Technological Forecasting & Social Change 88 (2014) 98–105

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

A simple selection test between the Gompertz and Logistic growth models☆

Pierre Nguimkeu *

Department of Economics, Andrew Young School of Policy Studies, Georgia State University, 14 Marietta Street NW, Suite 524, Atlanta, GA 30303, USA

ARTICLE INFO

Article history:
Received 24 July 2013
Received in revised form 25 June 2014
Accepted 26 June 2014
Available online 19 July 2014

Keywords:
Gompertz function
Logistic function
Model selection
t-test

ABSTRACT

This paper proposes a simple model selection test between the Gompertz and the Logistic growth models based on parameter significance testing in a comprehensive linear regression. Simulations studies are provided to show the accuracy of the method. Two real-data examples are also provided to illustrate the implementation of the proposed method in practice.

1. Introduction

Let Y_t be a time series taking nonnegative values. The Gompertz trend curve for Y_t is given by

$$Y_t = \alpha_1 exp\left(-\beta_1 e^{-\gamma_1 t}\right), \quad (1)$$

and the Logistic trend curve for Y_t is given by

$$Y_t = \alpha_2 \left(1 + \beta_2 e^{-\gamma_2 t}\right)^{-1}, \quad (2)$$

where t represents time and α_i, β_i, and γ_i, $i = 1, 2$, are positive parameters. Model (1) and Model (2), together with their multi-response and multivariate generalizations, are now widely used in applied research work for modeling and forecasting the behavior of many diffusion processes like the adoption rate of technology based products (Chu et al., 2009; Gamboa and Otero, 2009), population growth (Nguimkeu and Rekkas, 2011; Meade, 1988), and marketing development (Mahajan et al., 1990; Meade, 1984). In fact, the Gompertz and Logistic curves both share the interesting property that their "S-shaped" feature is suitable to describe processes that consist of a slow early adoption stage, followed by a phase of rapid adoption which then tails off as the adopting population becomes saturated. However, despite these visual and numerical similarities there are fundamental differences between the two curves and one of the most important is that the Gompertz function is symmetric whereas the Logistic function is asymmetric. Failing to account for these differences and choosing an inappropriate growth curve for inference can lead to seriously misleading forecasts (see Chu et al., 2009; Yamakawa et al., 2013 for some empirical illustrations). The need to develop a reliable selection procedure to discriminate between the two models in practice is therefore salient.

Unfortunately, despite the important request to select between these models in practice, there rarely exists a framework for statistical test between the two. The selection is usually made in an ad hoc basis using criteria based on forecasting errors, on the plausibility of the estimated saturation levels, or on visual evidence obtained from plotting the data in a special way, see for example, Gregg et al. (1964). A notable exception is the approach of Franses (1994) who proposed a selection based on statistical significance testing in an auxiliary regression which we briefly discuss in Section 2. Other approaches used are based on criteria of fitness that

☆ I would like to thank two anonymous referees, Associate Editor Yuya Kajikawa, and Daisuke Satoh, for useful comments.

* Tel.: +1 404 413 0162.
E-mail address: nnguimkeu@gsu.edu.

http://dx.doi.org/10.1016/j.techfore.2014.06.017

require to actually estimate the two models and then compare their fits with historical data through measures like R^2, root mean squared errors (*RMSE*), mean absolute percentage error (*MAPE*), root mean squared prediction errors (*RMSPE*) (see Chu et al., 2009; Yamakawa et al., 2013). Such a procedure is however not attractive as it requires to estimate both models by nonlinear regression methods involving numerical optimization which is usually computer expensive and time consuming. There is thus a clear need for selection methods between Gompertz and Logistic models which are easy to understand and inexpensive to compute. In this context, it seems natural to investigate the use of statistical tests that require simple estimation and easy computation.

This paper proposes a model selection test based on one linear regression and the significance test of one parameter. Our approach is therefore similar in spirit to the one proposed by Franses (1994) who also based their method to a single parameter significance testing. However, whereas the Franses (1994) method requires to primarily impute the original data in order to get only strictly positive increments of Y_t, our approach is based on the original responses themselves regardless of their values. Thus, there is no loss or distortion of information that could possibly undermine the result of our test which at the same time is more straightforward to compute. We examine the empirical size and power performance of the proposed test through Monte Carlo simulations and also provide real data examples to illustrate its usefulness in practice. The results show that the proposed test performs reasonably well in finite samples and could be a better alternative to the Franses' test.

In Section 2 we discuss the transformations of the Gompertz and Logistic curves leading to our selection procedure as well as the difference between our test and the Franses (1994) method. Section 3 provides numerical studies including Monte Carlo simulations and two real-data examples. Some concluding remarks are given in Section 4.

