

# THE JENSEN EFFECT AND FUNCTIONAL SINGLE INDEX MODELS: ESTIMATING THE ECOLOGICAL IMPLICATIONS OF NONLINEAR REACTION NORMS

BY ZI YE<sup>1,\*</sup>, GILES HOOKER<sup>1,†</sup> AND STEPHEN P. ELLNER<sup>2</sup>

<sup>1</sup>*Department of Statistical Science, Cornell University, \*zy234@cornell.edu; †gjh27@cornell.edu*

<sup>2</sup>*Department of Ecology and Evolutionary Biology, Cornell University, spe2@cornell.edu*

This paper develops tools to characterize how species are affected by environmental variability, based on a functional single index model relating a response such as growth rate to environmental conditions. In ecology the curvature of such responses are used, via Jensen's inequality, to determine whether environmental variability is harmful or beneficial, and differing nonlinear responses to environmental variability can contribute to the coexistence of competing species.

Here, we address estimation and inference for these models with observational data on individual responses to environmental conditions. Because nonparametric estimation of the curvature (second derivative) in a nonparametric functional single index model requires unrealistic sample sizes, we instead focus on directly estimating the effect of the nonlinearity by comparing the average response to a variable environment with the response at the expected environment, which we call the *Jensen Effect*. We develop a test statistic to assess whether this effect is significantly different from zero. In doing so we reinterpret the SiZer method of Chaudhuri and Marron (*J. Amer. Statist. Assoc.* **94** (1999) 807–823) by maximizing a test statistic over smoothing parameters. We show that our proposed method works well both in simulations and on real ecological data from the long-term data set described in Drake (*Proc. R. Soc. Lond., B Biol. Sci.* **272** (2005) 1823–1827).

**1. Introduction.** In natural ecosystems, environmental conditions are highly variable over time and space (e.g., Vasseur and McCann (2007)), and many classical questions in ecology and evolution are, therefore, concerned with the potential consequences of this variation. Two important topics have been how species' traits and life histories evolve so that species can persist in environments that may be favorable at some times and unfavorable at others (see Cohen (1966) and Koons, Metcalf and Tuljapurkar (2008)). Differing nonlinear responses to environmental variability can contribute to maintaining the biodiversity of competing species, allowing them to coexist stably (see, e.g., Chesson (1994), Chesson (2000a, 2000b), Chesson and Warner (1981), Ellner (1987), Hutchinson (1961)). Nonlinear responses to environmental conditions are also important for forecasting responses to climate change, as environmental variability can either increase population growth rate (Drake (2005), Koons et al. (2009)) or decrease it (Lewontin and Cohen (1969)), depending on the shape of the norm of reaction between environmental variables and the components of population growth rate (survival, growth and reproduction).

The goal of this paper is to develop methods for determining the effect of environmental variability on some component of population growth. Within mathematical biology this is a result of the curvature of the model; from Jensen's inequality convex functions result in higher growth under a variable environment, than when the environment is held constant at its mean and *vice versa* for concave functions. However, common statistical models for species

growth make parametric assumptions that predetermine these effects. For example, exponential models will always return convex results, while Michaelis–Menten saturation effects produce concave relationships. Recent statistical research in semiparametric or nonparametric methods inspires us to understand the effect of environmental variation via nonparametric models that do not impose these assumptions. Specifically, we consider spline-based methods (see Wood (2000), Ramsay (2006), Ramsay, Hooker and Graves (2009) and Ruppert, Wand and Carroll (2003)) to predict nonlinear responses under environmental fluctuation.

The nonparametric model considered here is

$$(1) \quad \mathbf{G} = g(\mathbf{E}) + \epsilon,$$

where  $\mathbf{G}$  and  $\mathbf{E}$  are the growth (or future size) and environment of an organism. The function  $g$  is the link function to be estimated, and  $\epsilon$  is random error. We assume that the environmental  $\mathbf{E}$  is described by a climate history, such as temperature or precipitation observed at a fine time-scale, such as daily resolution. Following Teller et al. (2016), these are thought of as functional covariates, leading to representation of  $\mathbf{E}$  as a functional linear term

$$(2) \quad \mathbf{E} = \int X(t)\beta(t) dt,$$

where  $\beta(t)$  is the coefficient function to be estimated and  $X(t)$  is the observed climate history.

Combining (1) and (2), a functional model for observed species growth is

$$(3) \quad Y = g\left(\int X(t)\beta(t) dt\right) + \epsilon.$$

This is the functional single index model introduced in Chen, Hall and Müller (2011) and Ma (2016). The functional single index model is an extension of the single index model to a functional covariate via  $\int X(t)\beta(t) dt$ . This model allows a nonlinear relationship between response  $Y$  and covariate function  $X(t)$ . In addition, it improves stability in estimating the link function  $g$  through imposing smoothness on  $\beta(t)$ .

With estimates of  $g$  and  $\mathbf{E}$ , we wish to assess the impact of environmental variability. Within mathematical biology this is done by an analysis of  $g''$ ; an always positive second derivative corresponds to  $g$  being convex and increased growth under environmental variability due to Jensen's inequality:  $Eg(\mathbf{E}) > g(E\mathbf{E})$ . Ye and Hooker (2018) investigate estimating  $g''$  in a functional single index model using a local quadratic estimate for  $g$  and demonstrates convergence rates but finds that the resulting estimators are impractical for finite sample size. A brief demonstration of this same effect using our estimators is given in Section A of the Supplementary Material (Ye, Hooker and Ellner (2020)).

The methods in this paper bypass estimating curvature and, instead, we estimate the consequences of environmental variation directly. That is, we define the quantity

$$\delta = g[E(\mathbf{E})] - E[g(\mathbf{E})]$$

directly and conduct inference about its sign. We have titled this the ‘‘Jensen Effect,’’ as being the result of Jensen's inequality. However, we note that  $g$  need not be strictly convex for  $\delta$  to be positive (or, conversely, for it to be negative) and observe that the ecological interest lies in  $\delta$  rather than  $g''$ . Indeed, in our real-world examples in Section 5 we observe that  $g$  is often not estimated to have the same curvature over the whole range; nonetheless the interaction of curvature and the distribution of covariates does produce a consistent Jensen Effect.

We argue that this target is also better suited for inference: it avoids estimating derivatives and, by averaging, we expect to gain stability relative to pointwise estimates of  $g$ . By contrast, direct inference about curvature requires a test of  $\min_s g''(s) > 0$  or  $\max_s g''(s) < 0$  which is substantially less stable. While bounding curvature does allow statements about consequences of variability to be independent of the distribution of  $\mathbf{E}$ , we observe that using

a nonparametric estimate of  $g$  already constrains the range of the argument  $s$  at which we can make statements about  $g''(s)$  and our tests can be employed using any given alternative distribution for  $E$  that is of interest.

In order to conduct inference, we take inspiration from the SiZer method of [Chaudhuri and Marron \(1999\)](#). Our estimates  $\hat{g}_\lambda(\cdot)$  and  $\hat{\beta}_\lambda(\cdot)$  both employ smoothing parameters. Rather than choosing these parameters, we instead examine the resulting  $t_\lambda = \hat{\delta}_\lambda / \text{sd}(\hat{\delta}_\lambda)$  as a function of  $\lambda$  and use  $\max_\lambda |t_\lambda|$  as a test statistic. We can assess the significance of this statistic by treating  $t_\lambda$  as a Gaussian process over  $\lambda$  and simulating from a null distribution in which  $E t_\lambda = 0$ . This avoids the need to account for the selection of  $\lambda$  as well as allowing us to detect relationships that may not be significant at smoothing parameters chosen by GCV or other criteria.

In the presentation of our methods, we have used  $\max |t_\lambda|$  as a statistic in order to include both positive and negative Jensen effects in the space of alternative hypotheses and allowing the sign to be deduced later. These approaches do have the potential to find significant deviations occur in both directions at different values of  $\lambda$ ; however, where we have encountered this in practice, it has been relatively easy to distinguish which represents a more plausible model for the data. One-sided tests can be performed using either  $\max t_\lambda$  or  $\min t_\lambda$  employing the same methods. Doing so removes the potential ambiguity above; we have chosen to retain and illustrate it here.

