HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2016 December 10.

Published in final edited form as:

Stat Med. 2015 December 10; 34(28): 3724–3749. doi:10.1002/sim.6728.

The Power Prior: Theory and Applications

Joseph G. Ibrahim[a,*], **Ming-Hui Chen**[b], **Yeongjin Gwon**[b], and **Fang Chen**[c]

[a] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, U.S.A.

[b] Department of Statistics, University of Connecticut, Storrs, Connecticut 06269, U.S.A.

[c] SAS Institute Inc., Cary, North Carolina 27513, U.S.A.

Abstract

The *power prior* has been widely used in many applications covering a large number of disciplines. The power prior is intended to be an informative prior constructed from historical data. It has been used in clinical trials, genetics, health care, psychology, environmental health, engineering, economics, and business. It has also been applied for a wide variety of models and settings, both in the experimental design and analysis contexts. In this review article, we give an A to Z exposition of the power prior and its applications to date. We review its theoretical properties, variations in its formulation, statistical contexts for which it has been used, applications, and its advantages over other informative priors. We review models for which it has been used, including generalized linear models, survival models, and random effects models. Statistical areas where the power prior has been used include model selection, experimental design, hierarchical modeling, and conjugate priors. Prequentist properties of power priors in posterior inference are established and a simulation study is conducted to further examine the empirical performance of the posterior estimates with power priors. Real data analyses are given illustrating the power prior as well as the use of the power prior in the Bayesian design of clinical trials.

Keywords

Bayesian design; Borrowing; Clinical trials; Discounting; Historical data; Informative prior

1. Introduction

Informative prior elicitation is one of the biggest and most important topics in Bayesian inference. Bayesian inference using informative priors is becoming more widely used in an age of massive datasets and prior information including settings such as clinical trials and observational studies. Informative prior elicitation is typically not an easy task since it is typically not easy to quantify and synthesize prior information into a suitable prior. Thus, techniques and methods for synthesizing and quantifying prior information are highly needed. In the presence of historical data, informative prior elicitation can proceed in a

* Correspondence to: Joseph G. Ibrahim, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, U.S.A. ibrahim@bios.unc.edu.

much more systematic fashion, and in such cases, the quantification of prior information is more straightforward, and even "objective" in some sense.

The power prior discussed in [1] has emerged as a useful class of informative priors for a variety of situations in which historical data are available. The first paper to discuss the formalization of the power prior as a general prior for various classes of regression models is [1]. Several applications to clinical trial design and analysis as well as epidemiological studies using historical data in prior elicitation have appeared in the literature. Examples of papers discussing the use of historical data in prior elicitation include [2, 3, 4, 5, 6, 7]. Papers using the power prior and its variations include [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24]. Books illustrating the use of the power prior in epidemiological studies and clinical trials contexts include [25, 26].

One of the reasons that the power prior has become such a powerful tool in the last decade as an informative prior in both design and analysis settings is because of its ease in construction and its natural form for incorporation of historical data, its attractiveness in interpretation, its relative ease in computation, its attractive theoretical properties and uses in model selection problems, and because of the relatively few hyperparameters that need to be specified. It is an ideal tool as an informative prior in settings where historical data are available and is arguably the most widely used informative prior in such settings. The power prior is a useful general class of priors that can be used for arbitrary classes of regression models, including generalized linear models, generalized linear mixed models, survival models with censored data, frailty models, multivariate models, and nonlinear models. The power prior specification for the regression coefficients focuses on observable quantities in that the elicitation is based on historical data D_0 and a scalar parameter a_0 quantifying the heterogeneity between the current data D and the historical data D_0. The power prior distribution is then constructed by raising the likelihood function of the historical data to the power a_0, where $0 \leq a_0 \leq 1$. Such constructions of prior distributions have been discussed by [27, 1]. The power prior provides, in some sense, an "objective" and historical data-driven approach to informative prior elicitation. It is objective in the sense that the degree of informative-ness of the prior is driven by the information contained in the ("objective") historical data, not from expert opinion elicited on parameters in the model. The only hyperparameter that requires subjective elicitation in the power prior is the discounting parameter a_0, for which we highly recommend several sensitivity analyses, including analyses with $a_0 = 0$ (non-informative prior) and $a_0 = 1$ (full borrowing).

A formal justification of the power prior is given in [15] where it is shown to be an optimal class of informative priors in the sense that it minimizes a convex sum of Kullback-Leibler (KL) divergences between two specific posterior densities, in which one density is based on no incorporation of historical data, and the other density is based on pooling the historical and current data. This result provides a strong motivation for using the power prior as an informative prior in Bayesian inference. In addition, a formal relationship between this convex sum of KL divergences and the information processing rules proposed by [27, 28] is derived. Specifically, Ibrahim, Chen, and Sinha [15] showed that the power prior is the 100% efficient information processing rule in the sense defined by [27]. The power prior also has close connections with hierarchical modeling as shown in [16]. Chen and Ibrahim

[16] showed that the parameter a_0 has a direct analytic connection with the variance hyperparameter in the prior for the mean function in a normal hierarchical model.

The rest of the paper is organized as follows. In Section 2, we introduce the basic formulation of the power power and review various variations of the power prior, including the full power prior, the normalized power prior, the commensurate power prior, the partial discounting power prior, and the partial borrowing power prior. The issue of fixed or random power parameters, extensions to multiple historical datasets, and the power prior for generalized linear models are all reviewed and presented in Section 2. An in-depth review of the theory of the power prior and its properties, including the theoretical justification, connections to hierarchical models, the power prior as a conjugate prior, the property of matching predictives, and the power priors in variable selection, is given in Section 3. Section 4 includes entirely new development. In Section 4, we examine frequentist properties of the posterior estimates using power priors in linear models as well as generalized linear models. In Section 5, the determination of a guide value of the power parameter is reviewed and new derivations for the linear model are obtained for estimating the guide value along with a new simulation study to examine the empirical performance of the power prior. Section 6 reviews seven applications of the power prior in various research fields. The use of the power prior for survival analysis in the context of cancer clinical trials as well as in the Bayesian design of non-inferiority clinical trials is demonstrated in Sections 7 and 8 with new analyses. We conclude the paper with a brief discussion in Section 9.

2. The Power Prior

2.1. Basic Formulation of the Power Prior

The power prior can be constructed as follows. Let the data for the current study be denoted by D and denote the corresponding likelihood function by $L(\theta|D)$, where θ is a vector of parameters. Suppose we have historical data D_0 from a similar previous study. Let $L(\theta|D_0)$ denote the likelihood function for the historical data D_0. Here, $L(\theta|D)$ and $L(\theta|D_0)$ are general likelihood functions for arbitrary models, such as normal linear models, generalized linear models, random effects models, nonlinear models, or survival models with censored data.

The basic formulation of the power prior, as discussed in [1], is

$$\pi(\boldsymbol{\theta}|\mathbf{D_0}, \mathbf{a_0}) \propto L(\boldsymbol{\theta}|\mathbf{D_0})^{a_0} \pi_0(\boldsymbol{\theta}), \quad (2.1)$$

where $0 \leq a_0 \leq 1$ is a scalar parameter and $\pi_0(\theta)$ is the *initial prior* for θ before the historical data D_0 is observed. In many applications, $\pi_0(\theta)$ is taken to be an improper prior. Using the power prior in (2.1), the corresponding posterior distribution of θ is given by

$$\pi(\boldsymbol{\theta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0}) \propto L(\boldsymbol{\theta}|\mathbf{D}) L(\boldsymbol{\theta}|\mathbf{D_0})^{a_0} \pi_0(\boldsymbol{\theta}). \quad (2.2)$$

We see from (2.2) that a_0 weights the historical data relative to the likelihood of the current study, and thus the parameter a_0 controls the influence of the historical data on $L(\theta|D)$. The parameter a_0 can be interpreted as a precision parameter for the historical data. Since D_0 is

historical data, it is unnatural in many applications such as clinical trials to weight the historical data more than the current data; thus it is scientifically more sound to restrict the range of a_0 to be between 0 and 1, and thus we take $0 \leq a_0 \leq 1$. One of the main roles of a_0 is that it controls the heaviness of the tails of the prior for $\boldsymbol{\theta}$. As a_0 becomes smaller, the tails of (2.1) becomes heavier. Setting $a_0 = 1$, (2.1) corresponds to the update of $\pi_0(\boldsymbol{\theta})$ using Bayes theorem. That is, with $a_0 = 1$, (2.1) corresponds to the posterior distribution of $\boldsymbol{\theta}$ based on the historical data. When $a_0 = 0$, then the prior does not depend on the historical data D_0; in this case, $\pi(\boldsymbol{\theta}|D_0, a_0 = 0) \equiv \pi_0(\boldsymbol{\theta})$. Thus, $a_0 = 0$ is equivalent to a prior specification with no incorporation of historical data. Therefore, (2.1) can be viewed as a generalization of the usual Bayesian update of $\pi_0(\boldsymbol{\theta})$. The parameter a_0 allows the investigator to control the influence of the historical data on the current study. Such control is important in cases where there is heterogeneity between the previous and current studies, or when the sample sizes of the two studies are quite different. One of the most useful applications of the power prior is for model selection problems since it inherently automates the informative prior specification for all possible models in the model space (see [1, 12, 13]). The power prior given by (2.1) will be proper if the initial prior $\pi_0(\boldsymbol{\theta})$ is proper. The propriety issue of the power prior arises when an improper initial prior, such as $\pi_0(\boldsymbol{\theta}) \propto 1$, is specified. If $\int L(\boldsymbol{\theta}|\mathbf{D_0})^{a_0} d\boldsymbol{\theta} = \infty$, the initial prior plays a dominant role in the power prior since in this case, the historical data may not contain much information about the parameters $\boldsymbol{\theta}$. However, even in this case, we are still able to evaluate the extent of information contained in the power prior about $\boldsymbol{\theta}$ by taking a proper initial prior. The role of the initial prior in the power prior for model selection was extensively discussed and carefully examined in [12].

