Bayesian Analysis (2008) **3**, Number 1, pp. 121–160

An asymptotic viewpoint on high-dimensional Bayesian testing

Dan J. Spitzner*

Abstract. The Bayesian point-null testing problem is studied asymptotically under a high-dimensional normal-means model. A noninformative prior structure is proposed for general problems, and then refined for the specialized contexts of goodness-of-fit testing and functional data analysis. The associated tests are demonstrated on existing data sets and shown to provide a cornerstone for a toolbox of detailed analysis tools. The conceptual approach is to allow the prior null probability to vary with dimension and with prior dispersion parameters, then to guide its parametrization so that the posterior null probability behaves in accordance with Bayesian asymptotic-consistency concepts. Among the theoretical issues studied are the objectivity of setting the prior null probability to one-half, the Jeffreys-Lindley paradox, and the influence of smoothness constraints.

Keywords: Bayesian testing, high-dimensional testing, rates of testing, functional data analysis, goodness-of-fit testing

1 Introduction

This article studies the problem of testing a point-null hypothesis in high dimensions. The model under consideration is defined componentwise according to

$$Y_{n,j} \quad = \quad \theta_{n,j} + \sigma_{n,j} e_{n,j}, \tag{1}$$

for its j'th component, where $j = 1, \ldots, p_n$, n is an index parameter akin to "sample size," and p_n is "dimensionality," representing the maximum number of observable components at a given n. The variables $e_{n,1}, \ldots, e_{n,p_n}$ are standard-normal errors, which are independent across j (at each n), $\sigma_{n,1}, \ldots, \sigma_{n,p_n}$ are error-variance parameters, and $\theta_{n,1}, \ldots, \theta_{n,p_n}$ are mean parameters. The objective is to test $\mathrm{H}_0 : \theta_{n,j} = 0$ for all $j = 1, \ldots, p_n$ against a general alternative H_1. The investigation studies the asymptotic behavior of Bayesian test procedures as n and p_n increase simultaneously and without bound. Specific attention is given to the context where $n^{-1}\sum_{j=1}^{p_n} j^{2s}(\theta_{n,j}/\sigma_{n,j})^2$ is uniformly bounded across n for some, possibly unknown, $s > 1/2$, and the $\sigma_{n,j}$ each shrink at the rate $1/\sqrt{n}$. The former condition represents a "smoothness" assumption.

Motivation for studying this problem stems primarily from interest in goodness-of-fit (GOF) testing and functional data analysis (FDA). However, the Bayesian testing problem is challenging even in more basic scenarios, especially if prior information is vague or absent, and further challenges arise if the dimensionality is high. Because the priors used in testing place mass on a point-null hypothesis, the standard techniques

*Department of Statistics, University of Virginia, Charlottesville, VA, mailto:spitzner@virginia.edu

used in estimation to construct noninformative priors lead to test procedures that are not sensible. For instance, flat, improper priors, which are commonly used in estimation, lead to test procedures that are sensitive to arbitrary normalizing constants. Though various authors, *e.g.*, Aitkin (1991) and Spiegelhalter and Smith (1982), have proposed specific normalizing constants, and others such as Robert (1993) have proposed clever routes around this problem, in high dimensions those solutions can lead to tests that are insensitive to patterns in the data. An additional challenge arises in GOF-testing and FDA in that the model (1) is there often formulated as a discrete transformation of a "smooth" functional model. For instance, in Section 2, below, this is illustrated in an example application which uses Fourier decomposition to discretize a set of densely measured functions, putting it in the form of (1). The challenge this presents is how to incorporate smoothness assumptions into the testing procedure.

In this article, a new methodology for Bayesian high-dimensional testing is proposed for general analysis under (1) and for analyses specific to the GOF-testing and FDA contexts. Its formulation exploits high-dimensionality by parametrizing the prior so as to connect the prior null probability to prior dispersion parameters, then applies an asymptotic-consistency principle discussed in Diaconis and Freedman (1986) to avoid the problems noted above. The resulting tests are sensitive to the data in high-dimensions and meaningful in the sense of being proper Bayes or limits of proper Bayes procedures. Moreover, the proposed priors are noninformative, but avoid the need to specify arbitrary normalizing constants. To address the GOF-testing and FDA contexts, the asymptotic setup is further refined to accommodate smoothness assumptions. A Bayesian "rates of testing" theory is developed, through which, by adapting ideas from an established frequentist analogue, recommendations are made for configuring the prior for testing under a smooth model.

The proposed tests are demonstrated on two example data sets, in which the testing approach is shown to play a central role in a broader analysis methodology. Various theoretical issues are also investigated in detail. Among them are the objectivity of setting the prior null probability to one-half, the Jeffreys-Lindley paradox in high-dimensions, and the impact of conditioning on prior information provided by a smoothness assumption. Interestingly, the latter investigation leads to a recommendation to *not* condition directly on the smoothness constraints, but instead weight the prior so as induce favorable rates-of-testing properties.

1.1 Related work

There is extensive frequentist literature on GOF-testing and FDA, which includes the book on "smooth" GOF testing by Rayner and Best (1989), and articles by Eubank and Hart (1992), Eubank and LaRiccia (1992), Fan (1996), Inglot and Ledwina (1996), Fan and Lin (1998), Eubank (2000), Aerts, Claeskens, and Hart (2000), Fan and Huang (2001), Fan, Zhang, and Zhang (2001), Abramovich *et al.* (2002), Claeskens & Hjort (2004), and Spitzner (2006). See also Eubank (1999) and Brockwell and Davis (1991) for further discussion of Fourier-series decomposition and related techniques. Frequentist rates of testing theory, whose Bayesian reformulation will be developed below in Section

5.3, is discussed in Ingster (1993), Spokoiny (1996), Fan, Zhang, and Zhang (2001), and Spitzner (2008), among others.

There is a growing Bayesian literature on the use of Bayes factors for GOF testing, especially in the context of testing for a hypothesized distribution. Relevant papers include Verdinelli and Wasserman (1998) (in which a stated goal is to formulate a Bayesian version of smooth GOF testing), Ghosal (2001), Petrone and Wasserman (2002), and Robert and Rousseau (2002).

Models similar to (1) are used in other discussions to study asymptotic properties of Bayes factors. Berger, Ghosh, and Mukhopadhyay (2003) examine details of the Laplace approximation for purposes of establishing consistency of the BIC criterion. Among other insights, that paper describes how the relative rates between n and p_n play a critical role on the behavior of BIC. (Rates between n and p_n will play a critical role here as well.) Other papers study Bayes factors under (possibly non-normal) models analogous to (1) with $\sigma_{n,j}$ shrinking at the rate $1/\sqrt{n}$, but with p_n fixed. In this context, Johnson (2005) considers a multinomial model for GOF testing on p_n fixed categories, for which Bayes factors are derived from the asymptotic distribution of standard contingency-table statistics. Andrews (1994) and Efron and Gous (2001) use similar asymptotics to identify close limiting connections between Bayes factors and frequentist p-values. The frequentist comparisons discussed below in Section 4.6 have a similar flavor as these investigations, but their setup is markedly different in that $p_n \to \infty$ and the errors need not shrink.

The use of improper priors in the Bayesian testing problem is considered in Aitkin (1991) and Spiegelhalter and Smith (1982). A related solution is developed in Robert (1993). These will be discussed in some detail in Section 3. Berger and Sellke (1987) discuss a test procedure based on lower bounds on the posterior null probabilities given by proper priors. Berger and Pericchi (1996) propose a data-splitting approach for the construction of “intrinsic” priors, which are proper. See also Berger (1985 sec. 4.3.3), Berger and Delampady (1987), and Robert (2001, sec. 5.2.5) for further discussion of noninformative priors for testing.

1.2 Organization

The article is organized as follows. Section 2 describes three example applications which motivate the investigation and which will be used later to demonstrate the proposed procedures. Section 3 sets up the Bayesian testing framework, and discusses in detail the challenges of testing in high dimensions mentioned above. Main results are given in Sections 4 and 5. Section 4 motivates and lays out guidelines for asymptotic analysis, then applies them to develop a test procedure for general use under the model (1). The objectivity of setting the prior null probability to one-half and the Jeffreys-Lindley paradox are studied in that section as well. Section 5 formulates a Bayesian rates of testing theory, and configures the proposed high-dimensional procedures to accommodate smoothness assumptions. The impact of prior conditioning on smoothness assumptions is also studied there. Conclusions and discussion are given in Section 6, and technical arguments appear in the appendices.

1.3 Notation

Our notation will use boldface to indicate finite vectors and matrices. Basic quantities are $\boldsymbol{Y}_n = [Y_{n,1}, \ldots, Y_{n,p_n}]^T$, $\boldsymbol{\theta}_n = [\theta_{n,1}, \ldots, \theta_{n,p_n}]^T$, $\boldsymbol{e}_n = [e_{n,1}, \ldots, e_{n,p_n}]^T$, and $\boldsymbol{\Sigma}_n = \text{diag}(\sigma^2_{n,1}, \ldots, \sigma^2_{n,p_n})$. It will be convenient to also define the scaled vectors $\boldsymbol{Y}^{SB}_n = [Y^{SB}_{n,1}, \ldots, Y^{SB}_{n,p_n}]^T$ and $\boldsymbol{\theta}^{SB}_n = [\theta^{SB}_{n,1}, \ldots, \theta^{SB}_{n,p_n}]^T$, for which $Y^{SB}_{n,j} = Y_{n,j}/\sigma_{n,j}$ and $\theta^{SB}_{n,j} = \theta_{n,j}/\sigma_{n,j}$, where the superscript "SB" means "scaled for the basic model" (An updated notation which is "scaled for the smooth model" will be defined in Section 5.) Corresponding arrays are $Y = \{Y_{n,j} : (n,j) \in \mathcal{I}_n\}$, $\theta = \{\theta_{n,j} : (n,j) \in \mathcal{I}_n\}$, $e = \{e_{n,j} : (n,j) \in \mathcal{I}_n\}$, $\Sigma = \{\sigma_{n,j} : (n,j) \in \mathcal{I}_n\}$, $Y^{SB} = \{Y^{SB}_{n,j} : (n,j) \in \mathcal{I}_n\}$, and $\theta^{SB} = \{\theta^{SB}_{n,j} : (n,j) \in \mathcal{I}_n\}$, where the index set is $\mathcal{I}_n = \{(n,j) : j = 1, \ldots, p_n; n = 1, 2, \ldots\}$.

Asymptotic analysis will hold θ and Σ fixed but treat Y through the distribution of the standardized-error array e. See Section 4.1, below, for clarification of the mode of convergence. Associated probabilities, expectations, *etc.*, shall be denoted $P[Y \in A]$, $E[h(Y)]$, *etc.* When θ or Σ are to follow a specified prior, associated probabilities and conditional probabilities shall be indicated using boldface and the vertical bar symbol, *e.g.*, $\boldsymbol{P}_n[\theta \in A]$, $\boldsymbol{P}_n[\theta \in A|\boldsymbol{Y}_n]$,$\boldsymbol{E}_n[h(\theta)]$, $\boldsymbol{E}_n[h(\theta)|\boldsymbol{Y}_n]$, *etc.* (The subscript on $\boldsymbol{P}_n$ and $\boldsymbol{E}_n$ is to reflect that the prior may depend on n.)

The notation $\|\boldsymbol{c}_n\| = \{\sum_j c^2_{n,j}\}^{1/2}$ identifies the standard Euclidean norm of the vector $\boldsymbol{c}_n = [c_{n,1}, \ldots, c_{n,p_n}]^T$; for two vectors, $\boldsymbol{c}_n$ and $\boldsymbol{d}_n = [d_{n,1}, \ldots, d_{n,p_n}]^T$, write $\|\boldsymbol{c}_n \boldsymbol{d}_n\| = \{\sum_j c^2_{n,j} d^2_{n,j}\}^{1/2}$, with obvious generalization to multiple vectors. For arrays $\{a_{n,j} : (n,j) \in \mathcal{I}_n\}$ and $\{b_{n,j} : (n,j) \in \mathcal{I}_n\}$, the statement $a_{n,j} = O(b_{n,j})$ will here mean that $|a_{n,j}/b_{n,j}|$ is bounded; $a_{n,j} \asymp b_{n,j}$ means that both $a_{n,j} = O(b_{n,j})$ and $b_{n,j} = O(a_{n,j})$, and $a_{n,j} = o(b_{n,j})$ means that for each $\epsilon > 0$ there are indices n_* and j_* such that $\sup\{|a_{n,j}/b_{n,j}| : n \geq n_*, j \geq j_*\} < \epsilon$. In addition, $a_{n,j} \propto b_{n,j}$ means $a_{n,j}/b_{n,j} = c$ for some constant c and $a_n \approx b_n$ means $a_n/b_n = 1 + o(1)$. The analogues for sequences and other asymptotic contexts will be denoted in a parallel manner.

2 Relevant applications

This section describes three example applications which motivate and help to illustrate the concepts discussed in the later sections. Novel analyses of the data in Examples 2 and 3 will be presented later in Sections 4.5 and 5.3, in which the proposed tests and supplemental analysis tools will be demonstrated.

The first example of a simple GOF-testing scenario illustrates an asymptotic setup which is of special interest to this investigation.

Example 1. (Fourier-series regression) *Suppose n measurements $X_1, \ldots, X_n$ of a noisy regression function are taken along an equally spaced grid, $t_k = -\pi + 2k\pi/n$ for $k =$*

$1, \ldots, n$, and modeled according to $X_k = \mu(t_k) + \sigma\epsilon_k$, where μ is an underlying square-integrable regression function on $(-\pi, \pi]$, $\sigma > 0$, and $\epsilon_1, \ldots, \epsilon_n$ are independent standard-normal errors. The interest is a test for "no effect," $H_0 : \mu(t) = 0$ across t against a general alternative.

Though $X_k = \mu(t_k) + \sigma\epsilon_k$ itself has the structure of (1) an interesting "coefficient model" arises as follows. First, define the Fourier basis functions $\psi_1, \ldots, \psi_n$ according to: $\psi_1(t) = 1$, $\psi_{2j}(t) = c_j \sin(jt)$, and $\psi_{2j+1}(t) = c_j \cos(jt)$, where $c_1, \ldots, c_{\lfloor (n-1)/2 \rfloor} = \sqrt{2}$ and, if needed, $c_{n/2} = 1$. ($\lfloor x \rfloor$ is the largest integer not to exceed x.) It follows that $n^{-1} \sum_{k=1}^{n} \psi_{j_1}(t_k)\psi_{j_2}(t_k) = I_{j_1,j_2}$ and $n^{-1} \sum_{j=1}^{n} \psi_j(t_{k_1})\psi_j(t_{k_2}) = I_{k_1,k_2}$ where $I_{j,k} = 1$ if $j = k$ and 0 otherwise. (To see this, write each ψ_j in polar coordinates and apply the geometric formula, as in Brockwell and Davis, 1991, for. 2.11.3.) Thus it becomes a simple exercise to show that the associated Fourier coefficients $Y_{n,j} = n^{-1} \sum_{k=1}^{n} X(t_k)\psi_j(t_k)$ follow the model (1), with parameters $p_n = n$, $\theta_{n,j} = n^{-1} \sum_{k=1}^{n} \mu(t_k)\psi_j(t_k)$, $e_{n,j} = n^{-1/2} \sum_{k=1}^{n} \epsilon_k \psi_j(t_k)$, and $\sigma_{n,j} = \sigma/\sqrt{n}$. Moreover, we have the inversion formula $X(t_k) = \sum_{j=1}^{n} Y_{n,j}\psi_j(t_k)$. Now $\theta_{n,j} \to \int_{-\pi}^{\pi} \mu(t)\psi_j(t)dt$ for each j as $n \to \infty$, which will not all vanish if $\mu(t) \neq 0$ on a nontrivial subset of $(-\pi, \pi]$. Hence the null hypothesis $H_0 : \mu(t) = 0$ across t translates to $H_0 : \theta_{n,j} = 0$ for every $j = 1, \ldots, p_n$. Also, $\sigma_{n,j} \to 0$ for each j as $n \to \infty$; i.e., *the magnitude of noise associated with any specific component shrinks as p_n increases.*

More complicated GOF scenarios will often translate to a coefficient model in a manner roughly paralleling the construction of this example. Such scenarios include tests for lack of structure in the residuals of a parametric regression (*cf.* Fan and Huang, 2001), or might substitute an alternative set of basis function for the Fourier functions (see, *e.g.*, Eubank, 1999). Still more complicated scenarios involve measurements taken on an uneven or random grid, and these may still give rise to a coefficient model, although independence may hold in the limit only. One shared feature of these coefficient models is an asymptotic setup similar to that of Example 1, in which increases in p_n are coupled with shrinking error-variances. In fact, Brown and Low (1996) establish that (1), with error-variance parameters shrinking at the rate $1/\sqrt{n}$, forms a canonical asymptotic model for non-parametric regression in general. Moreover, Nussbaum (1996) uses discrete approximation (in similar spirit as Fourier decomposition) to establish the same canonical asymptotic model for the distribution context of GOF-testing: the null hypothesis is that an independent and identically distributed random sample $X_1, \ldots, X_n$ follows a specified probability density.

The next example uses similar concepts in an applied FDA testing problem.

Example 2. (Functional data analysis) *Figure 1 displays scatterplots of "vertical density profiles" (VDPs) from twenty-four newly manufactured particleboards. The data set has been divided into profiles corresponding to $n_1 = 20$ boards manufactured during daytime shifts (top-left panels), and $n_2 = 4$ boards manufactured by a different team of operators in overnight shifts (bottom-left panels). Each particleboard has a 0.628-inch thickness, along which the corresponding VDP measures board-densities at $P = 314$ fixed locations on an even grid, $t_k = 0.628k/P$ for $k = 1, \ldots, P$. (The parameter P is distinct from p_n, which is defined below.) These were obtained by a laser device at increments*

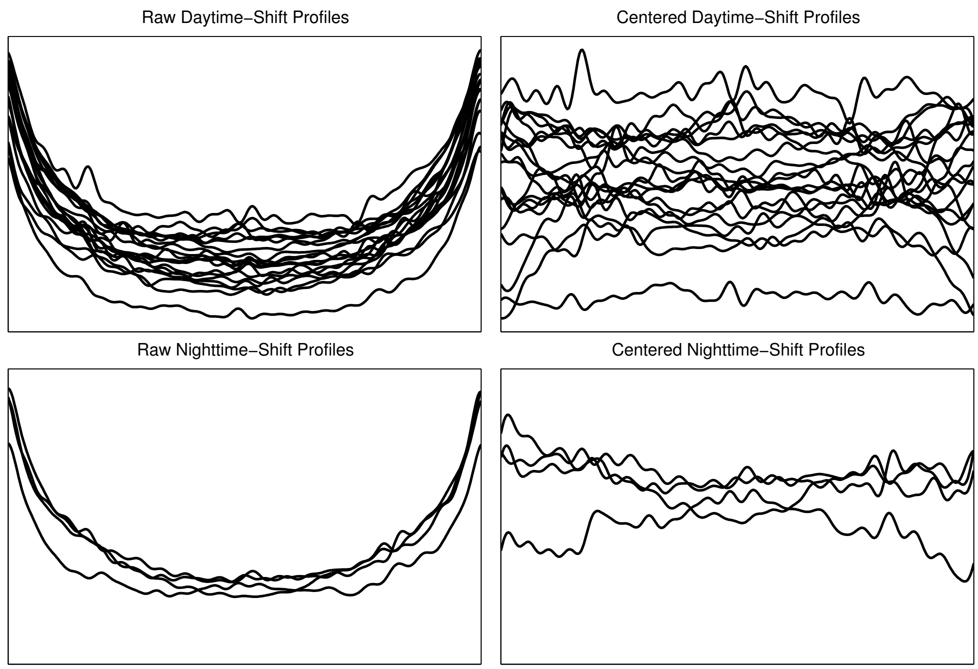

Figure 1: *Raw and centered density profiles for daytime and nighttime shifts. The horizontal axis indicates location along the 0.628-inch thickness of the boards, and the vertical axis indicates board-density. The top panels graph the density measurements of $n_1 = 20$ boards manufactured in daytime shifts, while the bottom panels graph those of $n_2 = 4$ boards manufactured in nighttime shifts. The left panels show raw density measurements, while the right panels differences from the cross-sectional average of the $n_1 = 20$ daytime-shift profiles. The model assumes the left panels are stationary Gaussian functions.*

of 0.002 inches. Let $g = 1, 2$ indicate the respective daytime-shift and nighttime-shift groups and denote by $X_{g,i}(t)$ the density measurement of the i'th board in group g at location t along its thickness. These raw profiles, $X_{g,i}(t)$, are shown in the left panels of Figure 1, with t forming the horizontal axis. The centered profiles shown in the right panels are $X^C_{g,i}(t) = X_{g,i}(t) - \bar{X}_1(t)$ where $\bar{X}_1(t) = n_1^{-1} \sum_{i=1}^{n_1} X_{1,i}(t)$.

