



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Computational Statistics & Data Analysis ■■■ (■■■■) ■■■–■■■

COMPUTATIONAL
STATISTICS
& DATA ANALYSISwww.elsevier.com/locate/csda

Performing hypothesis tests on the shape of functional data

Gareth M. James*, Ashish Sood

Marshall School of Business, University of Southern California, USA

Received 28 September 2004; accepted 25 February 2005

Abstract

We explore different approaches for performing hypothesis tests on the shape of a mean function by developing general methodologies both, for the often assumed, i.i.d. error structure case, as well as for the more general case where the error terms have an arbitrary covariance structure. The procedures work by testing for patterns in the residuals after estimating the mean function and are extremely computationally fast. In the i.i.d. case, we fit a smooth function to the observed residuals and then fit similar functions to the permuted residuals. Under the null hypothesis that the curve comes from a particular functional shape, the permuted residuals should have a similar distribution to the unpermuted ones. So the fitted curves will have the same distribution thus allowing significance levels to be computed very efficiently. In the more general case, when several curves are observed, one can directly estimate the covariance structure and incorporate this into the analysis. However, when only one curve is observed, we adopt a graphical approach where one plots the p -value for differing levels of potential complexity in the covariance structure. This allows one to judge the degree of deviation from the assumed null distribution. We demonstrate the power of these methods on an extensive set of simulations. We also illustrate the approach on a data set of technology evolution curves which current theory suggests should have an underlying S shape. The developed techniques have wide potential applications in empirical testing of the shape of functional data.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Functional hypothesis test; Permutation; Bootstrap.

* Corresponding author. Tel.: 213 740 9696; fax: 213 740 7313.

E-mail address: gareth@usc.edu (G.M. James).

1. Introduction

Suppose we observe curves or functions $Y_1(t), \dots, Y_N(t)$, and wish to perform hypothesis tests on the shape of their mean function, $\mu(t) = EY(t)$, when either one, or multiple curves are observed. In this paper, we develop general methodologies for performing these hypothesis tests.

First, we describe an application of the method that we will discuss in this paper. Consider the case of a manager faced with the issue of detecting common patterns from prior information on the evolution of a technology, and devising investment strategies for the future. The theory of S-curves from the technology management literature (Foster, 1986; Sahal, 1981; Utterback, 1994) suggests that all technologies evolve in the shape of an S-curve, i.e. when plotted against time, the performance of a technology exhibits an initial period of slow growth, followed by one of fast growth culminating in a plateau. The manager would be interested in finding out whether their particular technology evolves as expected i.e. in the form of an S-shaped curve. In Fig. 1, we plot the evolution of three technologies over time (Sood and Tellis, 2005). Fig. 1(a) plots the evolution of incandescent lighting over a 123 year period along with the best fitting S-curve. Though the monotonic growth in performance observed seems to roughly fit the expected S-shaped growth, a number of deviations are observed, and it is unclear how best to quantify the departure from the hypothesized shape. Similarly, Figs. 1(b) and (c), plot the evolution curves for cathode ray tube (CRT) display monitors and dot-matrix printers. While these plots also indicate some departures from the S-curve, the disparities do not appear to be nearly as large as those for incandescent lighting. In addition, the curves have been observed over a significantly shorter period of time, making it more difficult to assess whether the deviations are statistically significant. This example highlights the need to develop techniques to perform formal hypothesis tests on the shape of functional data. Some other examples where functional hypothesis testing is useful include testing the average shape of human growth curves, product life cycle, sales growth and new product diffusion curves, stock market returns or even weather patterns.

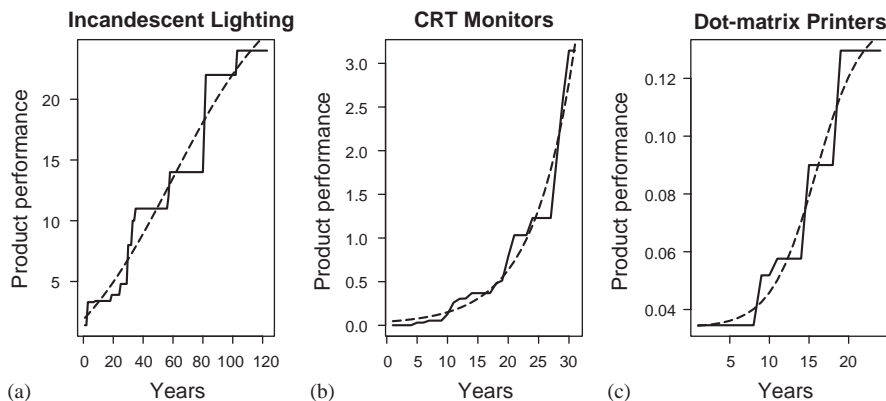


Fig. 1. Plots of the evolutions of three different technologies over time along with the best fitting “S” curves (dashed lines).

We develop a set of methods in this paper that allow one to answer these questions with a minimal set of general assumptions.

In the standard finite dimensional setting performing a hypothesis test on the mean of a population is a well-studied problem. However, the infinite dimensional functional case possesses additional difficulties. Dealing with simple hypotheses, where the null is completely specified, is comparatively easier to handle and offers many alternatives like Pearson's χ^2 test of goodness of fit (refer [Darling \(1957\)](#) and [Johnson and Kotz \(1970\)](#) for a survey of this field). However, the more common situation involving composite hypotheses, where some or all of the parameters of the functional form are left unspecified, is more challenging, and the existing methods pose many limitations. One option is to modify the classical Pearson χ^2 test of goodness-of-fit to this problem by estimating the unspecified parameters. However, the assumed asymptotic χ^2 distribution of the test statistic becomes questionable when the sample size is small. Other options like using the Kolmogorov–Smirnov test statistic pose additional computational difficulties because of the use of Monte-Carlo techniques ([Srinivasan, 1970](#)). Moreover, it may not be possible to adapt the test to some functional forms. Past research has also proposed semi-parametric forms of tests for such hypotheses ([Robinson, 1989](#)). The advantage of these tests is that they can be extended to cases where the functional form is unspecified even under the null, and hence are quite general in nature. However, the tests are dependent on a restrictive set of assumptions like stationarity of time series data, and pose serious limitations like lack of parametric estimates, dimensionality problems and sample size problems. Still other researchers have approached the problem from a Bayesian perspective ([Bewley and Griffiths, 2001](#); [Montgomery and Bradlow, 1999](#)). While these methods allow the inclusion of prior information especially when the sample size is small, it is difficult to assess the confidence that may be placed on the assumed model for the observed data.

When dealing with functional data, one must generally use a finite number of observations to try to make inferences about an infinite dimensional curve. One must necessarily assume some form of finite dimensional representation for the curves. Often this is achieved by placing a smoothness constraint on each curve ([James and Silverman, 2005](#)). A significant additional difficulty with testing the mean of functions such as those in [Fig. 1](#) involves the estimation of a covariance structure. One option often taken in the literature involves assuming i.i.d. measurement errors at the observed time points, e.g. [Bewley and Griffiths \(2001\)](#). This assumption simplifies the problem considerably and we begin by exploring this model. However, often in practice a more realistic model needs to allow for correlations in the error terms. We first develop an approach that assumes a general covariance structure and is applicable when several curves are observed from the same mean function. We then extend this approach to the more difficult, but rather common, situation where only one curve is observed. All three methods utilize permutation tests on the residuals between the observed curve and the estimated mean function. The i.i.d. approach computes a test statistic on the observed residuals and then permutes the residuals multiple times to compute a sample distribution under the null hypothesis. The general correlation approach works in a somewhat similar manner but makes adjustments for the estimated covariance structure. Our methods address the common difficulty of computational feasibility with functional data, which often contain a large number of observations. All three methods are extremely fast but have also demonstrated high power levels in simulation studies.