Since the power prior is basically a likelihood function raised to a power, it shares all of the properties that likelihood functions have, and therefore has several advantages over other priors. Some of these advantages include

- **(i)** Propriety: techniques for showing propriety of $\pi(\boldsymbol{\theta}|D_0, a_0)$ are exactly the same as those for showing propriety for a posterior distribution based on a dataset D_0 with likelihood function $L(\pi|D_0)$, and prior $\pi 0(\boldsymbol{\theta})$.
- **(ii)** A semi-automatic prior elicitation scheme for variable subset selection and general model selection problems.
- **(iii)** Asymptotics: Since the power prior is a likelihood raised to a power, all of the asymptotic results for likelihood theory carry over to the power prior.

Propriety results for the power prior are not difficult to characterize since one can use results from likelihood theory to obtain necessary and sufficient conditions for obtaining propriety of the power prior for a wide class of models. Ibrahim, Ryan, and Chen [10] and Chen, Ibrahim, and Yiannoutsos [12] developed such results for the power prior based on logistic regression, Chen, Ibrahim, and Shao [14] established general results for GLMs, Chen et al. [11, 29] examined propriety results for the cure rate model and piecewise exponential model. For the asymptotic property, it can be shown that for many classes of models that as $n_0 \to \infty, \pi(\boldsymbol{\theta}|\mathbf{D_0}, \mathbf{a_0}) \approx N\left(\hat{\boldsymbol{\theta}}_0, a_0^{-1} H^{-1}\left(\hat{\boldsymbol{\theta}}\right)\right)$, where $\hat{\boldsymbol{\theta}}_0$ is the mode of the power prior and

$H(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} log\left(\pi\left(\boldsymbol{\theta}|\mathbf{D_0}, \mathbf{a_0}\right)\right)$. This was demonstrated in many applications and models by Ibrahim and his co-workers.

2.2. Variations of the Power Prior

In many applications of the power prior, one may take a_0 to be fied and then do several sensitivity analysis using different values of a_0. However, one can also develop the hierarchical prior specifiation by taking a_0 random and specifying a beta distribution for it, for example. In this case, the full prior specification becomes

$$\pi\left(\boldsymbol{\theta}, \mathbf{a_0}|\mathbf{D_0}\right) \propto \pi^*\left(\boldsymbol{\theta}, \mathbf{a_0}|\mathbf{D_0}\right) \equiv L(\boldsymbol{\theta}|\mathbf{D_0})^{a_0}\pi_0\left(\boldsymbol{\theta}\right)\pi_0\left(a_0\right), \quad (2.3)$$

where $\pi_0(\boldsymbol{\theta})$ and $\pi_0(a_0)$ are the initial priors. We call (2.3) the *joint power prior*.

Another modification of the power prior when a_0 is random, which was introduced by [19, 30], is called the *normalized power prior*, and is given by

$$\pi\left(\boldsymbol{\theta}, \mathbf{a_0}|\mathbf{D_0}\right) = \pi\left(\boldsymbol{\theta}|\mathbf{D_0}, \mathbf{a_0}\right)\pi\left(a_0\right) = \frac{L(\boldsymbol{\theta}|\mathbf{D_0})^{a_0}\pi_0\left(\boldsymbol{\theta}\right)}{\int L(\boldsymbol{\theta}|\mathbf{D_0})^{a_0}\pi_0\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta}}\pi\left(a_0\right), \quad (2.4)$$

where $\pi 0(\boldsymbol{\theta})$ and $\pi_0(a_0)$ are the initial prior.

The main difference between (2.3) and (2.4) is that (2.3) specifies a joint prior distribution directly for $(\boldsymbol{\theta}, a_0)$ while (2.4) first specifies a conditional prior distribution for $\boldsymbol{\theta}$ given a_0 and then specifies a marginal distribution for a_0. For the normalized power prior in (2.4), we must have $\int L(\boldsymbol{\theta}|\mathbf{D_0})^{a_0}\pi_0\left(\boldsymbol{\theta}\right)d\boldsymbol{\theta} < \infty$ for $0 < a_0 \leq 1$. The joint power prior in (2.3) may or may not need to be proper as long as the resulting posterior is proper. However, when $\int \pi^*\left(\boldsymbol{\theta}, \mathbf{a_0}|\mathbf{D_0}\right)d\boldsymbol{\theta}\mathbf{da_0} < \infty$, the joint power prior in (2.3) is proper and can be rewritten as

$$\pi\left(\boldsymbol{\theta}, \mathbf{a_0}|\mathbf{D_0}\right) = \frac{L(\boldsymbol{\theta}|\mathbf{D_0})^{a_0}\pi_0\left(\boldsymbol{\theta}\right)\pi_0\left(a_0\right)}{\int \pi^*\left(\boldsymbol{\theta}, \mathbf{a_0}|\mathbf{D_0}\right)d\boldsymbol{\theta}\mathbf{da_0}}. \quad (2.5)$$

In this case, the joint power prior can also be viewed as a normalized power prior with the normalizing constant free of a_0. An in-depth examination of the propriety of the joint power prior in (2.3) can be found in [14] for GLMs and in [31] for generalized linear mixed Models (GLMMs).

An extension of the power prior introduced by [22] allows for different parameters for the historical and current data. Hobbs et al. [22] called such a prior the *commensurate power prior*. To illustrate this idea, we consider θ and θ_0 as the one-dimensional parameters for the current and historical data, respectively. A vague initial prior is chosen for θ_0 and the prior for θ depends on θ_0 and τ, where τ parameterizes commensurability between θ and θ_0. The information on τ is used to guide the prior on a_0. Assuming a uniform improper initial prior for θ_0, the commensurate power prior is given by

$$\pi(\theta, \theta_0, a_0, \tau | D_0) \propto \frac{L(\theta_0|D_0)^{a_0}}{\int L(\theta_0|D_0)^{a_0} d\theta_0} \pi(\theta|\theta_0, \tau)\, \pi(a_0|\tau)\, \pi_0(\tau), \quad (2.6)$$

where $\pi(\theta|\theta_0, \tau) \propto exp\left\{-\frac{\tau}{2}(\theta - \theta_0)^2\right\}$, which is the commensurate prior, $\pi(a_0|\tau) \propto a_0^{g(\tau)-1}$, $g(\tau) > 0$ is a function of the commensurability parameter that is small for τ close to zero and large for large values of τ, and $\pi_0(\tau)$ is an initial prior of the commensurate parameter. The variations of the commensurate power prior have been recently developed and discussed in [23, 26].

The other formulation of the power prior, which is now called the "*partial discounting power prior*", is especially useful and most easily motivated in latent variable models, where one wishes to discount the likelihood function of the historical data but not discount the distribution of the latent variables. The partial discounting power prior is formulated as

$$\pi(\boldsymbol{\theta}|\mathbf{D_0}, \mathbf{a_0}) \propto [\int L(\boldsymbol{\theta}|\mathbf{D_0}, \boldsymbol{\xi})^{a_0} g(\boldsymbol{\xi})\, d\boldsymbol{\xi}]\, \pi_0(\boldsymbol{\theta}), \quad (2.7)$$

where $\boldsymbol{\xi}$ is a vector of latent variables in the model, and $g(\boldsymbol{\xi})$ is the distribution of the latent variables. We see in this formulation that the discounting occurs only in the likelihood function of $\boldsymbol{\theta}$ based on the historical data, conditional on the latent variables, and the latent variable distribution $g(\boldsymbol{\xi})$ is not discounted. Chen, Dey, and Shao [32] used (2.7) in the context of skewed link models for dichotomous response data, where $g(\boldsymbol{\xi})$ denotes a skewed distribution of latent variables $\boldsymbol{\xi}$. The partial discounting power prior is attractive in the sense that information in the historical data is typically available on the regression parameters but not directly on the distribution of the latent variables in the model. In addition, (2.7) is more computationally advantageous than the full discounting power prior, which is defined as $\pi(\boldsymbol{\theta}|\mathbf{D_0}, \mathbf{a_0}) \propto [\int L(\boldsymbol{\theta}|\mathbf{D_0}, \boldsymbol{\xi})\, g(\boldsymbol{\xi})\, d\boldsymbol{\xi}]^{a_0} \pi_0(\boldsymbol{\theta})$. The partial discounting power prior is not restricted to latent variable models. This idea can easily be extended to models with random effects in which $g(\boldsymbol{\xi})$ may depend on an additional unknown variance parameter τ of random effects (e.g., [24]). The variations of the partial discounting power priors have also been developed in the literature, including [32, 33, 34, 35, 36].

In addition, a more recent variation of the power prior is called the *partial borrowing power prior* formally-introduced by [24]. The idea of the partial borrowing power prior can be traced back to [8] in analyzing human twin data, in which only summary statistics from the historical studies were available and consequently, the prior information was available only for certain parameters. Shao [37] also discussed the partial borrowing power prior method in toxicity study design and benchmark dose estimation. Chen et al. [21] used the partial borrowing power prior to borrow the historical data only for the control device from previous medical device trials. The key idea of the partial borrowing power prior is that the historical data are borrowed only through the common parameters shared in the models for the historical data and the current data. Thus, strength from the historical data is borrowed through those common parameters and at the same time, the parameters in the power prior are allowed to be different than those in the likelihood function for the current data. This attractive and flexible feature of the partial borrowing power prior allows the historical data

to have different forms (e.g., summary statistics versus individual-level data) or different models than the current data. Moreover, the partial borrowing power prior can be adapted to the fixed-a_0, random a_0, normalized, and commensurate settings. Chen et al. [21], Ibrahim et al. [24], and Chen et al. [38, 39] applied the partial borrowing power prior for clinical trial design.