In a previous analysis, Walker and Wright (2002) applied techniques for generalized additive models to measure variability in these data due to such sources as changes in profile-specific smoothed fits and changes across shift groups. The interest here is to formally test for differences across shift groups. It shall be assumed that for each $g = 1, 2$ there is a common "mean profile" $\mu_g(t) = E[X_{g,i}(t)]$, so that the null hypothesis to be tested is $H_0 : \mu_1(t) = \mu_2(t)$ for every t, against general alternatives.

To carry out the test, Fourier-basis decomposition will first be used to translate the *(centered) functional data* $X^C_{g,i}(t)$ *to a coefficient model. Define* $\psi_1, \ldots, \psi_P$ *as in Example 1 (with* P *replacing* n*) and set* $\psi^{VDP}_j(t) = \psi_j(-\pi + 2\pi t/0.628)$*. Fourier coefficients of the individual centered curves are* $X_{g,ij} = P^{-1}\sum_{k=1}^{P} X^C_{g,i}(t_k)\psi^{VDP}_j(t_k)$*. Next write* $\bar{X}_{g,j} = n_g^{-1}\sum_{i=1}^{n_g} X_{g,ij}$*,* $\mu_{g,j} = E[X_{g,ij}]$*,* $V_j = V[X_{g,ij}]$*, and* $n = n_1 + n_2$*. The components of model (1) are now given by* $Y_{n,j} = \bar{X}_{1,j} - \bar{X}_{2,j}$*,* $\theta_{n,j} = \mu_{1,j} - \mu_{2,j}$*,* $\sigma^2_{n,j} = (n_1^{-1} + n_2^{-1})V_j$*, and* $e_{n,j} = (Y_{n,j} - \theta_{n,j})/\sigma_{n,j}$*. The null hypothesis* $H_0 : \mu_1(t) = \mu_2(t)$ *for every* t *translates to* $H_0 : \theta_{n,j} = 0$ *for every* j*.*

Justification for the distributional properties of the $e_{n,j}$ *derives from an assumption that each* $X_{g,i}(t)$ *is an observation of a Gaussian random process such that the centered profiles* $X^C_{g,i}(t)$ *are stationary with an absolutely summable covariance function. It follows that the coefficients* $X_{g,ij}$ *are normal and asymptotically independent* (cf., *Chapter 10 of Brockwell and Davis, 1991). Nevertheless, a diagnostic check will indicate that independence among the* $X_{g,ij}$ *across* j *is corrupted for very large* j*, which is likely due to inaccuracies of the model at profile end-regions. A conservative assessment is that independence is preserved for indices* $j \leq 51$*. (These correspond to the flat basis function* ψ^{VDP}_1 *along with the first 25 pairs of sine and cosine basis-functions.) One possible remedy is to attempt to find a basis other than* $\psi^{VDP}_1, \ldots, \psi^{VDP}_P$ *which better de-correlates the data. Instead we will take* $p_n << P$*, for which here it suffices to set* $p_n = 51$*. Refer to Spitzner, Marron, and Essick (2003) for more discussion of this type of modeling complication.*

This second example involves similar techniques as in Example 1, specifically in its use of Fourier-series decomposition, but differs in its asymptotic setup. In Example 2 there are multiple samples sizes, and they are unrelated to the dimensionality parameter p_n. Nevertheless, multiple sample sizes are accommodated asymptotically by writing $\sigma_{n,j} = \tilde{\sigma}_{n,j}/\sqrt{n}$, for which $\tilde{\sigma}^2_{n,j} = V_j\{\gamma(1-\gamma)\}^{-1}$ where $\gamma = n_1/(n_1 + n_2)$, and $n = n_1 + n_2$. Then treating γ as a quantity independent of n (or at any rate asymptotically constant as $n \to \infty$), the rate at which the errors shrink is $1/\sqrt{n}$, the same rate as in the GOF problem. As for the parameter p_n, consider that dimensionality is for the most part determined by the resolution of the digitizing instrument (along with empirical checks on the suitability of the model). Since a finer grid is always plausible this makes p_n an *arbitrarily* large quantity, for which $p_n \to \infty$ represents a suitable conceptualization. There is no explicit connection between n and p_n, however, so the rate at which p_n diverges relative to n is left hypothetical.

A characteristic shared by both Examples 1 and 2 is their involvement of functional parameters (the "mean functions" μ in Example 1 and the μ_g in Example 2) which might reasonably be considered continuous and "smooth." After translating to a coefficient model, a suitable technique for imposing a smoothness assumption is to require the $\theta_{n,j}$ to satisfy $n^{-1}\sum_{j=1}^{p_n} j^{2s}(\theta_{n,j}/\sigma_{n,j})^2 \leq M$ across $n = 1, 2, \ldots$, for fixed "smoothness parameters" $s > 1/2$ and $M > 0$. It is usually possible to translate this restriction to a constraint on the original functional representation, requiring that the underlying mean function has a certain number (depending on s) of derivatives. In other words, the mean function is constrained to an element of a "Sobolev space." (This connection is a consequence of Parseval's identity. For details, see Adams and Fournier, 2003.) More

School	A	B	C	D	E	F	G	H
X_j	28	8	-3	7	-1	1	18	12
S_j	15	10	16	11	9	11	10	18

Table 1: *Aptitude-test coaching data. The X_j are "effects" of a school's coaching program for a standardized aptitude test. The S_j are associated standard errors, which are taken as fixed.*

intuitively, it is seen that uniform boundedness of $n^{-1}\sum_{j=1}^{p_n} j^{2s}(\theta_{n,j}/\sigma_{n,j})^2$ restricts high-indexed entries of the scaled mean parameters more heavily than low-indexed entries. From this viewpoint, the parameters s and M play the following roles: s, controls the strength of restrictions placed on θ and M controls the size of the space.

Among the GOF-testing and FDA literature, smoothness assumptions are sometimes alternatively formulated as a restriction to a "Besov space," in place of a Sobolev space. This is especially true in regard to statistical methodology based on wavelet basis functions. (See, *e.g.*, Abramovich *et al.*, 2002). Nevertheless, the Sobolev constraint, expressed through θ as above, is suitable across a wide range of applications.

The next example is simpler than the previous two, but will serve to illustrate issues at the core of the high-dimensional testing problem.

Example 3. (Multi-sample testing) *Gelman* et al. *(2004, sec. 5.5 and p. 185-186) describe a study of the effect of "coaching" high-school students on a certain standardized aptitude test. The data are laid out in $g = 8$ groups, each representing a distinct high school that administers a coaching program for students preparing to take the test. Subsequent to the receipt of test scores, those of students who completed the coaching programs were compiled and passed through a statistical processing step to produce estimated "effects," $X_1, \ldots, X_g$, of the schools' coaching programs, along with estimates of standard deviations, $S_1, \ldots, S_g$. These are listed in Table 1. Here, as in Gelman* et al. *(2004), the X_j will be assumed independent and to follow normal distributions, and the S_j will be taken as fixed assessments of the $\sqrt{V[X_j]}$. (Uncertainty in each S_j is small as the number of students involved at each school was quite large.) That is, it shall be assumed $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{V})$ where $\boldsymbol{X} = [X_1, \ldots, X_g]^T$, $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_g]^T$ with $\mu_j = E[X_j]$, and $\boldsymbol{V} = diag(S_1^2, \ldots, S_g^2)$. The interest is to test whether the coaching programs of all eight high schools have identical effects, $H_0 : \mu_j = \mu_k$ for all $j, k = 1, \ldots, g$ versus a general alternative.*

Translation to a relevant version of model (1) is carried out according to $\boldsymbol{Y}_n = [Y_{n,1}, \ldots, Y_{n,p_n}]^T = \boldsymbol{C}^{-1/2}\boldsymbol{X}^D$, where $p_n = g - 1$, $\boldsymbol{X}^D = [X_1 - X_2, \ldots, X_{g-1} - X_g]^T$, and $\boldsymbol{C}$ is the covariance matrix of $\boldsymbol{X}^D$, which has (j,k) entry $S_j^2 + S_{j+1}^2$ if $j = k$, $-S_j^2$ if $k = j - 1$, $-S_{j+1}^2$ if $k = j + 1$, and 0 otherwise. It follows that the $Y_{n,j}$ are independent, normally distributed, and each has $\sigma_{n,j}^2 = V[Y_{n,j}] = 1$. There is also an inversion formula, $\boldsymbol{X}^D = \boldsymbol{C}^{1/2}\boldsymbol{Y}_n$, from which it is seen that the null hypothesis translates appropriately to $H_0 : \theta_{n,j} = 0$ for all $j = 1, \ldots, p_n$, where $\theta_{n,j} = E[Y_j]$. A relevant asymptotic setup for this example is formulated as increases in the number of *groups, irrespective of number of students participating in the coaching programs. Thus, n is never defined in this example; it is a spurious parameter and asymptotic analysis is defined directly through increasing p_n.*

3 Challenges of high-dimensional Bayesian testing

The testing concepts applied in this investigation will follow a standard Bayesian setup (*cf.* Berger, 1985, or Robert, 2001): a prior mass $\rho_{0,n} \in (0,1)$ is placed on the null hypothesis and a continuous distribution $(1-\rho_{0,n})\boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\mathrm{H}_1)$ is placed on the alternative, where $\boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\mathrm{H}_1)$ is a specified density. In much of the evaluation below the $\sigma_{n,j}$ are treated as known, but in general contexts it shall be assumed there is an array $\hat{\Sigma} = \{\hat{\sigma}_{n,j} : (n,j) \in \mathcal{I}_n\}$, for which each $\hat{\boldsymbol{\Sigma}}_n = \mathrm{diag}(\hat{\sigma}^2_{n,1}, \ldots, \hat{\sigma}^2_{n,p_n})$ is independent of $\boldsymbol{Y}_n$ and follows a distribution that depends on $\boldsymbol{\Sigma}_n = \mathrm{diag}(\sigma^2_{n,1}, \ldots, \sigma^2_{n,p_n})$ but not $\boldsymbol{\theta}_n$. The error-variances would then be treated with their own prior specification, $\boldsymbol{\pi}_n(\boldsymbol{\Sigma}_n)$; specific forms and are discussed in Section 4.3. The evidence provided by $\boldsymbol{Y}_n$ and $\hat{\boldsymbol{\Sigma}}_n$ about H_0 is reported as the posterior probability of H_0, $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$, or $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ when the $\sigma_{n,j}$ are known. One might instead report a Bayes factor $\boldsymbol{B}_n(\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n) = \{\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]/\rho_{0,n}\}/\{(1-\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n])/(1-\rho_{0,n})\}$, as recommended by Kass and Raftery (1995); however, for most of our exposition it is simpler to focus on $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$, noting that equivalent interpretations are possible through transformation. Denoting by $m(\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n|\boldsymbol{\theta}_n, \boldsymbol{\Sigma}_n)$ a density for the model, the posterior probability is calculated according to the formula

$$\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n] \quad = \quad \left[1 + \left(\rho_{0,n}^{-1} - 1\right) \frac{\boldsymbol{m}_n(\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n|\mathrm{H}_1)}{\boldsymbol{m}_n(\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n|\mathrm{H}_0)}\right]^{-1}, \tag{2}$$

where

$$\begin{aligned}
\boldsymbol{m}_n(\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n|\mathrm{H}_0) &= \int m(\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n|\boldsymbol{0}, \boldsymbol{\Sigma})\boldsymbol{\pi}_n(\boldsymbol{\Sigma}_n)d\boldsymbol{\Sigma}_n, \\
\boldsymbol{m}_n(\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n|\mathrm{H}_1) &= \int\int_{\boldsymbol{\theta}_n \neq \boldsymbol{0}} m(\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n|\boldsymbol{\theta}_n, \boldsymbol{\Sigma}_n)\boldsymbol{\pi}_n(\boldsymbol{\theta}_n, \boldsymbol{\Sigma}_n|\mathrm{H}_1)d\boldsymbol{\theta}_n d\boldsymbol{\Sigma}_n.
\end{aligned}$$

When the $\sigma_{n,j}$ are known, these reduce to $\boldsymbol{m}_n(\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n|\mathrm{H}_0) = m(\boldsymbol{Y}_n|\boldsymbol{0})$ and $\boldsymbol{m}_n(\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n \mid \mathrm{H}_1) = \boldsymbol{m}_n(\boldsymbol{Y}_n|\mathrm{H}_1) = \int_{\boldsymbol{\theta}_n \neq \boldsymbol{0}} m(\boldsymbol{Y}_n|\boldsymbol{\theta}_n)\ \boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\mathrm{H}_1)d\boldsymbol{\theta}_n$, where the model and conditional prior densities are now denoted $m(\boldsymbol{Y}_n|\boldsymbol{\theta}_n)$ and $\boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\mathrm{H}_1)$, respectively.

Our investigation shall take as its starting point the case where the $\sigma_{n,j}$ are known and the prior on $\boldsymbol{\theta}_n$ is specified according to $\boldsymbol{\theta}_n|\mathrm{H}_1 \sim N(\boldsymbol{\xi}_n, \tau_n^2\boldsymbol{W}_n\boldsymbol{\Sigma}_n)$, where $\boldsymbol{\xi}_n = [\xi_1, \ldots, \xi_{p_n}]^T$ is a mean vector, τ_n is an overall scale parameter, and $\boldsymbol{W}_n = \mathrm{diag}(w_{n,1}, \ldots, w_{n,p_n})$ is a diagonal matrix of weights. The model density, under (1), follows $\boldsymbol{Y}_n|\boldsymbol{\theta}_n \sim N(\boldsymbol{\theta}_n, \boldsymbol{\Sigma}_n)$ so that, unconditionally with respect to $\boldsymbol{\theta}_n$ in H_1, $\boldsymbol{Y}_n|\mathrm{H}_1 \sim N(\boldsymbol{\xi}_n, \boldsymbol{\Sigma}_n + \tau_n^2\boldsymbol{W}_n\boldsymbol{\Sigma}_n)$, which defines $\boldsymbol{m}_n(\boldsymbol{Y}_n|\mathrm{H}_1)$. Also, the full posterior distribution is such that $\boldsymbol{\theta}_n|\mathrm{H}_1, \boldsymbol{Y}_n \sim N(\{\boldsymbol{I} + (\tau_n\boldsymbol{W}_n)^{-1}\}^{-1}\boldsymbol{Y}_n, \{\boldsymbol{I} + (\tau_n\boldsymbol{W}_n)^{-1}\}^{-1}\boldsymbol{\Sigma}_n)$. In the analogous estimation problem, this setup reflects a standard starting point for the construction of noninformative priors. As stated, the prior is, of course, informative as it places the bulk of its mass in H_1 near the point $\boldsymbol{\xi}_n$. To modify it so as to become noninformative, two standard approaches (in the estimation problem) are first to set the overall scale parameter τ_n arbitrarily large at each fixed n, and second to place noninformative hyperpriors on the parameters $(\boldsymbol{\xi}_n, \tau_n)$. A third, unconventional approach introduced in Robert (1993) will also be examined, which is specific to testing. None admit a satisfactory test procedure for the high-dimensional context, but following their logic illustrates some of the challenges at hand, and routes to suitable solutions.

3.1 Disconnect from "noninformative" priors for estimation

Let us first take $\boldsymbol{\xi}_n = 0$, so that the required posterior probability (2) is

$$\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \quad = \quad \left[1 + \left(\rho_{0,n}^{-1} - 1\right) \exp\left\{\frac{1}{2}\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SB}\|^2 + \frac{1}{2}\sum_{j=1}^{p}\log(1 - v_{n,j,1})\right\}\right]^{-1} \quad (3)$$

where $\boldsymbol{Y}_n^{SB}$ is defined in Section 1.3, and $\boldsymbol{v}_{n,k} = [v_{n,1,1}, \ldots, v_{n,p_n,1}]$ has j'th entry $v_{n,j,k} = \{1 + 1/(\tau_n^2 w_{n,j})\}^{-k}$. (Here $k = 1$, but other values will be used subsequently.) If one näively sends $\tau_n \to \infty$ (independently of n), the following complication is readily seen: for fixed n and $\rho_{0,n}$, each $v_{n,j,1} \to 1$ as $\tau_n \to \infty$ and so also $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 1$ (assuming a fixed data-array, Y). The resulting test is therefore unsatisfactory as it is completely insensitive to the patterns in the data.

The alternative of placing flat hyperpriors on prior parameters illustrates the sensitivity of the testing procedure to arbitrary normalizing constants. Consider the hyperprior density on $(\boldsymbol{\xi}_n, \tau_n)$ which is specified according to $\boldsymbol{\pi}_n(\boldsymbol{\xi}_n, \tau_n) = \boldsymbol{\pi}_n(\boldsymbol{\xi}_n|\tau_n)\boldsymbol{\pi}_n(\tau_n) = c_n\boldsymbol{\pi}_n(\tau_n)$, so that $\boldsymbol{\pi}_n(\boldsymbol{\xi}_n|\tau_n) \propto 1$ defines an improper conditional distribution $\boldsymbol{\xi}_n|\tau_n$, with c_n its explicit normalizing constant. In the estimation problem, posterior computations are insensitive to normalizing constants such as c_n. However, in the present case, writing $\boldsymbol{m}_n(\boldsymbol{Y}_n|\mathrm{H}_1) = \boldsymbol{m}_n(\boldsymbol{Y}_n|\mathrm{H}_1, \boldsymbol{\xi}_n, \tau_n)$ to indicate explicit dependence on $\boldsymbol{\xi}_n$ and τ_n, one has $\boldsymbol{m}_n(\boldsymbol{Y}_n|\mathrm{H}_1) = \int\int \boldsymbol{m}_n(\boldsymbol{Y}_n|\mathrm{H}_1, \boldsymbol{\xi}_n, \tau_n)\boldsymbol{\pi}_n(\boldsymbol{\xi}_n, \tau_n)d\boldsymbol{\xi}_n d\tau_n = c_n \int \boldsymbol{\pi}_n(\tau_n)d\tau_n$, and the posterior null probability (2) is

$$\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] = \quad (4)$$

$$\left[1 + c_n\left(\rho_{0,n}^{-1} - 1\right)\left\{\int \boldsymbol{\pi}_n(\tau_n)d\tau\right\}\exp\left\{\frac{p_n}{2}\left(\frac{1}{p_n}\|\boldsymbol{Y}_n^{SB}\|^2 + \frac{1}{p_n}\sum_{j=1}^{p_n}\log(2\pi\sigma_{n,j}^2)\right)\right\}\right]^{-1}$$

which clearly depends on c_n. If $\boldsymbol{\pi}_n(\tau_n)$ is improper the problem worsens, for then $\int \boldsymbol{\pi}_n(\tau_n)d\tau_n = \infty$, so $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] = 0$ always, and the test is insensitive to patterns in the data. (Note also that if $\boldsymbol{\pi}_n(\tau_n)$ is proper, formula 4 may be derived alternatively by setting $\boldsymbol{\pi}_n(\theta|\mathrm{H}_1) = c_n$.)