In Section 2, we develop our methods for performing hypothesis tests on $\mu(t)$. The first involves the situation where errors are assumed to be i.i.d. at the observation times. We call this approach “Functional IID Tests of Shape” or FITS. The other two methods allow for a far more general covariance structure in the errors and are called “Functional Arbitrary Covariance Tests of Shape” or FACTS. Different procedures are required in the cases where multiple curves are observed (FACTS_N) and the more difficult situation where only one curve is sampled (FACTS₁). The rest of the paper is organized as follows. Section 3 provides a detailed simulation study where we evaluate the performance of these methods under a number of scenarios. In particular, we examine the significance levels of the methods under the null hypothesis and their power under alternatives with varying levels of signal. These methods are then used to perform an empirical study on a data set of 20 technology evolution curves (Sood and Tellis, 2005) in Section 4. Finally, a general discussion of the methodology and their potential future extensions is provided in Section 5.

2. Methodology

Let $Y_1(t), \dots, Y_N(t)$ represent the observed values of N separate curves where the i th curve is measured at times t_{i1}, \dots, t_{ini} . Suppose we wish to test the null hypothesis

$$H_0 : EY(t) = \mu_0(t) \quad \text{for all } t$$

versus the alternative

$$H_A : EY(t) \neq \mu_0(t) \quad \text{for at least one } t.$$

Then we begin by modeling each curve using

$$Y_i(t) = \mu(t) + \varepsilon_i(t), \quad i = 1, \dots, N, \quad (1)$$

where $E\varepsilon(t) = 0$. In Section 2.1, we explore the FITS methodology for the simpler situation where the ε_i 's are taken to be uncorrelated measurement errors so that

$$\text{Cov}(\varepsilon(s), \varepsilon(t)) = \begin{cases} \sigma^2, & s = t, \\ 0 & s \neq t. \end{cases} \quad (2)$$

In Sections 2.2 and 2.3 we extend these methodologies to the more general FACTS situation where the ε_i 's are assumed to consist of a combination of smooth systematic deviations from the mean function and uncorrelated measurement errors. In this case

$$\text{Cov}(\varepsilon(s), \varepsilon(t)) = \Gamma(s, t) + \begin{cases} \sigma^2, & s = t, \\ 0 & s \neq t, \end{cases} \quad (3)$$

where $\Gamma(s, t)$ is a suitably smooth function of s and t . Computer code for the FITS and FACTS procedures can be obtained from www-rcf.usc.edu/~gareth. It should be noted that none of the algorithms in this paper require that the curves be measured at the same time points.

2.1. Functional IID tests of shape (FITS)

Let $Y_{ij} = Y_i(t_{ij})$, $\mu_{ij} = \mu(t_{ij})$ and $e_{ij} = \varepsilon_i(t_{ij})$. Then under the restricted covariance structure given by (2), we can express (1) in the form

$$Y_{ij} = \mu_{ij} + e_{ij}, \quad i = 1, \dots, N, \quad (4)$$

where the e_{ij} 's are i.i.d. with $Ee_{ij} = 0$ and $\text{Var}(e_{ij}) = \sigma^2$. Our approach to performing a hypothesis test is based on the following heuristic argument. Let $r_{ij} = Y_{ij} - \hat{\mu}_{0ij}$ be the residual at time t_{ij} where $\hat{\mu}_{0ij}$ is the best estimate for $\mu(t_{ij})$ under the null hypothesis. Let $s(t) = \mathbf{b}(t)^T \boldsymbol{\theta}$ be a q -dimensional function where $\mathbf{b}(t)$ is the basis function and $\boldsymbol{\theta}$ are the corresponding coefficients. We utilize a cubic B-spline basis which partitions the interval into a series of cubic functions joined at ‘‘knot points’’ (Hastie et al., 2001, Chapter 5). However, in general we have found these types of methods to be robust to the exact choice of basis function. One can use any finite dimensional basis provided it contains enough flexibility to model whatever pattern the residuals may display. Suppose that we choose $s(t)$ so as to provide the best fit, in the least-squares sense, to the residuals, r_{ij} . Then under the null hypothesis it should be the case that $s(t) \approx 0$ or correspondingly that $T = \sum \theta_l^2 \approx 0$. In addition, under the null hypothesis, the residuals will be approximately i.i.d. Hence if we permute the ordering of the residuals, refit $s(t)$ and compute the sum of the squared basis coefficients, $T^{(b)}$, then it should be the case that T and $T^{(b)}$ will have approximately the same distribution. Alternatively, if the null hypothesis is false there will be a systematic pattern in the residuals, causing T to increase but having less effect on $T^{(b)}$ because of the randomizing of the residuals. Hence we reject H_0 if T is larger than a significant majority of the $T^{(b)}$'s. Below we formally outline the algorithm.

2.1.1. FITS algorithm

1. Compute $\hat{\mu}_0$, the least-squares estimate of μ under the null hypothesis.
2. Compute the residuals $r_{ij} = Y_{ij} - \hat{\mu}_{0ij}$.
3. Fit $s(t) = \mathbf{b}(t)^T \boldsymbol{\theta}$ to the residuals via least squares $\hat{\boldsymbol{\theta}} = (X^T X)^{-1} X^T \mathbf{r}$, where X is a basis matrix with rows corresponding to $\mathbf{b}(t_{ij})^T$.
4. Compute the test statistic $T = \sum_{l=1}^q \hat{\theta}_l^2$.
5. Pool the residuals over all curves, randomize the r_{ij} 's and refit $s(t)$ to the permuted residuals to obtain $\hat{\boldsymbol{\theta}}^{(b)}$.
6. Compute $T^{(b)} = \sum_{l=1}^q \hat{\theta}_l^{(b)2}$.
7. Repeat steps 5 and 6 to obtain $T^{(1)}, \dots, T^{(B)}$.
8. The estimated p -value corresponds to $(1/B) \sum_{b=1}^B I(T \leq T^{(b)})$.