*1.1. Related literature: Single index models.* There is a large literature on single index models covering both applied methodology and theoretical properties. The link function  $g$  and the coefficient vector  $\beta$  have been estimated by three different methods: (1) The most widely used is the Projection Pursuit Regression (PPR) approach introduced in [Härdle, Hall and Ichimura \(1993\)](#). This method is a nested estimation procedure, with the link function  $g$  estimated by local polynomial approximation and the coefficient function by minimizing the MSE. Theoretical properties were studied in [Härdle, Hall and Ichimura \(1993\)](#) and [Ichimura \(1993\)](#). (2) The average derivative approach was introduced in [Hristache, Juditsky and Spokoiny \(2001\)](#). (3) [Li \(1991\)](#) introduced the sliced inverse regression method which considered the estimation of the coefficient vector as a dimension-reduction problem.

In contrast, there are few studies of the functional single index model. A counterpart to the projection pursuit regression was introduced in [Chen, Hall and Müller \(2011\)](#), where the coefficient function  $\beta$  was approximated by a spline basis and the coefficient vector was estimated. In addition, a convergence rate was found for this method. [Ma \(2016\)](#) used two spline bases to approximate the coefficient function and the link function, respectively, and derived some asymptotic properties of the resulting estimate. Methods that use single-index models as an additive term for function-on-scalar regression have been developed and studied in [Li, Huang and Zhu \(2017\)](#).

*1.2. Related literature: SiZer.* The SiZer (Significant Zero crossings of derivative) method that we adopt to assess significance was introduced by [Chaudhuri and Marron \(1999\)](#) to assess the significance of peaks and other features in nonparametric smooths while bypassing the selection of smoothing parameters. [Chaudhuri and Marron \(1999\)](#) observe that smoothing parameters can have a substantial impact on the features observed, and, rather than base inference on a selected smoothing parameter, they examine a range of reasonable smooths.

Specifically, SiZer is particularly concerned with the assessment of sign changes of derivatives. Local modes, in particular, can be represented by sign changes in a first derivative. To assess the significance of these, [Chaudhuri and Marron \(1999\)](#) obtain a nonparametric smooth  $\hat{f}(x; \lambda)$  of a relationship over  $x$  indexed by smoothing parameters  $\lambda$ . The method

then constructs the  $t$ -statistic  $t(x; \lambda) = \hat{f}^{(k)}(x; \lambda) / \text{sd}(\hat{f}^{(k)}(x; \lambda))$ .  $\hat{f}^{(k)}(x; \lambda)$  is then plotted over both  $x$  and  $\lambda$  with regions in which  $|t(x; \lambda)| > C$  are indicated. In particular, changes from a positive value of  $\hat{f}^{(1)}(x; \lambda)$  to a negative value as  $x$  is changed indicate a local maximum. Chaudhuri and Marron (1999) present their methods in the context of local polynomial smoothing and kernel density estimation, but exactly the same tools can be employed for any nonparametric method, including the smoothing splines we employ here. See Sonderegger et al. (2009) for an example of the use of SiZer in ecological models when searching for points of rapid ecological change in response to environmental forcing.

The interpretation of the significance of these changes depends on the selection of the critical value  $C$ . Chaudhuri and Marron (1999) suggest a variety of choices, including using pointwise significance levels (in which a number of false positives may be expected when searching over both  $x$  and  $\lambda$ ), using a critical value that approximately controls the family-wise error rate in  $x$  for each  $\lambda$  and obtaining a uniform bound by finding a critical value of  $\max_{x, \lambda} |t(x; \lambda)|$  via a bootstrap. Marron and Zhang (2005) develop approximations for Gaussian processes that reduce the computational burden while improving coverage probabilities.

We borrow from these ideas and, particularly, repurpose the search over “scale-space” to examine the Jensen Effect over a range of smoothing parameters. In contrast to SiZer, which examines features over a range of  $x$ , the Jensen Effect averages over covariates. Similar to Marron and Zhang (2005), we select a significance threshold designed to control the family-wise error rate over smoothing parameters; however, we select this from an explicit Gaussian approximation which can be efficiently simulated.

In the remainder of this paper we provide details of our estimation procedure for a functional single index model in Section 2, and our assessment and test of the Jensen Effect in Section 3. Simulation studies to assess the efficacy of our test are conducted in Section 4. Section 5 provides a motivating example in which we examine the response of eight copepod species to water temperature variability in data obtained from the North Temperate Lakes Long Term Experimental Research station where we find evidence for positive adaptation to environmental variability in five species. Section 6 concludes.

**2. Estimation procedure.** Here, we provide details of our representation of  $\beta$  and  $g$  and our estimation procedure given particular smoothing parameters. Section A of the Supplementary Material (Ye, Hooker and Ellner (2020)) provides a demonstration of using these approaches to estimate curvature rather than the Jensen Effect we examine here.

We represent both  $\beta$  and  $g$  using smoothed basis expansions. Assume that  $n$  independent and identically distributed data pairs  $(X_1(t), Y_1), \dots, (X_n(t), Y_n)$  are observed where  $X_j(t)$  is a real-valued function on  $[0, 1]$ . The functional single index model is

$$Y = g\left(\int X(t)\beta(t) dt\right) + \epsilon,$$

where the coefficient function  $\beta(t)$  is, like  $X(t)$ , defined on the interval  $[0, 1]$ . We assume  $\epsilon$  to be Gaussian random error. This integral may need to be evaluated numerically, depending on the representations used for  $X_i$  and  $\beta$ , and we assume that this is done up to ignorable error throughout the calculations below. To ensure identifiability of the model, we require that  $\|\beta\| = \int \beta(t)^2 dt = 1$ .

We use a  $K_1$ -dimensional B-spline basis for the link function  $g$ . For any  $s$  in the range of possible  $\int X\beta$  values, the link function  $g$  can be written as

$$g(s) = \boldsymbol{\phi}^\top(s)\boldsymbol{d},$$

where  $\boldsymbol{d}$  is a  $K_1$ -dimensional column coefficient vector.

We use a  $K_2$ -dimensional basis for the coefficient function  $\beta$ , such that

$$\beta(t) = \boldsymbol{\psi}^\top(t)\mathbf{c},$$

where  $\mathbf{c}$  is a  $K_2$ -dimensional column coefficient vector, and  $t \in [0, 1]$ .

The coefficient vectors  $\mathbf{c}$  and  $\mathbf{d}$  are estimated by minimizing a penalized sum of squares

$$\begin{aligned} (4) \quad 2\text{PLS} &\doteq \sum_{i=1}^n (Y_i - g_i)^2 + \lambda_g \int (g^{(2)}(s))^2 ds + \lambda_\beta \int (\beta^{(2)}(t))^2 dt \\ &= \sum_{i=1}^n \left\{ Y_i - \boldsymbol{\phi}^\top \left[ \left( \int X_i \boldsymbol{\psi}^\top \right) \mathbf{c} \right] \mathbf{d} \right\}^2 + \lambda_g \mathbf{d}^\top \mathbb{P}_g \mathbf{d} + \lambda_\beta \mathbf{c}^\top \mathbb{P}_\beta \mathbf{c}, \end{aligned}$$

where  $g_i \doteq g(f X_i \beta)$  and the penalty matrices  $[\mathbb{P}_\beta]_{ij} = \int \psi_i^{(2)}(t) \psi_j^{(2)}(t) dt$  and  $[\mathbb{P}_g]_{ij} = \int \phi_i^{(2)}(t) \phi_j^{(2)}(t) dt$  are available, analytically, for most common choices of basis expansion.

Equation (4) specifies a nonlinear optimization problem, which we solve numerically using built-in optimizers in R (see below). Denoting the estimated coefficients as  $\hat{\mathbf{c}}$  and  $\hat{\mathbf{d}}$ , the estimates are

$$\hat{\beta}(t) = \boldsymbol{\psi}^\top(t)\hat{\mathbf{c}}$$

and

$$\hat{g}_i \doteq \hat{g} \left( \int X_i \hat{\beta} \right) = \boldsymbol{\phi}^\top \left[ \left( \int X_i(t) \boldsymbol{\psi}^\top(t) dt \right) \hat{\mathbf{c}} \right] \hat{\mathbf{d}},$$

where  $i = 1, \dots, n$ .