There is existing discussion that seeks justification for specific choices of c_n. For instance, Aitkin (1991) proposes a double-use of data in order to connect c_n to a relevant posterior probability. Spiegelhalter and Smith (1982) propose fixing c_n at the value that sets a Bayes factor to one when the data is "most favorable" to H_0 in an imaginary experiment of minimal sample-size. Yet even if non-arbitrary choices for c_n can be found (and $\boldsymbol{\pi}_n(\tau_n)$ is proper), formula (4) implies that the test may have an unreasonable dependency on the error-variance parameters in high-dimensions. To see this, suppose $c_n \asymp 1$, $\int \boldsymbol{\pi}_n(\tau_n)d\tau_n = 1$, $\rho_{0,n} \asymp 1/2$, and limits exist for both $p_n^{-1}\sum_{j=1}^{p_n}\log\sigma_{n,j}^2$ and $p_n^{-1}\|\boldsymbol{\theta}_n^{SB}\|^2$. It follows that $p_n^{-1}\|\boldsymbol{Y}_n^{SB}\|^2 \to 1+p_n^{-1}\|\boldsymbol{\theta}_n^{SB}\|^2$ (where convergence is "almost sure," as discussed below in Section 4.1); hence $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 1$ if $\lim_n p_n^{-1}\|\boldsymbol{\theta}_n^{SB}\|^2 < -1-\lim_n p_n^{-1}\sum_{j=1}^{p_n}\log(2\pi\sigma_{n,j}^2)$, and $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0$ if $\lim_n p_n^{-1}\|\boldsymbol{\theta}_n^{SB}\|^2 > -1-\lim_n p_n^{-1}\sum_{j=1}^{p_n}\log(2\pi\sigma_{n,j}^2)$. Thus, for instance, if each $\sigma_{n,j} > 1/\sqrt{2\pi}$ then $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0$ always and the test is insensitive to the patterns in the data.

3.2 Robert's procedure

Rather than mimicking what is done in the estimation problem, Robert (1993) opens an alternative perspective by exploiting the specific mechanisms of testing. Though we ultimately reject Robert's specific solution, we find his perspective extremely valuable as it plants the seeds of the approach taken here. (The description below has been extended slightly to the multidimensional context; Robert's original paper works exclusively with univariate testing.)

The key idea is to connect H_0 and H_1 through relevant prior parameters: $\rho_{0,n}$ is to be treated as a function of τ_n in such a way that $\rho_{0,n} \to 0$ as $\tau_n \to \infty$. Robert (1993) specifically proposes an "equiponderance device," setting

$$\rho_{0,n} \quad = \quad (1-\rho_{0,n}) \lim_{\boldsymbol{\theta}_n \to 0} \boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\mathrm{H}_1) \tag{5}$$

(assuming $\boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\mathrm{H}_1)$ is continuous in a region around H_0), which he motivates by suggesting that it provides $\boldsymbol{\theta}_n = 0$ equivalent probability under each alternative. The "equiponderance" concept is blurry, however, and, in our opinion, lacks a proper justification. For instance, one inherent flaw in (5) is that by comparing a point mass with a density it introduces an awkward dependency on the scale of measurement. Nevertheless, it is useful to examine the remainder of Robert's construction. Continuing on, under (5) the posterior probability (3) becomes

$$\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \quad = \quad \left[1+(2\pi)^{p_n/2}\exp\left\{\frac{1}{2}\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SB}\|^2+\frac{1}{2}\sum_{j=1}^{p_n}\log(\sigma_{n,j}^2 v_{n,j,1})\right\}\right]^{-1} \tag{6}$$

which has

$$\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \quad \to \quad \left[1+(2\pi)^{p_n/2}\exp\left\{\frac{1}{2}\|\boldsymbol{Y}_n^{SB}\|^2+\frac{1}{2}\sum_{j=1}^{p_n}\log\sigma_{n,j}^2\right\}\right]^{-1} \tag{7}$$

as $\tau_n \to \infty$. The limit in (7) is (the multivariate analogue of) Robert's proposed noninformative solution to the testing problem. However, an immediate problem is seen by noting that this solution is a special case of (4) with $c_n = 1$, $\int \boldsymbol{\pi}_n(\tau_n)d\tau_n = 1$, and $\rho_{0,n} = 1/2$. Thus, in addition to the difficulties of interpretation noted above, there is a problematic dependency on the error variances, and a possible insensitivity to patterns in the data.

The present investigation resolves these difficulties by extending Robert's idea to connect τ_n and $\rho_{0,n}$ so that p_n is also taken into account. Consider the following: Suppose $\sigma_{n,j} = 1$. The posterior probability (3) tends to 1 whenever $\rho_{0,n} \asymp 1/2$ and τ_n increases at sufficiently fast rate, even as p_n increases. On the other hand, $\|\boldsymbol{Y}_n^{SB}\|^2 \to \infty$ ("almost surely," see Section 4.1) and so the limit in (7) tends to 0 for increasing p_n. Thus, in order to make $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ sensitive to the patterns in the data, one would want to choose a rate of decrease for $\rho_{0,n}$ (with respect to τ_n) slower than that induced by the equiponderance device (5), and also choose a rate of increase for τ_n (with respect to p_n) while keeping the rate at which $\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}^{SB}\|^2$ grows in mind. A principal outcome of this article is the development of a precise framework for balancing the rates of $\rho_{0,n}$, τ_n, and p_n in this manner, which is described in Section 4.2.

In addition, one might consider the following version of the Jeffreys-Lindley paradox (*cf.* Lindley, 1957) in which each $Y_{n,j}^{SB} = Y_{n,j}/\sigma_{n,j}$ is regarded as a "z-score" in the Neyman-Pearson testing setup. Set $w_{n,j} = \tilde{w}_{n,j}/\sigma_{n,j}^2$, so as to make the prior independent of the $\sigma_{n,j}^2$. Now $v_{n,j,1} = \{1 + \sigma_{n,j}^2/(\tau_n^2\tilde{w}_{n,j})\}^{-1}$, for which $v_{n,j,1} \to 1$ as $\sigma_{n,j} \to 0$. Thus, from (3), it is seen that if the $Y_{n,j}^{SB}$ were to remain fixed, $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 1$ whenever the $\sigma_{n,j} \to 0$. (Note also that $\sigma_{n,j} \to 0$ reflects the situation of GOF-testing and FDA.) The "paradox" here (really a frequentist criticism) is that the frequentist might "reject" H_0 if the magnitude of $Y_{n,j}^{SB}$ is sufficiently large, even though the Bayesian observes overwhelming evidence in its favor. It is curious that such behavior is also exhibited in the limit in (7), suggesting that Robert's equiponderance device does not necessarily avoid the paradox either. The procedures developed here do avoid the paradox in high-dimensions, as will be shown in Section 4.6, and, in fact, do so regardless of whether $\sigma_{n,j} \to 0$.

Note we have not yet discussed the relevance of the prior weights $w_{n,1}, \ldots, w_{n,p_n}$. Under the type of smoothness considerations used in GOF-testing and FDA, the Bayesian rates-of-testing theory of Section 5 will highlight an important role the weights can play. However, it will be the case that the procedure proposed below for generic situations is insensitive to the choice of weights (this is true for Robert's solution as well; they drop out of the limit in 7).

4 High-dimensional analysis of the basic model

In this section, a set of guidelines for asymptotic analysis are established, then applied to formally develop the testing framework alluded to in Section 3.2.

4.1 Asymptotic criteria

Asymptotic analysis is not inherently sensible in a Bayesian context, as it is in frequentist analysis, and it is necessary to have a formal principle with which to guide our treatment of the high-dimensional testing problem. Diaconis and Freedman (1986) provide what is needed in their "what if" method, which scrutinizes the choice of prior by asking whether, given a particular data set, the posterior distribution makes a meaningful update of the prior. (The name of the method alludes to the question, "What if the data came out that way?") For present purposes, to check for "meaning" it is sufficient to consider the extreme limits of the posterior null probability: $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 1$ is here to mean "overwhelming evidence for H_0" and $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0$ is to mean "overwhelming evidence against H_0." The limits here are "almost sure" with respect to the model (1) for fixed θ; this mode of convergence is stronger than convergence "in probability" and is required by the "what if" method's focus on *data* rather than *probabilities*.

The statements of meaning lead to the following guidelines for how a test is to behave in high-dimensions: (*i.*) One would want to avoid the almost-sure limit $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0$ for θ consistent with H_0; (*ii.*) One would instead want $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 1$ for θ consistent with H_0; (*iii.*) One would want $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0$ for θ consistent with H_1. These three statements will together be referred to as "'what if' guidelines." Observe that (*i*) is a weaker requirement than (*ii*), whereas neither has a logical relation with (*iii*). In what follows, guideline (*i*) will be treated as the more basic consideration compared to (*ii*) and (*iii*). Moreover, Section 4.2 will show that simultaneous, strict achievement of (*ii*) and (*iii*) is not possible, but the best one can hope for is an intelligent trade-off between them.

Because the "what if" guidelines require almost sure limits, it is necessary to clarify a technical aspect of the model (1) and those of Examples 1 and 2. To simplify evaluation, it shall be assumed that $e_{n,j} = e_j$ for some common standardized-error sequence $\{e_j\}$, for it then follows that $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ tends "almost surely" to an extreme limit if and only if it tends "in probability" to the same extreme limit. To see this, consider the transformation $\boldsymbol{U}_n(\boldsymbol{Y}_n) = \log\big(\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]^{-1} - 1\big)$, which, in light of (3), may be written $\boldsymbol{U}_n(\boldsymbol{Y}_n) = \sum_{j=1}^{p_n} h_{n,j}(e_{n,j})$, for suitable $h_{n,j}$. The assertion then follows from Billingsley (1995, th. 22.7). The implication for the Examples 1 and 2 is that the original functional models (μ in Example 1 and μ_1, μ_2 in Example 2) are to be understood through the inverse transformation from the $Y_{n,j}$.

Frequentists (especially) and others may object to this formulation, on the grounds that it reflects an unrealistic scheme for repeated sampling. For instance, the inversion formula $X(t) = \sum_{j=1}^{n} Y_{n,j}\psi_j(t)$ in Example 1 would allow $X(t)$ to vary with n. To this objection, let us offer several comments: The first is that the asymptotic viewpoint taken here, being intended for evaluation of Bayesian procedures, need not be interpreted in the context of frequentist repeated sampling. Rather, it is a tool with which to conceptualize the high-dimensional context. Second, in some applications the Fourier coefficients, $Y_{n,j}$, are the natural focus of analysis and the inversion formula is just a "processing step" for data visualization. For instance, see Spitzner, Marron, and Essick (2003) for an application in which functional data are measured on grids of varying sizes. Direct comparisons between functional measurements are impossible, since the grid points do not line up, and so it is natural to treat the Fourier coefficients as central, and the transformation to them merely as part of the measurement process. Third, it would take us far off track to establish almost-sure limits under less objectionable repeated-sampling schemes. Such detailed assessment would be useful to develop modeling intuition for GOF-testing and FDA contexts, especially for applications in which the model (1) serves only as an asymptotic approximation, but this will be left for follow-up investigations.

4.2 Considerations for prior selection

Our evaluation will focus on tests derived from the prior $\boldsymbol{\theta}|\mathrm{H}_1 \sim N(0, \tau_n^2 \boldsymbol{W}_n \boldsymbol{\Sigma}_n)$, treating $\boldsymbol{\Sigma}_n$ as a known model parameter, for which the associated posterior null probability is given by formula (3). The approach adapts Robert's (1993) idea to impose a dependency structure among the prior parameters, but now, rather than invoking an equiponderance device (5), the correct form of the dependencies is to be determined by the "what if" guidelines. As it turns out, the precise connection between the choice of prior and the "what if" guidelines is made through the following parametrization in which $\rho_{0,n}$ is written in terms of the remaining prior parameters:

$$\log\left(\rho_{0,n}^{-1} - 1\right) \quad = \quad -\frac{1}{2}\left\{\sum_{j=1}^{p} \log\left(1 - v_{n,j,1}\right) + \|\boldsymbol{v}_{n,1/2}\|^2 + r_n \|\boldsymbol{v}_{n,1}\|\right\}, \tag{8}$$

where r_n depends neither on Y nor θ. Note this parametrization does not constrain the prior in any way.

To observe the critical role played by (8), and the sequence $\{r_n\}$ in particular, first observe

$$E\left[\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SB}\|^2\right] \quad = \quad \|\boldsymbol{v}_{n,1/2}\|^2 + \|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SB}\|^2 \tag{9}$$

$$V\left[\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SB}\|^2\right] \quad = \quad 2\|\boldsymbol{v}_{n,1}\|^2 + 4\|\boldsymbol{v}_{n,1}\boldsymbol{\theta}_n^{SB}\|^2, \tag{10}$$

and define the transformation $\boldsymbol{U}_n(\boldsymbol{Y}_n) = \log\left(\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]^{-1} - 1\right)$ of the posterior null probability. The quantity $\boldsymbol{U}_n(\boldsymbol{Y}_n)$ is the posterior log-odds on H_1, and is more convenient to work with than $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ directly. Note that $\boldsymbol{U}_n(\boldsymbol{Y}_n) \to \infty$ is equivalent to $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0$ and $\boldsymbol{U}_n(\boldsymbol{Y}_n) \to -\infty$ is equivalent to $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 1$. Incorporating

(8) into (3), the posterior null probability may be understood through the expressions

$$\boldsymbol{U}_n(\boldsymbol{Y}_n) = \frac{1}{2}\left\{\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SB}\|^2 - \left(\|\boldsymbol{v}_{n,1/2}\|^2 + r_n\|\boldsymbol{v}_{n,1}\|\right)\right\} \tag{11}$$

$$= \|\boldsymbol{v}_{n,1}\|\left[\left\{\frac{1}{2} + \left(\frac{\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SB}\|}{\|\boldsymbol{v}_{n,1}\|}\right)^2\right\}^{1/2} T_n + \frac{1}{2}\left(\frac{\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SB}\|^2}{\|\boldsymbol{v}_{n,1}\|} - r_n\right)\right] \tag{12}$$

$$= \|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SB}\|^2\left[\left\{\frac{1}{2}\left(\frac{\|\boldsymbol{v}_{n,1}\|}{\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SB}\|^2}\right)^2 + \left(\frac{\alpha(\boldsymbol{\theta}_n)}{\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SB}\|}\right)^2\right\}^{1/2} T_n \right. \tag{13}$$

$$\left. + \frac{1}{2}\left(1 - r_n\frac{\|\boldsymbol{v}_{n,1}\|}{\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SB}\|^2}\right)\right],$$

where T_n is $\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SB}\|^2$ standardized to have mean zero and variance one (for fixed θ), and $\alpha(\boldsymbol{\theta}_n^{SB}) = \|\boldsymbol{v}_{n,1}\boldsymbol{\theta}_n^{SB}\|/\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SB}\|$, which has $0 \leq \alpha(\boldsymbol{\theta}_n^{SB}) \leq 1$ (since $v_{n,j,1} \leq v_{n,j,1/2}$).

Considering the essential uniform boundedness of T_n, expression (12) identifies conditions on r_n by which $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ behaves favorably under H_0: observe that terms in (12) involving $\boldsymbol{\theta}_n^{SB}$ can be ignored (since H_0 is $\boldsymbol{\theta}_n^{SB} = 0$); so it is clear that $\boldsymbol{U}_n(\boldsymbol{Y}_n) \to \infty$ (hence $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0$) for θ consistent with H_0 if and only if $r_n \to -\infty$. Following guideline (*i*), this shows that one would want $\limsup_n r_n > -\infty$ under the parametrization (8). Similarly, (12) shows that $\boldsymbol{U}_n(\boldsymbol{Y}_n) \to -\infty$ (hence $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 1$) for θ consistent with H_0 if and only if $r_n \to \infty$. This means $r_n \to \infty$ is required to achieve guideline (*ii*). Expression (13) addresses guideline (*iii*) by identifying θ^{SB} consistent with H_1 for which $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0$: observe from (13) that $\boldsymbol{U}_n(\boldsymbol{Y}_n) \to \infty$ for $\boldsymbol{\theta}_n^{SB} \neq 0$ if and only if both

$$\frac{\|\boldsymbol{v}_{n,1}\|}{\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SB}\|^2} \to 0 \quad \text{and} \quad \lim_n r_n\frac{\|\boldsymbol{v}_{n,1}\|}{\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SB}\|^2} < 1$$

(which follows since then $\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SB}\|^2 \to \infty$). This means there are θ^{SB} in a region around H_0 for which $\limsup_n \boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] > 0$, making clear that guideline (*iii*) cannot be achieved across all of H_1. These exceptional θ^{SB} will be referred to as "indistinguishable," and the collection of them the "indistinguishable region." The size of the indistinguishable region is influenced by the parameter r_n, for if $r_n \to \infty$ there are fewer indistinguishable θ^{SB} when r_n diverges at a slower rate. Together, these observations suggest that to satisfy (*i*) and balance guidelines (*ii*) and (*iii*), one would want $r_n \to \infty$, but diverging as slowly as possible.

There remains a fair bit of ambiguity in how to actually specify r_n, since for each setting there is always another that diverges at a slower rate (*e.g.*, $\log r_n$). Perspectives on the possible subjective elicitation of this parameter are discussed in Section 6. However, it is also possible to recommend some "default" settings for r_n on the basis of model simplicity. One basic simplification arises upon considering that there is no apparent purpose to any direct dependence between r_n and the other prior parameters. In other words, nothing is lost in restricting r_n to a sequence of constants that depends neither on τ_n nor $\boldsymbol{W}_n$. A more drastic simplification is to specify r_n so that $r = \lim_n r_n < \infty$, which is justified as placing r_n "on the lower boundary" of sequences that diverge (slowly) to infinity. This sacrifices the "what if" guideline (*ii*), and moreover leaves ambiguous an appropriate specification of the limit r. Yet, the former issue is not critical, and the latter can be resolved by setting $r_n = 0$ for all p, which is justified as a further simplification achieved by reducing the number of parameters. With this in mind, the setting $r_n = 0$ will be used as a default setting in the application examples below.

For theoretical analysis, the guideline by which r_n is set to diverge "as slowly as possible" creates no difficulty: any property deduced will be qualified with an assumption that r_n diverges more slowly than some given critical rate.

A possible criticism is that the precise form of the parametrization (8) might be seen as *ad hoc* since its role is only illuminated in the limit. For instance, one could modify the individual terms of (8) by introducing leading factors that tend asymptotically to one, without substantially affecting the previous evaluation of $\boldsymbol{U}_n(\boldsymbol{Y}_n)$ or its conclusions. Our answer to such criticism is that (8) is not *ad hoc*, but is the simplest relevant parameterization that reflects this author's subjective understanding of the behavior of asymptotic phenomena. Moreover, such asymptotically negligible modifications would have little impact on the numerical values of $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ if the model is truly high-dimensional. On the other hand, should some principle arise which clarifies or explains the properties of $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ observed in this investigation, and in doing so suggests specific leading factors, it would be sensible to modify (8) accordingly.

4.3 Impact of placing a prior on the error-variances

For the case where the error-variances are unknown, the prior structure may be extended to accommodate a prior on $\boldsymbol{\Sigma}_n$. Suppose the array $\hat{\Sigma}$, introduced in Section 3, is available and is such that each $\hat{\boldsymbol{\Sigma}}_n = \mathrm{diag}(\hat{\sigma}_{n,1}, \ldots, \hat{\sigma}_{n,p_n})$ is independent of $\boldsymbol{Y}_n$ and $\nu_n \hat{\sigma}^2_{n,j}/\sigma^2_{n,j}|\sigma^2_{n,j} \sim \chi^2_{\nu_n}$, independently across j, for some sequence $\{\nu_n\}$. Assume that $\nu_n \approx n$, which is typical, in which case $\hat{\sigma}^2_{n,j}/\sigma^2_{n,j} \to 1$, where convergence is "in probability." The "what if" guidelines require us to also assume the stronger property that such convergence also holds "almost surely."

