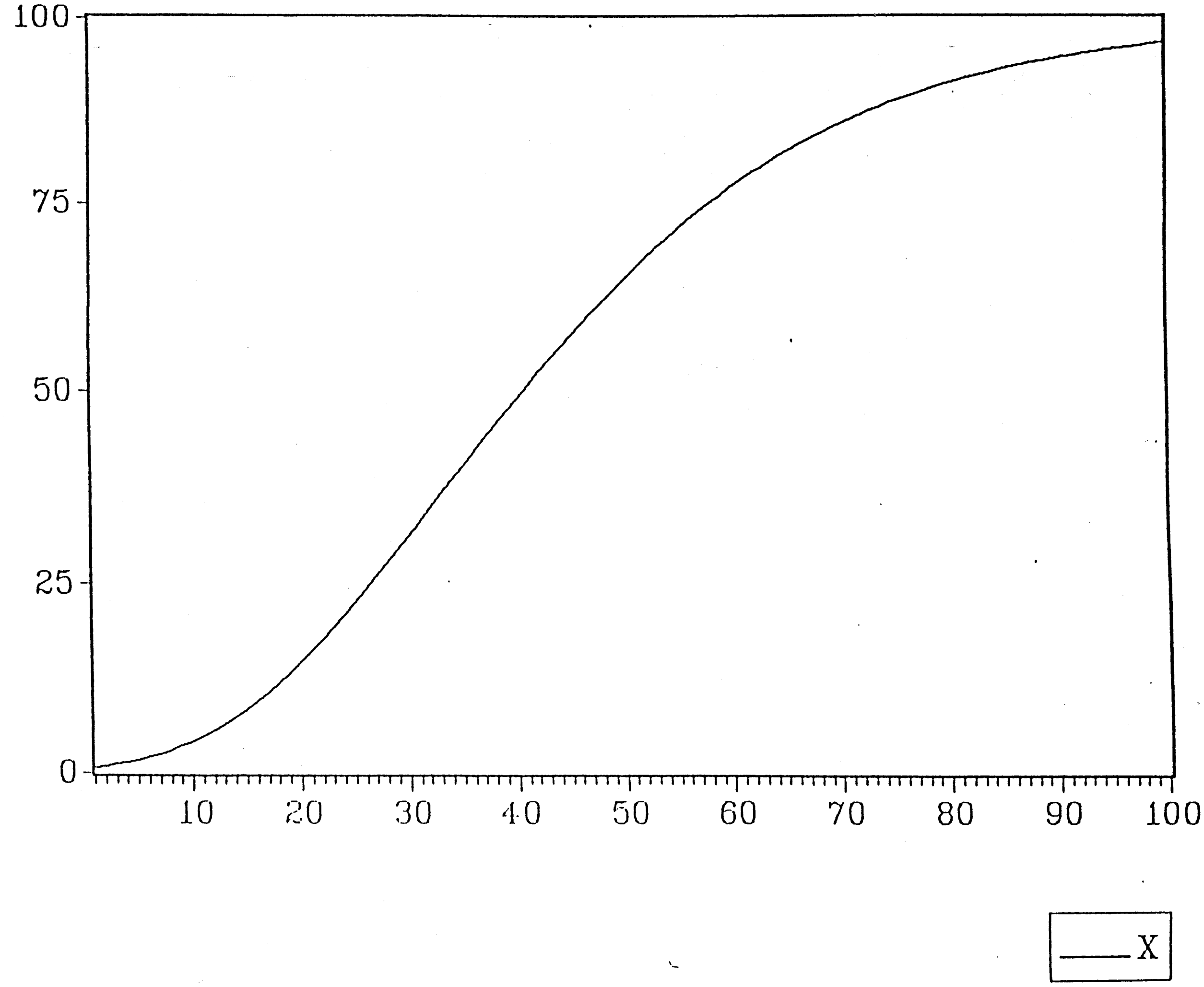

Figure 1. The typical shape of a Gompertz curve

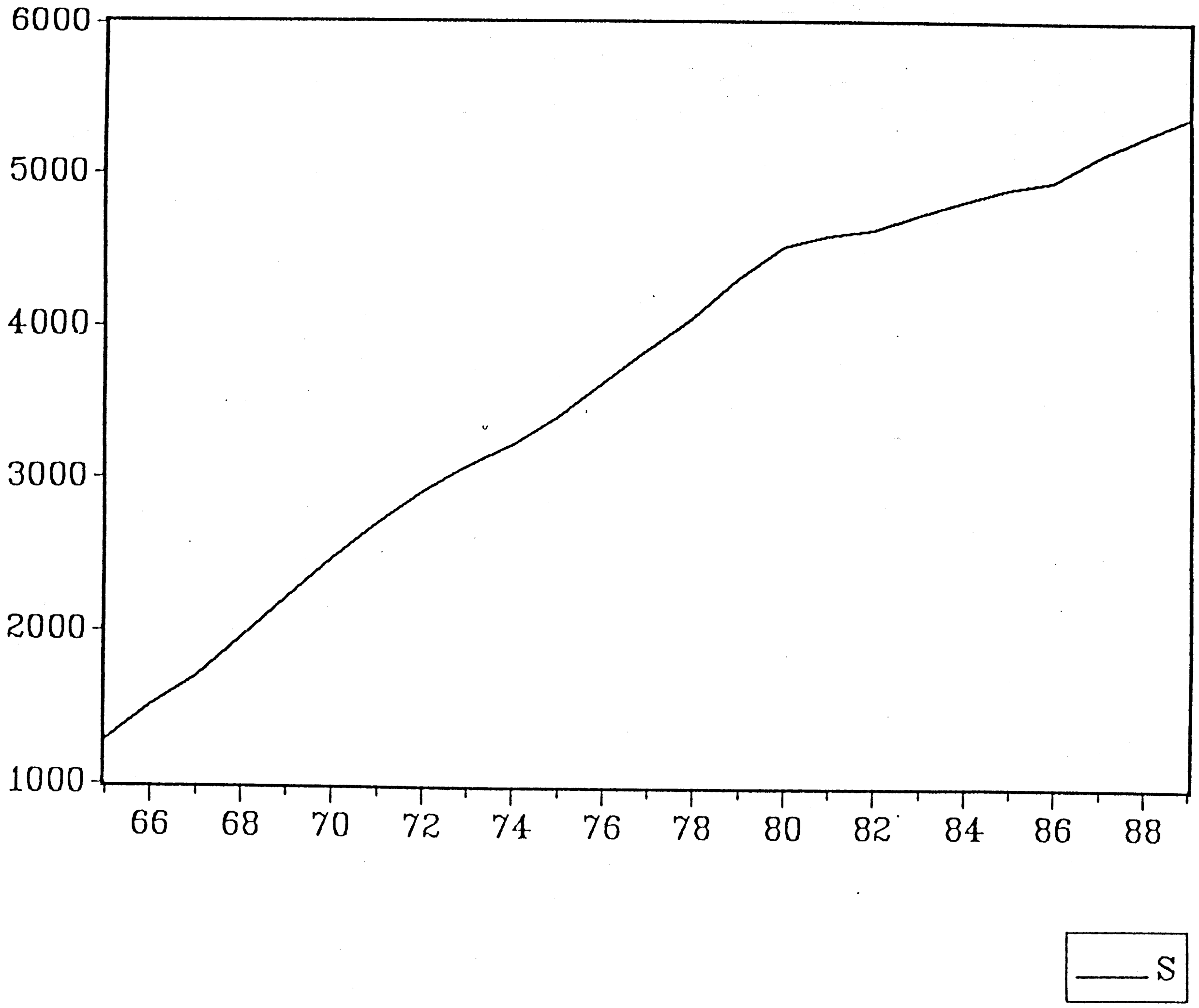

Figure 2. The stock of cars in the Netherlands, 1965–1989

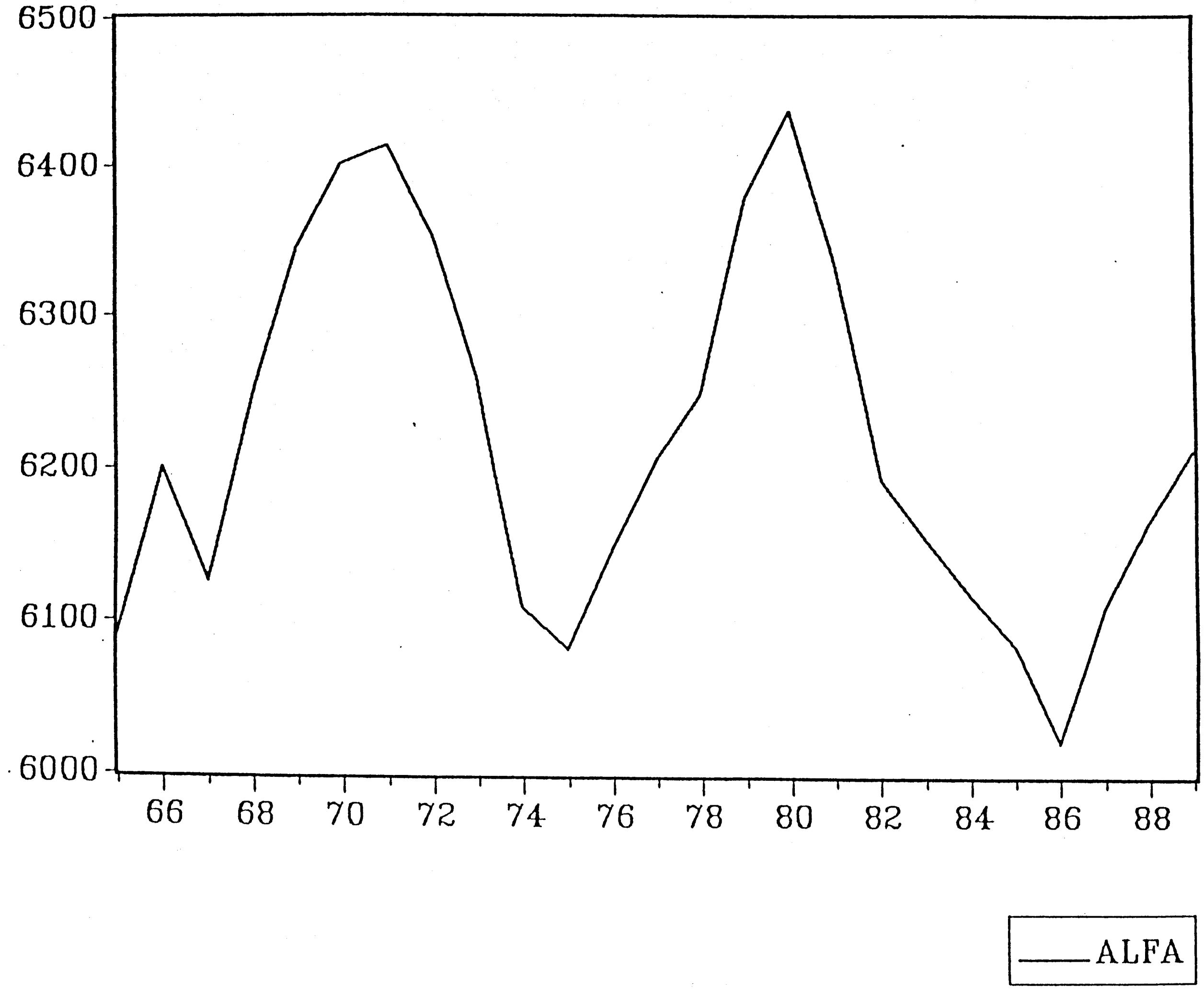

Figure 3. The estimated saturation level, 1965–1989

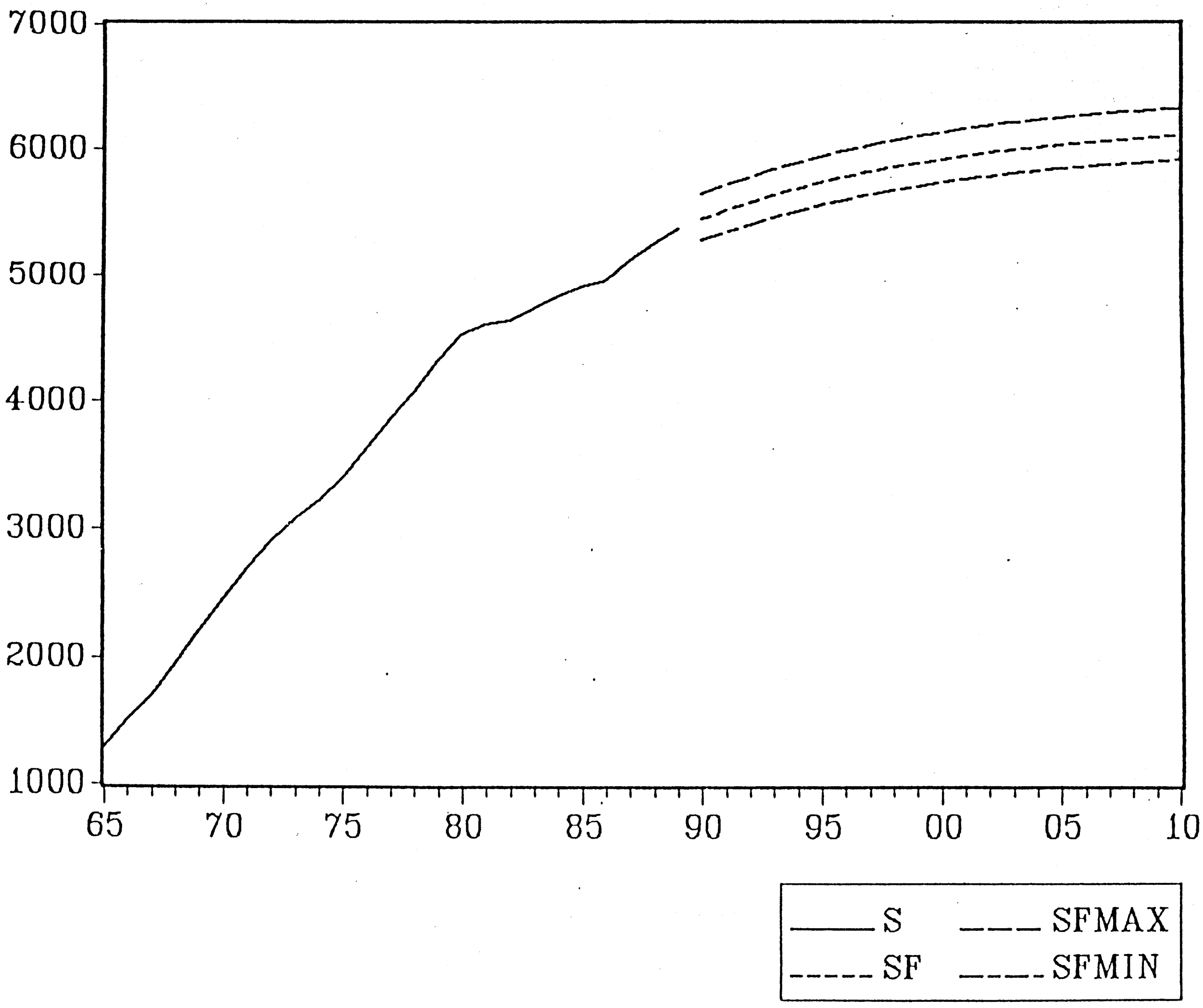

Figure 4. Forecasting the stock of cars 1990–2010

REFERENCES