*2.1. Notes on implementation and model selection.* Our objective criterion (4) requires nonlinear numerical optimization. In our experiments below we have used the R function `optim` with some additional modifications. The simulations reported in Section 4 used the BFGS gradient-based optimizer. However, at large values of  $\lambda$  we find that very tight convergence criteria are needed to reduce the numerical error to below that of the estimated noise. This was mitigated with three strategies:

1. We initialize our optimization with  $\mathbf{d}$  chosen so that  $\hat{g}$  is exactly linear and  $\mathbf{c}$  is obtained from functional linear regression.
2. We reinitialize BFGS once it converges and run it a second time. BFGS uses a sequentially-calculated approximate Hessian and can stop early due to poor estimation of this Hessian. Reinitialization resets the approximate Hessian to the identity so that optimization is restarted with a steepest descent step.
3. In our motivating data example we additionally attempted reinitializing from the solutions found at surrounding *eight* combinations of  $\lambda_g$  and  $\lambda_\beta$  and chose the best optima among these. This was repeated until the maximum relative improvement in optima was less than 0.01. This strategy was employed in order to ensure a smooth relationship between smoothing parameters and estimated effects but can substantially increase computational costs.

To maintain identifiability of our model, we normalize our estimate of  $\beta(t)$  within each evaluation of the objective function and multiply by  $\text{sign}(\beta(0))$ .

In order to represent  $\hat{g}$  with a basis expansion, we need to control the range of its arguments. Throughout our estimates below, we have used the identifiability requirement that  $\|\beta\| = 1$  to use a range of  $[-S, S]$  where  $S$  is the largest score for the maximum eigenvalue from a principal components decomposition of the  $X_i$ . If  $|\int X_i(t) \beta(t) dt| > S$  for some  $i$ , we replace the argument with the corresponding endpoint of the range and add a penalty of

$|\int X_i(t)\beta(t) dt - S|$  to the objective (4). In practice, while we find that this exceedance can occur during optimization, it never appears in the final result.

We find that these procedures are sufficient to provide reliable inference for the Jensen Effect. However, estimates of  $g''(\cdot)$  can be highly sensitive to initial conditions and optimization strategies. Section A of the Supplementary Material (Ye, Hooker and Ellner (2020)) provides a brief example of the sensitivity of curvature to both noise and initial conditions. We speculate that this sensitivity is a result of a complex optimization landscape in which there are many good estimates of  $g$ , but these can vary substantially for  $g''$ .

While our assessment of statistical significance avoids selecting  $\lambda$ , it will be useful to have a value for visualization and for an estimate of residual variance. To choose  $\lambda$ , we define a smoother matrix  $\mathbb{S}_\lambda$  associated with  $\lambda$ , and the GCV value for selecting  $\lambda$  is calculated from

$$\text{GCV}(\lambda) \doteq \frac{\frac{1}{n} \|(\mathbb{I} - \mathbb{S}(\lambda))\mathbf{Y}\|^2}{[\frac{1}{n} \text{tr}(\mathbb{I} - \mathbb{S}(\lambda))]^2},$$

where  $\mathbb{I}$  is the  $(n \times n)$ -dimensional identity matrix. We derive  $\mathbb{S}_\lambda$  from a Taylor expansion in (18) below.

**3. Jensen effect.** The ecological interest in  $g''$  is in the comparison of  $g[E(\mathbf{E})]$  and  $E[g(\mathbf{E})]$ . Because reliable estimation of  $g''$  requires unrealistic sample sizes, we instead compare these quantities directly to estimate what we call the ‘‘Jensen Effect.’’

We define a difference statistic by

$$\delta = \frac{1}{n} \sum_{i=1}^n g\left(\int X_i \beta\right) - g\left(\int \bar{X} \beta\right),$$

where  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ . In mathematical analyses, if the link function  $g$  is convex, then  $\delta > 0$  which indicates better growth with a varying environment; otherwise, the difference  $\delta < 0$  and a constant environment is better for growth. However, this estimate still depends on the smoothing parameters  $\lambda_g$  and  $\lambda_\beta$ . Inspired by the SiZer method of Chaudhuri and Marron (1999), we examine the difference  $\delta$  over a range of  $\lambda$  values for  $g$  and  $\beta$  and generate hypothesis tests using the maximum or minimum value of  $\delta$  as a function of  $\lambda$ .

3.1. *Hypothesis test 1: Nonparametric smoothing.* We begin by briefly developing our SiZer-inspired test for a standard smoothing spline (treating the environment  $\mathbf{E}$  as known) before developing the test for a functional single index model. Here, defining  $\Phi$  to be matrix of evaluations,  $\Phi_{ij} = \phi_j(E_i)$  and  $\mathbb{P}$  to be the second derivative penalty matrix, the standard smoothing spline estimate is

$$(5) \quad \hat{g}_\lambda(e) = \phi(e)^\top (\Phi^\top \Phi + \lambda \mathbb{P})^{-1} \Phi^\top \mathbf{Y}.$$

Define the  $(n + 1)$ -dimensional column vector  $\mathbf{a} \doteq (\frac{1}{n}, \dots, \frac{1}{n}, -1)^\top$  and the augmented set of evaluation points  $\mathbf{e} = (E_1, \dots, E_n, \bar{E})^\top$ , where  $E_i = \int X_i \hat{\beta}$  (at each observed environment value) and  $\bar{E} = \frac{1}{n} \sum_{i=1}^n \int X_i \hat{\beta}$  (averaged across all environment values) with corresponding evaluation matrix  $\Phi^+$ . We can write

$$(6) \quad \delta_\lambda = \frac{1}{n} \sum \hat{g}_\lambda(E_i) - \hat{g}_\lambda(\bar{E}) = \mathbf{a}^\top \Phi^+ (\Phi^\top \Phi + \lambda \mathbb{P})^{-1} \Phi^\top \mathbf{Y} = \mathbf{u}_\lambda \mathbf{Y},$$

which we can standardize to obtain the  $t$ -statistic

$$(7) \quad t_\lambda = \frac{\mathbf{u}_\lambda \mathbf{Y}}{\hat{\sigma} \sqrt{\mathbf{u}_\lambda \mathbf{u}_\lambda^\top}}$$



in which  $\hat{\sigma}$ , the estimate of the standard deviation of the random error  $\epsilon$ , is obtained from the value of  $\lambda$  selected by GCV (see details below).

Since the response variable  $Y$  is Gaussian, the test statistic  $t_\lambda$  is also a Gaussian process with the covariance function

$$(8) \quad \Sigma(\lambda_1, \lambda_2) = \frac{\mathbf{u}_{\lambda_1} \mathbf{u}_{\lambda_2}}{\|\mathbf{u}_{\lambda_1}\| \|\mathbf{u}_{\lambda_2}\|}$$

which involves no unknown parameters. We can thus use  $\max_\lambda |t_\lambda|$  as a test statistic, obtaining critical values by simulating from the Gaussian process  $\mathbb{G}\mathbb{P}(\mathbf{0}, \Sigma)$ . Under the null hypothesis  $\delta = 0$ ,  $t_\lambda$  is a Gaussian process with mean  $\mathbf{0}$ .

An important consideration here is that we expect  $\delta_\lambda$  to inherit smoothing bias, but this should result in underestimation of the Jensen Effect because it will shrink the estimated second derivative. In analogy to SiZer, by examining  $t_\lambda$  over the whole range of  $\lambda$  we can assess this effect at various levels of smoothing; our use of  $\max |t_\lambda|$  as a test statistic allows to maintain a conservative test. We do still need to choose  $\lambda$  by GCV in our estimate  $\hat{\sigma}^2$ , because we use the same  $\hat{\sigma}^2$  when calculating the covariance matrix  $\Sigma$ . We expect  $\hat{\sigma}^2$  to be relatively insensitive to the specific  $\lambda$  chosen, so long as we do not oversmooth (see arguments in Ruppert, Wand and Carroll (2003)); maintaining a constant  $\hat{\sigma}^2$  in the  $t$ -statistic removes the need to account for changes in  $\hat{\sigma}^2$  across  $\lambda$ .