Note that the FITS algorithm can be applied to either a single curve or to multiple curves. Fig. 2 illustrates the various components of this procedure, using a null hypothesis that $\mu(t)$ corresponds to a logistic curve, on three different simulated data sets. The first column relates to data generated from the hypothesized logistic curve plus i.i.d. error sampled at 200 time points. Fig. 2(a) plots the raw data along with the true mean that they were

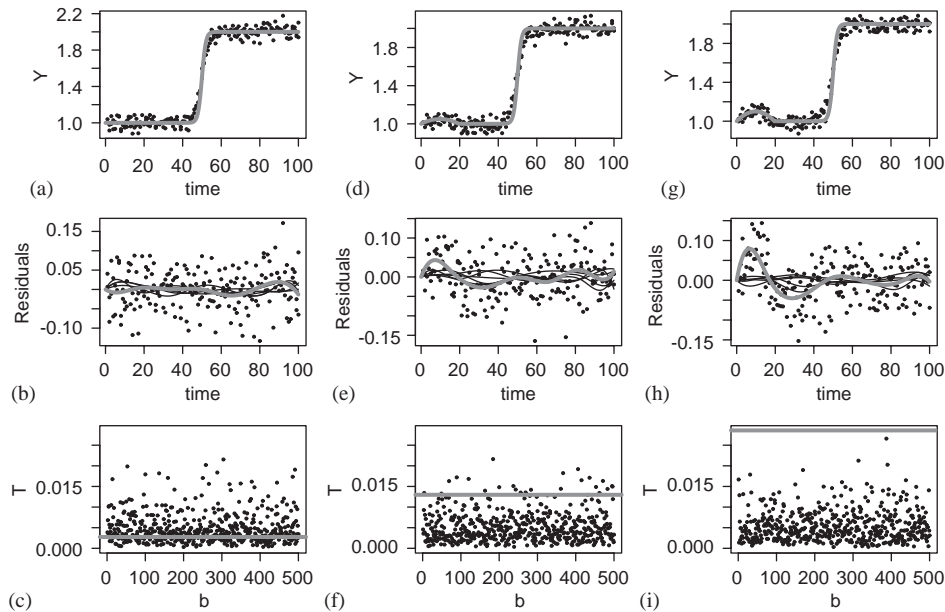


Fig. 2. Plots of FITS methodology applied to three different simulated data sets.

generated from. Fig. 2(b) shows a plot of the residuals. The grey line through the centre is the least-squares fit of a smooth regression spline to the residuals while the five black lines are the corresponding fits resulting from permuting the ordering of the residuals five times. Notice that the curves are all close to zero. Fig. 2(c) plots the values of $T^{(b)}$ generated from 500 random permutations of the residuals along with a grey horizontal line which corresponds to the observed value of T . As we would hope, T is near the middle of the $T^{(b)}$'s with a p -value of approximately 70%. There is no evidence here to reject the null hypothesis. The second column contains data that has been very slightly perturbed from a logistic curve between times 1 and 20. Even though the perturbation is slight, the fitted spline in Fig. 2(e) contains significantly more pattern than most of those generated from the permuted residuals. This difference is evident in Fig. 2(f) where T is significantly increased from Fig. 2(c) giving a p -value around 5%. Finally, the last column further increases the perturbation, which induces a value of T larger than all 500 $T^{(b)}$'s. Notice that the perturbations have no noticeable effect on the $T^{(b)}$'s but cause T to increase rapidly. Ultimately, the usefulness of a hypothesis test depends on its ability to correctly specify the significance level under the null hypothesis and to produce high power under the alternative. In Section 3, we demonstrate that the FITS method fulfills both criteria.

2.1.2. Adjustment for correlation in residuals

The FITS algorithm does not account for the correlation in the residuals that may exist even if the underlying error terms are independent. For example, in a standard linear

regression the residuals are generally correlated with each other as a consequence of the least-squares fitting procedure. Hence, the permuted and unpermuted residuals may have slightly different distributions even if the null hypothesis is correct. A simple addition can be made to the algorithm to adjust for this effect. We replace step 8 in the FITS algorithm by the following:

- 8a. Generate pseudoerrors according to an i.i.d. Gaussian distribution with mean zero and standard deviation estimated from the observed residuals.
- 8b. Fit a simple linear regression to the pseudoerrors and record the resulting pseudo-residuals.
- 8c. Calculate the test statistic T_p based on fitting a spline to the pseudo-residuals.
- 8e. Randomize the pseudo-residuals to calculate $T_p^{(b)}$ and record the difference $d^{(b)} = T_p - T_p^{(b)}$.
- 8f. Repeat steps 8a through 8e B times and let $\bar{d} = \sum_b d^{(b)} / B$.
- 8g. The estimated p -value corresponds to $(1/B) \sum_{b=1}^B I(T - \bar{d} \leq T^{(b)})$.

This addition estimates the average increase in T over $T^{(b)}$ i.e., \bar{d} , that one might expect as a result of correlation in the residuals. The observed value of T is then reduced by \bar{d} to adjust for any correlation effect. We induce correlation in the residuals using a simple linear regression because it can be computed extremely efficiently. However, one could also generate pseudoobservations and re-estimate the entire mean curve to produce the residuals. In practice, we have found that the correlation effect is too low to cause problems unless the function is sparsely observed.

2.2. Functional arbitrary covariance tests of shape : multiple curves (FACTS_N)

The i.i.d. error model given by (2) provides a simple and often effective means to test for a given mean structure. However, an i.i.d. assumption may be unrealistic in a number of real-world situations because functions observed over time tend to have correlated error structures. A few examples include the evolution of technologies as discussed in the introduction, product life cycle, sales growth and new product diffusion curves as well as human growth curves. In this section, we develop an approach that assumes a general covariance structure and uses the observations from N different curves, all assumed to have common mean, to estimate the covariance effect. In Section 2.3, we deal with the situation where only one curve is observed. Let $\gamma_1(t), \dots, \gamma_N(t)$ be a sequence of random curves with $E\gamma(t) = 0$ and $Cov(\gamma(s), \gamma(t)) = \Gamma(s, t)$. Then under the general covariance structure given by (3), we can express (1) in the form

$$Y_{ij} = \mu_{ij} + \gamma_{ij} + e_{ij}, \quad i = 1, \dots, N, \quad (5)$$

where $\gamma_{ij} = \gamma_i(t_{ij})$ and the e_{ij} 's are i.i.d. with $Ee_{ij} = 0$ and $Var(e_{ij}) = \sigma^2$. When using this model, the FITS approach where the residuals are permuted makes little sense because even under the null hypothesis the covariance structure is not preserved. Instead, we use an

alternative test statistic, namely,

$$T = \int (\hat{\mu}_0(t) - \hat{\mu}(t))^2 dt,$$

where $\hat{\mu}_0(t)$ is the least-squares estimate of μ under the null hypothesis and $\hat{\mu}(t)$ is also a smooth estimate of μ but without the constraint of the null hypothesis on its shape. To test the significance of T we use a bootstrap approach (Efron and Tibshirani, 1993). We first estimate the γ_i 's and compute $r_{ij} = Y_{ij} - \hat{\mu}_{ij} - \hat{\gamma}_{ij}$. To preserve the covariance structure in our bootstrapped sample we then take a sample, $\hat{\gamma}_1^{(b)}, \dots, \hat{\gamma}_N^{(b)}$, permute the residuals, r_{ij}^* , and let $Y_{ij}^{(b)} = \hat{\mu}_{ij} + \hat{\gamma}_{ij}^{(b)} + r_{ij}^*$. Under the null hypothesis the $Y_{ij}^{(b)}$ should have a similar distribution to the observed data. The bootstrapped test statistics are then computed using

$$T^{(b)} = \int (\hat{\mu}(t) - \hat{\mu}^{(b)}(t))^2 dt,$$

where $\hat{\mu}^{(b)}(t)$ is a smooth estimate of μ based on the $Y_{ij}^{(b)}$'s. Under the null hypothesis, $\hat{\mu}(t)$ will be similar to $\hat{\mu}_0(t)$ and $T^{(b)}$ will have approximately the same distribution as T . Finally, we compare $T^{(1)}, \dots, T^{(B)}$ to T to obtain a p -value. Below we provide details of the algorithm.