Granger, C.W.J. and P. Newbold (1986), *Forecasting Economic Time Series*, San Diego: Academic Press.ess.

Harrison, P.J. and S.F. Pearce (1972), The Use of Trend Curves as an Aid to Market Forecasting, *Industrial Marketing Management*, 2, 149–170.

Harvey, A.C. (1984), Time Series Forecasting based on the Logistic Curve, *Journal of the Operational Research Society*, 35, 641–646.

Mar–Molinero, C. (1980), Tractors in Spain: A Logistic Analysis, *Journal of the Operational Research Society*, 31, 141–152.

Meade, N. (1984), The Use of Growth Curves in Forecasting Market Development–A Review and Appraisal, *Journal of Forecasting*, 3, 429–451.

Meade, N. (1985), Forecasting Using Growth Curves – An Adaptive Approach, *Journal of Operational Society*, 36, 1103–1115.

Meade, N. (1988), Forecasting with Growth Curves: the Effect of Error Structure, *Journal of Forecasting*, 7, 235–244.

Oliver, R.M. (1987), A Bayesian Model to predict Saturation and Logistic Growth, *Journal of the Operational Research Society*, 38, 49–56.

A METHOD TO SELECT BETWEEN GOMPERTZ AND LOGISTIC TREND CURVES

ABSTRACT

In this paper a simple method is proposed to select between two often applied trend curves, i.e. the Gompertz and the logistic curve. The method is based on one auxiliary regression. Two applications illustrate its merits.