3.2. *Hypothesis test 2: Functional single index model.* The functional single index model complicates the process described above by including two smoothing parameters and nonlinear effects of  $\hat{\beta}$ , necessitating a Taylor expansion to approximate the recipe above. For each pair of smoothing parameters  $(\lambda_g, \lambda_\beta)$ , we obtain an estimate of  $\beta$ ,  $g$ , and  $\delta$ , denoted as  $\hat{\delta}(\lambda_g, \lambda_\beta)$ .

Defining

$$(9) \quad \mathbf{i} = \left( \int X_1 \hat{\beta}, \dots, \int X_n \hat{\beta}, \int \bar{X} \hat{\beta} \right)^\top,$$

$$(10) \quad \mathbf{i}_{-1} = \left( \int X_1 \hat{\beta}, \dots, \int X_n \hat{\beta} \right)^\top,$$

$$(11) \quad \mathbf{v} = \left( \hat{g} \left( \int X_1 \hat{\beta} \right), \dots, \hat{g} \left( \int X_n \hat{\beta} \right), \hat{g} \left( \int \bar{X} \hat{\beta} \right) \right)^\top,$$

the estimated difference function given  $(\lambda_g, \lambda_\beta)$  is

$$(12) \quad \begin{aligned} \hat{\delta}(\lambda_g, \lambda_\beta) &= \mathbf{a}^\top \mathbf{v} = \mathbf{a}^\top \boldsymbol{\phi}(\mathbf{i}) \hat{\mathbf{d}} \\ &= \mathbf{a}^\top \boldsymbol{\phi}(\mathbf{i}) (\boldsymbol{\phi}(\mathbf{i}_{-1})^\top \boldsymbol{\phi}(\mathbf{i}_{-1}) + \lambda_g \mathbb{P}_g)^{-1} \boldsymbol{\phi}(\mathbf{i}_{-1})^\top \mathbf{Y}. \end{aligned}$$

To construct a  $t$ -statistic to test the significance of  $\delta$ , an estimate of the variance of the difference function  $\hat{\delta}$  is needed. The estimated difference function  $\hat{\delta}(\lambda_g, \lambda_\beta)$  is defined on an estimate of  $\hat{\mathbf{c}}$  and  $\hat{\mathbf{d}}$  which are the coefficients of  $\beta$  and  $g$ , respectively. Therefore, we need to calculate the covariance of the estimated  $\hat{\mathbf{c}}$  and  $\hat{\mathbf{d}}$ .

Recall (4), the penalized least squares criterion to be minimized, and define the matrices of linear basis effects  $\Psi_{ij} = \int X_i(t) \psi_j(t) dt$  and evaluations of the link function bases and derivatives  $\Phi_{ij}^{(k)} = \phi_j^{(k)}(\Psi_i \cdot \mathbf{c})$  with  $\mathbf{c}$  taken at its expected estimate. We derive gradients of PLS as

$$(13) \quad \begin{pmatrix} \nabla \mathbf{d} \\ \nabla \mathbf{c} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi}^\top \{\mathbf{Y} - \boldsymbol{\Phi} \mathbf{d}\} + \lambda_g \mathbb{P}_g \mathbf{d} \\ \boldsymbol{\Psi}^\top \text{diag}\{\boldsymbol{\Phi}^{(1)} \mathbf{d}\} \{\mathbf{Y} - \boldsymbol{\Phi}^\top \mathbf{d}\} + \lambda_\beta \mathbb{P}_\beta \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbb{Z}_g \\ \mathbb{Z}_\beta \end{pmatrix} (\mathbf{Y} - \boldsymbol{\Phi} \mathbf{d}) + \begin{pmatrix} \lambda_g \mathbb{P}_g \mathbf{d} \\ \lambda_\beta \mathbb{P}_\beta \mathbf{c} \end{pmatrix}$$

and expected Hessian

$$(14) \quad \mathbb{H} = \begin{pmatrix} \mathbb{Z}_g^\top \mathbb{Z}_g + \lambda_g \mathbb{P}_g & \mathbb{Z}_g^\top \mathbb{Z}_\beta \\ \mathbb{Z}_\beta^\top \mathbb{Z}_g & \mathbb{Z}_\beta^\top \mathbb{Z}_\beta + \lambda_\beta \mathbb{P}_\beta \end{pmatrix}.$$

We can now obtain the sandwich covariance

$$(15) \quad \text{cov} \begin{pmatrix} \mathbf{d} \\ \mathbf{c} \end{pmatrix} = \hat{\sigma}^2 \mathbb{H}^{-1} \begin{pmatrix} \mathbb{Z}_g \\ \mathbb{Z}_\beta \end{pmatrix}^\top \begin{pmatrix} \mathbb{Z}_g \\ \mathbb{Z}_\beta \end{pmatrix} \mathbb{H}^{-1},$$

where we estimate  $\sigma^2$  from

$$(16) \quad \hat{\sigma}^2 = \frac{1}{\text{df}_{\text{res}}} \sum_{i=1}^n \left[ Y_i - \hat{g} \left( \int X_i \hat{\beta} \right) \right]^2,$$

where, following [Ruppert, Wand and Carroll \(2003\)](#), the residual degree of freedom is defined as

$$(17) \quad \text{df}_{\text{res}} = n - 2 \text{tr}(\mathbb{S}) + \text{tr}(\mathbb{S}\mathbb{S}^\top),$$

where

$$(18) \quad \mathbb{S} \doteq \mathbb{S}(\lambda_g, \lambda_\beta) = \begin{pmatrix} \mathbb{Z}_g \\ \mathbb{Z}_\beta \end{pmatrix} \mathbb{H}(\mathbb{Z}_g^\top, \mathbb{Z}_\beta^\top)$$

is an approximate smoother matrix in which we use the values of  $(\lambda_g, \lambda_\beta)$  selected by GCV.

We now define a  $t$ -statistic for  $\delta$  as a function of  $\lambda$ ,

$$(19) \quad t \doteq t(\lambda_g, \lambda_\beta) \doteq \frac{\hat{\delta}(\lambda_g, \lambda_\beta)}{\text{sd}[\hat{\delta}(\lambda_g, \lambda_\beta)]},$$

where  $\text{sd}[\hat{\delta}(\lambda_g, \lambda_\beta)]$  is given by

$$(20) \quad \text{sd}[\hat{\delta}(\lambda_g, \lambda_\beta)] = \{ [\mathbf{a}^\top \boldsymbol{\phi}(\mathbf{i})] \text{cov}(\hat{\mathbf{d}}) [\mathbf{a}^\top \boldsymbol{\phi}(\mathbf{i})]^\top \}^{\frac{1}{2}}.$$

Defining the  $n$ -dimensional row vector  $\mathbf{u}_\lambda$  as

$$(21) \quad \mathbf{u}_\lambda = \mathbf{a}^\top \boldsymbol{\phi}(\mathbf{i}) (\boldsymbol{\phi}(\mathbf{i}_{-1}))^\top \boldsymbol{\phi}(\mathbf{i}_{-1}) + \lambda \mathbb{P}^{-1} \boldsymbol{\phi}(\mathbf{i}_{-1})^\top,$$

the estimated covariance matrix of  $\hat{\delta}_\lambda$  is  $\hat{\sigma}^2 \mathbf{u}_\lambda \mathbf{u}_\lambda^\top$ .  $\delta_\lambda$  is, therefore, approximately a Gaussian process indexed by  $\lambda$ , and we can get the estimated variance of  $t_\lambda$  from  $\frac{\mathbf{u}_\lambda \mathbf{u}_\lambda^\top}{\|\mathbf{u}_\lambda\|^2}$ .