2.2.1. FACTS_N algorithm

1. Compute $\hat{\mu}_0$, the least-squares estimate of μ under the null hypothesis, using all N curves Y_1, \dots, Y_N .
2. Estimate $\hat{\mu}$ using a single smooth fit to Y_1, \dots, Y_N . Note that H_0 is not assumed to calculate $\hat{\mu}$.
3. Calculate $T = \int (\hat{\mu}_0(t) - \hat{\mu}(t))^2 dt$.
4. Estimate the functions $\hat{\gamma}_i(t)$ using smooth fits to $Y_{ij} - \hat{\mu}_{ij}$ for $i = 1, \dots, N$, where $\hat{\mu}_{ij} = \hat{\mu}(t_{ij})$.
5. Estimate the residuals $r_{ij} = Y_{ij} - \hat{\mu}_{ij} - \hat{\gamma}_{ij}$, where $\hat{\gamma}_{ij} = \hat{\gamma}_i(t_{ij})$.
6. Resample the $\hat{\gamma}_i$'s with replacement to produce the bootstrapped sample $\hat{\gamma}_1^{(b)}, \dots, \hat{\gamma}_N^{(b)}$.
7. Permute the residuals to produce r_{ij}^* .
8. Let $Y_{ij}^{(b)} = \hat{\mu}_{ij} + \hat{\gamma}_{ij}^{(b)} + r_{ij}^*$.
9. Calculate $T^{(b)} = \int (\hat{\mu}(t) - \hat{\mu}^{(b)}(t))^2 dt$, where $\hat{\mu}^{(b)}(t)$ is a smooth fit to the $Y_{ij}^{(b)}$'s.
10. Repeat steps 6–9 B times to obtain $T^{(1)}, \dots, T^{(B)}$.
11. The estimated p -value corresponds to $(1/B) \sum_{b=1}^B I(T \leq T^{(b)})$.

Note that in step 9 we use $\hat{\mu}(t)$ rather than $\hat{\mu}_0(t)$ because otherwise when H_0 is false $T^{(b)}$ would still have a similar value to T and hence the test would have no power. To implement this algorithm in practice, we approximate the integrals in steps 3 and 9 by evaluating the integrands over a fine grid of time points ranging from the smallest to the largest t under

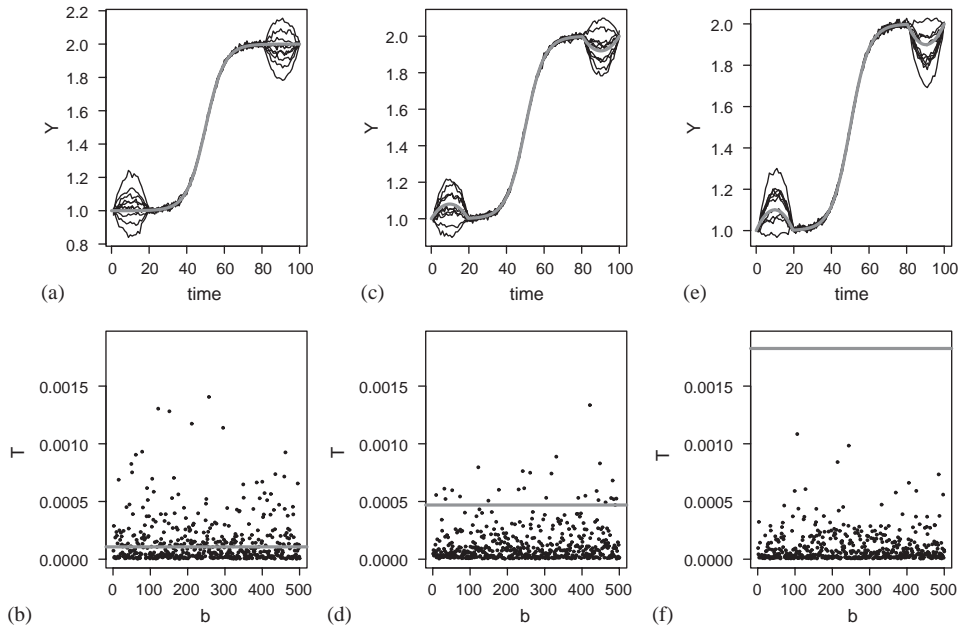


Fig. 3. Plots of FACTS_N methodology applied to three different simulated data sets.

consideration. Both $\hat{\mu}(t)$ and the $\hat{\gamma}_i(t)$'s are estimated using smoothing splines (Hastie et al., 2001, Chapter 5) which are high dimensional cubic splines that have been regularized using a penalty term on the squared second derivative.

Fig. 3 illustrates this approach on three simulated data sets with non-independent covariance structures. Each data set contains ten curves sampled from a common mean function plus curve-specific γ_i 's and i.i.d. measurement error. All three data sets have the same covariance structure. The curves in Fig. 3(a) were generated from a common logistic mean function which is shown in grey. The covariance structure induced by the γ_i 's means that each of the individual curves differs in shape from a logistic. However, the average over all curves does have the correct form so the null hypothesis is satisfied for this data. Fig. 3(b) illustrates the value of the test statistic in grey along with 500 values of $T^{(b)}$. As we would hope, given the null hypothesis is true, there is no evidence to reject H_0 with a p -value of only 0.414. Alternatively, Fig. 3(c) illustrates data generated from a mean function that has been slightly perturbed from a logistic, as shown by the grey curve. Here the null hypothesis is false but the signal is fairly low relative to the covariance in the γ_i 's. Despite the low signal-to-noise (SN) ratio the test statistic, T , illustrated by the grey line in Fig. 3(d), has increased markedly and the corresponding p -value has declined to 0.05. Finally, Fig. 3(e) and (f) illustrate a slightly higher SN ratio. The p -value for this data is zero. Notice that in all three cases the distribution of the $T^{(b)}$'s is very similar, but the value of T increases with larger departures from the logistic mean curve.

2.3. Functional arbitrary covariance tests of shape: one curve (FACTS₁)

The methodology outlined in the previous section works well when there are enough curves that one can reasonably hope to produce an accurate estimate of the underlying covariance structure using a bootstrap analysis. However, for certain types of data one may wish to infer the underlying shape of the mean function based on observing only one curve. An example of this situation is the technology evolution curves illustrated in Fig. 1. In this case a bootstrap analysis of the γ_i 's must fail because one can estimate at most one covariance term. One option involves the approach taken in Section 2.1 where we assumed an i.i.d. covariance structure but, as mentioned previously, such an assumption may be overly simplistic. Of course when only one curve is observed there is a fundamental identifiability problem between the mean curve, μ , and the covariance, γ . Consider, for example, the data in Fig. 2(g). This data clearly suggests a deviation from a logistic curve for the early time periods. However, based on one curve we cannot identify whether this effect is caused by the mean curve or whether it is simply a result of the covariance structure as in Fig. 3. In other words, if we had observed multiple curves perhaps, the difference from a logistic curve would have evened out.

Nevertheless, one can still ask “How complicated would the covariance structure need to be to explain this large a deviation from a logistic curve?” For example, it may be the case that for a given curve using, for example, the i.i.d. FITS methodology, it is possible to reject a logistic curve hypothesis. But after including a very simple covariance structure in the model it is no longer possible to reject the null hypothesis. In this situation, one could not state with a high level of confidence that the mean curve differed from a logistic. Alternatively, if you were still able to reject the null hypothesis even after allowing for a complicated covariance structure one would state with more confidence that the mean curve seemed to differ from a logistic.