2.3. Fixed or Random a_0

When a_0 is fixed, we know exactly how much historical data D_0 are incorporated into the analysis of the current data D, and also how the type I error and power are related to a_0 in Bayesian design of clinical trials, which is discussed in detail in Section 8. In addition, there is a theoretical connection between the power prior formulation and the hierarchical prior specification, as established in [16]. Some useful comments on the fixed a_0 case can be found in [18]. When a_0 is random, we need to compute the prior normalizing constant, given by

$$m^*(a_0) = \int L(\boldsymbol{\theta}|\mathbf{D_0})^{a_0} \pi_0(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \quad (2.8)$$

This normalizing constant is often analytically intractable except for normal linear regression models.

2.4 Computations of the Power Prior

The computational properties of the power prior were discussed in many papers, including [9, 10, 11, 12, 13, 14, 29]. The power prior for variable subset selection was demonstrated in [12, 13, 31, 33, 40].

When a_0 is fixed, the implementation of Markov Chain Monte Carlo (MCMC) sampling from the posterior distribution becomes straightforward, especially for complex models such as generalized linear models, random effects models, nonlinear models, or survival models with censored data. The joint power prior formulation is more computationally intensive than the a_0 fixed case. The normalized power prior formulation is even more computationally extensive than the joint power prior formulation for models other than normal linear regression models since for most non-normal models, an analytical evaluation of the integral, $\int L(\boldsymbol{\theta}|\mathbf{D_0})^{a_0} \pi_0(\boldsymbol{\theta})\, d\boldsymbol{\theta}$ in (2.4), is not available, which poses a huge challenge in sampling from the resulting posterior distribution and computing the posterior quantities of interest. To circumvent the computational issues that arise from the normalized power prior, one may extend the Monte Carlo method developed in [41] to compute the posterior quantities and the computational algorithms developed in [42, 43, 44] to sample from the posterior distribution.

2.5. Extension to Multiple Historical Datasets

Multiple historical datasets often arise in clinical trials, observational studies, carcinogenicity studies, and environmental studies. For example, in phase II and phase III clinical trials, a particular treatment is tested several times under various conditions within a

certain population. Suppose we have K_0 historical datasets, D_{0k}, $k = 1,..., K_0$. Write $D_0 = (D_{01}, \cdots, D_{0K_0})$. By extending (2.1) to the K_0 historical datasets, we have

$$\pi(\boldsymbol{\theta}|\boldsymbol{D_0}, \boldsymbol{a_0}) \propto \prod_{k=1}^{K_0} L(\boldsymbol{\theta}|\mathbf{D_{0k}})^{a_{0k}} \pi_0(\boldsymbol{\theta}). \quad (2.9)$$

where $\pi_0(\boldsymbol{\theta})$ is the initial prior for $\boldsymbol{\theta}$, $a_0 = (a_{01}, \cdots, a_{0K_0})$, and $0 \leq a_{0k} \leq 1$ for $k = 1,..., K_0$. The prior in (2.9) is attractive since it allows for diffierent a_{0k}'s for diffrent historical datasets, providing a flexible degree of discounting for each historical dataset. The theoretical and computational properties of (2.9) are similar to those of the single historical dataset case, and (2.9) can also be extended to the variations of the normalized power prior discussed in Section 2.2. The power prior for multiple historical datasets has been discussed in the literature and used in several applications, including [1, 15, 16, 19, 21, 31, 30, 45, 46, 47].

2.6. The Power Prior for Generalized Linear Models

Let y_i he the response variable and also let $\boldsymbol{x}_i$ be a p-dimensional vector of covariates for $i = 1,... ,n$ for the current study. Write $D = \{(y_i, \boldsymbol{x}_i), i = 1,...,n\} = (n, \boldsymbol{y}, X)$, where $\boldsymbol{y} = (y_1,..., y_n)'$ and $X = \left(x_1', x_2', \ldots, x_n'\right)$. Throughout the paper, we assume a generalized linear model (GLM) for y_i given $\boldsymbol{x}_i$, which has a density in the exponential class

$f(y_i|\boldsymbol{x}_i, \theta_i, \tau) = exp\left\{\alpha_i^{-1}(\tau)(y_i\theta_i - \psi(\theta_i)) + \phi(y_i, \tau)\right\}, \quad i=1, \ldots, n$, indexed by the canonical parameter θ_i and the scale parameter τ. The functions ψ and φ determine a particular family in the class, such as the binomial, normal, Poisson, etc.. The function $a_i(\tau)$ is commonly of the form $\alpha_i(\tau) = \tau^{-1} w_i^{-1}$ where the w_i's are known weights. Further suppose the θ_is satisfy the equations: $\theta_i = h(\eta_i)$ and $\eta_i = x_i'\beta$, where h is a monotone differentiable function, often referred to as the link function and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a p-dimensional vector of regression coefficients.

We assume that τ is known and denote $a_i \equiv a_i(\tau)$ and $\varphi(y_i) \equiv \varphi(y_i, \tau)$throughout the remainder of the paper. For the binomial and Poisson regression models, τ is intrinsically equal to 1. We further rewrite $f(y_i|\boldsymbol{x}_i, \theta_i, \tau)$ under the GLM as $f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}) = exp\left[\alpha_i^{-1}\left\{y_i h\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) - \psi\left(h\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)\right)\right\} + \phi(y_i)\right]$, $i = 1,..., n$. In the special case of the normal linear regression model, we have

$$f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{1}{2\sigma^2}\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)^2\right\}. \quad (2.10)$$

The likelihood function of the current data D is given by $L(\boldsymbol{\beta}|\mathbf{D}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta})$

Similarly, let $D_0 = \{(y_{0i}, \boldsymbol{x}_{0i}), i=1, 2, \ldots, n_0\} \equiv (n_0, \boldsymbol{y}_0, X_0)$ denote the historical data, where $\boldsymbol{y}_0 = (y_{01}, \ldots, y_{0n_0})'$ and $X_0 = \left(\boldsymbol{x}_{01}', \boldsymbol{x}_{02}', \ldots, \boldsymbol{x}_{0n_0}'\right)'$. We assume the GLM for y_{0i},

given by $f(y_{0i}|\boldsymbol{x}_{0i},\boldsymbol{\beta}) = exp\left[\alpha_{0i}^{-1}\left\{\left(y_{0i}h\left(\boldsymbol{x}_{0i}'\boldsymbol{\beta}\right) - \psi\left(h\left(\boldsymbol{x}_{0i}'\boldsymbol{\beta}\right)\right)\right\} + \phi(y_{0i})\right.\right.$, $i = 1,\ldots, n_0$, where $\alpha_{0i}=\tau^{-1}w_{0i}^{-1}$ and x_{0i} is a p-dimensional vector of covariates in the historical data. In the special case of normal linear regression, we have

$$f(y_{0i}|\boldsymbol{x}_{0i},\boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{1}{2\sigma^2}\left(y_{0i} - \boldsymbol{x}_{0i}'\boldsymbol{\beta}\right)^2\right\}. \quad (2.11)$$

The likelihood function of the historical data D_0 is given by $L(\boldsymbol{\beta}|\mathbf{D_0}) = \prod_{i=1}^{n_0} f(y_{0i}|\boldsymbol{x}_{0i},\boldsymbol{\beta})$, and the power prior in (2.1) with a fixed a_0 for the GLM is given by $\pi(\boldsymbol{\beta}|\mathbf{D_0},\mathbf{a_0}) \propto \{L(\boldsymbol{\beta}|\mathbf{D_0})\}^{a_0}\pi_0(\boldsymbol{\beta})$, where $0 \le a_0 \le 1$ and $\pi_0(\boldsymbol{\beta})$ is the initial prior of $\boldsymbol{\beta}$. In the normal linear regression case, the power prior reduces to $\pi(\boldsymbol{\beta}|\mathbf{D_0},\mathbf{a_0}) \propto \frac{1}{(\sigma^2)^{\frac{a_0 n_0}{2}}} exp\left\{-\frac{a_0}{2\sigma^2}\sum_{i=1}^{n_0}\left(y_{0i} - \boldsymbol{x}_{0i}'\boldsymbol{\beta}\right)^2\right\}\pi_0(\boldsymbol{\beta})$. Assume that we take an improper uniform initial prior for $\boldsymbol{\beta}$, i.e., $\pi_0(\boldsymbol{\beta}) \propto 1$. Then we have

$$\boldsymbol{\beta}|\mathbf{D_0},\mathbf{a_0} \sim \mathbf{N}\left(\left(\mathbf{X_0}'\mathbf{X_0}\right)^{-1}\mathbf{X_0}'\mathbf{y_0}, \frac{\sigma^2}{\mathbf{a_0}}\left(\mathbf{X_0}'\mathbf{X_0}\right)^{-1}\right). \quad (2.12)$$

Similarly, the normalized power prior (2.4) with a random a_0 for the GLM takes the form $\pi(\boldsymbol{\beta},\mathbf{a_0}|\mathbf{D_0}) = \pi(\boldsymbol{\beta}|\mathbf{D_0},\mathbf{a_0})\,\pi_0(a_0)$, where $\pi_0(a_0)$ is the initial prior for a_0,

$$\pi(\boldsymbol{\beta}|\mathbf{D_0},\mathbf{a_0}) = \frac{\{L(\boldsymbol{\beta}|\mathbf{D_0})\}^{a_0}\pi_0(\boldsymbol{\beta})}{\int\{L(\boldsymbol{\beta}|\mathbf{D_0})\}^{a_0}\pi_0(\boldsymbol{\beta})\,d\boldsymbol{\beta}}; \quad (2.13)$$

and $\pi_0(\boldsymbol{\beta})$ is the initial prior for $\boldsymbol{\beta}$. We may simply consider $\pi_0(a_0) \propto 1$, i.e., $a_0 \sim$ beta(1,1). We note that a closed form expression of (2.13) under the GLM is not available except for the linear model, in which we have

$$\pi(\boldsymbol{\beta}|\mathbf{D_0},\mathbf{a_0}) = \frac{a_0^{p/2}|X_0'X_0|^{1/2}}{(2\pi\sigma^2)^{p/2}} exp\left\{-\frac{a_0}{2\sigma^2}\left[\boldsymbol{\beta} - \left(\mathbf{X_0}'\mathbf{X_0}\right)^{-1}\mathbf{X_0}'\mathbf{y_0}\right]'\left(X_0'X_0\right)\left[\boldsymbol{\beta} - \left(\mathbf{X_0}'\mathbf{X_0}\right)^{-1}\mathbf{X_0}'\mathbf{y_0}\right]\right\}$$

.