We want to test if  $\delta \equiv 0$ . Denote the number of the smoothing parameters  $\lambda_g$  as  $m$ , we test  $H_0: (\hat{\delta}_{\lambda_1}, \dots, \hat{\delta}_{\lambda_m})^\top = \mathbf{0}_m$ . Under  $H_0$ ,  $(t_{\lambda_1}, \dots, t_{\lambda_m})^\top \sim N(\mathbf{0}_m, \mathbb{A}_{mm})$ , where  $\mathbf{0}_m$  is a  $m$ -dimensional column vector and the covariance matrix  $\mathbb{A}$  is  $(m \times m)$ -dimensional, with the  $(i, j)$  term equals to  $\frac{\mathbf{u}_{\lambda_i} \mathbf{u}_{\lambda_j}^\top}{\|\mathbf{u}_{\lambda_i}\| \|\mathbf{u}_{\lambda_j}\|}$ . The test statistic that we examine is  $T = \max\{|t_{\lambda_1}|, \dots, |t_{\lambda_m}|\}$ .

In order to obtain a critical value for this statistic, we repeatedly simulate  $t_\lambda$  from  $N(\mathbf{0}_m, \mathbb{A}_{mm})$  and obtain a distribution for  $\max_\lambda |t_\lambda|$ .

Throughout our experiments, we have used  $\max_\lambda |t_\lambda|$  as a test statistic in order to be agnostic about the direction of the alternative hypothesis. This combines both positive and negative rejection regions, allowing us to identify the sign of the resulting Jensen Effect post hoc. We note that, when a specific direction is of interest, either  $\min t_\lambda$  or  $\max t_\lambda$  could be employed as a test statistic following the same procedures.

We note the potential for  $\hat{\delta}_\lambda$  to change signs over the range of  $\lambda$ . For example, if the underlying  $g$  is strongly convex in a very narrow region but its general shape is moderately concave, we might find a positive effect at small values of  $\lambda$  and a negative effect at large



values as the convex portion of  $g$  is smoothed over. We would regard this as good reason to examine the resulting estimates of  $g$  with an eye to plausibility at both values of  $\lambda$ . In our motivating data in Section 5, we found a couple of examples in which  $\delta_\lambda$  was declared significant at two values of  $\lambda$  where  $\delta_\lambda$  had opposite signs. One of these could be dismissed easily as occurring only at one set of extreme values. The case of *Diacyclops Thomasi* required further investigation which we defer to Section 5. Prespecifying the direction of interest would avoid these issues, but in many ecological applications (such as our empirical case study) both possible directions are equally plausible and equally of interest. Therefore, we believe that the  $\max_\lambda |t_\lambda|$  test statistic will generally be most useful in practice.

**4. Simulation study.** In this section we use simulated data to explore the power of our test for both single index and functional single index models. Computation time for these models depends on numerous properties of the model: the number of smoothing parameter values to try, the nonlinearities of the underlying  $g(\cdot)$ , the size of the error variance and the data set size. Because differing settings could result in substantially different computing times for exactly the same problem, we have not given exact timings here but note that our single index models run within a few minutes per simulation, while functional versions can require 10 to 30 minutes. Our real-world examples, with larger data sets and a finer mesh of smoothing parameters, required more than half an hour on a recently purchased laptop.

**4.1. Single index model.** We test for a Jensen Effect by calculating the difference function  $\delta$  over a range of smoothing parameters. If the link function  $g$  is convex, the  $\delta$  function will be positive for most of  $\lambda$  values, although it may have high variance at low  $\lambda$  and high bias at high  $\lambda$ . For each simulation we conduct the hypothesis test introduced in previous section.

Our simulation study starts with the single index model with  $p = 5$  covariates, generated independently and uniformly on  $[-0.5, 0.5]$ , and the coefficient  $\beta = \frac{1}{\sqrt{p}} \mathbf{1}_p$  so that  $\|\beta\| = 1$ .

To illustrate the Jensen Effect, we choose three different link functions: (1)  $g(s) = e^s$ , (2)  $g(s) = -s^2$ , (3)  $g(s) = s$ . We represented  $g$  by a 25-dimensional quintic B-spline basis. For each link function we simulated 1000 data sets of size 100, with error standard deviation 0.1. We obtained critical values for our test by simulating 5000 normal samples from the null distribution. Figure 1 presents a sample of  $\delta_\lambda$  and  $t_\lambda$  functions vs.  $\log(\lambda)$  for  $g(s) = e^s$ ; plots for the other link functions are in Section B of the Supplementary Material (Ye, Hooker and Ellner (2020)). The rejection rates for these functions are: 99.2%, 99.3% and 5.7%, respectively.

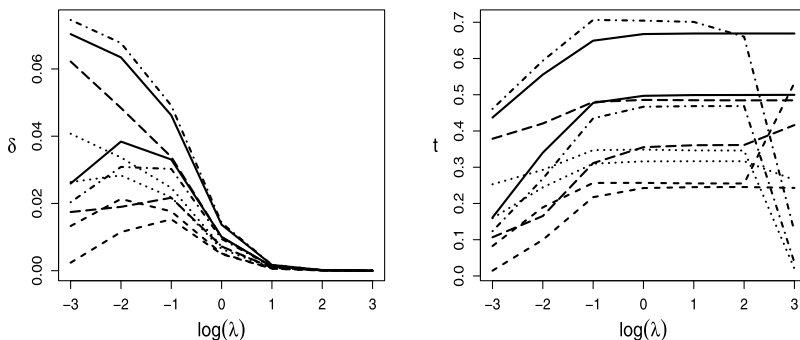


FIG. 1. Left: a sample of  $\delta_\lambda$  as a function of  $\lambda$  in a single index model with link function  $g(s) = e^s$ . Right: the corresponding  $t_\lambda$  functions.

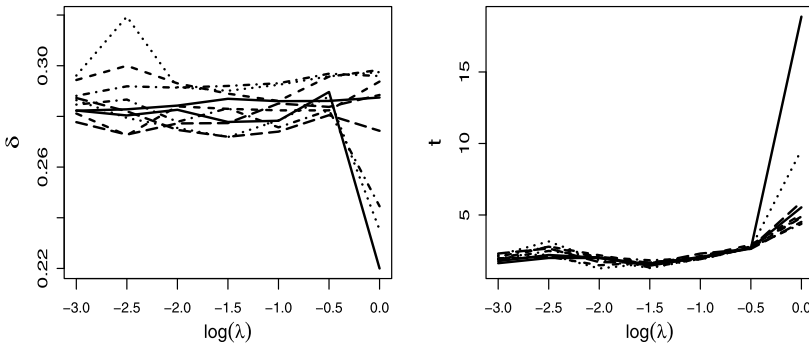


FIG. 2. Left: a sample of  $\delta_\lambda$  as a function of  $\lambda$  in a functional single index model with link function  $g(s) = e^s$ . Right: the corresponding  $t_\lambda$  functions.

4.2. *Functional single index model.* To define a distribution for the functional covariates, we use a 25-dimensional Fourier basis  $\psi(t)$ , where  $t \in [0, 1]$ . The covariate functions  $X(t)$  are generated as

$$(22) \quad X(t) = \sum_{i=1}^{25} \xi_i \psi_i(t),$$

where  $\xi_i \sim N(0, e^{-(i-1)/12})$ . The coefficient function is

$$(23) \quad \beta(t) = \mathbf{c}^\top \psi(t),$$

where  $\mathbf{c} = (0, 1, 1, 0.5, 0, \dots, 0)^\top$ .

Again, we used the three link functions  $g(s) = e^2$ ,  $g(s) = -s^2$ ,  $g(s) = s$ . We represented  $g$  by a 25-dimensional quintic B-spline basis. For each link function we generated 1000 simulated data sets of size 100 with error standard deviation 0.1, and for each such data set we generated 5000 normal samples from the null distribution to obtain critical values.

A plot of the  $\delta_\lambda$  and  $t_\lambda$  functions for  $g(s) = e^s$  is presented in Figure 2. We have placed equivalent plots for  $g(s) = -s^2$  and  $g(s) = s$  in Section C of the Supplementary Material (Ye, Hooker and Ellner (2020)). The rejection rates for the three link functions were 100%, 100% and 7.3%, showing very good power with a reasonable sample size and close to nominal rate when the null hypothesis is true (no Jensen Effect).