We use the following approach to implement this strategy. We first compute $\hat{\mu}_0$, the least-squares estimate of μ under H_0 . We then obtain a “low complexity” estimate of the covariance by fitting a smooth function, γ , with given “equivalent degrees of freedom” (edf) (Green and Silverman, 1994) to $Y_j - \hat{\mu}_{0j}$. (Note we have dropped the subscript i because here $N = 1$.) The concept of edf was developed to measure the flexibility of a given functional form. It is easy to measure the flexibility of simple functions such as an n th degree polynomial which has $n + 1$ parameters and hence $n + 1$ degrees of freedom (dof). In this case, the edf and the dof coincide. However, directly calculating dof of more complicated functions such as a smoothing spline (which involves many parameters but also a penalty term to ensure smoothness) is more difficult. In these cases, we use edf which gives the same answer as dof in simple cases but generalizes to more complicated situations. We calculate the residuals by subtracting the estimated values of μ and γ from the observed data. Then we calculate a p -value by using the FITS procedure applied to the residuals we have just computed. This entire procedure is repeated for increasingly complex covariance structures. Finally, we plot the different p -values versus the equivalent dof of γ . A curve that has low p -values even for high dof in γ strongly suggests a true deviation from the null hypothesis. However, one that does not have any low p -values or only low values for very simple γ curves provides no significant evidence of a deviation from the null hypothesis. Below we provide details of the algorithm.

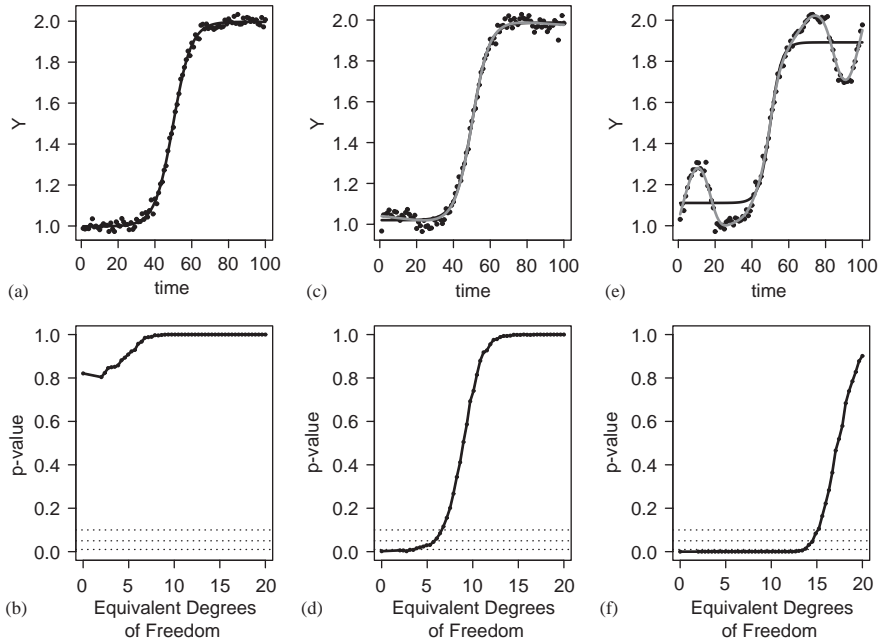


Fig. 4. Three different simulated data sets and their corresponding p -value curves for different levels of flexibility in the covariance structure. The black lines in (a), (c) and (e) correspond to the best fitting logistic function. The grey lines illustrate the least flexible γ curve that would explain the deviations from a logistic curve.

2.3.1. FACTS₁ algorithm

1. Compute $\hat{\mu}_0$, the least-squares estimate of μ under the null hypothesis.
2. Select a possible range of equivalent degrees of freedom for γ , $\text{edf}^1, \dots, \text{edf}^K$.
3. Estimate $\hat{\gamma}^{(\text{edf}^l)}$ by fitting a smooth function with equivalent degrees of freedom no greater than edf^l to $Y_j - \hat{\mu}_{0j}$.
4. Compute the residuals $r_j = Y_j - \hat{\mu}_{0j} - \hat{\gamma}_j^{(\text{edf}^l)}$
5. Perform steps 3–8 of the FITS algorithm to obtain the p -value, p^l .
6. Repeat steps 3–5 for $l = 1, \dots, K$ to obtain p^1, \dots, p^K .
7. Plot p versus edf .

As with the FACTS_N algorithm, we estimate $\hat{\gamma}$ using a smoothing spline. We illustrate the algorithm on three simulated data sets in Fig. 4. The data in Fig. 4(a) was generated from a logistic curve plus i.i.d. errors. In Fig. 4(b) we have plotted the p -values as a function of the various different equivalent degrees of freedom (edf) used to generate the γ curves. Here zero edf corresponds to the FITS procedure from Section 2.1 while two edf corresponds to a fit involving a γ curve that is restricted to be linear. As we would hope, given that this data was generated from a logistic curve with i.i.d. errors, there is no evidence to reject the null hypothesis for any value of edf. The data in Fig. 4(c) is generated from a mean

curve with a slight perturbation from a logistic. We now see in Fig. 4(d) that there is clear evidence to reject the null hypothesis when edf equals zero (the i.i.d. case). However, the p -value increases rapidly even with relatively little flexibility in the γ curve. The grey line in Fig. 4(c) illustrates the mean curve plus the least flexible γ such that we can no longer reject the null at the 5% significance level. Notice that only a very slight adjustment to the mean curve is required to explain the deviation from a logistic. All this suggests that even a relatively simple covariance structure accounts for the observed effect. In other words, the evidence to reject a logistic distribution is weak. Finally, the data in Fig. 4(e) exhibits significant deviation from a logistic function. Correspondingly, Fig. 4(f) shows very low p -values even for complicated covariance structures which suggests strong evidence that the data does not have a logistic mean function.

3. Simulation study

In this section, we present results from an extensive set of simulations designed to test the true significance levels, under the null, and power, under the alternative, of the three methodologies from Section 2.

3.1. FITS results

For the FITS simulation, we generated data sets using the i.i.d. error distribution illustrated in Fig. 2. These data come from a mean function which is a logistic curve of the form

$$\mu_L(t) = a + \frac{b}{1 + \exp(-c(t - d))}, \quad (6)$$

where $a = b = 1$, $c = 0.5$ and $d = 50$ plus a perturbation of the form $At(20 - t)$, $0 \leq t \leq 20$. For $A = 0$ this gives a pure logistic, while as A increases, the deviation from a logistic curve becomes more pronounced. In addition, Gaussian noise with a standard deviation of 0.05 was added to each measurement. A total of 200 observations from 100 equally spaced time points between 1 and 100 were generated for each data set. To perform the simulation, we chose a range of values for A starting at zero. For each value of A , 200 data sets were generated. We then applied the FITS procedure from Section 2.1 to each of the data sets to calculate a p -value. An important part of this approach is the choice of the basis for $s(t)$, the curve that is fit to the residuals. We used three different B-spline bases with 5, 8 and 14 degrees of freedom. Larger degrees of freedom correspond to more knots and hence greater flexibility.