3. Theory of the Power Prior and its Properties

3.1. Theoretical Justification of the Power Prior

The power prior in (2.1) has attractive theoretical properties. First, the power prior is an optimal class of informative priors in the sense that it minimizes a convex sum of Kullback-Leibler (KL) divergences between two posterior densities $f_0(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|\mathbf{D},\mathbf{D_0},\mathbf{a_0}=\mathbf{0})$ (no borrowing) and $f_1(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|\mathbf{D},\mathbf{D_0},\mathbf{a_0}=\mathbf{1})$ (full borrowing), where $\pi(\boldsymbol{\theta}|\mathbf{D},\mathbf{D_0},\mathbf{a_0})$ is given in (2.2). Mathematically, Ibrahim, Chen, and Sinha [15] showed that $\pi(\boldsymbol{\theta}|\mathbf{D},\mathbf{D_0},\mathbf{a_0}) = \underset{g\in\mathscr{G}}{argmin} D_{KL}(g)$, where $\mathscr{G} = \{g: g \text{ is a density function, i.e., } \int g(\boldsymbol{\theta})\,d\boldsymbol{\theta} = \mathbf{1}\}$ and $D_{KL}(g) = (1-a_0)\int log\left(\frac{g(\boldsymbol{\theta})}{f_0(\boldsymbol{\theta})}\right) g(\boldsymbol{\theta})\,d\boldsymbol{\theta} + \mathbf{a_0}\int \mathbf{log}\left(\frac{\mathbf{g(\boldsymbol{\theta})}}{\mathbf{f_1(\boldsymbol{\theta})}}\right)\mathbf{g}(\boldsymbol{\theta})\,\mathbf{d\boldsymbol{\theta}}$. Second, the power prior in (2.1) is a 100% efficient processing rule in the sense that $\pi(\boldsymbol{\theta}|\mathbf{D},\mathbf{D_0},\mathbf{a_0})$ minimizes the

weighted information-processing rule (Zellner [28], [48]) defined by Δ[g] = output information – weighted input information = $\int g(\boldsymbol{\theta})\ log\, g(\boldsymbol{\theta})\, d\boldsymbol{\theta} + \int g(\boldsymbol{\theta})\ log\, m(D, D_0)\, d\boldsymbol{\theta} - \int g(\boldsymbol{\theta})\ log\, \pi_0(\boldsymbol{\theta})\, d\boldsymbol{\theta} - \int g(\boldsymbol{\theta})\ log\, L(\boldsymbol{\theta}|\mathbf{D})\, d\boldsymbol{\theta} - a_0 \int g(\boldsymbol{\theta})\ log\, L(\boldsymbol{\theta}|\mathbf{D_0})\, d\boldsymbol{\theta}$, where $g \in \mathcal{G}$, $0 \le a_0 \le 1$, and $m(D, D_0) = \int L(\boldsymbol{\theta}|\mathbf{D})\, L(\boldsymbol{\theta}|\mathbf{D_0})\, \pi_0(\boldsymbol{\theta})\, d\boldsymbol{\theta}$. Similar results have been established in [30] for the normalized power prior in (2.4) based on Shannon's mutual information theory. Extensions of these results to multiple historical datasets can be found in [15, 30].

3.2. Connections to Hierarchical Models

Hierarchical modeling is a common method for combining several datasets or incorporating prior information. Chen and Ibrahim [16] established a formal connection between the power prior and hierarchical models for the class of generalized linear models via an approximate relationship between the power parameter a_0 and the variance components of the hierarchical model. This connection facilitates a direct interpretation and estimation of a_0 and unifies these two different approaches for incorporating prior information.

For the GLMs, the power prior is given in Section 2.6. Under the hierarchical GLM specification, We take $(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\Omega}) \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Omega})$ for the current data and $(\boldsymbol{\beta}_0|\boldsymbol{\mu}, \boldsymbol{\Omega}) \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Omega})$ for the historical data. In order to establish the connection between the power prior and the hierarchical model for GLMs, we further assume that Ω is fixed and specify an improper uniform initial prior for $\boldsymbol{\mu}$, i.e., $\pi_0(\boldsymbol{\mu}) \propto 1$. Under the hierarchical formulation, the posterior distribution of $\boldsymbol{\beta}$ give D and D_0 is given by $\pi_h(\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}) = \int \pi_h(\boldsymbol{\beta}, \boldsymbol{\beta_0}, \boldsymbol{\mu}|\mathbf{D}, \mathbf{D_0})\, d\boldsymbol{\beta}_0 d\boldsymbol{\mu}$, where $\pi_h(\boldsymbol{\beta}, \boldsymbol{\beta_0}, \boldsymbol{\mu}|\mathbf{D}, \mathbf{D_0})$ is the joint posterior distribution of $\boldsymbol{\beta}$, $\boldsymbol{\beta}_0$ and $\boldsymbol{\mu}$. Chen and Ibrahim [16] obtained an asymptotic approximation to $\pi_h(\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0})$ similar to [49], which is given by

$$\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0} \sim \mathbf{N_p}\left(\hat{\mathbf{A}}_\mathbf{h}^{-1}\hat{\mathbf{B}}_\mathbf{h}, \hat{\mathbf{A}}_\mathbf{h}^{-1}\right), \quad (3.1)$$

where $\hat{A}_h = \hat{\Sigma}^{-1} + \Omega^{-1} - \Omega^{-1}\left\{2\Omega^{-1} - \Omega^{-1}\left(\Omega^{-1} + \hat{\Sigma}_0^{-1}\right)^{-1}\Omega^{-1}\right\}^{-1}\Omega^{-1}$,

$\hat{B}_h = \hat{\Sigma}^{-1}\hat{\boldsymbol{\beta}} + \left[\Omega^{-1}\left\{2\Omega^{-1} - \Omega^{-1}\left(\Omega^{-1} + \hat{\Sigma}_0^{-1}\right)^{-1}\Omega^{-1}\right\}^{-1}\Omega^{-1}\left(\Omega^{-1} + \hat{\Sigma}_0^{-1}\right)^{-1}\hat{\Sigma}_0^{-1}\hat{\boldsymbol{\beta}}_0\right]$, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_0$ are the maximum likelihood estimates (MLEs), $\hat{\Sigma}^{-1}$ and $\hat{\Sigma}_0^{-1}$ are the Hessian matrices of $log\, L(\boldsymbol{\beta}|\mathbf{D})$ and $log\, L(\boldsymbol{\beta}|\mathbf{D_0})$ evaluated at the respective MLEs of $\boldsymbol{\beta}$, and $L(\boldsymbol{\beta}|D)$ and $L(\pi|D_0)$ are the likelihood functions under the GLMs. Similarly, under the power prior formulation, the posterior distribution $\pi(\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0})$ corresponding to the power prior $\pi(\boldsymbol{\beta}|\mathbf{D_0}, \mathbf{a_0})$ given in (2.12) with $\pi_0(\boldsymbol{\beta}) \propto 1$ can be approximated by

$$\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0} \sim \mathbf{N_p}\left(\hat{\mathbf{A}}_\mathbf{p}^{-1}\hat{\mathbf{B}}_\mathbf{p}, \hat{\mathbf{A}}_\mathbf{p}^{-1}\right), \quad (3.2)$$

where $\hat{A}_p=\hat{\Sigma}^{-1}+a_0\hat{\Sigma}_0^{-1}$ and $\hat{B}_p=\hat{\Sigma}^{-1}\hat{\boldsymbol{\beta}}+a_0\hat{\Sigma}_0^{-1}\hat{\boldsymbol{\beta}}_0$. We note that (3.1) and (3.2) are the exact posterior distributions in the special case of normal linear regression. Chen and Ibrahim [16] showed that the approximate posterior distributions in (3.1) and (3.2) match, i.e., $\hat{A}_h=\hat{A}_p$ and $\hat{B}_h=\hat{B}_p$ if and only if $a_0\left(I_p+2\Omega\hat{\Sigma}_0^{-1}\right)=I_p$, where I_p is the $p\times p$ identity matrix. Chen and Ibrahim [16] proposed a guide value for a_0 based on this analytical connection and discussed extensions to multiple historical datasets as well as the case in which Ω is unknown.

3.3. The Power Prior as a Conjugate Prior

For GLMs, it turns out that a special case of the power prior is a conjugate prior. That is, if we take $D_0 = (n, \boldsymbol{y}_0, X)$, where $\boldsymbol{y}_0$ is a prior elicitation of the response vector for the current data $D = (n,\boldsymbol{y},X)$ ($\boldsymbol{y}_0$ is not historical data here), then the power prior reduces to a conjugate prior. Prior elicitation of observable quantities has been examined in detail by [50, 51, 52, 53]. If $\pi_0(\boldsymbol{\beta}) \propto 1$, the conjugate prior is given by

$$\pi(\boldsymbol{\beta}|\mathbf{D_0},\mathbf{a_0}) \propto \prod_{i=1}^{n} exp\left\{a_0\left[\alpha_{0i}^{-1}\left\{y_{0i}h\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)-\psi\left(h\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)\right)\right\}+\phi(y_{0i})\right]\right\}, \quad (3.3)$$

where $a_0 > 0$ is a scalar prior parameter, and $\boldsymbol{y}_0 = (y_{01},..., y_{0n})'$ is an $n \times 1$ vector of hyperparameters. In (3.3), ($\boldsymbol{y}_0$, a_0) have different meanings and interpretations than that of the historical data case of the earlier subsections. Now $\boldsymbol{y}_0$ is a prior prediction for $\boldsymbol{y}$, and a_0 reflects the degree of confidence in that prediction. Thus, *a0* is no longer restricted to be between 0 and 1 in this conjugate prior situation, and we only need $a_0 \geq 0$. We denote the conjugate prior by $(\boldsymbol{\beta}|\mathbf{D_0},\mathbf{a_0}) \sim \mathscr{D}(\boldsymbol{y}_0, a_0)$. As shown in [53], the resulting posterior distribution takes of the form $(\boldsymbol{\beta}|\mathbf{D},\mathbf{D_0},\mathbf{a_0}) \sim \mathscr{D}\left(\frac{1}{a_0+1}\boldsymbol{y}+\frac{a_0}{a_0+1}\boldsymbol{y}_0, a_0+1\right)$.