4.3. *Power analysis.* To investigate the power of our test in more detail we consider a series of increasingly nonlinear link functions

$$(24) \quad g(s) = s + \eta e^{-s},$$

with  $0 \leq \eta \leq 1.2$  for the single index model and  $0 \leq \eta \leq 0.8$  for the functional single index model. As  $\eta$  increases,  $g$  becomes strongly convex. For each  $\eta$  we generate 1000 simulated data sets and again used 5000 normal samples under the null distribution to obtain critical values. Due to the computational overhead associated with searching over another smoothing parameter, we used only 200 simulations for the functional single index model. We replicated each simulation changing  $n$  to 200 and changing  $\sigma$  to 0.2 to test for the expected loss of power with increasing noise and gain of power with sample size.

Figure 3 presents the rejection rate plotted against  $\eta$ . We observe a sharp increase as  $\eta$  increases, as expected: as the link function  $g$  becomes more and more convex, the rejection rate will converge to 1. The expected patterns of decreasing power with increasing  $\sigma$  and increasing power with  $n$  are observed, although a smaller simulation size makes this less clear in the functional single index model.

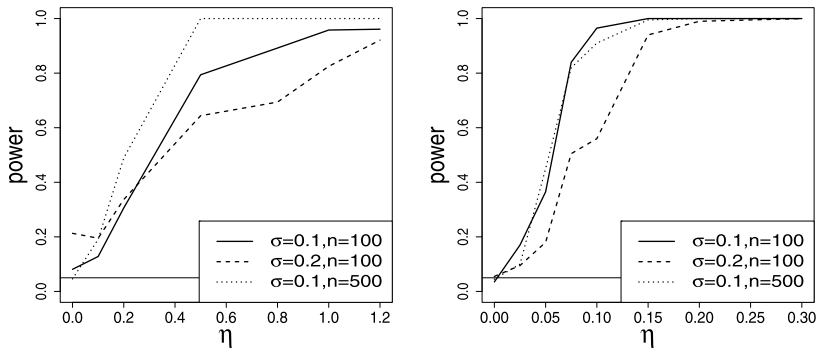


FIG. 3. The power function of the Jensen Effect test plotted against  $\eta$  in the link function  $g(s) = s + \eta e^{-s}$  at sample sizes  $n \in \{100, 200\}$  and residuals standard errors  $\sigma \in \{0.1, 0.2\}$ . A single index model using the simulation settings from Section 4.1 produced the left plot, while the right-hand plot is obtained using a functional single index model with settings described in Section 4.2.

We also used this experiment as an opportunity to verify that our estimates of residual standard error,  $\sigma$ , performed well. See Section D of the Supplementary Material (Ye, Hooker and Ellner (2020)) for further details.

**5. Application to real ecological data.** To demonstrate application of our tests, we analyze the North Temperate Lakes LTER: Zooplankton–Trout Lake Area data set.<sup>1</sup> An earlier version of these data were analyzed by Drake (2005) to examine the temperature dependence of copepod populations. In our data set the density of the populations of eight species of copepods and rotifers, along with water temperature, were recorded from 1981 to 2015 in eight different lakes. Our choice of species was determined based on the number and length of observations available and differs from those studied in Drake (2005). Note that the growth response in these data is not the growth (in size) of an individual but the growth (in numbers) of a population; nevertheless, our functional single index model is still appropriate for this setting.

The values recorded in the original data set are:

1.  $\bar{d}$ : Species' density at a specific time and lake.
2.  $\tau$ : Record of water temperature collected at the same time as  $\bar{d}$ .

Both measurements were recorded on irregular time points among different years and lakes, so in order to obtain functional covariates, we preprocessed the temperature data by fitting a smoothing spline; see details in Section E of the Supplementary Material (Ye, Hooker and Ellner (2020)).

Our response is change in density between successive observed time points normalized by the time change

$$Y_i = (\delta_{s_{i+1}} - \delta_{s_i}) / (s_{i+1} - s_i)$$

recorded so long as  $s_{i+1} - s_i < 100$ , where  $s_i$  is the time (in days) of the  $i$ th sample.

At the time of each observed response, we used temperature values over the 60 preceding days as the climate history covariate  $X(t)$ . For each species we fit a penalized spline functional single index model for the growth in population density as a function of temperature

<sup>1</sup>Zooplankton records from <https://portal.edirepository.org/nis/mapbrowse?scope=knb-lter-ntl&identifier=37&revision=29> water temperature from <https://portal.edirepository.org/nis/mapbrowse?scope=knb-lter-ntl&identifier=129>.

history in each lake; details of these procedures are given in Section E of the Supplementary Material (Ye, Hooker and Ellner (2020)). We represented  $\beta$  with 12 order-6 B-splines covering the 60 day history and reinterpolated the smoothed  $X(t)$  processes onto this basis. We used a 25-dimensional cubic B-spline basis to represent  $g$  because we wanted a linear function to fall in the span of our basis and set its range to be  $\pm \max \int \phi_j(t) X_i(t) dt$  to ensure that the single index values fell within it. We searched over values of  $\log_{10}(\lambda)$  in the range  $[-6, 2]$  for  $g$  and  $[-2, 6]$  for  $\beta$ . These were chosen to cover the range of values selected by GCV for most data sets while avoiding spurious significance due to optimization errors.

The Jensen Effect  $\delta$  was estimated to be positive at all smoothing parameters in six out of the eight species; in *Keratella Earlinae* only one extreme combination of smoothing parameters produced a negative  $\delta$ . However, statistically significant values of  $\delta$  were only found for five out of these seven species (*Kellicottia Longispina*, *Keratella Cochlearis*, *Keratella Earlinae*, *Polyarthra Remata*, and *Polyarthra Vulgaris*). Nonetheless, we conclude from this that a majority of the copepod species in this data set are evolved to take advantage of environmental variability. Examining estimates of  $\beta(t)$ , we find some of these are undersmoothed at GCV values, but they tend to represent gradients corresponding to either warming or cooling water temperatures, likely associated with seasonal abundance trends. The use of water temperature as a sole covariate means that its effects are conflated with other environmental variables that change seasonally, such as the availability of nutrient sources. Thus, we cannot conclude a causal relationship but note that the same analysis can be undertaken while accounting for other covariates if and when they are available.

Figure 4 provides a canonical set of plots for the species *Polyarthra vulgaris*; equivalent plots for the remaining species are presented in Section E of the Supplementary Material (Ye, Hooker and Ellner (2020)). The top four panels provide  $g(s)$ ,  $g''(s)$  and  $\beta(t)$  estimated at the smoothing parameters selected by GCV as well as a contour plot of  $\delta$  over the smoothing parameters with a shaded area that indicates values at which the effect was found not to be significant. In this case  $\delta$  is positive and significant at all smoothing parameters. In order to explore the shape of the response further, we also plot  $g(s)$  and  $\beta(t)$  at the smoothing parameters that result in the maximum value of  $\delta$ . Note that maximizing  $\delta$  here results in particularly large edge effects which should be treated with caution.

One species, *Diacyclops Thomasi*, was an exception to the general pattern in our analyses, producing areas in the  $(\lambda_g, \lambda_\beta)$  plane where the estimated  $\delta$  has large positive values and other areas where it has large negative values. Figure 5 provides plots of  $g(s)$  at three smoothing parameter values as well as  $\delta(\lambda_g, \lambda_\beta)$  along with indicators of where it is significant in order to visualize the effects of smoothing parameters. We note that  $\delta$  is not significant at the values of the smoothing parameters selected by GCV (black square), but it is declared to be significant in both the negative (middle of the plot given by the black circle) and positive (bottom indicated by the triangle) directions. Figure 5 provides plots of  $g$  at each of these two points for comparison. The positive estimate is associated with an inflection of  $g$ , at the low end of the range of  $\int X_i(t) \hat{\beta}(t) dt$  values, and is likely the result of undersmoothing. Therefore, we conclude, tentatively, that *Diacyclops Thomasi* apparently differs from the other species in being harmed by temperature variability, but we feel that further experiments are warranted in this case. We note that this is the only species in which our signal is ambiguous (as opposed to inconclusive). In *Keratella Earlinae* (Figure 15) a single significant negative value is found at the bottom corner of the contour plot, which we feel can be dismissed; see an equivalent but positive corner effect in *Kellicottia Longispina* in Figure 13.