Key words: Trend curves; Model selection

INTRODUCTION

The Gompertz and logistic trend curves are often applied in forecasting market development, see Meade (1984) for an extensive overview. Although these curves can describe similar behaviour in some phases of this development, one of the most important differences is that the growth towards the saturation level is less rapid for the Gompertz process than for the logistic process. Therefore, using an inappropriate growth curve can have a large impact on forecasting. Though the selection of an appropriate curve appears to be important, the choice between the two models is usually made using criteria based on forecasting errors and on the plausibility of the estimated saturation levels. In this paper, I propose an alternative selection method which is based on one auxiliary regression, and on a significance test for one parameter.

In the next section, some aspects of the Gompertz and logistic curve are discussed, and the simple selection method is explained. In section 2, it will be applied to two empirical series to illustrate its merits. The third section concludes this paper.

1. A SELECTION METHOD

The Gompertz trend curve for a time series X_t is given by

$$X_t = a.\exp(-b.\exp(-ct)), \tag{1}$$

where t represents time, and where a is the saturation level and $a,b,c > 0$. A logistic curve for X_t can be written as

$$X_t = a.(1 + b.\exp(-ct))^{-1}, \tag{2}$$

where $a,b,c > 0$. Typical graphs of these curves are displayed in exhibit 1.

Recently, in Franses (1991), a simple estimation method for the Gompertz curve has been developed and applied. This method uses the fact that the model in (1) can be rewritten as

$$\log(\Delta \log X_t) = d - ct, \tag{3}$$

where log denotes the natural logarithm, where Δ is the differencing filter defined by $\Delta z_t = z_t - z_{t-1}$, and where d is a nonlinear function of b and c.

The model for $\log(\Delta \log X_t)$ in (3) is linear, and it seems worthwhile to investigate whether the logistic model in (2) can be rewritten analogously. A first step is to take logs of both sides of (2), and to apply the differencing filter Δ, which results in

$$\begin{aligned} \Delta \log X_t &= \log[(1 + b\exp(-ct+c))/(1 + b\exp(-ct))] \\ &\approx b\exp(-ct).(\exp c-1)/(1 + b\exp(-ct)) \end{aligned} \tag{4}$$

since the latter expression only obtains values between 0 and 1. Taking logs of both sides of (4) and some rewriting yields

$$\log(\Delta \log X_t) \approx d - ct + (\log X_t - \log a) \tag{5}$$

where d is a nonlinear function of b and c.

Typical graphs of the $\log(\Delta \log X_t)$ series for the Gompertz and logistic curves are depicted in exhibit 2. It is obvious that the $(\log X_t - \log a)$ element of (5) ensures that $\log(\Delta \log X_t)$ is a nonlinear function of time for the logistic curve. A simple selection method between (1) and (2) may therefore be given by the auxiliary regression

$$\log(\Delta \log X_t) = \delta + \gamma t + \tau t^2, \tag{6}$$

and a test for the significance of the τ parameter based on its t ratio. Of course, one may also want to consider variables like t^{-1} or $t^{-1/2}$ instead of t^2.

2. APPLICATIONS

The illustrate the merits of the proposed model selection method, I consider two examples. The first is taken from Mar–Molinero (1980), and it concerns the tractors in Spain data series. In that paper, and also in Meade (1984), it has been argued that a logistic curve fits these data best. This conjecture can be verified by looking at the graph of the $\log(\Delta\log X_t)$ series in exhibit 3. This figure is similar to that for the logistic curve in exhibit 2. Additionally, the t ratio of the τ parameter in the regression as (6) obtains a value of –3.740.

The second example is given by the Dutch annual stock of cars series, as it is analysed in Franses (1991), where a Gompertz curve has been fitted to this series. The graph of the $\log(\Delta\log X_t)$ series is depicted in exhibit 4, and there is some visual evidence that a model like (3) is indeed appropriate. The t value of the τ parameter is 0.633, when a dummy variable for an outlying observation in 1982 is included in (6). Deleting this dummy does not change the result that a Gompertz curve seems indeed appropriate for the stock of cars series.

When $\Delta\log X_t$ obtains a negative value for a certain observation, the log transformation can not be taken. In that case, this observation can be treated as a missing observation, and the above method can be applied to the remaining data points.