We now briefly discuss the elicitation of $\boldsymbol{y}_0$ for the conjugate prior. There are two ways of eliciting $\boldsymbol{y}_0$. First, in the case of direct elicitation, one can use expert opinion or case-specific information on each subject. We can also elicit $\boldsymbol{y}_0$ from forecasts or predictions obtained from a theoretical prediction model. In this case, we could obtain a point prediction $\boldsymbol{y}_0$ based on a previous similar study. Second, in the case of indirect elicitation, we can specify $\boldsymbol{y}_0$ indirectly through a prior specification for the prior mode $\boldsymbol{\mu}_0$ of β. As shown in [53], $\boldsymbol{y}_0=\left(\dot{\psi}\left(h\left(\boldsymbol{x}_1'\boldsymbol{\mu}_0\right)\right),\ldots,\dot{\psi}\left(h\left(\boldsymbol{x}_n'\boldsymbol{\mu}_0\right)\right)\right)'$ yields a prior mode of $\boldsymbol{\beta}$ equal to $\boldsymbol{\mu}_0$, where $\dot{\psi}(h)=\frac{\partial\psi(h)}{\partial h}$. In the context of binary regression, $\boldsymbol{y}_0=\left(F\left(\boldsymbol{x}_1'\boldsymbol{\mu}_0\right),\ldots,F\left(\boldsymbol{x}_n'\boldsymbol{\mu}_0\right)\right)'$, where F is a cdf. When $\boldsymbol{\mu} = \mathbf{0}$ and F is the cdf corresponding to a symmetric distribution, then $\boldsymbol{y}_0 = (1/2,..., 1/2)'$. Many other interesting special cases for the specification of $\boldsymbol{y}_0$ as well as the elicitation of a_0 for GLMs can be found in [53].

3.4. Matching Predictives

Another attractive feature of the power prior examined in [54] is that it has the property of *matching predictives*. In variable selection or model selection problems, many authors have advocated the notion that the priors for the parameters should somehow "match" across the

models in the model space. For example, for two nested models, the prior specification for the parameters "in common" between these two models should be consistent in some sense.

Suppose we have two nested models $\mathscr{M}_1 \subset \mathscr{M}_2$ and take the priors $\pi(\boldsymbol{\theta}|\mathscr{M}_i)$ $i = 1, 2$, where the prior predictive densities $f(\boldsymbol{y}|\mathscr{M}_1)$ and $f(\boldsymbol{y}|\mathscr{M}_2)$ are given by $f(\boldsymbol{y}|\mathscr{M}_i) = \int f(\boldsymbol{y}|\boldsymbol{\theta}, \mathscr{M}_i)\, \pi(\boldsymbol{\theta}|\mathscr{M}_i)\, d\boldsymbol{\theta}$ for $i = 1,2$. For variable subset selection in linear models, it turns out that the power prior is the class of priors that minimizes the discrepancy between $f(\boldsymbol{y}|\mathscr{M}_1)$ and $f(\boldsymbol{y}|\mathscr{M}_2)$, when the discrepancy measure is the symmetric KL divergence, defined by

$$D_{SKL} = \int log\left(\frac{f(\boldsymbol{y}|\mathscr{M}_1)}{f(\boldsymbol{y}|\mathscr{M}_2)}\right) f(\boldsymbol{y}|\mathscr{M}_1)\, d\boldsymbol{y} + \int log\left(\frac{f(\boldsymbol{y}|\mathscr{M}_2)}{f(\boldsymbol{y}|\mathscr{M}_1)}\right) f(\boldsymbol{y}|\mathscr{M}_2)\, d\boldsymbol{y}. \quad (3.4)$$

To illustrate this, we first consider the linear model with no covariates, i.e., the intercept model, $\boldsymbol{y} = \boldsymbol{\mu} + \varepsilon$, where $\epsilon \sim \mathbf{N}\left(\mathbf{0}, \tau^{-1}\mathbf{I_n}\right)$. Take the prior for $\boldsymbol{\mu}$ as $\boldsymbol{\mu}|\tau, y_0 \sim \mathbf{N}\left(y_0, \tau^{-1}\mathbf{a_0^{-1}}\mathbf{I_n}\right)$. Then the prior predictive distribution of $\boldsymbol{y}$ is given by

$$\boldsymbol{y}|\tau, y_0 \sim N\left(y_0, \tau^{-1}\left(1 + a_0^{-1}\right) I_n\right). \quad (3.5)$$

Now consider the linear model with covariates given by $\boldsymbol{y} = \boldsymbol{X\beta} + \varepsilon$, where $\epsilon \sim \mathbf{N}\left(\mathbf{0}, \tau^{-1}\mathbf{I_n}\right)$. We take the prior for $\boldsymbol{\beta}$ as $\boldsymbol{\beta}|\tau, \boldsymbol{\beta_0}, \mathbf{\Sigma_0} \sim \mathbf{N}\left(\boldsymbol{\beta_0}, \tau^{-1}\mathbf{\Sigma_0}\right)$. Then we have

$$\boldsymbol{y}|\tau, X, \boldsymbol{\beta}_0, \Sigma_0 \sim N\left(\boldsymbol{X\beta}_0, \tau^{-1}\left(I_n + \boldsymbol{X}\Sigma_0\boldsymbol{X}'\right)\right). \quad (3.6)$$

As shown in [54], as a function of $(\boldsymbol{\beta}_0, \Sigma_0)$, D_{SKL} in (3.4) is minimized when $\boldsymbol{\beta}_0 = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}_0$ and $\Sigma_0 = a_0^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$. That is, the power prior is the prior that minimizes the discrepancy between the prior predictive distribution with no covariates (i.e., (3.5)) and the prior predictive distribution with covariates (i.e., (3.6)).

3.5. Variable Selection Problems

The power prior is "semi-automatic" in the sense that once one identifies the likelihood function of the historical data, then the kernel of the power prior is immediately determined with minimal prior elicitation. One only has to elicit a single scalar a_0. This type of prior elicitation scheme is very powerful in variable selection problems, since by the mere specification of likelihood function of the historical data, the hyperparameters of the power prior for all possible subset models in the model space are automatically determined. A detailed discussion of this semi-automated elicitation scheme based on the power prior for variable selection in linear models can be found in [51] and [55]. Chen, Ibrahim, and Yiannoutsis [12] developed the power prior for variable selection and computation in the GLM, Chen et al. [31] extended these results to generalized linear mixed models, Ibrahim and Chen [1] and Ibrahim, Chen, and MacEachern [13] developed the variable selection methodology and computation for power priors in the Cox regression model, and Ibrahim,

Chen, and Ryan [40] developed the power prior for variable selection and its computational properties for time series models.

For the GLM, Chen, Ibrahim, and Yiannoutsos [12] used the power prior to specify both a prior for the regression coefficients for all subset models in the model space and prior probabilities for all models in the model space. To do this, they let p denote the number of the regression coeffiients including an intercept for the full model and let $\mathscr{M}$ denote the model space. They enumerate the models in $\mathscr{M}$ by $k = 1,..., \mathscr{K}$, where $\mathscr{K}$ is the dimension of $\mathscr{M}$ and model $\mathscr{K}$ denotes the full model. Also, they let $\boldsymbol{\beta}^{(\mathscr{K})}=(\beta_0, \beta_1, \ldots, \beta_{p-1})'$ denote the regression coefficients for the full model including an intercept, and let $\boldsymbol{\beta}^{(k)}$ denote a $p_k \times 1$ vector of regression coefficients for model k with an intercept, and a specific choice of $p_k - 1$ covariates. They take $D_0^{(k)}=\left(n_0, \boldsymbol{y}_0, X_0^{(k)}\right)$ as the historical data for model k, where $X_0^{(k)}$ is an $n_0 \times p_k$ design matrix. Under model k, Chen, Ibrahim, and Yiannoutsos [12] proposed the following form of the power prior based on $D_0^{(k)}$ for $\boldsymbol{\beta}^{(k)}$:

$$\pi\left(\boldsymbol{\beta}^{(k)}|D_0^{(k)}, a_0\right) \propto \prod_{i=1}^{n_0} exp\left\{a_0\left[\alpha_{0i}^{-1}\left\{y_{0i}h\left(x_{0i}^{(k)'}\boldsymbol{\beta}^{(k)}\right) - \psi\left(h\left(x_{0i}^{(k)'}\boldsymbol{\beta}^{(k)}\right)\right)\right\} + \phi\left(y_{0i}\right)\right]\right\} \pi_0\left(\boldsymbol{\beta}^{(k)}|c_0\right), \quad (3.7)$$

where $\pi_0\left(\boldsymbol{\beta}^{(k)}|c_0\right)$ is the initial prior for $\boldsymbol{\beta}^{(k)}$, c_0 is a fixed hyperparameter, and a_0 is the discounting parameter. The parameter c_0 controls the impact of $\pi_0\left(\boldsymbol{\beta}^{(k)}|c_0\right)$ on the entire prior, and the p arameter a_0 controls the influence of the historical data on $\pi\left(\boldsymbol{\beta}^{(k)}|D_0^{(k)}, a_0\right)$. From (3.7), we see how the prior distribution of $\boldsymbol{\beta}^{(k)}$ is automatically determined from the historical data for all models in the model space. All one needs to do is just to specify the historical data $D_0^{(k)}=\left(n_0, \boldsymbol{y}_0, X_0^{(k)}\right)$ and elicit the discounting parameter a_0 and the hyperparameter c_0.