Consider prior specifications on $\boldsymbol{\Sigma}_n$ such that each $\lambda/\sigma^2_{n,j} \sim \chi^2_\kappa$ for $\kappa, \lambda > 0$, for which densities are $\boldsymbol{\pi}_n(\sigma^2_{n,j}) \propto (\sigma^2_{n,j})^{-(\kappa/2+1)} \exp\{-\lambda/(2\sigma^2_{n,j})\}$. Also consider the improper priors of the same form but indexed with $\kappa = 0$ or $\lambda = 0$, or both. (The setting $\kappa = \lambda = 0$ is a standard noninformative prior which is equivalent to a flat prior on $\log \sigma$.) The same prior on $\boldsymbol{\Sigma}_n$ is assumed for each hypothesis, so that $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$ is well-behaved even in the improper limit. With $\boldsymbol{Y}_n|\mathrm{H}_1, \boldsymbol{\theta}_n, \boldsymbol{\Sigma}_n \sim N(\boldsymbol{\theta}_n, \boldsymbol{\Sigma}_n)$ and

$\boldsymbol{\theta}_n|\mathrm{H}_1, \boldsymbol{\Sigma}_n \sim N(0, \tau_n^2 \boldsymbol{W}_n \boldsymbol{\Sigma}_n)$, as before, the posterior null probability becomes

$$\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n] = \tag{14}$$
$$\left[1 + (\rho_0^{-1} - 1) \prod_{j=1}^{p_n} \left\{ (1 - \upsilon_{n,j,1}) \left(\frac{Y_{n,j}^2 + \nu_n \hat{\sigma}_{n,j}^2 + \lambda}{(1 - \upsilon_{n,j,1})\, Y_{n,j}^2 + \nu_n \hat{\sigma}_{n,j}^2 + \lambda} \right)^{\nu_n + \kappa + 1} \right\}^{1/2} \right]^{-1}.$$

A derivation of (14) is given in Appendix 1. There, it is also shown that the marginal posterior distribution has $\theta_{n,j}|\mathrm{H}_1, Y_{n,j}, \hat{\sigma}_{n,j}^2 \sim \upsilon_{n,j,1} Y_{n,j} + \eta_{n,j} t_{\nu_n+\kappa+1}$, independently across j, where $\eta_{n,j}^2 = \upsilon_{n,j,1}\{(1-\upsilon_{n,j,1})Y_{n,j}^2 + \nu_n \hat{\sigma}_{n,j}^2 + \lambda\}/(\nu_n + \kappa + 1)$ and t_ν denotes Student's t distribution with ν degrees of freedom.

The factor in (14) involving the $Y_{n,j}$ may be written

$$\frac{Y_{n,j}^2 + \nu_n \hat{\sigma}_{n,j}^2 + \lambda}{(1 - \upsilon_{n,j,1})\, Y_{n,j}^2 + \nu_n \hat{\sigma}_{n,j}^2 + \lambda} \quad = \quad 1 + B_{n,j} \frac{\upsilon_{n,j,1}(Y_{n,j}/\hat{\sigma}_{n,j})^2}{\nu_n},$$

where $B_{n,j} = 1/[1 + \{(1-\upsilon_{n,j,1})Y_{n,j}^2 + \lambda\}/(\nu_n \hat{\sigma}_{n,j}^2)]$ for which $0 < B_{n,j} < 1$. Writing $\sigma_{n,j} = \tilde{\sigma}_{n,j}/\sqrt{n}$, note that $(Y_{n,j}/\hat{\sigma}_{n,j})^2/\nu_n \approx (\theta_{n,j}/\tilde{\sigma}_{n,j})^2$ and $B_{n,j} \approx 1/[1 + \{(1 - \upsilon_{n,j,1})\theta_{n,j}^2 + \lambda\}/\tilde{\sigma}_{n,j}^2]$. Thus, since $n \log(1 + 1/n) < 1$ and $n \log(1 + 1/n) \approx 1 - 1/(2n)$, this means

$$\log \left[\prod_{j=1}^{p_n} \left\{ (1 - \upsilon_{n,j,1}) \left(\frac{Y_{n,j}^2 + \nu_n \hat{\sigma}_{n,j}^2 + \lambda}{(1 - \upsilon_{n,j,1})\, Y_{n,j}^2 + \nu_n \hat{\sigma}_{n,j}^2 + \lambda} \right)^{\nu_n + \kappa + 1} \right\}^{1/2} \right]$$
$$= \left\{ \frac{1}{2} + O(\nu_n^{-1}) \right\} \sum_{j=1}^{p_n} \{B_{n,j} \upsilon_{n,j,1} (Y_{n,j}/\hat{\sigma}_{n,j})^2 - C_{n,j}\} + \frac{1}{2} \sum_{j=1}^{p_n} \log(1 - \upsilon_{n,j,1}),$$

where each $C_{n,j} > 0$ and $C_{n,j} \asymp B_{n,j} \upsilon_{n,j,1} (\theta_{n,j}/\tilde{\sigma}_{n,j})^2$.

It has therefore been shown that the posterior probability (14) is asymptotically no smaller than its analogue (3) for the case of known error-variances. Treating (14) as an approximation to (3), its accuracy is degraded somewhat if $\theta_{n,j}^2$ is small or $\tilde{\sigma}_{n,j}^2$ is large. More seriously, however, the term $\sum_{j=1}^{p_n} C_{n,j}$ will typically explode if p_n diverges at a rate faster than n, opening the possibility of substantial inaccuracy. To accommodate, a detailed analysis of (14) might suggest a beneficial adjustment to the parametrization (8) in terms of the "what if" guidelines. Nevertheless, this would be quite complicated and will be investigated elsewhere. Let us instead note that assessments based on (14), with $\rho_{0,n}$ parametrized according to (8), are conservative in the sense that if one were to look beyond the priors specified with $\limsup_n r_n > -\infty$, some with $r_n \to -\infty$ might be found which would shrink the test's indistinguishable region while still making $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n] \to 1$ for θ consistent with H_0.

4.4 Objectivity of $\rho_{0,n} = 1/2$

In low-dimensional testing it is often argued that $\rho_{0,n} = 1/2$ defines an "objective" or "noninformative" setting, but careful examination of the parametrization (8) will show that label may be inappropriate in high dimensions, as is now described. The key insights derive from the following theorem.

Theorem 1. *Assume (8) with $r_n \to \infty$. (*i.*) For fixed n, $\rho_{0,n} \to 0$ as $\tau_n \to \infty$. (*ii.*) $\limsup_n \tau_n^2 \sum_{j=1}^{p_n} w_{n,j} < \infty$ implies $\lim_n \rho_{n,0} = 1$. (*iii.*) Suppose $\{j_n\}$ is a sequence for which $1 \le j_n \le p_n$ and $r_n p_n^{1/2}/(p_n - j_n) \to 0$. If $\liminf_n \{\inf_{j \ge j_n} w_{n,j}\} > 0$ then $\limsup_n \tau_n > 0$ implies $\liminf_n \rho_{0,n} = 0$.*

Proof. Rewrite (8) as $-\{p_n A_n + r_n \|\boldsymbol{v}_{n,1}\|\}/2$ where $A_n = p_n^{-1} \sum_{j=1}^{p_n} \{\log(1 - v_{n,j,1}) + v_{n,j,1}\}$, and note the inequality $\log(1 - v) + v < 0$ for $0 < v < 1$ implies $A_n < 0$. Statement (*i*) follows immediately from $v_{n,j,1} \to 1$, which is easily deduced. To prove statement (*ii*), set $B_n = p_n^{-1} \sum_{j=1}^{p_n} \log(1 - v_{n,j,1})$ so that $B_n < A_n < 0$. Next observe that $\log(1 + x) < x$ for $x > 0$ implies $\log(1 - v_{n,j,1}) = -\log(1 + \tau_n^2 w_{n,j}) > -\tau_n^2 w_{n,j}$. Thus $-C < p_n B_n < p_n A_n < 0$ for any $C > 0$ such that $\tau_n^2 \sum_{j=1}^{p_n} w_{n,j} < C$, hence $-\{p_n A_n + r_n \|\boldsymbol{v}_{n,1}\|\}/2 > \{C - r_n \|\boldsymbol{v}_{n,1}\|\}/2 \to -\infty$. To prove statement (*iii*), observe there is some $\epsilon > 0$ for which $\inf_{j \ge j_{n_k}} v_{n_k,j,1} > \epsilon$ along a subsequence $\{n_k\}$; hence $\sup_{j \ge j_{n_k}} \log(1 - v_{n_k,j,1}) + v_{n_k,j,1} < D$ where $D = \log(1 - \epsilon) + \epsilon < 0$. Therefore $p_{n_k} A_{n_k} < (p_{n_k} - j_{n_k}) D$ and $-\{p_{n_k} A_{n_k} + r_{n_k} \|\boldsymbol{v}_{n_k,1}\|\}/2 \to \infty$, since $\|\boldsymbol{v}_{n,1}\| = O(p_n^{1/2})$. □

Theorem 1 addresses the possibility of setting $\lim_n \rho_{0,n} = 1/2$ by solving (8) for a suitable sequence $\{\tau_n\}$. If $\limsup_n \sum_{j=1}^{p_n} w_{n,j} < \infty$ (which implies $\liminf_n \{\inf_{j \ge j_n} w_{n,j}\} = 0$), then statements (*i*) and (*ii*) imply that $\rho_{0,n} \to 0$ if $\tau_n \to \infty$ at a sufficiently fast rate, and $\rho_{0,n} \to 1$ if $\tau_n \to \infty$ at a sufficiently slow rate. Thus, in this case it is at least plausible there is a setting for τ_n which achieves $\lim_n \rho_{0,n} = 1/2$. However, statement (*iii*) implies that $\lim_n \rho_{0,n} = 1/2$ is impossible if $\liminf_n \{\inf_{j \ge j_n} w_{n,j}\} > 0$ and $\tau_n \to \infty$.

The main consequence, therefore, of Theorem 1 is that $\rho_{0,n} = 1/2$ might be achieved with $\tau_n \to \infty$ only if it is possible to specify $w_{n,1}, \dots w_{n,p_n}$ in such a way that $\liminf_n \{\inf_{j \ge j_n} w_{n,j}\} = 0$. To do this requires information, for it is necessary to first specify a suitable ordering of the dimension indices, then, to set the $w_{n,1}, \dots w_{n,p_n}$ precisely, elicit the relative certainties of $\theta_{n,j} \approx 0$ for smaller *versus* larger j. In generic high-dimensional settings, this clearly compromises the "noninformative" character of the prior. Left unexplored by Theorem 1 are prior structures for which $\tau_n \to 0$. However, these are immediately seen as informative since prior mass becomes increasingly concentrated near H_0.

Thus, each condition $\liminf_n \{\inf_{j \ge j_n} w_{n,j}\} = 0$ and $\tau_n \to 0$ reflects a use of prior information, and so we are led to scrutinize whether $\rho_{0,n} = 1/2$ is a necessary property of noninformative priors. Indeed, a counterargument is provided in an observation by Robert (1993), by which, as $\tau_n \to 0$, prior mass becomes stretched increasingly thin across fixed regions in H_1. From this perspective, $\rho_{0,n} = 1/2$ represents increasing imbalance between H_0 and the region in H_1 surrounding it. In contrast, $\rho_{0,n} \to 0$ describes a sensible reallocation of mass. (Such reallocation is even more imperative here, since, with both $\tau_n \to 0$ and $p_n \to 0$, mass thins out within H_1 even more dramatically.) Thus, though we have rejected Robert's equiponderance device, his suggestion to allow $\rho_{0,n} \to 0$ is reflected in the implications of Theorem 1.

As an additional note, the property $\liminf_n\{\inf_{j\geq j_n} w_{n,j}\} = 0$ reflects a common feature of priors used in hypothesis testing, which is that the bulk of prior mass is allocated to alternatives "near" H_0. The reason usually given for such allocations is that H_0 would not be under test without specific interest in the surrounding region of the parameter space. (See, *e.g.*, Berger and Sellke, 1987, rejoiner, for discussion.) The property $\liminf_n\{\inf_{j\geq j_n} w_{n,j}\} = 0$ can be interpreted similarly as placing "nearest" to H_0 the alternatives expressed mainly through lower-indexed dimensions. This property will be exploited in Section 5 to accommodate smoothness assumptions.

4.5 Diffuse priors

A proposed generic procedure for high-dimensional testing is formulated as follows. Set $w_{n,j} = 1$, fix each r_n independently of τ_n, and, as $n \to \infty$, take $\tau_n \to \infty$ diverging at an arbitrarily fast rate. The fast divergence of τ_n spreads mass evenly (to an arbitrarily degree) across H_1 and makes the prior noninformative. A consequence for the posterior is that each $v_{n,j,1}$ is approximated by 1 to an arbitrary accuracy, hence $\rho_{0,n} \to 0$ *via* (8), and the posterior null probability is approximately

$$\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n] \quad = \quad \left[1 + \exp\left\{\frac{1}{2}\left(\|\boldsymbol{Y}_n^{SB}\|^2 - p_n - r_n p_n^{1/2}\right)\right\}\right]^{-1}. \tag{15}$$

At any n, $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n]$ is an arbitrarily close approximation to $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$, the formal posterior calculation derived from a proper prior. Thus, it provides a meaningful (in the sense of approximating a Bayesian procedure) assessment of H_0 in high-dimensions. For the case where the $\sigma_{n,j}$ are unknown, the same approach will produce an analogous limiting quantity $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$, which gives an arbitrarily close approximation to $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$, as in (14), again with $\rho_{0,n}$ understood through (8).

Note further that $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n]$ in (15) avoids the pitfalls discussed in Section 3.2 of Robert's (1993) noninformative solution (7). Observe that for each fixed p_n, $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n]$ is alternatively deduced as the limit of $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ as $\tau_n \to \infty$, in a similar manner as Robert's solution. The difference is that here τ_n and $\rho_{0,n}$ are connected through the criterion (8), rather than the equiponderance device (5); hence $\rho_{0,n} \to 0$ converges at a (necessary) slower rate. Next write $\|\boldsymbol{Y}_n^{SB}\|^2 - p_n - r_n p_n^{1/2} = \{T_n - r_n\} p_n^{1/2}$, where, using (9) and (10), T_n is $\|\boldsymbol{Y}_n^{SB}\|^2$ standardized to have mean zero and variance one for fixed $\theta = 0$. Considering how this expression appears in (15) it is clear that $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0$ for θ consistent with H_0 and $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n] \to 1$ for θ outside of an indistinguishable region; the slower $r_n \to \infty$ diverges, the smaller the indistinguishable region. Thus, the assessment $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n]$ is sensitive to the data in the manner described in Section 4.2 of tests consistent with the "what if" guidelines.

	Gelman *et al. (2004)*					Proposed				
School	2.5%	25%	50%	75%	97.5%	2.5%	25%	50%	75%	97.5%
A	-2	7	10	16	31	-0	6	9	14	47
B	-5	3	8	12	23	-5	4	8	11	21
C	-11	2	7	11	19	-23	3	7	10	18
D	-7	4	8	11	21	-7	4	8	11	21
E	-9	1	5	10	18	-12	3	7	10	16
F	-7	2	6	10	28	-13	3	7	10	17
G	-1	7	10	15	26	-0	6	9	13	31
H	-6	3	8	13	33	-11	5	8	12	35

Table 2: *Posterior quantiles for the aptitude-test coaching data. The left side lists quantiles calculated in Gelman* et al. *(2005), and the right side lists those of the posterior distribution (16).*

It is also possible to construct this solution using hyperpriors, as in Section 3.1. Observe that $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n]$ is a special case of (4) with $c_n = (2\pi)^{-p_n/2}|\boldsymbol{\Sigma}_n|^{-1/2}\exp\{-(p_n + r_n p_n^{1/2})/2\}$, $\int \boldsymbol{\pi}_n(\tau_n)d\tau_n = 1$, and $\rho_{0,n} = 1/2$. As we have seen, this c_n avoids the problems associated with settings for which $c_n \asymp 1$, and it is easily verified to be of a different character, for if $\sigma_{n,j} \asymp 1$, *e.g.*, and $\liminf_n r_n > -\infty$, then $c_n \to 0$.

A conceptual difficulty of our precise construction of $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n]$ is that it does not yield a sensible Bayes factor: with τ_n arbitrarily large, the parametrization (8) makes $\rho_{0,n}$ arbitrarily close to zero at *every* n, hence the approximate Bayes factor, $\boldsymbol{B}_n^*(\boldsymbol{Y}_n) = \{\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n]/\rho_{0,n}\}/\{(1-\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n])/(1-\rho_{0,n})\}$, is arbitrarily large. Nevertheless, the connection noted above to a hyperprior construction suggests an alternative calculation. If one treats $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n]$ as a special case of (4) with $c_n = (2\pi)^{-p_n/2}|\boldsymbol{\Sigma}_n|^{-1/2}\exp\{-(p_n + r_n p_n^{1/2})/2\}$, $\int \boldsymbol{\pi}_n(\tau_n)d\tau_n = 1$, and $\rho_{0,n} = 1/2$, the corresponding Bayes factor is $\boldsymbol{B}_n(\boldsymbol{Y}_n) = \boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n]/(1-\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n])$. Though reasonable, this does essentially make an arbitrary choice of $\rho_{0,n}$ (for one could always revise c_n to keep $c_n(\rho_{0,n}^{-1} - 1)$ in formula 4 constant). Yet since $\rho_{0,n} = 0$ is disqualified the setting $\rho_{0,n} = 1/2$ is the next sensible choice, absent other prior information, and we recommend associating this Bayes factor with $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n]$. The same transformation of $\boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$ provides a Bayes factor for the case when $\hat{\Sigma}_n$ is unknown.

Example 3. (Multi-sample testing, continued) *In analyzing the aptitude-test "coaching" data, Gelman* et al. *(2005) argue that one would want to use a prior which finds a suitable balance between estimating μ_j at the observed X_j and at the "pooled" effect, $\bar{X} = g^{-1}\sum_{j=1}^{g} X_j$, observed across all eight schools. That is, they argue to avoid having to choose "between complete pooling and none at all" by considering estimators of the form $\hat{\mu}_j = \lambda_j\bar{X} + (1-\lambda_j)X_j$, for weights $0 \le \lambda_j \le 1$. With this in mind, they specify a (continuous) hierarchical prior by which the $\mu_1, \ldots, \mu_g$ are conditionally independent with each $\mu_j \sim N(\mu, \tau^2)$, given hyperparameters μ and τ; a flat prior, with density $\boldsymbol{\pi}_n(\mu, \tau) \propto 1$, is specified for the hyperprior parameters. Posterior quantiles are computed by simulation, and they list those appearing on the left side of Table 2.*

An alternative treatment which also strikes a balance between complete pooling and none at all begins by approaching the analysis as a testing problem. Our proposed prior on $\boldsymbol{\theta}_n$ *has* $\boldsymbol{\theta}_n|H_1 \sim N(0, \tau_n^2 \boldsymbol{W}_n)$ *(recalling that* $\boldsymbol{\Sigma}_n = I$ *in this example), which is parametrized according to (8), with* $r_n = 0$ *and* $\tau_n \to \infty$ *at a very fast rate. This gives rise to the approximate posterior null probability (15), and also* $N(\boldsymbol{Y}_n, I)$ *for the approximate conditional posterior distribution of* $\boldsymbol{\theta}_n|H_1, \boldsymbol{Y}_n$.

Next observe that our construction of model (1) makes $\boldsymbol{Y}_n$ *independent of a modified version of the pooled effect,* $\tilde{X} = \{\sum_{j=1}^g X_j/S_j^2\}/\{\sum_{j=1}^g 1/S_j^2\}$. *To see this, define the matrix* $\boldsymbol{A}$ *to have* (j,k) *entry:* 1 *if* $k = j$ *and* $j < g$; -1 *if* $k = j+1$ *and* $j < g$; 0 *otherwise if* $j < g$; *and* $\{1/S_k^2\}/\{\sum_{j=1}^g 1/S_j^2\}$ *if* $j = g$. *Then* $\boldsymbol{X}^D$ *and* $\tilde{X}$ *are the first* p_n *and last 1 entries, respectively, of the vector* $\boldsymbol{A}X$, *whose covariance matrix is block-diagonal with blocks* $\boldsymbol{C}$ *and* $\tilde{S}^2 = 1/\{\sum_{j=1}^g 1/S_j^2\}$.