The results are summarized in Fig. 5. Figs. 5(a)–(c) are plots of the proportion of p -values less than 1%, 5% and 10%, respectively, as a function of the SN ratio. We calculated the SN ratio by taking the average deviation over time of the mean curve from a logistic function divided by the standard deviation of the error terms after adjusting for the fact that there are multiple observations at each time. A SN ratio of zero corresponds to the null hypothesis so, ideally the curves should respectively take on values around 1%, 5% or 10% at this point, but then rapidly increase towards one for larger ratios. In fact, we see that, for all three basis choices, at zero the methodology is if anything producing slightly conservative p -values.

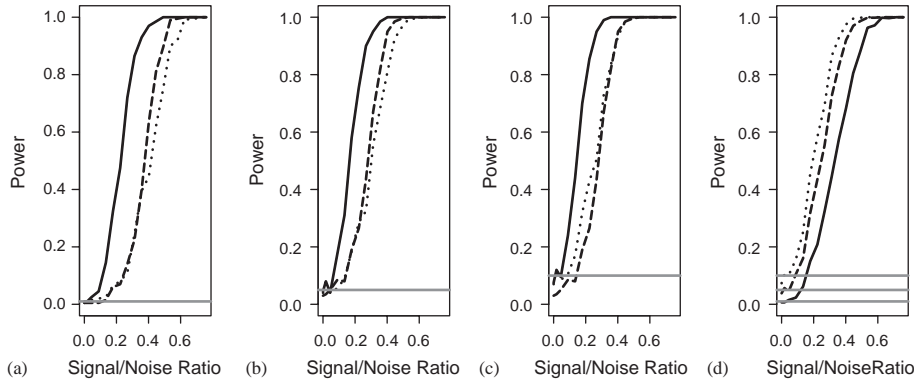


Fig. 5. Plots of power as a function of signal to noise ratio using the FITS methodology. (a)–(c) respectively illustrate 1%, 5% and 10% significance levels corresponding to fits using bases with 5 (—), 8 (---), and 14 (···) degrees of freedom. (d) illustrates the average power for the 1% (—), 5% (---), and 10% (···) significance levels. Grey horizontal lines plot the corresponding 1%, 5% or 10% values.

However, the power increases rapidly as the signal increases with, for example, the power with a basis of five degrees of freedom and significance level of 5% reaching 1 with a signal to noise ratio of only 0.4. All three bases exhibit a similar pattern, but it is clear that the basis with 5 dof dominates those with 8 and 14. It is interesting to note that the increased power is not at the expense of a misspecified significance level under the null. The other two bases gave similar performance to each other. Fig. 5(d) plots the average power over the three bases for each of the three different significance levels to provide a direct comparison of the effect of significance on power.

3.2. FACTS_N results

To test the performance of the FACTS_N approach, we generated data from the distribution illustrated in Fig. 3. The data consisted of a logistic curve given by (6) with $a = b = 1$, $c = 0.2$ and $d = 50$ plus perturbations of the form $At(20 - t)$, $0 \leq t \leq 20$ and $-At(100 - t)$, $81 \leq t \leq 100$. In addition, Gaussian noise with standard deviation 0.01 was added to each measurement. For each data set we generated ten curves with a different, random, value of A for each one. The A 's were generated from a Gaussian distribution with mean m and a variance of one. Data sets with $m = 0$ corresponded to the null hypothesis because the average over all curves had a logistic shape. Alternatively, as m grew, there was an increasing deviation from a logistic mean curve. One key component of the FACTS_N approach is the flexibility of the unconstrained estimate of the mean function. We used smoothing splines with edf of 8, 10 and 12 on each data set to test the sensitivity of the results to this parameter. As with the previous simulation, we generated 200 data sets for each one of a range of values for m , beginning with $m = 0$.

The results are summarized in Fig. 6. Figs. 6(a)–(c) are plots of the proportion of p -values less than 1%, 5% and 10%, respectively, as a function of the SN ratio. The SN ratio was calculated in a similar fashion to that for the previous FITS simulation. Note that the SN

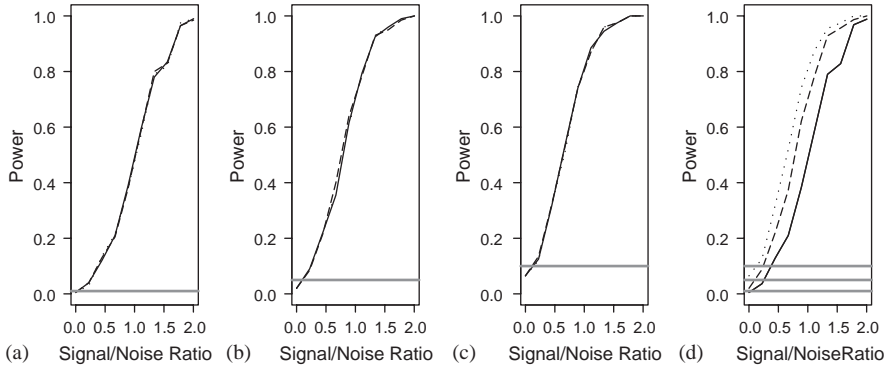


Fig. 6. Power plots using the FACTS_N methodology. (a)–(c), respectively, illustrate 1%, 5% and 10% significance levels corresponding to fits using smoothing splines with equivalent degrees of freedom of 8 (—), 10 (---), and 12 (\cdots). (d) illustrates the average power for the 1% (—), 5% (---), and 10% (\cdots) significance levels.

ratio is measuring signal per curve rather than total signal over all N curves. This makes the results comparable to those for FITS and FACTS_1 . The first point we note is that there is very little difference between the fits using 8, 10 or 12 effective degrees of freedom with all three curves almost identical. This is comforting because it suggests a low sensitivity of the results to any reasonable fit for the mean function. Secondly, the true significance levels under the null hypotheses, i.e. when the SN ratio is zero, are equal to the nominal 1% significance level and slightly conservative for the 5% and 10% levels. As with the FITS methodology considered in the previous simulation, there is a steady increase in power as the SN ratio increases. It is noticeable that the FACTS_N SN ratio needs to be approximately four times as large as the corresponding FITS SN ratio to achieve the same level of power. This difference is not surprising given that the general case requires estimation of the γ 's i.e. the covariance structure while the i.i.d. case does not. However, it does provide useful information about the increased strength in the signal that may be needed before one can remove the i.i.d. assumption and still reasonably hope to obtain significant results. Finally, Fig. 6(d) gives a direct comparison of the power levels for the three different significance levels.