**6. Conclusion.** Environmental variability is ubiquitous and projected to change significantly over the next century as one aspect of global climate change. Therefore, projecting how biological populations will respond to climate change requires methods to estimate how they will respond to changes in the variance of conditions, not just the mean.

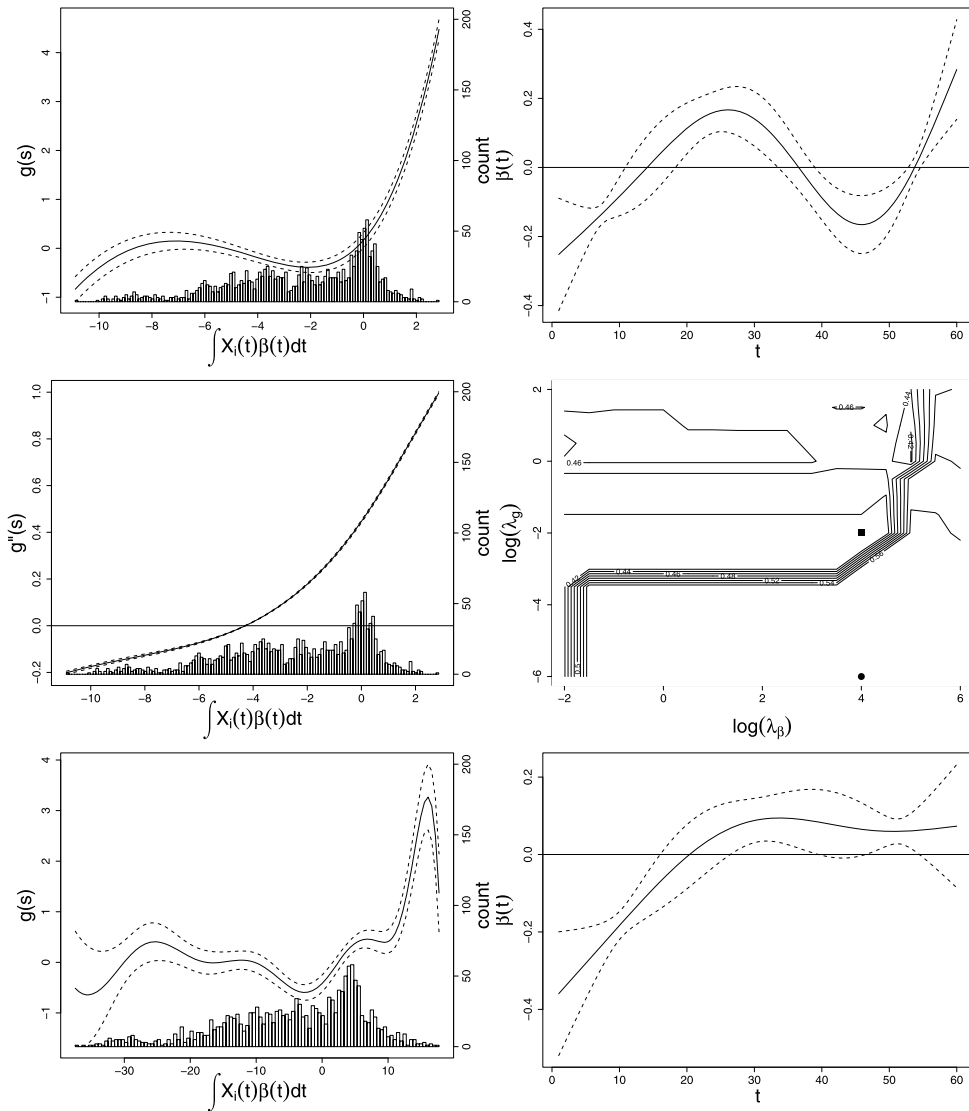


FIG. 4. Diagnostic plots for *Polyarthra Vulgaris*,  $n = 1577$ . The estimated  $g(s)$  (top left) along with  $\beta(t)$  (top right), and  $g''(s)$  (middle left) along with pointwise confidence intervals. Histograms give the distribution of estimated  $\int X_i(t)\hat{\beta}(t) dt$ . All plots are given at the values which minimize GCV. Middle right:  $\delta$  as a function of both  $\lambda_g$  and  $\lambda_\beta$ . Regions where  $t_{\lambda_g, \lambda_\beta}$  exceed the critical value are indicated by a white background (the whole plot in this case), and a black square gives the minimizing value of GCV; black circle indicates values at which  $\delta$  is maximized. Bottom row:  $g$  and  $\beta$  at the smoothing parameter values that maximize  $\delta$ .

With the goal of estimating the net effect of environmental variability on components of population growth rate, we first attempted to estimate the curvature of the link function  $g$  in a functional single index model. In our penalized spline based method, we found that unrealistic sample sizes were needed to obtain accurate estimates. Instead, we directly investigated the effect of environmental variability by comparing the expected response (averaged across the environmental variation) to the response at the expected environment. We termed this the “Jensen Effect” since it describes the effect of Jensen’s inequality, but we note that it operates far more broadly than on just convex or concave functions. Inspired by the SiZer method, our test for a nonzero Jensen Effect is based on maximizing a test statistic across a wide range of smoothing parameters encompassing all plausible values, thus avoiding the need for

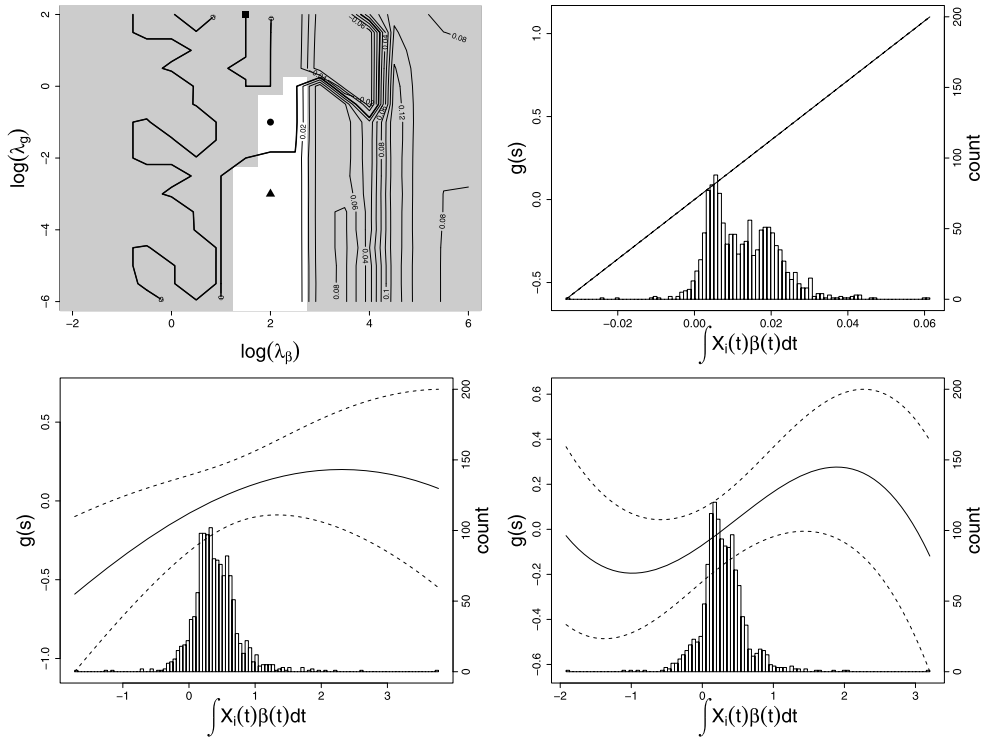


FIG. 5. Comparison of estimated reaction norm  $g$  at different smoothing parameters for *Diacyclops thomasi*. Top left: contour of  $\delta$  as a function of smoothing parameters with values minimizing GCV (solid square), maximizing  $\delta$  (triangle) and minimizing  $\delta$  (circle). The corresponding plots of  $g$  are given in the top right (minimizing GCV), bottom left (minimizing  $\delta$ ) and bottom right (maximizing  $\delta$ ). We judge the positive values of  $\delta$  to be due to large edge effects.

smoothing parameter selection. We have shown that our proposed procedures work well, on both simulated and real data.