In addition, the inversion formulas for this example may be extended to write $\boldsymbol{X}$ *in terms of* $\tilde{X}$ *and* $\boldsymbol{Y}_n$, *and* $\boldsymbol{\mu}$ *in terms of* $\tilde{\mu} = \{\sum_{j=1}^g \mu_j/S_j^2\}/\{\sum_{j=1}^g 1/S_j^2\}$ *and* $\boldsymbol{\theta}_n$. *Subsequently, by placing a suitable prior on* $\tilde{\mu}$, *the approximate posterior distribution on* $\boldsymbol{\theta}_n$ *may be extended to account for the whole of* $\boldsymbol{\mu}$, *admitting calculation of approximate posterior quantiles for the* μ_j. *For this, observe the matrix-inverse of* $\boldsymbol{A}$ *is given by the partition* $\boldsymbol{A}^{-1} = [\boldsymbol{B}\,|\boldsymbol{1}]$, *where* $\boldsymbol{1}$ *denotes a column of ones and* $\boldsymbol{B}$ *is the* $g \times p_n$ *matrix with* (j,k) *entry:* $\{\sum_{l=k+1}^g 1/S_l^2\}/\{\sum_{l=1}^g 1/S_l^2\}$ *if* $j \le k$; *and* $-\{\sum_{l=1}^k 1/S_l^2\}/\{\sum_{l=1}^g 1/S_l^2\}$ *if* $j > k$. *It follows that* $\boldsymbol{X} = \tilde{X}\boldsymbol{1} + \boldsymbol{B}\boldsymbol{C}^{1/2}\boldsymbol{Y}_n$ *and* $\boldsymbol{\mu} = \tilde{\mu}\boldsymbol{1} + \boldsymbol{B}\boldsymbol{C}^{1/2}\boldsymbol{\theta}_n$. *The independence between* $\tilde{X}$ *and* $\boldsymbol{Y}_n$ *implies that the covariance matrix of* $\boldsymbol{X}$ *may be written* $\boldsymbol{V} = \tilde{S}^2\boldsymbol{1}\boldsymbol{1}^T + \boldsymbol{B}\boldsymbol{C}\boldsymbol{B}^T$.

We are now ready to carry out our analysis. Let us specify prior independence between $\tilde{\mu}$ *and* $\boldsymbol{\theta}_n$, *and place a flat prior on* $\tilde{\mu}$, *having density* $\boldsymbol{\pi}_n(\tilde{\mu}) \propto 1$. *An approximate posterior distribution for* $\boldsymbol{X}$ *now is given as the mixture*

$$\boldsymbol{\mu}|\boldsymbol{X} \quad \dot{\sim} \quad \boldsymbol{P}_n^*[H_0|\boldsymbol{Y}_n] N(\tilde{X}\boldsymbol{1}, \tilde{S}^2\boldsymbol{1}\boldsymbol{1}^T) + (1 - \boldsymbol{P}_n^*[H_0|\boldsymbol{Y}_n]) N(\boldsymbol{X}, \boldsymbol{V}). \tag{16}$$

Noting that $\tilde{\mu}\boldsymbol{1}|\tilde{X} \sim N(\tilde{X}\boldsymbol{1}, \tilde{S}^2\boldsymbol{1}\boldsymbol{1}^T)$, *and if a flat prior were placed on* $\boldsymbol{\mu}$ *one would have* $\boldsymbol{\mu}|\boldsymbol{X} \sim N(\boldsymbol{X}, \boldsymbol{V})$, *this mixture provides a sought-after balance between complete pooling and none at all. Yet here the balance is struck between the full posterior distributions of* $\tilde{\mu}|\tilde{X}$ *and* $\boldsymbol{\mu}|\boldsymbol{X}$, *not just the means, and* $\boldsymbol{P}_n^*[H_0|\boldsymbol{Y}_n]$, *serving as a common value for the* λ_j, *ascribes some further meaning to the weights.*

For the data of Table 1, the modified pooled effect is $\tilde{X} = 7.6856$ *and the posterior null probability is* $\boldsymbol{P}_n^*[H_0|\boldsymbol{Y}_n] = 0.7589$, *which suggests a leaning toward complete pooling. The associated Bayes factor is* $\boldsymbol{B}_n(\boldsymbol{Y}_n) = \boldsymbol{P}_n^*[H_0|\boldsymbol{Y}_n]/(1 - \boldsymbol{P}_n^*[H_0|\boldsymbol{Y}_n]) = 3.1470$, *which, on the scales for weight of evidence discussed in Kass and Raftery (1995), is categorized as indicating "positive" evidence for* H_0, *but "worth [just slightly more] than a bare mention." Corresponding posterior quantiles are listed in the right side of Table 2. Comparing these with Gelman's* et. al *hierarchical solution, one observes a widespread consistency and comparable degree of shrinkage toward complete pooling. Noting there is less variability among the 50% quantiles of the present solution, one would conclude it has emphasized complete pooling slightly more strongly. However, there are also dis-*

crepancies in the extreme quantiles, specifically between the 2.5% quantiles of Schools C, F, and H and the 97.5% quantiles of Schools A and G, that suggest the opposite interpreation. Nevertheless, excepting the notable inconsistency between the 97.5% quantiles of School A, the discrepancies between the two solutions are quite small.

4.6 Jeffreys-Lindley paradox in high dimensions

The Jeffreys-Lindley paradox is investigated in high-dimensions by comparing $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ with the frequentist p-value under the "what if" guidelines. For this, let us suppose the frequentist takes $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ as a "test statistic" and normalizes it as $\hat{T}_n = (\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SB}\|^2 - \|\boldsymbol{v}_{n,1/2}\|^2) \,/\, (\sqrt{2}\|\boldsymbol{v}_{n,1}\|)$. Assuming $\|\boldsymbol{v}_{n,1/2}\|^2 \to \infty$, $\hat{T}_n$ is approximately standard normal under H_0 for large p_n (as discussed in Jensen and Solomon, 1972, for example). A corresponding p-value is therefore $\hat{p}_n \approx 1 - \Phi(\hat{T}_n) \approx \hat{T}_n^{-1}\phi(\hat{T}_n)$, where ϕ and Φ are the standard-normal density and distribution functions, respectively, the latter approximation being valid for large $|\hat{T}_n|$. Assuming the parametrization (8) with $r_n \to \infty$, and writing $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] = [1 + \exp\{\|\boldsymbol{v}_{n,1}\|(\hat{T}_n - r_n)/2\}]^{-1}$, the desired comparison is made through the approximation

$$\begin{aligned}
\frac{\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]}{\sqrt{2\pi}\hat{p}_n} &\approx \frac{\hat{T}_n \exp\left\{\hat{T}_n^2/2\right\}}{1 + \exp\left\{\|\boldsymbol{v}_{n,1}\|\left(\hat{T}_n - r_n/\sqrt{2}\right)/\sqrt{2}\right\}} \\
&= \frac{\hat{T}_n \exp\left\{\hat{T}_n^2\left(1 - \sqrt{2}\|\boldsymbol{v}_{n,1}\|/\hat{T}_n\right)/2 + r_n/2\right\}}{1 + \exp\left\{-\|\boldsymbol{v}_{n,1}\|\left(\hat{T}_n - r_n/\sqrt{2}\right)/\sqrt{2}\right\}}.
\end{aligned} \tag{17}$$

Taking $p_n \to \infty$, the middle expression in (17) shows that $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]/\hat{p}_n \to \infty$ for data such that $\hat{T}_n \to -\infty$. Thus, for p-values near one, both the Bayesian and frequentist would conclude very little evidence against H_0, but $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ would be closer to one than $\hat{p}_n$. A similar, though slightly more complex, comparison is made for small p-values. For data such that $\hat{T}_n \to \infty$, both $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0$ and $\hat{p}_n \to 0$, but the comparison is different depending on the rate at which $\hat{T}_n$ grows. The last expression in (17) shows that $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]/\hat{p}_n \to \infty$ if $\lim_n \|\boldsymbol{v}_{n,1}\|/\hat{T}_n \le 1/\sqrt{2}$ and $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]/\hat{p}_n \to 0$ otherwise. Thus it is seen there is a cutoff point at which the comparison changes, and its presence indicates a region of possible data for which *some* evidence is exhibited against H_0 and for which $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ and $\hat{p}_n$ take very similar values. (An exact description of this region depends on r_n, in order that (17) takes an intermediate limit.) The nature of the Jeffreys-Lindley paradox is therefore changed in high-dimensions, alleviating much of what makes it paradoxical.

Robert (1993) observes a similar phenomenon in the univariate case, noting that for $p_n = 1$ and $\sigma_{n,1} = 1$, the frequentist p-value closely matches his solution, the limit in (7), at the critical range of values near 0.05. (It is even suggested this phenomenon accounts for the historical survival of the p-value, since it behaves like a Bayesian procedure when evidence about H_0 is most difficult to discern.) However, recall from Section 3 that Robert's solution does not avoid the version of the paradox that fixes $Y_{n,j}^{SB}$, sets

$w_{n,j} = \tilde{w}_{n,j}/\sigma_{n,j}^2$, and takes $\sigma_{n,j} \to 0$. Curiously, solutions parametrized *via* (8) do, provided each r_n is independent of the $\sigma_{n,j}$. To see this, fix $Y_{n,j}^{SB}$ and use (11) to deduce, under (8), that $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to \boldsymbol{P}_n^*[\mathrm{H}_0|\boldsymbol{Y}_n]$ as each $\sigma_{n,j} \to 0$.

5 High-dimensional analysis of the smooth model

This section treats the high-dimensional testing problem in a context that accommodates the approaches to GOF-testing and FDA outlined in Section 2.

5.1 Updated notation for the smooth model

Asymptotic analysis will here make critical use of the "sample size" parameter n. The magnitudes of error are parametrized as $\sigma_{n,j} = \tilde{\sigma}_{n,j}/\sqrt{n}$, for an array $\tilde{\Sigma} = \{\tilde{\sigma}_{n,j} : (n,j) \in \mathcal{I}_n\}$ such that $\tilde{\sigma}_{n,j} \asymp 1$ uniformly across n and j. Other elements of the notation defined in Section 3 are updated accordingly: Set $\tilde{\boldsymbol{\Sigma}}_n = \mathrm{diag}(\tilde{\sigma}_{n,1}, \ldots, \tilde{\sigma}_{n,p_n})$, and write $\boldsymbol{Y}_n^{SS} = [Y_{n,1}^{SS}, \ldots, Y_{n,p_n}^{SS}]^T$ and $\boldsymbol{\theta}_n^{SS} = [\theta_{n,1}^{SS}, \ldots, \theta_{n,p_n}^{SS}]^T$, with $Y_{n,j}^{SS} = Y_{n,j}/\tilde{\sigma}_{n,j}$ and $\theta_{n,j}^{SS} = \theta_{n,j}/\tilde{\sigma}_{n,j}$. The superscript "SS" means "scaled for the smooth model." Corresponding arrays are $Y^{SS} = \{Y_{n,j}^{SS} : (n,j) \in \mathcal{I}_n\}$ and $\theta^{SS} = \{\theta_{n,j}^{SS} : (n,j) \in \mathcal{I}_n\}$.

5.2 Incorporating a smoothness assumption

Recall from Section 2 that a central issue in GOF-testing and FDA is how to incorporate the smoothness assumption of an underlying functional model. From the discussion following Example 2, a suitable means of imposing this assumption is to restrict the model so that each $\boldsymbol{\theta}_n^{SS}$ is an element of the space

$$\mathcal{B}_{n,s,M} = \left\{ \boldsymbol{\theta} = [\theta_1, \ldots, \theta_{p_n}]^T : \sum_{j=1}^{p_n} j^{2s}\theta_j^2 \leq M \right\}, \tag{18}$$

where $s > 1/2$ and $M > 0$. To reflect this assumption in the prior, $\boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\mathrm{H}_1)$ might be reformulated so that $\boldsymbol{\theta}_n^{SS}|\mathrm{H}_1 \sim N(0, \tau_n^2 \boldsymbol{W}_n \boldsymbol{\Sigma}_n)$, as before, but conditional on $\mathcal{B}_{n,s,M}$. Under the parametrization (8), the posterior null probability (2) then becomes

$$\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \mathcal{B}_{n,s,M}] = \left[1 + \left(\frac{\boldsymbol{P}_n[Q_{n,2} \leq M|\boldsymbol{Y}_n]}{\boldsymbol{P}_n[Q_{n,1} \leq M]}\right) \exp\left\{\frac{1}{2}\left(n\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SS}\|^2 - \|\boldsymbol{v}_{n,1/2}\|^2 - r_n\|\boldsymbol{v}_{n,1}\|\right)\right\}\right]^{-1} \tag{19}$$

where

$$Q_{n,1} = \frac{1}{n}\sum_{j=1}^{p_n} j^{2s} v_{n,j,1} Z_{n,j}^2/(1 - v_{n,j,1}) \tag{20}$$

and

$$Q_{n,2} \quad = \quad \frac{1}{n}\sum_{j=1}^{p_n} j^{2s} v_{n,j,1}\left\{Z_{n,j} + v_{n,j,1/2}\sqrt{n}Y_{n,j}^{SS}\right\}^2, \tag{21}$$

for an array $\{Z_{n,j}\}$ of independent standard normal random variables. A derivation of (19) is provided in Appendix 2, where it is shown that $\boldsymbol{P}_n[Q_{n,1} \leq M] = \boldsymbol{P}_n[\boldsymbol{\theta}_n^{SS} \in \mathcal{B}_{n,s,M}|\mathrm{H}_1]$ and $\boldsymbol{P}_n[Q_2 \leq M|\boldsymbol{Y}_n] = \boldsymbol{P}_n[\boldsymbol{\theta}_n^{SS} \in \mathcal{B}_{n,s,M}|\mathrm{H}_1, Y]$.

The quantity $SF(\boldsymbol{Y}_n) = \boldsymbol{P}_n[Q_{n,2} \leq M|\boldsymbol{Y}_n]/\boldsymbol{P}_n[Q_{n,1} \leq M]$ in (19) is the sole factor that distinguishes (19) from the formulas for $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ considered in Section 4. It therefore determines the impact of prior conditioning on $\mathcal{B}_{n,s,M}$, and so will be referred to as the "smoothness factor." The present evaluation of the smooth model will consider two cases. First, the smoothness factor is ignored, and the evaluation treats the formula (19) as if $SF(\boldsymbol{Y}_n) = 1$, which is the correct expression under the unconditional-prior formulation of Section 4. The restriction to $\boldsymbol{\theta}_n \in \mathcal{B}_{n,s,M}$ is nevertheless assumed in asymptotic analysis, and the goal to identify favorable settings of the $w_{n,1}, \ldots, w_{n,p_n}$ under the "what if" guidelines. This is carried out in Section 5.3. The second part of the evaluation takes both $SF(\boldsymbol{Y}_n)$ and the assumption $\boldsymbol{\theta}_n \in \mathcal{B}_{n,s,M}$ fully into account. Section 5.4 treats this case, and shows that, while conditioning on $\mathcal{B}_{n,s,M}$ is appealing from the point of view of eliciting assumptions, any advantage under the "what if" guidelines is absent.

5.3 Rates of testing

Recall from Section 4.2 the notion of an indistinguishable region, which collects the θ consistent with H_1 which do not yield $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0$ almost surely. Previously, the objective to keep the indistinguishable region small led to the guideline that r_n should diverge as slowly as possible. The same objective will now be considered for the purpose of selecting $w_{n,1}, \ldots, w_{n,p_n}$ under a smooth model. One consequence of this approach is to set up a mathematical framework similar to the frequentist rates of testing theory developed in Ingster (1993) and Spokoiny (1996). Much of the following discussion draws from that theory and its associated terminology.

To begin, let us refine and update the description of the indistinguishable region given in Section 4.2. Consider sequences $\delta_n \to 0$ satisfying

$$\sup_{\boldsymbol{\theta}_n^{SS} \in \mathcal{H}_{1,n}(\delta_n/\delta_n^*)} \boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 0 \text{ for every } \delta_n^* \to 0, \tag{22}$$

where $\mathcal{H}_{1,n}(\delta) = \{\boldsymbol{\theta} \in \mathcal{B}_{n,s,M} : \|\boldsymbol{\theta}\| \geq \delta\}$. Associated with this criterion is the following optimality concept. Suppose (22) holds for some sequence $\hat{\delta}_n \to 0$, and furthermore every $\delta_n \to 0$ for which (22) also holds satisfies $\hat{\delta}_n = O(\delta_n)$. The rate of $\hat{\delta}_n$ will then be referred to as a "rate of testing" of a given prior specification. These rates provide the kind of description we want: $\hat{\delta}_n$ gives the boundary of the indistinguishable region, and so a faster rate of testing identifies a smaller indistinguishable region. Thus, the most favorable settings for $w_{n,1}, \dots, w_{n,p_n}$ are those for which the associated rates of testing are as fast as possible.

Next let us ignore the smoothness factor $SF(\boldsymbol{Y}_n)$ of formula (19) (or set it to one) and follow deductions paralleling those of Section 4.2. A reanalysis of expression (13) will show, noting $\boldsymbol{\theta}_n^{SS} = \sqrt{n}\boldsymbol{\theta}_n^{SB}$, that the criterion determining (22) is

$$\sup_{\boldsymbol{\theta}_n^{SS} \in \mathcal{H}_{1,n}(\delta_n/\delta_n^*)} \frac{\|\boldsymbol{v}_{n,1}\|}{n\|\boldsymbol{v}_{n,1/2}\theta^{SS}\|^2} \to 0 \text{ for every } \delta_n^* \to 0, \tag{23}$$

provided r_n diverges so slowly that

$$\limsup_n \left\{ \sup_{\boldsymbol{\theta}_n^{SS} \in \mathcal{H}_{1,n}(\delta_n/\delta_n^*)} r_n \frac{\|\boldsymbol{v}_{n,1}\|}{n\|\boldsymbol{v}_{n,1/2}\theta^{SS}\|^2} \right\} < 1 \text{ for every } \delta_n^* \to 0.$$

This criterion makes a fortunate connection to the mathematics of frequentist rates of testing theory, specifically to the properties of frequentist tests derived from weighted quadratic forms, which are studied in Spitzner (2008). The detailed route of this connection is provided in Appendix 3. What follows below is a summary of its implications to the present problem.

It will be helpful to first remark on general performance bounds implied by the results of Ingster (1993) and Spokoiny (1996). These are relevant not just under the criterion (23), but under any prior specification subject to the "what if" guidelines (*i*) and (*iii*). First, it is implied from Ingster (1993) that the fastest rate of testing that can be achieved by any prior is $\hat{\delta}_n = n^{-2s/\tilde{s}}$, where $\tilde{s} = 4s + 1$, and s is the parameter specified in $\mathcal{B}_{n,s,M}$. However, an associated result is that to achieve this rate it is necessary that the prior depends on the parameter s. Unfortunately, it is often difficult to elicit a precise value of s and so Ingster's result is usually irrelevant for applications. To avoid this problem, Spokoiny (1996) formulates an "adaptive" rates of testing framework in which performance is considered simultaneously for s across a finite range of values, $s_* < s < s^*$ say. In the present context, the adaptive framework amounts to the goal of finding a single prior, which may depend on s only through s_* and s^*, for which no other prior induces uniformly faster rates of testing across $s_* < s < s^*$. Spokoiny's results imply that such adaptive rates of testing can be no faster than $\hat{\delta}_n = \{n^2(\log\log n)^{-1}\}^{-s/\tilde{s}}$.