To specify prior probabilities for all models on the model space $\mathscr{M}$ using the historical data D_0. Chen, Ibrahim, and Yiannoutsos [12] first specified the prior for $\boldsymbol{\beta}^{(k)}$ as

$$\pi_0^*\left(\boldsymbol{\beta}^{(k)}|D_0^{(k)}, d_0\right) = \prod_{i=1}^{n_0} exp\left[\alpha_{0i}^{-1}\left\{y_{0i}h\left(x_{0i}^{(k)'}\boldsymbol{\beta}^{(k)}\right) - \psi\left(h\left(x_{0i}^{(k)'}\boldsymbol{\beta}^{(k)}\right)\right)\right\} + \phi\left(y_{0i}\right)\right] \pi_0\left(\boldsymbol{\beta}^{(k)}|d_0\right), \quad (3.8)$$

where $\pi_0\left(\boldsymbol{\beta}^{(k)}|d_0\right)$ is the same density as that in (3.7) with c_0 replaced by d_0, and then defined the prior probability for mode k as

$$\pi\left(k|D_0, d_0\right) = \frac{\int \pi_0^*\left(\boldsymbol{\beta}^{(k)}|D_0^{(k)}, d_0\right) d\boldsymbol{\beta}^{(k)}}{\sum_{k\in\mathscr{M}} \int \pi_0^*\left(\boldsymbol{\beta}^{(k)}|D_0^{(k)}, d_0\right) d\boldsymbol{\beta}^{(k)}}. \quad (3.9)$$

In (3.8), the parameter d_0 is a scalar prior parameter that controls the impact of $\pi_0\left(\boldsymbol{\beta}^{(k)}|d_0\right)$ on the prior model probability $\pi(k|D_0,d_0)$ in (3.9). The prior model probability $\pi(k|D_0,d_0)$

defined by (3.8) and (3.9) has several nice properties. First, $\pi(k|D_0,d_0)$ in (3.9) corresponds to the posterior probability of model k based on the historical data D_0 using a uniform initial prior $\pi_0(k) = 2^{-p}$ for $k \in \mathscr{M}$. Second, as $d_0 \to 0$, then $\pi(k|D_0,d_0)$ reduces to a uniform prior on the model space. Therefore, as $d_0 \to 0$, the historical data D_0 have minimal impact in determining $\pi(k|D_0,d_0)$. On the other hand, with a large value of d_0, $\pi_0\left(\boldsymbol{\beta}^{(k)}|d_0\right)$ plays a minimal role in determining $\pi(k|D_0, d_0)$, and in this case, the historical data plays a larger role in determining $\pi(k|D_0, d_0)$. Thus as $d_0 \to \infty$, $\pi(k|D_0, d_0)$ will be regulated by the historical data. The parameter d_0 plays the same role as c_0 and thus serves as a tuning parameter to control the impact of D_0 on the prior model probability $\pi(k|D_0, d_0)$.

For the GLM, the prior $\pi\left(\boldsymbol{\beta}^{(k)}|D_0^{(k)}, a_0\right)$ and the prior model probability $\pi(k|D_0, d_0)$ given in (3.7) and (3.9) lead to convenient and efficient computation of the prior and posterior model probabilities. The computational algorithms developed in [12] only require two Gibbs samples, one from the prior and another from the posterior under the full model, to compute the prior and posterior model probabilities for all possible models in $\mathscr{M}$.

4. The Role of Power Priors in Posterior Inference

In Subsections 4.1 and 4.2, we consider new developments. In particular, we wish to theoretically examine the behavior of the posterior variance of $\boldsymbol{\beta}$ and the marginal variance of the posterior mean of $\boldsymbol{\beta}$ as the discounting parameter a_0 is varied between 0 and 1. Studying these properties is important since it shows how the marginal variance is reduced or maximized as a function of the discounting parameter a_0. We show that for the linear model and generalized linear model that the marginal variance of the posterior mean of $\boldsymbol{\beta}$ is always less than or equal to the posterior variance of $\boldsymbol{\beta}$ and that equality is only attained when $a_0 = 0$ and $a_0 = 1$ and the maximum discrepancy between the two variances is attained at $a_0 = 0.5$.

4.1. The Normal Case

Assume that we take an improper uniform initial prior for β, i.e., $\pi_0(\boldsymbol{\beta}) \propto 1$. Suppose that $L(\boldsymbol{\beta}|D)$ and $L(\boldsymbol{\beta}|D_0)$ are the likelihood functions in (2.10) and (2.11). Then, the posterior distribution of $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0} \sim \mathbf{N}\left(\left[\mathbf{X}'\mathbf{X}+\mathbf{a_0}\mathbf{X}_0'\mathbf{X_0}\right]^{-1}\left[\mathbf{X}'\boldsymbol{y}+\mathbf{a_0}\mathbf{X}_0'\boldsymbol{y_0}\right], \sigma^2\left[\mathbf{X}'\mathbf{X}+\mathbf{a_0}\mathbf{X}_0'\mathbf{X_0}\right]^{-1}\right), \quad (4.1)$$

for $0 \leq a_0 \leq 1$. From (4.1), we see that the posterior mean and variance of $\boldsymbol{\beta}$ are given by $\bar{\boldsymbol{\beta}} = E[\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0}] = \left[X'X+a_0X_0'X_0\right]^{-1}\left[X'\boldsymbol{y}+a_0X_0'\boldsymbol{y}_0\right]$ and

$$Var(\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0}) = \sigma^2\left[X'X+a_0X_0'X_0\right]^{-1}. \quad (4.2)$$

Theorem 4.1 *Let Var($\bar{\boldsymbol{\beta}}$) denote the varianee of $\bar{\boldsymbol{\beta}}$ with respect to the marginal distribution of* ($\mathbf{y}$, $\boldsymbol{y}_0$) defined by (2.10) and (2.11). Assume that X and X_0 are of full rank. Then, we have

$$Var\left(\bar{\boldsymbol{\beta}}\right) \leq Var\left(\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0}\right) \quad (4.3)$$

for $0 \leq a_0 \leq 1$, where $Var\left(\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0}\right)$ is the posterior variance of $\boldsymbol{\beta}$ and " $\leq$" denotes that $Var\left(\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0}\right) - Var\left(\bar{\boldsymbol{\beta}}\right)$ is a positive semi-definite definite matrix. In addition, the equality in (4.3) holds if and only if $a_0 = 0$ *or* $a_0 = 1$ *and the maximum difference between* $Var(\boldsymbol{\beta}|D, D_0, a_0)$ *and* $Var\left(\bar{\boldsymbol{\beta}}\right)$ *is reached at* $a_0 = 0.5$.

The proof of Theorem 4.1 is given in Appendix A.

REMARK In view of the frequentist properties of the posterior estimates of $\boldsymbol{\beta}$, the results established in Theorem 4.1 imply that for $1 \leq j \leq p$, (i) the $100(1 - a)\%$ HPD interval of β_j has exact coverage probability of $1 - a$ when $a_0 = 0$ or $a_0 = 1$; (ii) the coverage probability of the $100(1 - a)\%$ HPD interval of β_j is greater than $1 - a$ when $0 < a_0 < 1$; (iii) the highest coverage probability of the $100(1 - a)\%$ HPD interval of β_j is attained at $a_0 = 0.5$.

4.2. The General Case

Using the GLMs, the posterior distribution of $\boldsymbol{\beta}$ is given by $\pi\left(\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0}\right) \propto L\left(\boldsymbol{\beta}|\mathbf{D}\right) L(\boldsymbol{\beta}|\mathbf{D_0})^{a_0} \pi_0\left(\boldsymbol{\beta}\right)$. Following [49, 16] and ignoring constants that are free of the parameters, we have $L\left(\boldsymbol{\beta}|\mathbf{D}\right) \approx exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)' \hat{\Sigma}^{-1}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right\}$ and $L\left(\boldsymbol{\beta}|\mathbf{D_0}\right) \approx exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_\mathbf{0}\right)' \hat{\Sigma}_0^{-1}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_\mathbf{0}\right)\right\}$ where $L(\boldsymbol{\beta}|D)$ and $L(\boldsymbol{\beta}|D_0)$ are the likelihood functions,

$\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_0$ are the respective MLEs of $\boldsymbol{\beta}$ based on D and D_0 under the GLMs, and $\hat{\Sigma}^{-1}$ and $\hat{\Sigma}_0^{-1}$ are the Hessian matrices of $\log L(\boldsymbol{\beta}|D)$ and $\log L(\boldsymbol{\beta}|D_0)$ evaluated at the respective MLEs of $\boldsymbol{\beta}$. Then, it is straightforward to show that under the GLMs, $\hat{\Sigma}^{-1} = X' \hat{\Delta}^2 \hat{V} X$ and $\hat{\Sigma}_0^{-1} = X_0' \hat{\Delta}_0^2 \hat{V}_0 X_0$, where $\hat{\Delta}$ and $\hat{V}$ are $n \times n$ diagonal matrices with i^{th} diagonal elements $\delta_i = \delta_i\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) = dh_i/d\eta_i$ and $v_i = v_i\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) = \alpha_i^{-1} d^2\psi\left(h_i\right)/dh_i^2$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, where $h_i = h(\eta_i)$ and $\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}$, and $\hat{\Delta}_0$ and $\hat{V}_0$ are $n_0 \times n_0$ diagonal matrices with i^{th} diagonal elements $\delta_{0i} = \delta_{0i}\left(\boldsymbol{x}_{0i}'\boldsymbol{\beta}\right) = dh_{0i}/d\eta_{0i}$ and $v_{0i} = v_{0i}\left(\boldsymbol{x}_{0i}'\boldsymbol{\beta}_0\right) = \alpha_{0i}^{-1} d^2\psi\left(h_{0i}\right)/dh_{0i}^2$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_0$, where $h_{0i} = h\left(\eta_{0i}\right)$ and $\eta_{0i} = \boldsymbol{x}_{0i}'\boldsymbol{\beta}$. The above approximations are valid for large n and large n_0, respectively. Assuming $\pi_0\left(\boldsymbol{\beta}\right) \propto 1$, we obtain