3.3. FACTS_1 results

Finally, we tested the performance of the FACTS_1 methodology where a general covariance structure is assumed but only a single curve is observed. The data sets were generated in an almost identical fashion to those illustrated in Fig. 4 and described for the previous simulation. The only differences were that one rather than ten curves was generated, a standard deviation of 0.02 was used for the measurement error and A was chosen as a fixed rather than a random value. As in the previous simulations, a grid of values for A was chosen and 200 data sets were generated for each value. The FACTS_1 methodology was then applied to each data set. As with the first simulation, three different bases were tested for $s(t)$ corresponding to splines, with $q = 5, 8$ and 14 degrees of freedom. For each data

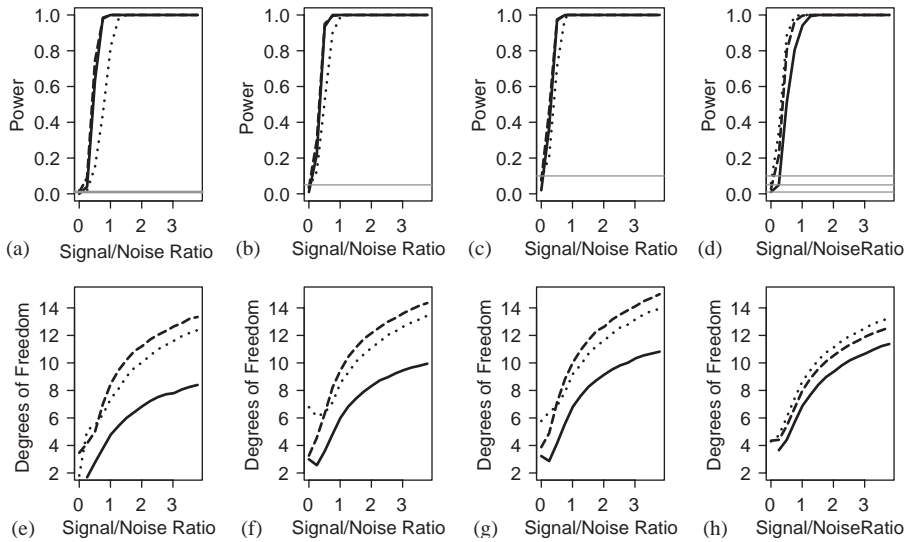


Fig. 7. (a)–(d) Power plots using the FACTS_1 methodology and (e)–(h) corresponding plots of degrees of freedom. (a)–(c) and (e)–(g), respectively, illustrate 1%, 5% and 10% significance levels corresponding to fits using bases with 5 (—), 8 (---), and 14 (···) degrees of freedom. (d) and (f) illustrate overall averages for the 1% (—), 5% (---), and 10% (···) significance levels.

set, we recorded whether there was significant evidence to reject the null at the 1%, 5% and 10% significance levels as well as the least flexible γ such that we could no longer reject the null at each significance level.

The results are summarized in Fig. 7. Figs. 7(a)–(c), respectively, give the fraction of data sets for which the null hypothesis was rejected for at least one choice of edf on the γ 's for significance levels of 1%, 5% and 10%. As with the other simulations, the observed significance levels are approximately the same as the nominal ones and the power increases rapidly with the SN ratio. For any particular SN ratio, the power is slightly lower than that for FITS but considerably higher than for FACTS_N . In addition, for this data, the $q = 5$ and $q = 8$ curves both give similar power while the $q = 14$ curve is clearly inferior. The relatively superior performance of the methodology with $q = 8$ is likely a result of the added complexity of the deviations of curves from a logistic in this data relative to that in the first simulation (compare Fig. 2 with Fig. 4). This suggests that ideally one should use more flexible bases for $s(t)$ when the underlying curves have a more complicated structure to them. Fig. 7(d) provides the average power level as a function of SN ratio for the three different significance levels.

Figs. 7(e) through (h) provide identical information to (a) through (d) except that they plot the average minimum degrees of freedom for the γ curve, among all data sets with a significant p -value, such that the null hypothesis was no longer rejected for each significance level. For example, a value of ten would indicate that a γ curve with at least 10 dof was required to explain the departure from a logistic curve. Even after the power levels

have reached 100%, the minimum degrees of freedom continue to rise with the SN ratio. Interestingly, the highest dof are obtained with $q = 8$ rather than $q = 5$ providing further evidence for the need for a somewhat more flexible spline $s(t)$ when the underlying curves are more complicated.

4. Empirical study

In this section we revisit the application problem discussed in Section 1 regarding testing whether technologies evolve in the shape of S-curves. Frequently, in firms from high-technology industries, managers are involved in making decisions regarding allocation of funds for research and development on various existing and emerging technologies. In such cases, they need to understand the patterns of technological evolution in order to predict the future growth of existing and new technologies to maximize their competitive advantage and return on investments. As stated in the introduction, theory from the technology management literature suggests that all technologies evolve in the shape of an S-curve. If true, this theory provides a good benchmark for estimating the ultimate decline of existing technologies, and for predicting the growth pattern of emerging technologies. However, in reality, some technologies seem to evolve through irregular step improvements instead of a continuous S-shape, and the problem is to test whether an observed pattern is from an S curve.

Sood and Tellis (2005) test this theory on 20 technologies from six markets—external lighting, desktop memory, display monitors, desktop printers, data transfer and analgesics. Using historical analysis techniques, they plotted the improvement in performance of all these technologies from the year they were first introduced in commercial applications. Their research cast doubt that these curves evolved according to a single S curve, instead suggesting that a series of irregular step functions may more accurately reflect the true structure. However, Sood and Tellis did not use formal hypothesis tests. Instead, they computed the MSE from a fits of a logistic function to a subset of the observations versus the MSE from a fit to all the observations and concluded that an S curve fits better over a subset rather than the full data if the latter value was greater. We are interested here in reexamining their data using our more formal hypothesis testing methodology. A logistic curve of the form given by (6) was used to model the S curve in Sood and Tellis (2005), where the parameters a , b , c and d were estimated from the data using non-linear least squares. We take the same approach so our null hypothesis is that $\mu(t) = \mu_L(t)$ with an alternative that $\mu(t) \neq \mu_L(t)$. Since a single curve was observed for each technology we applied the FACTS₁ approach to each of the 20 technologies. We wish to choose q so as to maximize the power of our hypothesis tests. In general, as q increases one can detect finer departures from the null hypothesis and hence increase the power. However, for curves sampled at a small number of time points using a large value of q will tend to overfit the data and can actually decrease the power. After some experimentation, we opted to use $q = 5$ for technologies measured at 30 or fewer points, $q = 12$ for those measured at 70 or more points and $q = 8$ for all others.

Using this approach, eight of the technologies showed significant departures from an S curve at 1% or 5% significance levels. We have summarized the results for these eight

Table 1

Average minimum equivalent degrees of freedom for γ required before we fail to reject the hypothesis of an S shaped mean curve for each the eight technology evolution curves that showed significant departures from the null hypothesis

| Market | Technology | p -values | | | |
|------------------|------------------|-------------|------|------|------|
| | | 0.01 | 0.05 | 0.1 | 0.2 |
| Lighting | Incandescent | 15.2 | 16.0 | 16.7 | 17.8 |
| | Arc-discharge | NA | 4.6 | 6.4 | 7.5 |
| | Gas-discharge | 5.7 | 10.5 | 10.8 | 11.9 |
| | LED | 5.7 | 7.9 | 8.3 | 8.6 |
| Display monitors | Cathode ray tube | 6.4 | 9.0 | 10.5 | 12.3 |
| Printers | Inkjet | NA | 4.0 | 5.0 | 5.9 |
| Data Transfer | Copper | NA | 2.0 | 4.9 | 6.0 |
| | Wireless | 5.3 | 6.0 | 7.2 | 7.5 |

technologies in Table 1. The values in the table correspond to the minimum degrees of freedom required in the γ curve, used to generate the covariance structure, such that the null hypothesis was no longer rejected at the given significance level. A value of NA indicates that the null hypothesis was not rejected for any level of dof. Notice that certain markets appear to exhibit stronger departures from an S curve than others. Four of the five lighting technologies and two of the three data transfer technologies showed strong evidence that they did not come from an S curve. However, only one each of the display monitor and printer technologies and none of the desktop memory or analgesics technologies showed significant departures from an S curve even at a 10% significance level. The incandescent lighting shows the most significant departures from the null hypothesis with p -values less than 1% for all edf up to 15. Fig. 8(a) illustrates the incandescent lighting data. The grey line indicates the least complicated γ curve for which we would no longer reject the null hypothesis at the 1% significance level. Fig. 8(b) plots the corresponding p -values vs. edf. Based on Figs. 8(a) and (b) there is strong evidence for a step function shape rather than a single S curve. In comparison the arc-discharge lighting data, illustrated in Fig. 8(c), while still significant for low edf, has p -values which climb rapidly towards 1 beginning at about 5 edf (see Fig. 8(d)). This indicates that the departures from an S curve could be explained using a moderately simple covariance structure. We also provide the data for the cathode ray tube monitors, Fig. 8(e), (moderate evidence) and dot-matrix printers, Fig. 8(g), (weak evidence) for comparison.