Spitzner (2008) studies these ideas in a restricted (frequentist) context where the only tests considered are those which use a quadratic form as a test statistic. As shown in Appendix 3, this is mathematically parallel to the present problem where the data appear in $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ through the quadratic form $\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SS}\|^2$. Relevant results are now stated as implications for the setup leading to (23).

The rate of p_n plays a critical role. For simplicity, let us assume that $\log p_n/\log n$ either has a limit or diverges to infinity, and set $L = \lim_n\{\tilde{s}\log p_n/(2\log n)\}$. From Spitzner (2008), it is known that if $L < 1$, the criterion (22) cannot hold for any $\hat{\delta}_n \to 0$. On the other hand, the setting $L = 1$ yields Ingster's optimal rate of testing whenever $v_{n,j,1} = j^{-\gamma_1/2}$ for any $0 < \gamma_1 < 1$ (also $\gamma_1 = 0$, but that setting is impossible here). However, observe that to set $L = 1$ requires a precise value of s, and this means the setting $L = 1$ is not valid in the adaptive context. Yet if $1 < L < B$ for some $B > 1$, the rate of testing $\hat{\delta}_n = \{n^2(\log n)^{-1}\}^{-s/\tilde{s}}$ is achieved by prior settings such that

$$v_{n,j,1} \quad = \quad j^{-1/2}(\log j)^{-(1-\gamma_2)/2}, \tag{24}$$

for $j > 1$ and some constant $\gamma_2 \geq 0$ (among which $\gamma_2 = 1$ might be preferred for simplicity). Moreover, Theorem 3 of Spitzner (2008) implies that this rate is adaptively optimal among tests based on quadratic forms. The property (24) therefore achieves the objective stated at the beginning of this section: it identifies favorable settings of the prior that, from an adaptive rates-of-testing viewpoint, reduce the indistinguishable region to the greatest extent possible.

It is easy to restate (24) in terms of the $w_{n,1}, \ldots, w_{n,p_n}$ rather than the entries of $\boldsymbol{v}_{n,1}$: if the setting $v_{n,j,1} = v_j^*$ is desired for some sequence $v_1^*, v_2^* \ldots$, then straightforward algebra will show an equivalence with the weight settings $w_{n,j} = \{\tau_n^2(v_j^{*-1} - 1)\}^{-1}$. For such $w_{n,j}$, the τ_n^2 cancels in $\tau_n^2 \boldsymbol{W}_n$, and the conditional prior may be written $\boldsymbol{\theta}_n|\mathrm{H}_1 \sim N(0, (\boldsymbol{V}_n^{*-1} - I)^{-1}\boldsymbol{\Sigma}_n)$, where $\boldsymbol{V}_n^* = \mathrm{diag}(\tilde{v}_1^*, \ldots, \tilde{v}_{p_n}^*)$. Also, the posterior distribution has $\boldsymbol{\theta}_n|\mathrm{H}_1, \boldsymbol{Y}_n \sim N(\boldsymbol{V}_n^*\boldsymbol{Y}_n, \boldsymbol{V}_n^*\boldsymbol{\Sigma}_n)$.

One technical note is that the setting $v_{n,1,1} = 1$, according to (24), must be treated as an approximation to a specification for which $\tau_n w_{n,j}$ is arbitrarily large. However, this can create a problem in calculating a Bayes factor, for then $\rho_{0,n}$ through the parametrization (8), is arbitrarily close to zero and the Bayes factor $\boldsymbol{B}_n(\boldsymbol{Y}_n) = \{\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]/\rho_{0,n}\}/\{(1 - \boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n])/(1 - \rho_{0,n})\}$ becomes arbitrarily large. To avoid this problem, an *ad hoc* adjustment is recommended by which the $v_{n,j,1}$ are set by (24) for $j \geq 2$ but $v_{n,1,1} = v_{n,2,1}$. If no Bayes factor is required, this adjustment is unnecessary.

It is interesting to consider (24) in the context of Section 4.4, where the concern is over the setting $\rho_{0,n} = 1/2$. It is relevant that $\lim_j w_{n,j} = 0$ under (24), for the discussion following Theorem 1 suggests the possibility of solving for τ_n so that $\lim_n \rho_{0,n} = 1/2$. Nevertheless, in fact, $\rho_{0,n} \to 0$ under the guideline that $r_n \to \infty$ diverges "as slowly as possible." This is easily deduced from the properties $v^{-1}\{1 + v^{-1}\log(1 - v)\} \to -1/2$ as $v \to 0$ and $\sum_{j=1}^{p} v_{n,j,2} \to \infty$, applied to (8) under (24), provided $r_n = O(\|\boldsymbol{v}_{n,1-t}\|^2/\|\boldsymbol{v}_{n,1}\|)$ for some $0 < t < 1/2$.

Example 2. (Functional data analysis, continued) *For analysis of the VDP data, there are statistics* $\hat{\boldsymbol{\Sigma}}_n = diag(\hat{\sigma}_{n,1}^2, \ldots, \hat{\sigma}_{n,p_n}^2)$ *available for use in specifying a prior on the* $\boldsymbol{\Sigma}_n$*. Set* $\hat{V} = \sum_{g=1}^{2}\sum_{i=1}^{n_g}(X_{g,ij} - \bar{X}_{g,j})^2$, $\nu_n = n - 2$ *(*$= n_1 + n_2 - 2$*) and define* $\hat{\sigma}_{n,j}^2 = (n_1^{-1} + n_2^{-1})\hat{V}_j$*. It is a well known property that* $\hat{\boldsymbol{\Sigma}}_n$ *and* $\boldsymbol{Y}_n$ *are independent and* $\nu_n\hat{\sigma}_{n,j}^2/\sigma_{n,j}^2 \sim \chi_{\nu_n}^2$*, as is required for the formulation discussed in Section 4.3. Our proposed prior on* $(\boldsymbol{\theta}_n, \boldsymbol{\Sigma}_n)$ *has* $\boldsymbol{\theta}_n|H_1, \boldsymbol{\Sigma}_n \sim N(0, (\boldsymbol{V}_n^{*-1} - I)^{-1}\boldsymbol{\Sigma}_n)$ *for* $\boldsymbol{V}_n^* = diag(1, 2^{-1/2}, \ldots, p_n^{-1/2})$*, which is parametrized* via *(8), with* $r_n = 0$*. An improper prior is selected for* $\boldsymbol{\Sigma}_n$*, with densities given by* $\boldsymbol{\pi}_n(\sigma_{n,j}^2) \propto (\sigma_{n,j}^2)^{-(\kappa/2+1)}\exp\{-\lambda/(2\sigma_{n,j}^2)\}$

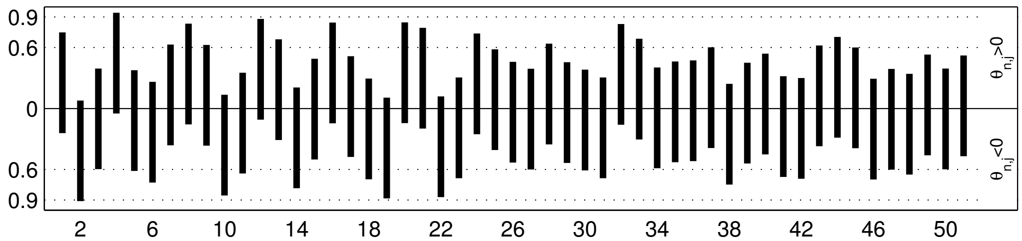

Figure 2: *Marginal probabilities $\boldsymbol{P}_n[\theta_{n,j} > 0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$ and $\boldsymbol{P}_n[\theta_{n,j} < 0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$ for the VDP data. The former are plotted above the reference line at zero, with larger values more toward the top, and the latter plotted below the reference line, with larger values more toward the bottom. Observe how the structure smooths out at the higher indices.*

Index	$\boldsymbol{P}_n[\theta_{n,j} < 0\|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$	$\boldsymbol{P}_n[\theta_{n,j} > 0\|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$	max.	rank
4	0.05	0.94	0.94	1
2	0.91	0.08	0.91	2
19	0.88	0.11	0.88	3
12	0.11	0.88	0.88	4
22	0.87	0.12	0.87	5
10	0.85	0.14	0.85	6
20	0.14	0.85	0.85	7
16	0.15	0.84	0.84	8
8	0.16	0.83	0.83	9
32	0.16	0.83	0.83	10

Table 3: *Indices j in $\theta_{n,j}$ with the largest values of $\boldsymbol{P}_n[\theta_{n,j} > 0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$ or $\boldsymbol{P}_n[\theta_{n,j} < 0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$. Only one index in these top ten is odd.*

for $\kappa = 0$ and $\lambda = 0$.

Under this specification, the posterior null probability (14), calculated on just the first $p_n = 51$ components, is $\boldsymbol{P}_n[H_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n] = 0.0094$. If $\boldsymbol{V}_n^*$ is adjusted so that its first diagonal entry is $2^{-1/2}$ (matching the second), the posterior null probability reduces slightly to $\boldsymbol{P}_n[H_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n] = 0.0090$, the prior null probability is $\rho_{0,n} = 0.1804$, by (8), and the Bayes factor is $\boldsymbol{B}_n(\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n) = \{\boldsymbol{P}_n[H_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]/\rho_{0,n}\}/\{(1 - \boldsymbol{P}_n[H_0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n])/(1 - \rho_{0,n})\} = 0.0411$. Kass and Raftery (1995) categorize this Bayes-factor value as indicating "strong" evidence for H_1.

To examine the apparent differences between shift groupings more closely, marginal posterior probabilities of $\boldsymbol{P}_n[\theta_{n,j} > 0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$ and $\boldsymbol{P}_n[\theta_{n,j} < 0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$ are calculated from the full posterior distribution. These probabilities are charted in Figure 2, in which a tendency is seen for the larger probabilities to appear at even index values. In addition, consider Table 3, which lists the index values of the ten largest $\boldsymbol{P}_n[\theta_{n,j} > 0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$ and $\boldsymbol{P}_n[\theta_{n,j} < 0|\boldsymbol{Y}_n, \hat{\boldsymbol{\Sigma}}_n]$; among these, only one index is odd. Noting that the Fourier basis-

functions with even indices correspond to $\psi_{2j}(t) = c_j \sin(jt)$, which are asymmetric on $(-\pi, \pi]$, our analysis therefore suggests that the greatest differences between shift groupings are exhibited as asymmetric attributes of the profiles.

Another interesting pattern in Figure 2 is that the charted values even out at the higher indices. This reflects a "smoothing" effect of the procedure by which the choice of $w_{n,1}, \ldots, w_{n,p_n}$ emphasizes the structure in the lower-indexed components.

5.4 Impact of prior conditioning

The analysis of the smooth model has thus far identified favorable settings for $w_{n,1}, \ldots, w_{n,p_n}$ while ignoring the smoothness factor, $SF(\boldsymbol{Y}_n)$; *i.e.*, the prior has been treated as if unconditional on $\mathcal{B}_{n,s,M}$. The impact of prior conditioning on $\mathcal{B}_{n,s,M}$ will now be considered by examining the influence of $SF(\boldsymbol{Y}_n)$ on the asymptotic behavior of $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$.

There are various conceptual and practical hurdles involved in calculating $SF(\boldsymbol{Y}_n)$ which limit the extent to which prior conditioning is even an option in practice. (Hence one might interpret the present discussion as seeking primarily to assess what might be missed by not conditioning.) The main conceptual problem is that $SF(\boldsymbol{Y}_n)$ depends s and M, which are usually difficult to pin down. The main practical problem, on the other hand, is that näive approximations of $SF(\boldsymbol{Y}_n)$ can be subject to substantial error. To see this, observe that both $Q_{n,1}$ and $Q_{n,2}|\boldsymbol{Y}_n$ are typically of order p_n and so $SF(\boldsymbol{Y}_n)$ involves probability calculations at these distributions' extreme lower tails. Thus, numerical simulation would be unstable, and the more elementary approximation formulas would be inaccurate. Instead, a high-accuracy method such as saddlepoint approximation would be required to calculate any reasonably reliable value of $SF(\boldsymbol{Y}_n)$. Accordingly, saddlepoint approximation will form the basis of the theoretical evaluation carried out here. To be clear, this means that $SF(\boldsymbol{Y}_n)$ will not be evaluated directly, but only through its saddlepoint approximation, and it is admitted that this relaxes the rigor of the argument. Nevertheless, such evaluation is still quite relevant given the likely possibility that no better approximation to $SF(\boldsymbol{Y}_n)$ would be available in practice. Refer to Barndorff-Nielsen and Cox (1989, ch. 4), Reid (1988), and Kuonen (1999), among others, for good, general discussion of this method, and to Jensen (1995) for detailed discussion of their accuracy.

The most relevant case has both $\boldsymbol{P}_n[Q_{n,2} \leq M|\boldsymbol{Y}_n] \to 0$ and $\boldsymbol{P}_n[Q_{n,1} \leq M] \to 0$, which occurs under the recommended settings (24) of the last section. In that case, $SF(\boldsymbol{Y}_n)$ takes an indeterminant form and L'Hospital's rule provides that it may be studied asymptotically as

$$\frac{\sqrt{\kappa_{n,2,2}(\boldsymbol{Y}_n)} f_{n,2}(M|\boldsymbol{Y}_n)}{\sqrt{\kappa_{n,1,2}} f_{n,1}(M)}, \tag{25}$$

where $f_{n,1}$ and $f_{n,2}(\cdot|\boldsymbol{Y}_n)$ are densities of $Q_{n,1}$ and $Q_{n,2}|\boldsymbol{Y}_n$, respectively, and $\kappa_{n,1,2}$ and $\kappa_{n,2,2}(\boldsymbol{Y}_n)$ are their corresponding variances. Saddlepoint approximations to $f_{n,1}(M)$

and $f_{n,2}(M|\boldsymbol{Y}_n)$ are given by

$$f_{n,1}(M) \approx \frac{\exp\{K_{n,1}(\hat{\lambda}_{n,1}) - \hat{\lambda}_{n,1}M\}}{\{2\boldsymbol{\pi}_n K''_{n,1}(\hat{\lambda}_{n,1})\}^{1/2}}$$

and

$$f_{n,2}(M|\boldsymbol{Y}_n) \approx \frac{\exp\{K_{n,2}(\hat{\lambda}_{n,2}(\boldsymbol{Y}_n)|\boldsymbol{Y}_n) - \hat{\lambda}_{n,2}(\boldsymbol{Y}_n)M\}}{\{2\boldsymbol{\pi}_n K''_{n,2}(\hat{\lambda}_{n,2}(\boldsymbol{Y}_n)|\boldsymbol{Y}_n)\}^{1/2}},$$

where $K_{n,1}$ and $K_{n,2}(\cdot|\boldsymbol{Y}_n)$ are the cumulant generating functions of $Q_{n,1}$ and $Q_{n,2}|\boldsymbol{Y}_n$, respectively, and $\hat{\lambda}_{n,1}$ and $\hat{\lambda}_{n,2}(\boldsymbol{Y}_n)$ are the respective values of t such that $K'_{n,1}(t) = M$ and $K'_{n,2}(t|\boldsymbol{Y}_n) = M$.

The assertion made at the beginning of this section is made precise by the following theorem, whose proof is provided in Appendix 4.

Theorem 2. *Suppose* $\lim_n \log p_n / \log n > 2/3$, $w_{n,j}$ *and* τ_n *are such that* $\gamma = \sup\{t : v_{n,j,1} = o(b_{n,j})$, *where* $b_{n,j} = j^{-t}\} > 0$, *and* r_n *is such that* $\boldsymbol{P}_n[H_0|\boldsymbol{Y}_n, \mathcal{B}_{n,s,M}] \to 0$ *is avoided for* θ *consistent with* H_0. *Suppose further that both* $\boldsymbol{P}_n[Q_{n,2} \leq M|\boldsymbol{Y}_n] \to 0$ *and* $\boldsymbol{P}_n[Q_{n,1} \leq M] \to 0$ *so that (25) approximates* $SF(\boldsymbol{Y}_n)$, *asymptotically. If* $\delta_n^4 = O(p_n/n^2)$ *and* $\delta_n^* \to 0$ *then there is a sequence of* $\boldsymbol{\theta}_n^{SS} \in \mathcal{H}_{1,n}(\delta_n/\delta_n^*)$ *for which* $\liminf_n \boldsymbol{P}_n[H_0|\boldsymbol{Y}_n, \mathcal{B}_{n,s,M}] > 0$.

Theorem 2 immediately dismisses the possibility of achieving Ingster's minimax rate, $\hat{\delta}_n = n^{-2s/\tilde{s}}$, since that rate has $n^2\hat{\delta}_n^4/p_n = n^{2/\tilde{s}}/p_n \to 0$ when $\lim_n \log p_n/\log n > 2/3$. The optimal adaptive rate for quadratic forms, $\hat{\delta}_n = \{n^2(\log n)^{-1}\}^{-s/\tilde{s}}$, is similarly disqualified. Thus it is seen that the indistinguishable region is made larger under the prior formulated conditionally on $\mathcal{B}_{n,s,M}$. Moreover, if $n^2/p_n \to 0$, Theorem 2 establishes that (22) cannot hold for any $\delta_n \to 0$, indicating a possible severe negative impact of prior conditioning. An interpretation of the latter case is that with p_n increasing so quickly the "size" of $\mathcal{B}_{n,s,M}$ relative to high-dimensional space is essentially that of a point. The conditional prior in effect reduces to a point mass, making it impossible to discern points within $\mathcal{B}_{n,s,M}$.

6 Conclusions and discussion

A new methodology for high-dimensional Bayesian testing has been developed, which is formulated in such a way that the underlying prior structure may be interpreted as noninformative. The proposed methodology is suitable for general high-dimensional testing problems, but a specialized version has also been proposed for the contexts of GOF-testing and FDA, where smoothness assumptions are an important consideration. Demonstrations on two interesting data sets have also been carried out and discussed, and new analysis tools based on the proposed tests have been introduced.

A number of interesting conceptual aspects of the problem have been observed.

Under the basic model, absent any smoothness assumption, the notion that $\rho_{0,n} = 1/2$ represents a noninformative setting has been weakened, since in high dimensions that requires one to order and weight dimensionality. The nature of the Jeffreys-Lindley paradox has also been weakened: it is less paradoxical, as the Bayesian posterior null probability and the frequentist p-value have been shown to behave consistently in high-dimensions. Investigation of the smooth model has led to a Bayesian formulation of rates of testing theory, and through this a favorable dimensionality-weighting configuration for the prior. In studying the effects of prior conditioning on $\mathcal{B}_{n,s,M}$, it has been shown that, surprisingly, the unconditional formulation yields the more favorable behavior.

Both the cases of known and unknown error-variance parameters have been explored, though the primary focus has been on the former. Scaled inverse-χ^2 priors on the $\sigma^2_{n,j}$ have shown to produce a test which is conservative, but not wholly inaccurate. Improvements in accuracy is a goal of future work.

Among issues that remain for discussion, there are questions of how $\rho_{0,n}$ is to be interpreted in a context where it is connected to other prior parameters and to dimensionality. Robert (1993) casts his arguments leading to the equiponderance device (5) as an interest in balancing the sizes of H_0 and H_1. In light of present results, however, it seems that high-dimensionality complicates whatever balance might be achieved, as indicated in particular by the role of the parameter r_n. What has become blurred is the traditional sense by which $\rho_{0,n}$ represents subjective belief in H_0. The results here suggest that any probabilistic interpretation of $\rho_{0,n}$ must be considered against the backdrop of geometry. Specifically, any program for eliciting $\rho_{0,n}$ subjectively would need to adjust for such issues as the total number and ordering of dimensions, the relative importance of individual dimensions, and possibly geometric constraints such as (18).

Much further exploration of the Bayesian rates-of-testing context remains to be done. In particular, Spokoiny's (1996) performance bound on adaptive rates is not achieved by the normal prior, and it remains an open question whether this gap can be closed by a Bayesian testing procedure.