$\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0} \overset{\text{approx.}}{\sim} \mathbf{N}\left(\bar{\boldsymbol{\beta}}, \left[\mathbf{X}'\hat{\boldsymbol{\Delta}}^{\mathbf{2}}\hat{\mathbf{V}}\mathbf{X} + \mathbf{a_0}\mathbf{X_0'}\hat{\boldsymbol{\Delta}}_{\mathbf{0}}^{\mathbf{2}}\hat{\mathbf{V}}_{\mathbf{0}}\mathbf{X_0}\right]^{-1}\right)$, where

$\bar{\boldsymbol{\beta}} = \left[X'\hat{\Delta}^2\hat{V}X + a_0X_0'\hat{\Delta}_0^2\hat{V}_0X_0\right]^{-1}\left[X'\hat{\Delta}^2\hat{V}X\hat{\boldsymbol{\beta}} + a_0X_0'\hat{\Delta}_0^2\hat{V}_0X_0\hat{\boldsymbol{\beta}}_0\right]$. Again, using the asymptotic variances of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_0$, it can be shown that the sample variance of the posterior mean $\bar{\boldsymbol{\beta}}$ is given by

$$Var\left(\bar{\boldsymbol{\beta}}\right) \approx \left[X'\hat{\Delta}^2\hat{V}X+a_0X_0'\hat{\Delta}_0^2\hat{V}_0X_0\right]^{-1}\left[X'\hat{\Delta}^2\hat{V}X+a_0^2X_0'\hat{\Delta}_0^2\hat{V}_0X_0\right]\left[X'\hat{\Delta}^2\hat{V}X+a_0X_0'\hat{\Delta}_0^2\hat{V}_0X_0\right]^{-1}$$

and the posterior variance of $\boldsymbol{\beta}$ is approximated by

$Var\left(\boldsymbol{\beta}|\mathbf{D},\mathbf{D_0},\mathbf{a_0}\right) \approx \left[X'\hat{\Delta}^2\hat{V}X+a_0X_0'\hat{\Delta}_0^2\hat{V}_0X_0\right]^{-1}$. Here, $Var\left(\bar{\boldsymbol{\beta}}\right)$ is taken with respect to the marginal distribution of ($\boldsymbol{y}$, $\boldsymbol{y}_0$). Thus, we are led to the following corollary.

Corollary 4.1 Theorem 4.1 holds approximately for the GLMs when n and n_0 *are large and* X *and* X_0 *are of full rank*.

5. The Choice of a_0

One of the most important issues in the use of the power prior is what value of a_0 to use in the analysis. There are several possible solutions to this issue. The easiest solution is to establish a hierarchical power prior by specifying a proper prior distribution for a_0, such as a beta prior, for example. A uniform prior on a_0, might be a good choice or a more informative prior would be to take a_0 ~ beta(c, c), where c is moderate to large, such as $c \geq 3$. Although a prior for a_0 is attractive, it is more computationally intensive than the a_0 fixed case and all closed forms are lost when taking this approach. The a_0 random case has been discussed in [1, 10, 12, 13, 15, 19, 22, 23, 30, 45, 46, 47, ?, 56, 57, ?].

Another approach is to take a_0 as fixed and elicit a specific value for it and conduct several sensitivity analyses about this value, or to take a_0 fixed and use a model selection criterion. To facilitate the choice of a_0, for the normal linear model, we derive here expressions for the penalized likelihood-type criterion, marginal likelihood criterion, deviance information criterion, and logarithm of the pseudo-marginal likelihood criterion for the linear model as well as present a new simulation study in Section 5.6. As discussed in [15], the guide values based on the criteria discussed below serve only as a starting point for the analysis, and several sensitivity analyses should be carried out in the range of the guide values.

5.1. The Penalized Likelihood-type Criterion

Ibrahim, Chen, and Sinha [15] proposed a penalized likelihood-type criterion (PLC) to determine a guide value of a_0. This criterion takes of the form

$G\left(a_0\right) = -2\,log\left[m^*\left(a_0\right)\right] + \frac{log(n_0)}{a_0}$, where

$$m^*\left(a_0\right) = \int L\left(\boldsymbol{\beta}|\mathbf{D}\right) L(\boldsymbol{\beta}|\mathbf{D_0})^{a_0}\pi_0\left(\boldsymbol{\beta}\right) d\boldsymbol{\beta}. \quad (5.1)$$

$L(\boldsymbol{\beta}|D)$ and $L(\boldsymbol{\beta}|D_0)$ are the likelihood functions under the GLMs, and $\pi_0(\boldsymbol{\beta})$ is the initial prior. Then, the guide value of a_0 based on the PLC is given by

$$a_{0,PLC}^{opt} = \underset{0<a_0<1}{arg\,min} G\left(a_0\right). \quad (5.2)$$

For the normal linear model, when $\pi_0\left(\boldsymbol{\beta}\right) \propto 1$, (5.1) reduces to

$$m^*\left(a_0\right) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{n-a_0n_0-p}{2}}} \times \frac{1}{\left|X'X+a_0X_0'X_0\right|^{\frac{1}{2}}} exp\left[-\frac{1}{2\sigma^2}\left\{\boldsymbol{y}'\boldsymbol{y}+a_0\boldsymbol{y}_0'\boldsymbol{y}_0-\left(X'y+a_0X_0'\boldsymbol{y}_0\right)'\left(X'X+a_0X_0'X_0\right)^{-1}\left(\right.\right.\right.$$

and the PLC in takes the form

$$\begin{aligned} G(a_0) = & (n + a_0 n_0 - p) \log\left(2\pi\sigma^2\right) \\ & + \log\left(\left|X'X + a_0 X_0' X_0\right|\right) \\ & + \frac{1}{\sigma^2}\left\{\boldsymbol{y}'\boldsymbol{y} + a_0 \boldsymbol{y}_0' \boldsymbol{y}_0 - \left(X'\boldsymbol{y} + a_0 X_0' \boldsymbol{y}_0\right)' \left(X'X + a_0 X_0' X_0\right)^{-1} \left(X'\boldsymbol{y} + a_0 X_0' \boldsymbol{y}_0\right)\right\} \\ & + \frac{\log(n_0)}{a_0} \end{aligned}.$$

5.2. The Marginal Likelihood Criterion

We take the power prior distribution of $\boldsymbol{\beta}$ given D_0 and a_0 in (2.13) with an initial prior $\pi_0(\boldsymbol{\beta})$. Then, the marginal likelihood is defined as

$$m(a_0) = \frac{\int L(\boldsymbol{\beta}|\mathbf{D}) L(\boldsymbol{\beta}|\mathbf{D_0})^{a_0} \pi_0(\boldsymbol{\beta}) d\boldsymbol{\beta}}{\int L(\boldsymbol{\beta}|\mathbf{D_0})^{a_0} \pi_0(\boldsymbol{\beta}) d\boldsymbol{\beta}} = \frac{m^*(a_0)}{\int L(\boldsymbol{\beta}|\mathbf{D_0})^{a_0} \pi_0(\boldsymbol{\beta}) d\boldsymbol{\beta}}. \quad (5.3)$$

The guide value of a_0 according to the marginal likelihood criterion is given by

$$a_{0,MLC}^{opt} = \underset{0<a_0<1}{arg\,min} \left[-2 \log\left\{m(a_0)\right\}\right]. \quad (5.4)$$

Note that when $a_0 > 0$ and X_0 is of full rank, $\pi(\boldsymbol{\beta}|\mathbf{D_0}, \mathbf{a_0})$ in (2.13) is still proper. For the normal linear model, when $\pi_0(\boldsymbol{\beta}) \propto 1$ and $a_0 > 0$, we have

$$\begin{aligned} & -2 \log\left\{m(a_0)\right\} \\ & = n \log\left(2\pi\sigma^2\right) \\ & + \log\left(\left|X'X + a_0 X_0' X_0\right|\right) \\ & - \log\left(\left|X_0' X_0\right|\right) \\ & - p \log(a_0) \\ & + \frac{1}{\sigma^2} \boldsymbol{y}'\boldsymbol{y} \\ & + \frac{1}{\sigma^2}\left\{a_0 \boldsymbol{y}_0' X_0 \left(X_0' X_0\right)^{-1} X_0' \boldsymbol{y}_0 - \left(X'\boldsymbol{y} + a_0 X_0' \boldsymbol{y}_0\right)' \left(X'X + a_0 X_0' X_0\right)^{-1} \left(X'\boldsymbol{y} + a_0 X_0' \boldsymbol{y}_0\right)\right\} \end{aligned}$$

.