In comparing with the results of Sood and Tellis (2005) we find that 12 of the 20 technologies produced identical conclusions. Seven of the remaining eight technologies differed in that Sood and Tellis concluded that an S curve did not fit the data while our FACTS₁ approach failed to reject this hypothesis. The differences are not surprising since our formal hypothesis test requires a stronger standard of evidence to reject the null than the more exploratory approach of Sood and Tellis. Most of the data sets for which differences in the two methods were noted contained relatively few years of observations, less than 30, making it more difficult to prove the statistical significance.

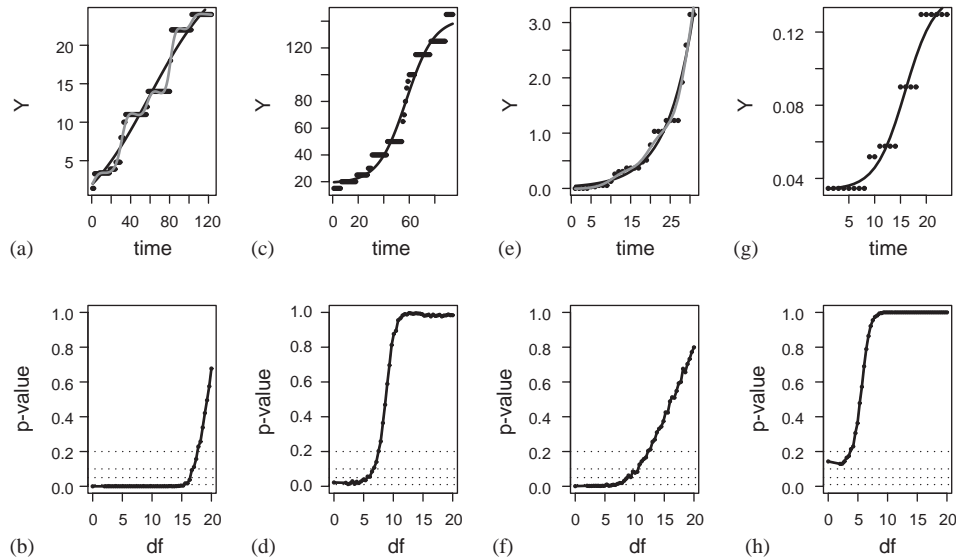


Fig. 8. Plots for (a,b) incandescent lighting, (c,d) arc-discharge lighting, (e,f) cathode ray tube display monitors and (g,h) dot-matrix desktop printers. The top row provides the technology evolution curves (black dots), the best fitting S curve (black line) and the minimum covariance structure that would need to be assumed to explain the deviation from an S curve (grey line). The second row gives the corresponding plots of p -value vs. df.

5. Discussion

We have developed three alternative but related methods to perform formal hypothesis tests on the shape of functional means. We see several advantages to our approach over those suggested previously. First, it allows one to place different levels of restriction on the covariance term. One can adopt either the FITS approach which imposes an i.i.d. restriction on the covariance function but has high power even for low signal to noise ratios or use the FACTS approach which imposes few assumptions on the underlying model. Second, even though we have concentrated on a null hypothesis of a logistic mean function, the methodology is completely general to any hypothesized form. Third, the permutation procedure we use is extremely computationally efficient which can be an important consideration when dealing with high dimensional data sets. Finally, in the case of a single curve with a general covariance structure, one can not only calculate statistical significance levels but also observe how complicated the covariance structure would need to be to explain the observed deviation from the null hypothesis.

There are several areas for future research. One involves the choice of q , the dimension of the basis for $s(t)$. The simulations in Section 3 suggest that, while any reasonable value for q gives the correct significance level, certain values provide higher power than others. In general it seems that the optimal value of q increases with the number of observations and complexity of the underlying curves. One promising solution would be to perform the tests for several values of q and then apply a correction to the p -value to account for

multiple testing. In addition, we have used regression splines both to model $s(t)$ and γ . Potentially, one may wish to use an automated procedure to adjust the basis that is used to maximize power for a given data set. Another possible extension involves the p -value plots illustrated, for example, in Fig. 8. Here the p -values are plotted as a function of the degrees of freedom of γ which we are using as a proxy for the complexity of Γ . However, there are other possible ways to measure this quantity which may also provide meaningful insights. Finally, we have only dealt with one-dimensional functional data. However, the methodology could be easily extended to perform tests on the shape of multidimensional data. In this situation, the computational advantages would become even more pronounced.

References

- Bewley, R., Griffiths, W.E., 2001. A forecasting comparison of classical and Bayesian methods for modelling logistic diffusion. *J. Forecasting* 20, 231–247.
- Darling, 1957. The Kolmogorov–Smirnov, Cramer–Von Mises tests. *Ann. Math. Statist.* 28, 823–888.
- Efron, B., Tibshirani, R., 1993. *An Introduction to the Bootstrap*. Chapman and Hall, London.
- Foster, R., 1986. *Innovation: The Attacker's Advantage*. Summit Books, New York.
- Green, P.J., Silverman, B.W., 1994. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, London.
- Hastie, T.J., Tibshirani, R.J., Friedman, J., 2001. *The Elements of Statistical Learning*. Springer, Berlin.
- James, G.M., Silverman, B.W., 2005. Functional adaptive model estimation. *J. Amer. Statist. Assoc.* 100 (470), 565–576.
- Johnson, N., Kotz, S., 1970. *Distributions in Statistics, vol. III: Continuous Univariate Distributions*. Houghton-Mifflin, New York.
- Montgomery, A.L., Bradlow, E.T., 1999. Why analyst overconfidence about the functional form of demand models can lead to overpricing. *Marketing Sci.* 18, 569–583.
- Robinson, P.M., 1989. Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Rev. Econom. Stud.* 56 (1), 511–534.
- Sahal, D., 1981. Alternative conceptions of technology. *Res. Policy* 10 (1), 2–24.
- Sood, A., Tellis, G.J., 2005. Technological evolution and radical innovation, *J. Marketing* 69 (July), forthcoming.
- Srinivasan, 1970. An approach to testing the goodness of fit of incompletely specified distributions. *Biometrika* 57, 605–611.
- Utterback, J.M., 1994. *Mastering the Dynamics of Innovation*. Harvard Business School Press,