3. CONCLUSION

In this paper a simple method is proposed to choose between a Gompertz and a logistic trend curve. Two examples indicate its practical use. Moreover, it seems possible to extend this method to other types of trend curves.

EXHIBITS

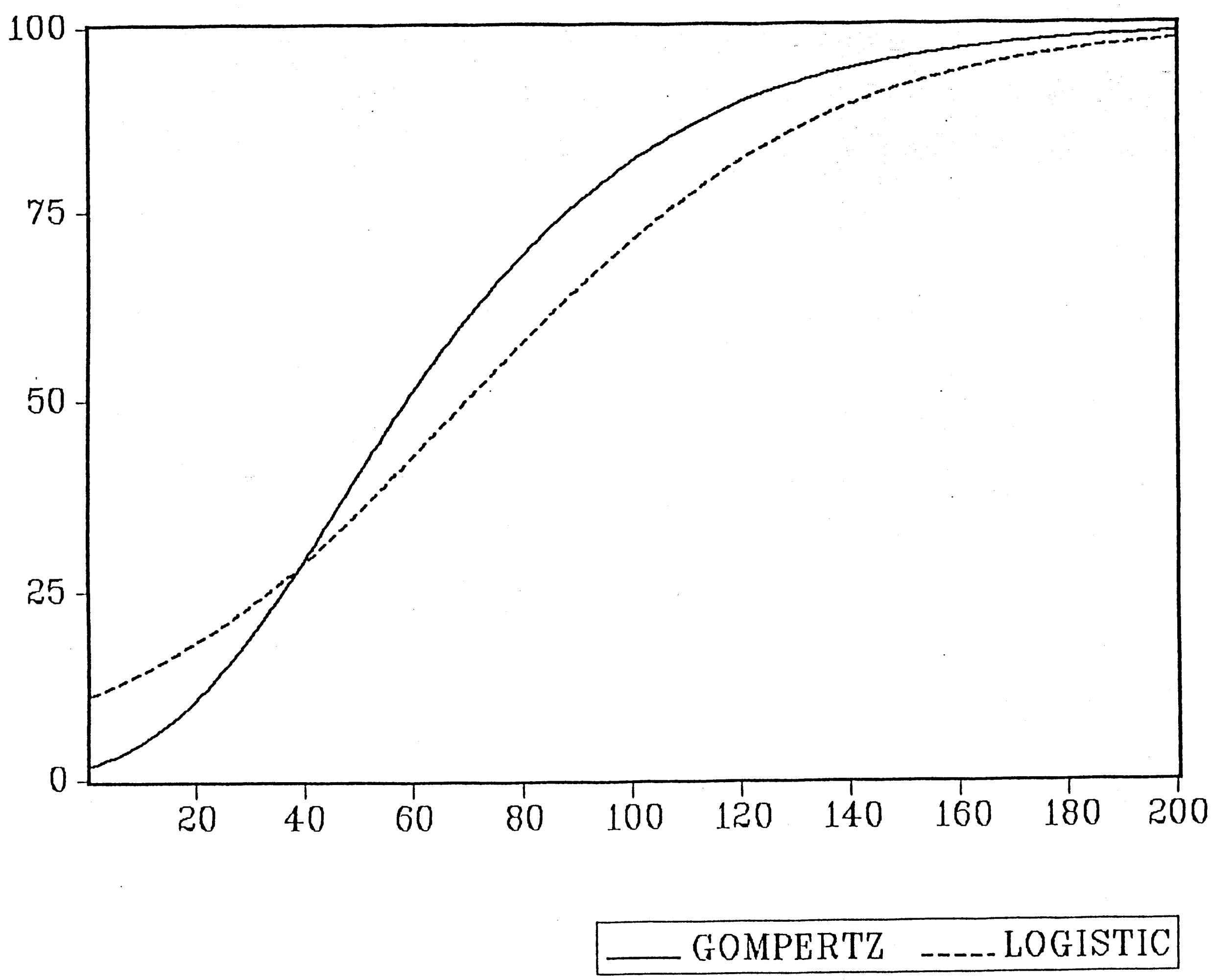

Exhibit 1. A Gompertz curve and a logistic curve

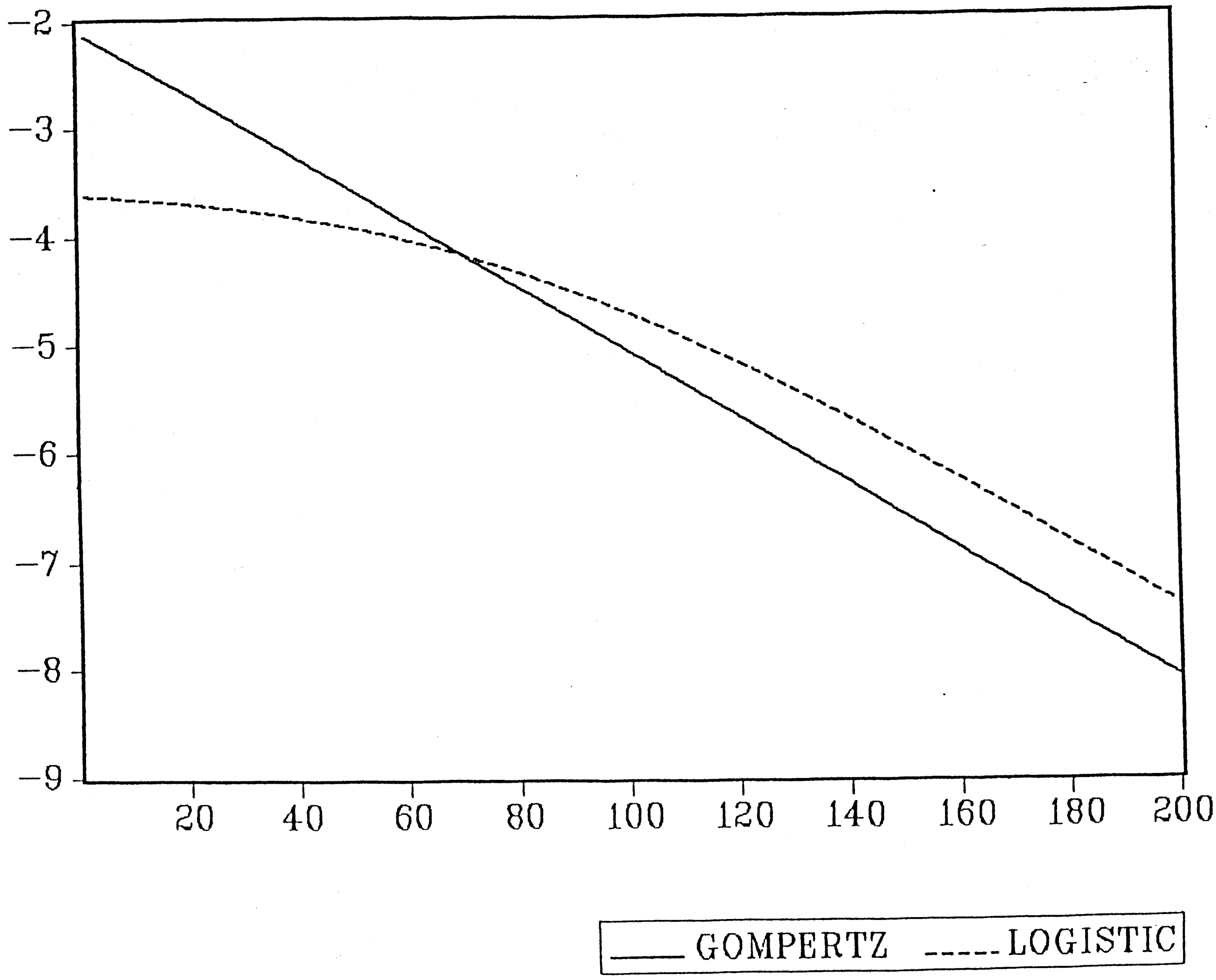

Exhibit 2. $\log(\log X_t - \log X_{t-1})$ for a Gompertz and a logistic process

Exhibit 3. log(logX_t-logX_{t-1}) for the tractors in Spain

Exhibit 4. log($\log X_t - \log X_{t-1}$) for the stock of cars in the Netherlands

REFERENCES

Franses, P.H., 'Fitting a Gompertz Curve', (1991), unpublished manuscript.

Mar-Molinero, C. 'Tractors in Spain: a Logistic Analysis', Journal of the Operational Research Society, **31** (1980), 141–152.

Meade, N., 'The Use of Growth Curves in Forecasting Market Development – a Review and Appraisal', Journal of Forecasting, **3** (1984), 429–451.