Finally, though the recommendations made here readily apply to FDA applications involving orthogonal basis decomposition, the details of how they apply to GOF-testing remain somewhat unclear. In particular, it is difficult to make comparisons with the many established GOF tests whose underlying models appear on the surface very different from (1). By viewing (1) as a canonical asymptotic model, however, as suggested in Section 2, it is hoped such translation and comparison becomes possible, opening new doors to further development of this interesting topic.

Appendices

1 Derivation of the posterior null probability with a prior on the error-variances

A joint model density is

$$m(\boldsymbol{Y}_n\hat{\boldsymbol{\Sigma}}_n|\boldsymbol{\theta}_n,\boldsymbol{\Sigma}) = A_n^{p_n}|\hat{\boldsymbol{\Sigma}}_n|^{\nu_n/2-1}|\boldsymbol{\Sigma}_n|^{-(\nu_n+1)/2} \times \exp\left\{-\sum_{j=1}^{p_n}\frac{(Y_{n,j}-\theta_{n,j})^2+\nu_n\hat{\sigma}^2_{n,j}}{2\sigma^2_{n,j}}\right\}.$$

and a joint conditional prior density is

$$\boldsymbol{\pi}_n(\boldsymbol{\theta}_n,\boldsymbol{\Sigma}_n|\mathrm{H}_1) = B_n^{p_n}\tau^{-p_n/2}|\boldsymbol{W}_n|^{-1/2}|\boldsymbol{\Sigma}_n|^{-(\kappa+3)/2}\exp\left\{-\sum_{j=1}^{p_n}\frac{\theta^2/(\tau_n w_{n,j})+\lambda}{2\sigma^2_{n,j}}\right\}.$$

where $A_n = (2\pi)^{-1/2}(\nu_n/2)^{\nu_n/2}/\Gamma(\nu_n/2)$ and $B_n = 2^{-(\kappa+1)/2}\pi^{-1/2}\lambda^{\kappa/2}/\Gamma(\kappa/2)$, with $\Gamma(\alpha)=\int_0^\infty t^{\alpha-1}\exp\{-t\}dt$. The identity $(Y_{n,j}-\theta_{n,j})^2+\theta^2/(\tau_n w_{n,j}) = (1-v_{n,j,1})Y^2_{n,j}+v_{n,j,1}\{\theta_{n,j}-(1-v_{n,j,1})Y_{n,j}\}^2$ along with $\int_{-\infty}^\infty \exp\{-(t-\alpha)^2/\beta\}dt = (\pi\beta)^{1/2}$ and $\int_0^\infty t^{-(\alpha+1)}\exp\{-\beta/t\}\,dt = \beta^{-\alpha}\Gamma(\alpha)$ provide

$$\boldsymbol{m}_n(\boldsymbol{Y}_n\hat{\boldsymbol{\Sigma}}_n|\mathrm{H}_1) = C_n^{p_n}|\hat{\boldsymbol{\Sigma}}_n|^{\nu_n/2-1}\prod_{j=1}^{p_n}\left[(1-v_{n,j,1})\left\{(1-v_{n,j,1})\,Y^2_{n,j}+\nu_n\hat{\sigma}^2_{n,j}+\lambda\right\}^{-(\nu_n+\kappa+1)}\right]^{1/2}, \tag{26}$$

where $C_n = \{\Gamma((\nu_n+\kappa+1)/2)\pi^{-1/2}\nu_n^{\nu_n/2}\lambda^{\kappa/2}\}/\{\Gamma(\nu_n/2)\Gamma(\kappa/2)\}$. A similar manipulation leads to

$$\boldsymbol{m}_n(\boldsymbol{Y}_n\hat{\boldsymbol{\Sigma}}_n|\mathrm{H}_0) = C_n^{p_n}|\hat{\boldsymbol{\Sigma}}_n|^{\nu_n/2-1}\prod_{j=1}^{p_n}\left\{Y^2_{n,j}+\nu_n\hat{\sigma}^2_{n,j}+\lambda\right\}^{-(\nu_n+\kappa+1)/2},$$

which, together with (26), yields (14). The identities leading to (26) also lead to the marginal posterior density

$$\boldsymbol{\pi}_n(\theta_{n,j}|\mathrm{H}_1,Y_{n,j},\hat{\sigma}^2_{n,j}) \propto \left\{1+\frac{1}{\nu_n+\kappa+1}\left(\frac{\theta_{n,j}-v_{n,j,1}Y_{n,j}}{\eta_{n,j}}\right)^2\right\}^{-(\nu_n+\kappa+2)/2},$$

where $\eta_{n,j}$ is as defined below (14). This defines the $v_{n,j,1}Y_{n,j}+\eta_{n,j}t_{\nu_n+\kappa+1}$ distribution.

2 Derivation of the smooth prior and posterior null probability

A prior density associated with $\boldsymbol{\theta}_n|\mathrm{H}_1 \sim N(0, \tau_n^2 \boldsymbol{W}_n \boldsymbol{\Sigma}_n)|\mathcal{B}_{n,s,M}$ is calculated as

$$\boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\mathrm{H}_1, \mathcal{B}_{n,s,M}) = \frac{(2\pi)^{-p_n/2} I_{\mathcal{B}_{n,s,M}}(\boldsymbol{\theta}_n^{SS})}{\boldsymbol{P}_n[Q_{n,1} \leq M]} \prod_{j=1}^{p_n} (\tau_n^2 w_{n,j} \sigma_{n,j}^2)^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{p_n} \theta_{n,j}^2 / (\tau_n^2 w_{n,j} \sigma_{n,j}^2) \right\}, \tag{27}$$

where $I_{\mathcal{B}_{n,s,M}}$ is an indicator function for $\mathcal{B}_{n,s,M}$. To see this, observe that by writing $Z_{n,j} = \{n/(\tau_n^2 w_{n,j})\}^{1/2} \theta_{n,j}^{SS}$, $Q_{n,1}$ in (20) is $\sum_j j^{2s} \{\theta_{n,j}^{SS}\}^2$, the norm defining $\mathcal{B}_{n,s,M}$. Its distributional properties are as claimed given that $\boldsymbol{\theta}_n|\mathrm{H}_1 \sim N(0, \tau_n^2 \boldsymbol{W}_n \boldsymbol{\Sigma}_n)$, unconditionally, since then the $Z_{n,j}$ are independent and standard normal. Thus, $\boldsymbol{P}_n[Q_{n,1} \leq M] = \boldsymbol{P}[\boldsymbol{\theta}_n^{SS} \in \mathcal{B}_{n,s,M}|\mathrm{H}_1]$, and (27) is calculated from $\boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\mathrm{H}_1, \mathcal{B}_{n,s,M}) = \boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\mathrm{H}_1) \times I_{\mathcal{B}_{n,s,M}}(\boldsymbol{\theta}_n^{SS}) \;/\; \boldsymbol{P}_n[\boldsymbol{\theta}_n^{SS} \in \mathcal{B}_{n,s,M}|\mathrm{H}_1]$.

The posterior probability $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \mathcal{B}_{n,s,M}]$ is calculated from the formula (2), but with $\boldsymbol{m}_n(\boldsymbol{Y}_n|\mathrm{H}_1, \mathcal{B}_{n,s,M}) = \int_{\boldsymbol{\theta}_n \neq \mathbf{0}} m(\boldsymbol{Y}_n|\boldsymbol{\theta}_n, \boldsymbol{\Sigma}_n) \boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\mathrm{H}_1, \mathcal{B}_{n,s,M}) d\boldsymbol{\theta}_n$ replacing $\boldsymbol{m}_n(\boldsymbol{Y}_n|\mathrm{H}_1)$. Under (1) and (27), an application of Bayes's theorem provides that

$$\boldsymbol{m}_n(\boldsymbol{Y}_n|\mathrm{H}_1, \mathcal{B}_{n,s,M}) \quad = \quad \frac{\boldsymbol{m}_n(\boldsymbol{Y}_n|\mathrm{H}_1)}{\boldsymbol{P}_n[Q_{n,1} \leq M]} \int_{\mathcal{B}_{n,s,M}} \boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\boldsymbol{Y}_n) d\boldsymbol{\theta}_n,$$

where $\boldsymbol{m}_n(\boldsymbol{Y}_n|\mathrm{H}_1)$ is as in the basic model, and $\boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\boldsymbol{Y}_n)$ is a multivariate normal density with mean vector $\{\boldsymbol{I} + (\tau_n \boldsymbol{W}_n)^{-1}\}^{-1} \boldsymbol{Y}_n$ and covariance matrix $\{\boldsymbol{I} + (\tau_n \boldsymbol{W}_n)^{-1}\}^{-1} \boldsymbol{\Sigma}_n$. Now $Q_{n,2}|\boldsymbol{Y}_n$ in (20) is $\sum_j j^{2s} \{\theta_{n,j}^{SS}\}^2$ for $Z_{n,j} = \sqrt{n}[v_{n,j,-1/2}\{\theta_{n,j}^{SS}\}^2 - v_{n,j,1/2}\{Y_{n,j}^{SS}\}^2]$. Conditioning on $\boldsymbol{Y}_n$, the $Z_{n,j}$ are independent and standard normal when $\boldsymbol{\theta}_n|\boldsymbol{Y}_n \sim \boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\boldsymbol{Y}_n)$, and so $\int_{\mathcal{B}_{n,s,M}} \boldsymbol{\pi}_n(\boldsymbol{\theta}_n|\boldsymbol{Y}_n) d\boldsymbol{\theta}_n = \boldsymbol{P}_n[Q_{n,2} \leq M|\boldsymbol{Y}_n]$, which establishes (19).

3 Connection to frequentist rates of testing

The following argument admits use of the frequentist rates of testing theory developed in Spitzner (2008), through criterion (23). Working in a manner similar to Section 4.6, let us take a frequentist perspective on the Bayesian testing problem by treating $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n]$ as a "test statistic." As before, with $SF(\boldsymbol{Y}_n)$ ignored, this is normalized as $\hat{T}_n = (n\|\boldsymbol{v}_{n,1/2} \boldsymbol{Y}_n^{SS}\|^2 - \|\boldsymbol{v}_{n,1/2}\|^2)/(\sqrt{2}\|\boldsymbol{v}_{n,1}\|)$. Frequentist rates of testing theory replaces (22) with

$$\inf_{\boldsymbol{\theta}_n^{SS} \in \mathcal{H}_{1,n}(\hat{\delta}_n/\delta_n)} P[\hat{T}_n > c] \to 1 \text{ for every } \delta_n \to 0,$$

provided there is an $\alpha < 1$ for which $\limsup_n P[\hat{T}_n > c] \leq \alpha$ when $\theta = 0$. The latter condition is identical to $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n] \to 1$ for any $r_n \to \infty$ for θ consistent with H_0. To see that $P[\hat{T}_n > c] \to 1$ is equivalent to (23), use (9,10) to write

$$P[\hat{T}_n > c] \quad = \quad P\left[T_n > -\frac{n\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SS}\|^2/\sqrt{2\|\boldsymbol{v}_{n,1}\|^2} + c}{\sqrt{1 + 2n\|\boldsymbol{v}_{n,1}\boldsymbol{\theta}_n^{SS}\|^2/\|\boldsymbol{v}_{n,1}\|^2}}\right],$$

where T_n is $\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SS}\|^2$ standardized to have mean zero and variance one. Thus, $P[\hat{T}_n > c] \to 1$ is equivalent to the conditions $n\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SS}\|^2/\|\boldsymbol{v}_{n,1}\| \to \infty$ and $n^{1/2} \times \|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SS}\|^2 \,/\, \|\boldsymbol{v}_{n,1}\boldsymbol{\theta}_n^{SS}\| \to \infty$. The former is (23) rewritten; it implies the latter since $\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SS}\| \geq \|\boldsymbol{v}_{n,1}\boldsymbol{\theta}_n^{SS}\|$ and so $n^{1/2}\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SS}\|^2/\|\boldsymbol{v}_{n,1}\boldsymbol{\theta}_n^{SS}\| \geq n^{1/2}\|\boldsymbol{v}_{n,1/2}\boldsymbol{\theta}_n^{SS}\|$, which diverges to infinity when (23) holds.

4 Proof of Theorem 2

It will be convenient to adopt the following vector notation for this section. With the exception of the data vectors $\boldsymbol{Y}_n^{SS}$ and parameter vectors $\boldsymbol{\theta}_n^{SS}$, every vector will have a double subscript and the entries of any vector will be triple-subscripted, in the manner of the previously-defined vector $\boldsymbol{v}_{n,k} = [v_{n,1,k}, \ldots, v_{n,p_n,k}]^T$, for which $v_{n,j,k} = \{1 + \tilde{\sigma}_{n,j}^2(\tilde{\tau}_n^2 w_{n,j})^{-1}\}^{-k}$. The first index of a vector-entry is n, the second shall indicate its place, and the third a power to which the vector-entry has been raised. All entries of the same vector shall share the same third-subscript value, matching that of the vector itself. Thus, for instance, one might define $\boldsymbol{c}_{n,1} = [c_{n,1,1}, \ldots, c_{n,p_n,1}]^T$ in which case $\boldsymbol{c}_{n,k} = [c_{n,1,k}, \ldots, c_{n,p_n,k}]^T$ is equivalent to $\boldsymbol{c}_{n,k} = [c_{n,1,1}^k, \ldots, c_{n,p_n,1}^k]^T$. Some special vectors which will be used in the proof are: $\boldsymbol{u}_{n,k}$, defined by entries $u_{n,j,1} = j^{2s}$; $\tilde{\boldsymbol{v}}_{n,k}$, defined by entries $\tilde{v}_{n,j,1} = v_{n,j,1}(1 - v_{n,j,1})^{-1}$; and $\boldsymbol{a}_{n,k}$ and $\tilde{\boldsymbol{a}}_{n,k}$ defined by respective entries $a_{n,j,1} = \{1 + (n/p_n)/(Cj^{2s}v_{n,j,1})\}^{-1}$ and $\tilde{a}_{n,j,1} = \{1 + (n_n/p)/(Cj^{2s}\tilde{v}_{n,j,1})\}^{-1}$, where C is a scalar argument yet to be given.

See Johnson and Kotz (1970) to verify that the cumulant generating functions of $Q_{n,1}$ and $Q_{n,2}|\boldsymbol{Y}_n$ are given by

$$\begin{aligned} K_{n,1}(t) &= \frac{1}{2}\left[-\sum_{j=1}^{p_n}\log\left(1 - 2n^{-1}j^{2s}\tilde{v}_{n,j,1}t\right)\right] \\ K_{n,2}(t|\boldsymbol{Y}_n) &= \frac{1}{2}\sum_{j=1}^{p_n}\left[-\log\left(1 - 2n^{-1}j^{2s}v_{n,j,1}t\right) + \frac{nv_{n,j,1}\{Y_{n,j}^{SS}\}^2}{1 - 2n^{-1}j^{2s}v_{n,j,1}t}\right] \\ &\quad -\frac{n}{2}\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SS}\|^2. \end{aligned}$$

Let us parametrize the relevant saddlepoints as $\hat{\lambda}_{n,1} = -C_{n,1}p_n/2$ and $\hat{\lambda}_{n,2}(\boldsymbol{Y}_n) =$

$-C_{n,2}p_n/2$, and furthermore write

$$K'_{n,1}(-Cp_n/2) = \frac{1}{C}\left\{1 - \frac{n}{Cp_n^2}\|\boldsymbol{u}_{n,-1/2}\boldsymbol{v}_{n,-1/2}\boldsymbol{a}_{n,1/2}\|^2\right\} + \frac{n}{C^2p_n^2}\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}_{n,1}\|^2 + \frac{n^2}{C^3p_n^3}\|\boldsymbol{u}_{n,-1}\boldsymbol{a}_{n,1}\tilde{\boldsymbol{a}}_{n,1/2}\|^2, \tag{28}$$

$$K'_{n,2}(-Cp_n/2|\boldsymbol{Y}_n) = \frac{1}{C}\left\{1 - \frac{n}{Cp_n^2}\|\boldsymbol{u}_{n,-1/2}\boldsymbol{v}_{n,-1/2}\boldsymbol{a}_{n,1/2}\|^2\right\} + \frac{n}{C^2p_n^2}\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}_{n,1}\|^2 + \frac{n^2}{C^2p_n^2}\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}_{n,1}\boldsymbol{\theta}_n^{SS}\|^2 + \frac{n}{C^2p_n^2}T_{n,1}\sqrt{2\|\boldsymbol{u}_{n,-1}\boldsymbol{a}_{n,2}\|^2 + 4n\|\boldsymbol{u}_{n,-1}\boldsymbol{a}_{n,2}\boldsymbol{\theta}_n^{SS}\|^2}, \tag{29}$$

where C defines $\boldsymbol{a}_{n,k}$ and $\tilde{\boldsymbol{a}}_{n,k}$, and $T_{n,1}$ is $\|u_{n,-1/2}a_{n,1}\boldsymbol{Y}_n^{SS}\|^2$ standardized to have mean zero and variance one, conditioning on C, and is calculated using (9,10). (Formula 28 follows from the identities $(a_{n,j,1}/v_{n,j,1}) - (\tilde{a}_{n,j,1}/\tilde{v}_{n,j,1}) = a_{n,j,1}\tilde{a}_{n,j,1}$ and $\tilde{a}_{n,j,1} - a_{n,j,1} = \{n/(Cp_n)\}u_{n,j,-1}a_{n,j,1}\tilde{a}_{n,j,1}$.) When C is replaced by $C_{n,1}$ or $C_{n,2}$, denote by $\boldsymbol{a}_{n,k}^{(1)}$ and $\tilde{\boldsymbol{a}}_{n,k}^{(1)}$ or $\boldsymbol{a}_{n,k}^{(2)}$ and $\tilde{\boldsymbol{a}}_{n,k}^{(2)}$, respectively, the corresponding values of $\boldsymbol{a}_{n,k}$ and $\tilde{\boldsymbol{a}}_{n,k}$.

Since $\gamma > 0$ and $s > 1/2$, one has $\|\boldsymbol{u}_{n,-k}\boldsymbol{a}_{n,l}\|^2 = O(1)$, $\|\boldsymbol{u}_{n,-k}\boldsymbol{v}_{n,-k}\boldsymbol{a}_{n,l}\|^2 = O(1)$, and $\|\boldsymbol{u}_{n,-k}\boldsymbol{a}_{n,l}\tilde{\boldsymbol{a}}_{n,l/2}\|^2 = O(1)$ for each $k \geq 1/2$ and $l \geq 0$, and $\boldsymbol{\theta}_n^{SS} \in \mathcal{B}_{n,s,M}$ implies $\|\boldsymbol{u}_{n,-k}\boldsymbol{a}_{n,l}\boldsymbol{\theta}_n^{SS}\|^2$ is bounded above by M. Moreover, the assumption $\lim_n \log p_n/\log n > 2/3$ implies $(n^2/p_n^3) \to 0$, hence also $(n/p_n^2) = (n^2/p_n^4)^{1/2} \to 0$.

Thus, it is clear from (28) that $K'_{n,1}(\hat{\lambda}_{n,1}) = M$ implies $C_{n,1} \to 1/M$. Similarly, (29) shows immediately that $K'_{n,2}(\hat{\lambda}_{n,1}|\boldsymbol{Y}_n) = M$ implies $C_{n,2} \to 1/M$ for the cases $n/p_n \to 0$ or $\theta = 0$. If $p_n/n = O(1)$ and $\boldsymbol{\theta}_n^{SS} \neq 0$, the $C_{n,2}$ for which $K'_{n,2}(\hat{\lambda}_{n,1}|\boldsymbol{Y}_n) = M$ is bounded above and $C_{n,2} \geq 1/M$. To see this, observe that $\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}_{n,1}\boldsymbol{\theta}_n^{SS}\|^2$ is maximized for $\boldsymbol{\theta}_n^{SS} \in \mathcal{B}_{n,s,M}$ when $\{\theta_{n,\hat{j}_n}^{SS}\}^2 = M\hat{j}_n^{-2s}$ for some $\hat{j}_n = 1, \ldots, p_n$. This means $\{n/(Cp_n)\}^2\|\boldsymbol{u}_{-1/2}\boldsymbol{a}_1\boldsymbol{\theta}^{SS}\|^2 \leq M\{C(p_n/n)\hat{j}_n^{2s} + \hat{j}_n^r\}^{-2} \leq M$, which, through (29), yields the assertion.