There are multiple potential extensions of this methodology. We have used observed data, as representative of the covariates of interest, to define the average and distribution of environmental variability. However, the test can be conducted for any assumed distribution of covariates, and it may be of interest to describe regions of single index values in which the estimated response function produces a Jensen Effect. A way to achieve this is to plot the  $a$  and  $b$  for which  $E(g(S)) - g(ES)$  is significantly different from 0, when our procedure is applied under the assumption that  $S \sim U[a, b]$ .

It may also be of interest to ask about Jensen Effects on different scales of measurement. For example, using size (rather than change in size) as a response in our empirical application to copepod growth rate results in heteroskedasticity in the response which can be ameliorated by a log transform. We would then employ the model  $E \log(Y) = g(\int X(t)\beta(t) dt)$  and the Jensen Effect of interest is  $E \exp(g(\int X(t)\beta(t) dt)) - \exp(g(E \int X(t)\beta(t) dt))$ . Smoothing  $\hat{g}_\lambda$  biases the Jensen Effect toward being positive, and our method would need modifications to take account of this.

The same challenge arises in extending our approach to other response structures, such as survival or count data, which also use nonlinear link functions for fitting within a generalized linear models framework. Any of the standard GLM links could be modified by placing a nonparametric  $g(s)$  within the GLM link. But the Jensen Effect then applies to the composition of the GLM link with  $g(s)$ , so, again, there would need to be a way of accounting for the bias introduced by smoothing.



**Acknowledgments.** This work was partially supported by NSF Grant DEB-1353039 and completed while the second author was visiting the Research School of Finance, Actuarial Science and Statistics at the Australian National University. The authors would like to thank the input of the Editors and referees which significantly strengthened the paper. We are particularly grateful to Peter Adler for discussion and assistance with the data.

## SUPPLEMENTARY MATERIAL

**Supplementary Material to “The Jensen effect and functional single index models: Estimating the ecological implications of nonlinear reaction norms”** (DOI: [10.1214/20-AOAS1349SUPPA](https://doi.org/10.1214/20-AOAS1349SUPPA); .pdf). A supplementary appendix providing further experimental details, some derivations and extended results on the copepod data.

**Code for Jensen effect** (DOI: [10.1214/20-AOAS1349SUPPB](https://doi.org/10.1214/20-AOAS1349SUPPB); .zip). Code and data to reproduce simulations and real world analyses contained in this paper.

## REFERENCES

- CHAUDHURI, P. and MARRON, J. S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.* **94** 807–823. [MR1723347 https://doi.org/10.2307/2669996](https://doi.org/10.2307/2669996)
- CHEN, D., HALL, P. and MÜLLER, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39** 1720–1747. [MR2850218 https://doi.org/10.1214/11-AOS882](https://doi.org/10.1214/11-AOS882)
- CHESSON, P. (1994). Multispecies competition in variable environments. *Theor. Popul. Biol.* **45** 227–276.
- CHESSON, P. (2000a). Mechanisms of maintenance of species diversity. *Ann. Rev. Ecol. Syst.* **31** 343–366.
- CHESSON, P. (2000b). General theory of competitive coexistence in spatially-varying environments. *Theor. Popul. Biol.* **58** 211–237.
- CHESSON, P. L. and WARNER, R. R. (1981). Environmental variability promotes coexistence in lottery competitive systems. *Amer. Nat.* **117** 923–943. [MR0647805 https://doi.org/10.1086/283778](https://doi.org/10.1086/283778)
- COHEN, D. (1966). Optimizing reproduction in a randomly varying environment. *J. Theoret. Biol.* **12** 119–129.
- DRAKE, J. M. (2005). Population effects of increased climate variation. *Proc. R. Soc. Lond., B Biol. Sci.* **272** 1823–1827.
- ELLNER, S. (1987). Alternate plant life history strategies and coexistence in randomly varying environments. In *Theory and Models in Vegetation Science* 199–208. Springer, Berlin.
- HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. [MR1212171 https://doi.org/10.1214/aos/1176349020](https://doi.org/10.1214/aos/1176349020)
- HRISTACHE, M., JUDITSKY, A. and SPOKOINY, V. (2001). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.* **29** 595–623. [MR1865333 https://doi.org/10.1214/aos/1009210681](https://doi.org/10.1214/aos/1009210681)
- HUTCHINSON, G. E. (1961). The paradox of the plankton. *Amer. Nat.* **95** 137–145.
- ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58** 71–120. [MR1230981 https://doi.org/10.1016/0304-4076\(93\)90114-K](https://doi.org/10.1016/0304-4076(93)90114-K)
- KOONS, D. N., METCALF, C. J. E. and TULJAPURKAR, S. (2008). Evolution of delayed reproduction in uncertain environments: A life-history perspective. *Amer. Nat.* **172** 797–805.
- KOONS, D. N., PAVARD, S., BAUDISCH, A., METCALF, J. E. et al. (2009). Is life-history buffering or lability adaptive in stochastic environments? *Oikos* **118** 972–980.
- LEWONTIN, R. C. and COHEN, D. (1969). On population growth in a randomly varying environment. *Proc. Natl. Acad. Sci. USA* **62** 1056–1060. [MR0269328 https://doi.org/10.1073/pnas.62.4.1056](https://doi.org/10.1073/pnas.62.4.1056)
- LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. With discussion and a rejoinder by the author. [MR1137117](https://doi.org/10.2307/2286619)
- LI, J., HUANG, C. and ZHU, H. (2017). A functional varying-coefficient single-index model for functional response data. *J. Amer. Statist. Assoc.* **112** 1169–1181. [MR3735368 https://doi.org/10.1080/01621459.2016.1195742](https://doi.org/10.1080/01621459.2016.1195742)
- MA, S. (2016). Estimation and inference in functional single-index models. *Ann. Inst. Statist. Math.* **68** 181–208. [MR3440219 https://doi.org/10.1007/s10463-014-0488-3](https://doi.org/10.1007/s10463-014-0488-3)
- MARRON, J. S. and ZHANG, J.-T. (2005). SiZer for smoothing splines. *Comput. Statist.* **20** 481–502. [MR2236619 https://doi.org/10.1007/BF02741310](https://doi.org/10.1007/BF02741310)
- RAMSAY, J. O. (2006). *Functional Data Analysis*. Wiley, New York.
- RAMSAY, J. O., HOOKER, G. and GRAVES, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer, Berlin.

- RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression. Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge Univ. Press, Cambridge. MR1998720 <https://doi.org/10.1017/CBO9780511755453>
- SONDEREGGER, D. L., WANG, H., CLEMENTS, W. H. and NOON, B. R. (2009). Using SiZer to detect thresholds in ecological data. *Front. Ecol. Environ.* **7** 190–195.
- TELLER, B. J., ADLER, P. B., EDWARDS, C. B., HOOKER, G. and ELLNER, S. P. (2016). Linking demography with drivers: Climate and competition. *Methods Ecol. Evol.* **7** 171–183.
- VASSEUR, D. A. and MCCANN, K. S. (2007). *The Impact of Environmental Variability on Ecological Systems*. Springer, Berlin.
- WOOD, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 413–428. MR1749600 <https://doi.org/10.1111/1467-9868.00240>
- YE, Z. and HOOKER, G. (2018). Local quadratic estimation of the curvature in a functional single index model. Preprint. Available at [arXiv:1803.09321](https://arxiv.org/abs/1803.09321).
- YE, Z., HOOKER, G. and ELLNER, S. P. (2020). Supplement to “The Jensen effect and functional single index models: Estimating the ecological implications of nonlinear reaction norms.” <https://doi.org/10.1214/20-AOAS1349SUPPA>, <https://doi.org/10.1214/20-AOAS1349SUPPB>