2. The selection procedure

Recall that Y_t is our variable of interest and denote by $y_t = (Y_t - Y_{t-1})/Y_{t-1}$ the relative increase in Y_t.[1] Let the Gompertz response function in Eq. (1) be denoted by $g(t) = \alpha_1 exp(-\beta_1 e^{-\gamma_1 t})$. Differentiating $g(t)$ and rearranging terms yield

$$\frac{g'(t)}{g(t)} = \gamma_1[ln\alpha_1 - lng(t)].$$

This suggests setting up a simple linear regression for the Gompertz model given in Eq. (1) with the form

$$y_t = \delta_1 + \rho_1 lnY_{t-1} + u_{1t}. \quad (3)$$

Likewise, if we denote by $h(t) = \alpha_2(1 + \beta_2 e^{-\gamma_2 t})^{-1}$ the Logistic response function in Eq. (2), a similar manipulation leads to the differential equation

$$\frac{h'(t)}{h(t)} = \gamma_2[\alpha_2 - h(t)].$$

Hence, a linear regression model of the form

$$y_t = \delta_2 + \rho_2 Y_{t-1} + u_{2t} \quad (4)$$

can be set up for the Logistic model given in Eq. (2). Testing Model (1) against (2) is therefore equivalent to testing Model (3) against (4). Models (1) and (2) as well as Models (3) and (4) are clearly nonnested in the sense of Cox (1961). For the latter models, it is desirable to use higher frequency data, if available, so that the first-derivative approximation by the difference score is more precise. However, regardless of the time frequency, the decision-rule provided by the test discussed below should not change, so long as one has enough data and one uses a definition of first-derivative that is consistent with the frequency of the data and is applied alike to both competitive models.

Following Davidson and MacKinnon (1981), an artificial comprehensive model can therefore be formulated as follows:

$$y_t = \delta + \gamma lnY_{t-1} + \theta Y_{t-1} + u_t, \quad (5)$$

where μ_t is an error term. It can be seen that when $\theta = 0$, Model (5) reduces to (3). Thus, it might seem that to test (3) against (4) we could simply estimate this model and test whether $\theta = 0$.[2] However, for the types of applications we consider here (i.e. growth curves), the series of interest, $\{Y_t\}$, will usually display an upward trend with no tendency of mean reversion, thus implying that they are non-stationary (Franses, 1998, pp. 67-68). Estimation of Model (5) using ordinary least squares might then lead to a spurious regression with an inconsistent estimate of θ (see, e.g., Hamilton, 1994, pp. 557-562, for a thorough discussion). In order to avoid spurious regressions, the simplest and most recommended way to base a test on Model (5) is to estimate a differenced version of it given by[3]

$$\Delta y_t = \mu + \gamma \Delta lnY_{t-1} + \theta \Delta Y_{t-1} + \epsilon_t, \quad (6)$$

where ϵ_t is the error term which can be assumed to be $NID(0, \sigma^2)$. We can estimate Model (6) by ordinary least squares and test the null hypothesis that $\theta = 0$ using an ordinary t-test for a desired significance level. This provides an easy and reliable way to test for Eq. (3). Alternatively, we could test for $\gamma = 0$, which would correspond to the logistic model given by Eq. (4). Since this can be done by simply interchanging the roles of the two models in all the following discussions, we focus on the former case in the rest of the paper, for brevity.[4] Note that the inclusion of the constant term μ is not strictly needed for the comprehensive specification of the

[1] One may instead consider $y_t = (Y_{t+1} - Y_t)/Y_t$ or the approximation $y_t = logY_t - logY_{t-1}$ and the results discussed would be similar.

[2] This idea is similar to the *J* test that was first suggested by Davidson and MacKinnon (1981) for nonnnested regressions.

[3] Although many researchers recommend routinely differencing non-stationary variables before estimating regressions, differencing may not be needed in the exceptional circumstance where the variables are cointegrated. In this case it is preferable to perform our selection test over Model (5), since differencing may cause a reduction of power (thanks to an anonymous referee for pointing this out). In our numerical applications, preliminary analysis has shown that they are not cointegrated. Cointegration tests are easy to perform and are available in most standard statistical software.

[4] Δ ln Y_{t-1} and Δ Y_{t-1} cannot be perfectly correlated (since one cannot be obtained as an affine transformation of the other). However, it is good practice to examine the magnitude of the squared correlation of the two regressors and verify that collinearity is not present in the model; otherwise simply testing individual coefficients might not be sufficient. In our empirical examples, no evidence of collinearity was found.