A saddlepoint approximation to $\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \mathcal{B}_{n,s,M}]$ is deduced as follows. Set $A_{n,1} = \log\{\kappa_{n,2,2}(\boldsymbol{Y}_n)/\kappa_{n,1,2}\} + \log\{K''_{n,1}(\hat{\lambda}_{n,1})/K''_{n,2}(\hat{\lambda}_{n,2}(\boldsymbol{Y}_n)|\boldsymbol{Y}_n)\}$ and write

$$K_{n,2}(\hat{\lambda}_{n,2}(\boldsymbol{Y}_n)|\boldsymbol{Y}_n) - K_{n,1}(\hat{\lambda}_{n,1}) - \hat{\lambda}_{n,2}(\boldsymbol{Y}_n)M + \hat{\lambda}_{n,1}M = \frac{1}{2}\left[A_{n,2} + A_{n,3} + A_{n,4} + n\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}_n^{SS}\|^2\right] \tag{30}$$

where

$$\begin{aligned} A_{n,2} &= \sum_{j=1}^{p_n} \log\left\{1 - \frac{(C_{n,2} - C_{n,1}) - C_{n,1}\tilde{v}_{n,j,1}}{C_{n,2} + (n/p_n)/(j^{2s} v_{n,j,1})}\right\}, \\ A_{n,3} &= M(C_{n,2} - C_{n,1})p_n, \qquad (31) \\ A_{n,4} &= \frac{n}{C_{n,2}p_n}\left\{\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}^{(2)}_{n,1/2}\|^2 + n\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}^{(2)}_{n,1/2}\boldsymbol{\theta}^{SS}_n\|^2 \right. \\ &\quad \left. + T_{n,2}\sqrt{2\|\boldsymbol{u}_{n,-1}\boldsymbol{a}^{(2)}_{n,1}\|^2 + 4n\|\boldsymbol{u}_{n,-1}\boldsymbol{a}^{(2)}_{n,1}\boldsymbol{\theta}^{SS}_n\|^2}\right\}, \end{aligned}$$

where $T_{n,2}$ is $\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}^{(2)}_{n,1/2}\boldsymbol{Y}^{SS}_n\|^2$ standardized to have mean zero and variance one, conditional on $C_{n,2}$. The term $n\|\boldsymbol{v}_{n,1/2}\boldsymbol{Y}^{SS}_n\|^2$ in (30) cancels in the formula (19), and the approximation becomes

$$\begin{aligned} &\boldsymbol{P}_n[\mathrm{H}_0|\boldsymbol{Y}_n, \mathcal{B}_{n,s,M}] \approx \qquad (32) \\ &\quad \left[1 + \exp\left\{\frac{1}{2}\left(A_{n,1} + A_{n,2} + A_{n,3} + A_{n,4} - \|\boldsymbol{v}_{n,1/2}\|^2 - r_n\|\boldsymbol{v}_{n,1}\|\right)\right\}\right]^{-1}. \end{aligned}$$

Under the "what if" guidelines, it is to be assumed the sequence r_n is such that $\boldsymbol{P}_n[\mathrm{H}_0|\, \boldsymbol{Y}_n, \mathcal{B}_{n,s,M}] \to 0$ is avoided for θ consistent with H_0. Subsequently, (32) indicates that a condition equivalent (under the saddlepoint approximation) to (22) is

$$\begin{aligned} \inf_{\boldsymbol{\theta}^{SS}_n \in \mathcal{H}_{1,n}(\delta_n/\delta^*_n)} &\left\{(A_{n,1} + A_{n,2} + A_{n,3} + A_{n,4})|\boldsymbol{\theta}^{SS}_n - (A_{n,1} + A_{n,2} + A_{n,3} + A_{n,4})|\mathbf{0}\right\} \quad (33) \\ &\to \infty \text{ for every } \delta^*_n \to 0, \end{aligned}$$

where "$\bullet|\boldsymbol{\theta}^{SS}_n$" indicates the term "$\bullet$" is to be considered at a particular value of $\boldsymbol{\theta}^{SS}_n$, and $\mathbf{0}$ is the vector with each entry zero. The criterion (33) will be evaluated by considering the differences $A_{n,k}|\boldsymbol{\theta}^{SS}_n - A_{n,k}|\mathbf{0}$ individually.

To evaluate $A_{n,1}|\boldsymbol{\theta}^{SS}_n - A_{n,1}|\mathbf{0}$, first write $A_{n,1} = \log\{C^2_{n,2}\|\tilde{\boldsymbol{a}}^{(1)}_{n,1/2}\|^2\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,1}\|^2 A_{11}\} - \log\{C^2_{n,1}\|\boldsymbol{u}_{n,1}\tilde{\boldsymbol{v}}_{n,1}\|^2 A_{12}\}$, where

$$\begin{aligned} A_{n,11} &= 1 + 2\frac{\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,3/2}\|^2}{\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,1}\|^2} + 2n\frac{\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,3/2}\boldsymbol{\theta}^{SS}_n\|^2}{\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,1}\|^2} \\ &\quad + \frac{T_{n,3}}{\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,1}\|^2}\sqrt{2\|\boldsymbol{u}_{n,2}\boldsymbol{v}_{n,3}\|^2 + 4n\|\boldsymbol{u}_{n,2}\boldsymbol{v}_{n,3}\boldsymbol{\theta}^{SS}_n\|^2}, \\ A_{n,12} &= \|\boldsymbol{a}^{(2)}_{n,1/2}\|^2 + 2\frac{n}{C_{n,2}p_n}\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}^{(2)}_{n,3/2}\|^2 + 2\frac{n^2}{C_{n,2}p_n}\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}^{(2)}_{n,3/2}\boldsymbol{\theta}^{SS}_n\|^2 \\ &\quad + 2\frac{n^2}{C_{n,2}p_n}T_{n,4}\sqrt{2\|\boldsymbol{u}_{n,-1}\boldsymbol{a}^{(2)}_{n,3}\|^2 + 4n\|\boldsymbol{u}_{n,-1}\boldsymbol{a}^{(2)}_{n,3}\boldsymbol{\theta}^{SS}_n\|^2} \end{aligned}$$

and $T_{n,3}$ and $T_{n,4}$ are respectively $n\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,3/2}\boldsymbol{Y}_n^{SS}\|^2$ and $\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}_{n,3/2}^{(2)}\boldsymbol{Y}_n^{SS}\|^2$ standardized to have mean zero and variance one, conditional on $C_{n,2}$. Thus it is seen that $A_{n,1}|\boldsymbol{\theta}^{SS} - A_{n,1}|\mathbf{0} \to \infty$ only if $n\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,3/2}\boldsymbol{\theta}_n^{SS}\|^2 \,/\, \|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,1}\|^2 \to \infty$, for any sequence of $\boldsymbol{\theta}_n^{SS} \in \mathcal{B}_{n,s,M}$. Now, $\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,3/2}\boldsymbol{\theta}_n^{SS}\|^2$ is maximized for $\boldsymbol{\theta}_n^{SS} \in \mathcal{B}_{n,s,M}$ when $\{\boldsymbol{\theta}_{n,\hat{j}_n}^{SS}\}^2 = M\hat{j}_n^{-2s}$ for some $\hat{j}_n = 1,\ldots,p_n$, which means $n\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,3/2}\boldsymbol{\theta}_n^{SS}\|^2 \le Mnp_n^{2s-3(\gamma-\epsilon)}$ for any $\epsilon > 0$ and sufficiently large p_n. Also $\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,1}\|^{-2} = O(p_n^{-4s+2(\gamma+\epsilon)-1})$ for any $\epsilon > 0$ and sufficiently large p_n. Setting $0 < \epsilon \le (2s+\gamma-1/2)/5$ then provides $n\|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,3/2}\boldsymbol{\theta}_n^{SS}\|^2 \,/\, \|\boldsymbol{u}_{n,1}\boldsymbol{v}_{n,1}\|^2 = O(n/p_n^{2s+\gamma+1-5\epsilon}) = O(n/p_n^{3/2}) \to 0$, hence $A_{n,1}|\boldsymbol{\theta}_n^{SS} - A_{n,1}|\mathbf{0} = O(1)$ for any sequence of $\boldsymbol{\theta}_n^{SS} \in \mathcal{H}_{1,n}(\delta_n/\delta_n^*)$.

To evaluate $A_{n,2}|\boldsymbol{\theta}_n^{SS} - A_{n,2}|\mathbf{0}$, first write $M(C_{n,2}-C_{n,1}) = C_{n,2}K'_{n,2}(-C_{n,2}p_n/2) - C_{n,1}K'_{n,1}(-C_{n,1}p_n/2)$, and consider (28) and (29) to see that

$$\begin{aligned} M(C_{n,2}-C_{n,1}) \quad &= \quad \frac{n^2}{C_{n,2}p_n^2}\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}_{n,1}^{(2)}\boldsymbol{\theta}_n^{SS}\|^2 - \frac{n^2}{C_{n,1}^2p_n^3}\|\boldsymbol{u}_{n,-1}\boldsymbol{a}_{n,1}^{(1)}\tilde{\boldsymbol{a}}_{n,1/2}^{(1)}\|^2 \\ &\quad + \frac{n}{C_{n,2}p_n^2}T_{n,1}\sqrt{2\|\boldsymbol{u}_{n,-1}\boldsymbol{a}_{n,2}^{(2)}\|^2 + 4n\|\boldsymbol{u}_{n,-1}\boldsymbol{a}_{n,2}^{(2)}\boldsymbol{\theta}_n^{SS}\|^2}. \end{aligned} \tag{34}$$

It follows that $\{C_{n,2}-C_{n,1}\}|\mathbf{0}$ is possibly negative, but $\{C_{n,2}-C_{n,1}\}|\boldsymbol{\theta}_n^{SS}$ is positive for sufficiently large n. Also (29) indicates that $C_{n,2}|0 \le C_{n,2}|\boldsymbol{\theta}_n^{SS}$, for sufficiently large n. Thus by considering the form of $A_{n,2}$ in (31), it is seen that $A_{n,2}|\boldsymbol{\theta}_n^{SS} - A_{n,2}|\mathbf{0} = O(1)$ for any sequence of $\boldsymbol{\theta}_n^{SS} \in \mathcal{H}_{1,n}(\delta_n/\delta_n^*)$.

It has been shown that (33) is satisfied only if

$$\inf_{\boldsymbol{\theta}_n^{SS}\in\mathcal{H}_{1,n}(\delta_n/\delta_n^*)} \Big\{(A_{n,3}+A_{n,4})|\boldsymbol{\theta}_n^{SS} - (A_{n,3}+A_{n,4})|\mathbf{0}\Big\} \to \infty \text{ for every } \delta_n^* \to 0. \tag{35}$$

The combination of (31) and (34), and the property

$$0 < \|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}_{n,1}^{(2)}\boldsymbol{\theta}_n^{SS}\|/\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}_{n,1/2}^{(2)}\boldsymbol{\theta}_n^{SS}\| \le 1,$$

provides that (35) holds only if $\{n^2/(C_{n,2}p_n)\}\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}_{n,1/2}^{(2)}\boldsymbol{\theta}_n^{SS}\|^2 \to \infty$ for each sequence of $\boldsymbol{\theta}_n^{SS} \in \mathcal{H}_{1,n}(\delta_n/\delta_n^*)$ and any $\delta_n^* \to 0$. Set $\hat{j}_n$ to the largest integer not to exceed $\{\delta_n/(\delta_n^*\sqrt{M})\}^{-1/s}$, and define $\boldsymbol{\theta}_n^{SS}$ according to $\{\theta_{n,\hat{j}}^{SS}\}^2 = M\hat{j}_n^{-2s}$ and $\{\theta_{n,j}^{SS}\}^2 = 0$ if $j \ne \hat{j}_n$. Therefore, $\boldsymbol{\theta}_n^{SS} \in \mathcal{H}_{1,n}(\delta_n/\delta_n^*)$ and

$$\frac{n^2}{C_{n,2}p_n}\|\boldsymbol{u}_{n,-1/2}\boldsymbol{a}_{n,1/2}^{(2)}\boldsymbol{\theta}_n^{SS}\|^2 \quad = \quad M\frac{nv_{n,\hat{j},1}}{C_{n,2}(p_n/n)\hat{j}_n^{4s}v_{n,\hat{j}_n,1} + \hat{j}_n^{2s}} = O\left(\frac{n^2}{p_n}\delta_n^4\right),$$

which completes the proof.

REFERENCES

Abramovich, F., Antoniadis, A., Sapatinas, T., and Vidakovic, B. (2002), Optimal testing in functional analysis of variance models, *International Journal of Wavelets, Multiresolution, and Information Processing*, 2:323–349.

Adams, R. A., and Fournier, J. J. F. (2003), *Sobolev Spaces*, Boston: Academic Press.

Aerts, M., Claeskens, G., and Hart, J. (2000), Testing lack of fit in multiple regression, *Biometrika,* 87:405-424.

Aitkin, M. (1991), Posterior Bayes factors, *Journal of the Royal Statistical Society B,* 53:111-142.

Andrews, D. (1994), The large sample correspondence between classical hypothesis tests and Bayesian posterior odds tests, *Econometrica,* 62:1207-1232.

Barndorff-Nielsen, O. E., and Cox, D. R. (1989), *Asymptotic Techniques for Use in Statistics,* London: Chapman and Hall.

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis,* New York: Springer.

Berger, J. O., and Delampady, M. (1987), Testing precise hypotheses (with discussion), *Statistical Science*, 2:317-352.

Berger, J. O., Ghosh, J. K., and Mukhopadhyay N. (2003), Approximations and consistency of Bayes factors as model dimension grows, *Journal of Statistical Planning and Inference*, 112:241-258.

Berger, J. O., and Pericchi, L. (1996), The intrinsic Bayes factor for linear models (with discussion), *Bayesian Statistics 5* (ed. J. M. Bernardo), pp. 25-44, Oxford: Oxford University Press.

Berger, J. O., and Sellke, T. (1987), Testing a point null hypothesis: the irreconcilability of P values and evidence, *Journal of the American Statistical Association,* 82:112-122.

Billingsley, P. (1995), *Probability and Measure* (3rd ed.), New York: Wiley.

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods* (2nd ed.), New York: Springer.

Brown, L. D., and Low, M. G. (1996), Asymptotic equivalence of nonparametric regression and white noise, *Annals of Statistics*, 24:2384-2398.

Claeskens, G., and Hjort, N. L. (2004), Goodness of fit via nonparametric likelihood ratios, *Scandinavian Journal of Statistics,* 31:487-513.

Diaconis, P., and Freedman, D. (1986), On the consistency of Bayes estimates, *Annals of Statistics*, 14:1-26.

Eubank, R. (1999), *Nonparametric Regression and Spline Smoothing*, New York:Marcel Dekker.

Eubank, R. (2000), Testing for no effects by cosine methods, *Scandinavian Journal of Statistics*, 27:747-763.

Eubank, R., and Hart, J. (1992), Testing goodness-of-fit in regression via order selection criteria, *Annals of Statistics*, 20:1412-1425.

Eubank, R., and LaRiccia, V. N., (1992), Asymptotic comparison of Cramér-von Mises and nonparametric function estimation techniques for testing goodness of fit, *Annals of Statistics*, 20:2071-2086.

Efron, B., and Gous, A., (2001), Scales of evidence for model selection: Fisher versus Bayes (with discussion), In: *Model Selection* (ed. P. Lahiri), 208-256, Beachwood, OH: Institute of Mathematical Statistics.

Fan, J. (1996), Tests of significance based on wavelet thresholding and Neyman's truncation, *Journal of the American Statistical Association*, 91:674-688.

Fan, J., and Huang, L.-S. (2001), Goodness-of-fit testing for parametric regression models, *Journal of the American Statistical Association*, 96:640-652.

Fan, J., and Lin, S.-K. (1998), Test of significance when data are curves, *Journal of the American Statistical Association*, 93:1007-1021.

Fan, J., Zhang, C., and Zhang, J. (2001), Generalized likelihood ratio statistics and Wilks phenomenon, *Annals of Statistics*, 29:153-193.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis* (2nd ed.), London: Chapman and Hall.

Ghosal, S. (2001), Convergence rates for density estimation with Bernstein polynomials, *Annals of Statistics*, 29:1264-1280.

Inglot, T., and Ledwina, T. (1996), Asymptotic optimality of data driven Neyman's tests, *Annals of Statistics*, 24:1982-2019.

Ingster, Y. I. (1993), Asymptotically minimax hypothesis testing for nonparametric alternatives I-III, *Mathematical Methods of Statistics*, 2:85-114; 3:171-189; 4:249-268.

Jensen, D. R., and Solomon, H. (1972), A Gaussian approximation to the distribution of a definite quadratic form, *Journal of the American Statistical Association*, 67:898-902.

Jensen, J. L. (1995), *Saddlepoint Approximations*, Oxford: Oxford University Press.

Johnson, N. L., and Kotz, S. (1970), *Continuous Univariate Distributions 2,* Boston: Houghton Mifflin.

Johnson, V. E. (2005), Bayes factors based on test statistics, *Journal of the Royal Statistical Society B,* 67:689-701.

Kass, R. E., and Raftery, A. E., (1995) Bayes factors, *Journal of the American Statistical Association,* 90:773-795.

Kuonen, D. (1999), Saddlepoint approximations for distributions of quadratic forms of normal variables, *Biometrika,* 86:929-935.

Lindley, D. V. (1957), A statistical paradox, *Biometrika,* 44:187-192.

Nussbaum, M. (1996), Asymptotic equivalence of density estimation and Gaussian white noise, *Annals of Statistics*, 24:2399-2430.

Petrone, S., and Wasserman, L. (2002), Consistency of Bernstein polynomial posteriors, *Journal of the Royal Statistical Society B,* 64:79-100.

Rayner, J. C. W., and Best, D. J. (1989), *Smooth Tests of Goodness-of-Fit,* Oxford: Oxford University Press.

Reid, N. (1988), Saddlepoint methods and statistical inference (with discussion), *Statistical Science*, 3:213-238.

Robert, C. P. (1993), A note on Jeffreys-Lindley paradox, *Statistica Sinica*, 3:603-608.

Robert, C. P. (2001), *The Bayesian Choice*, New York: Springer.

Robert, C. P., and Rousseau, J. (2002), A mixture approach to Bayesian goodness of fit, Technical report, Cahiers du CEREMADE, Université Paris Dauphine.

Spiegelhalter, D. J., and Smith, A. F. M. (1982), Bayes factors for linear and log-linear models with vague prior information, *Journal of the Royal Statistical Society B*, 44:377-387.

Spitzner, D. J. (2006), Use of goodness-of-fit procedures in high-dimensional testing, *Journal of Statistical Computation and Simulation*, 76:447-457.

Spitzner, D. J., (2008), Testing in functional data analysis using quadratic forms, *Under review.*

Spitzner, D. J., Marron, J. S., and Essick, G. (2003), Mixed-model functional ANOVA for studying human tactile perception, *Journal of the American Statistical Association*, 98:263-272.

Spokoiny, V. G. (1996), Adaptive hypothesis testing using wavelets, *Annals of Statistics,* 24:2477-2498.

Verdinelli, I., and Wasserman, L. (1998), Bayesian goodness-of-fit testing using infinite-dimensional exponential families, *Annals of Statistics,* 26:1215-1241.

Walker, E., and Wright, P. S. (2002), Comparing curves using additive models, *Journal of Quality Technology*, 34:118-129.

Acknowledgments

The author is grateful for the comments of Judith Rousseau, Dongchu Sun, the Editor, and two anonymous reviewers, to Amanda Wang for her careful reading of a draft, to William H. Woodall for introducing him to the VDP data set, and to Esteban Walker for making the data available on his website (http://bus.utk.edu/stat/walker/VDP/Allstack.TXT).