5.3. The Deviance Information Criterion

For the GLMs, the deviance function is defined as

$$Dev(\boldsymbol{\beta}) = -2\sum_{i=1}^{n} \log f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}) = -2\sum_{i=1}^{n} \left\{\alpha_i^{-1}\left(y_i h\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) - \psi\left(h\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)\right) + \phi(y_i)\right\}.$$

The Deviance Information Criterion (DIC) (Spiegelhalter et al. [58]) is given by

$$DIC(a_0) = Dev\left(\bar{\boldsymbol{\beta}}\right) + 2p_D(a_0), \quad (5.5)$$

where $\bar{\boldsymbol{\beta}} = E\left[\boldsymbol{\beta}|\mathbf{D}, \mathbf{D_0}, \mathbf{a_0}\right]$ and $p_D(a_0) = E\left[Dev(\boldsymbol{\beta})|D, D_0, a_0\right] - Dev\left(\bar{\boldsymbol{\beta}}\right)$. Using (5.5), the optimal value of a_0 according to DIC is given by

$$a_{0,DIC}^{opt} = \underset{0<a_0<1}{arg\,min} DIC\,(a_0)\,. \quad (5.6)$$

Now, we present a closed form expression of DIC for the normal linear model whose detailed derivation is given in Appendix A. In this special case, the deviance function reduces to

$$Dev\,(\boldsymbol{\beta}) = n\,log\left(2\pi\sigma^2\right) + \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i^{'}\boldsymbol{\beta}\right)^2 = n\,log\left(2\pi\sigma^2\right) + \frac{1}{\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta})^{'}(\boldsymbol{y} - X\boldsymbol{\beta}).$$

The DIC for the normal linear model is given by

$$\begin{aligned} DIC\,(a_0) =& n\,log\left(2\pi\sigma^2\right) \\ &+\frac{1}{\sigma^2}\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right)^{'}\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\right) \\ &+2tr\left(X^{'}X\left[X^{'}X + a_0X_0^{'}X_0\right]^{-1}\right) \\ &+\frac{a_0^2}{\sigma^2}\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_0\right)^{'}\left(X_0^{'}X_0\right)\left[X^{'}X + a_0X_0^{'}X_0\right]^{-1}X^{'}X\left[X^{'}X + a_0X_0^{'}X_0\right]^{-1}\left(X_0^{'}X_0\right)\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_0\right). \end{aligned} \quad (5.7)$$

It is interesting to see that when $a_0 = 0$, we have $p_D(a_0 = 0) = \text{tr}(X'X[X'X]^{-1}) = p$, which is exactly the same as the dimension of $\boldsymbol{\beta}$. In addition, $p_D(a_0)$ decreases when a_0 increases.

A Special Case: When $p = 1$, $X = (1,1,..., 1)'$, and $X_0 = (1,1,..., 1)'$, which corresponds to an intercept model, we have $\hat{\boldsymbol{\beta}} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \hat{\boldsymbol{\beta}}_0 = \bar{y}_0 = \frac{1}{n_0}\sum_{i=1}^{n_0} y_{0i}$ and

$$\begin{aligned} DIC\,(a_0) =& n\,log\left(2\pi\sigma^2\right) \\ &+\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \\ &+2\frac{n}{n + a_0n_0} + \frac{a_0^2}{\sigma^2}\frac{nn_0^2}{(n + a_0n_0)^2}\left(\bar{y} - \bar{y}_0\right)^2 \\ =& n\,log\left(2\pi\sigma^2\right) \\ &+\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \\ &+2b_0 + (1 - b_0)^2\frac{\left(\bar{y} - \bar{y}_0\right)^2}{\sigma^2/n} \end{aligned}$$

, where $b_0 = \frac{n}{n + a_0n_0}$ It is easy to show that $\frac{\partial^2 DIC(a_0)}{\partial b_0^2} \geq 0$. Therefore, we obtain $a_{0,DIC}^{opt} = 1$ if $|\bar{y} - \bar{y}_0| \leq \frac{\sigma}{\sqrt{n}}$ and

$$min\left\{\frac{n}{n_0}\left[\frac{1}{1 - \frac{\left(\sigma^2/n\right)}{\left(\bar{y} - \bar{y}_0\right)^2}} - 1\right], 1\right\}$$

if $|\bar{y} - \bar{y}_0| > \frac{\sigma}{\sqrt{n}}$. This result is quite interesting since if $|\bar{y} - \bar{y}_0| > \frac{\sigma}{\sqrt{n}}$ then $\sigma_{0,DIC}^{opt}$ decreases when n_0 increases.

5.4. The Logarithm of the Pseudo-Marginal Likelihood Criterion

Let $D_{(-i)}$ denote D with the i^{th} observation deleted. Then, for the i^{th} observation, the Conditional Predictive Ordinate (CPO) is defined as

$CPO_i\left(a_0\right)=f\left(y_i|\boldsymbol{x}_i,D_{(-i)}\right)=\int f\left(y_i|\boldsymbol{x}_i,\boldsymbol{\beta}\right)\pi\left(\boldsymbol{\beta}|\mathbf{D}_{(-\mathbf{i})},\mathbf{D_0},\mathbf{a_0}\right)d\boldsymbol{\beta}$, where

$\pi\left(\boldsymbol{\beta}|\mathbf{D}_{(-\mathbf{i})},\mathbf{D_0},\mathbf{a_0}\right)$ denotes the posterior density of $\boldsymbol{\beta}$ given $D_{(-i)}$, D_0, and a_0. Following Geisser [59] and Gelfand, Dey, and Chang [60], we have

$$CPO_i\left(a_0\right)=\left\{\int\frac{1}{f\left(y_i|\boldsymbol{x}_i,\boldsymbol{\beta}\right)}\pi\left(\boldsymbol{\beta}|\mathbf{D},\mathbf{D_0},\mathbf{a_0}\right)d\boldsymbol{\beta}\right\}^{-1}$$
$$=\left[\int exp\left\{-\alpha_i^{-1}\left(y_ih\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)-\psi\left(h\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)\right)-\psi\left(y_i\right)\right\}\pi\left(\boldsymbol{\beta}|\mathbf{D},\mathbf{D_0},\mathbf{a_0}\right)d\boldsymbol{\beta}\right]^{-1},\text{ where}$$

$\pi\left(\boldsymbol{\beta}|\mathbf{D},\mathbf{D_0},\mathbf{a_0}\right)$ is the posterior distribution under the GLM. Then, the logarithm of the Pseudo-marginal likelihood (LPML) in [61] is defined as

$$LPML\left(a_0\right)=\sum_{i=1}^{n}log\left\{CPO_i\left(a_0\right)\right\}. \quad (5.8)$$

Using (5.8), the optimal value of a_0 according to LPML is given by

$$a_{0,LPML}^{opt}=\underset{0<a_0<1}{arg,max}LPML\left(a_0\right). \quad (5.9)$$

For the normal linear model, the CPO reduces to

$CPO_i\left(a_0\right)=\left\{\int\left(2\pi\sigma^2\right)^{1/2}exp\left\{\frac{1}{2\sigma^2}\left(y_i-\boldsymbol{x}_i'\boldsymbol{\beta}\right)^2\right\}\times\pi\left(\boldsymbol{\beta}|\mathbf{D},\mathbf{D_0},\mathbf{a_0}\right)d\boldsymbol{\beta}\right\}^{-1}$. After some algdbra, we obtain the LPML for the normal linear mo el as

$$LPML\left(a_0\right)=-\frac{n}{2}log\left(2\pi\sigma^2\right)$$
$$+\frac{1}{2}\sum_{i=1}^{n}log\left\{1-h_{ii}\left(a_0\right)\right\}$$
$$+\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[-\left(y_i-\hat{y}_i\left(a_0\right)\right)^2+\hat{y}_i^2\left(a_0\right)-h_{ii}\left(a_0\right)y_i^2+\frac{\left\{\hat{y}_i\left(a_0\right)-h_{ii}\left(a_0\right)y_i\right\}^2}{1-h_{ii}\left(a_0\right)}\right]$$

Note that when $a_0 = 0$, using a first-order Taylor's series expansion, we have $\sum_{i=1}^{n}log\delta\left\{1-h_{ii}\left(a_0\right)\right\}\approx-\sum_{i=1}^{n}h_{ii}\left(0\right)=-p$. For a detailed derivation of the LPML for the normal linear model, see Appendix A.

5.5. Multiple Historical Datasets Case

For K_0 historical datasets, using the notation in Section 2.4 and (2.9), we first extend (5.1) to

$$m^*\left(\boldsymbol{a}_0\right)=\int L\left(\boldsymbol{\beta}|\mathbf{D}\right)\left[\prod_{k=1}^{K_0}L(\boldsymbol{\beta}|\mathbf{D_{0k}})^{a_{0k}}\right]\pi_0\left(\boldsymbol{\beta}\right)d\boldsymbol{\beta}. \quad (5.10)$$

and then define

$$G\left(\boldsymbol{a}_0\right)=-2\,log\left[m^*\left(\boldsymbol{a}_0\right)\right]+\sum_{k=1}^{K_0}\frac{log\left(n_{0k}\right)}{a_{0k}}. \quad (5.11)$$

Thus, the guide value of a_0 based on PLC in (5.11) is given by

$$\boldsymbol{a}_{0,PLC}^{opt} = \underset{\mathbf{0}<\boldsymbol{a_0}<\mathbf{1}}{arg\,min} G(\boldsymbol{a}_0), \quad (5.12)$$

where **0** (**1**) is a K_0-dimensional vector with all elements equal to 0 (1). Similarly, we can extend (5.4), (5.6), and (5.9) to obtain the guide values of a_0 for multiple historical datasets. For brevity, the details are omitted here.

5.6. A Simulation Study

We carry out a simulation study to examine the empirical performance of the power prior with fixed a_0 and random a_0 for the normal linear regression model. The model for the current data is $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ and $\epsilon_i \sim N(0,1)$ for $i = 1,\ldots, n$, where the ε_i's are independent, and the model for the historical data is $y_{i0} = \beta_{00} + \beta_{10} x_{i0} + \epsilon_{i0}$ and $\epsilon_{i0} \sim N(0,1)$ for $i = 1,\ldots, n_0$, where the ε_{i0}'s are independent. We further assume that the ε_i's and the ε_{i0}'s are independent. In all simulated datasets, we assume that the x_i's and x_{i0} are independently generated from a $N(0,1)$ distribution. We consider two scenarios: (i) the historical and current data are similar and (ii) the historical and current data are different. In Scenario I, we set $\beta_0 = \beta_{00} = 1$, $\beta_1 = \beta_{10} = 2$ $n = 400$, and $n_0 = 200$; and in Scenario II, we set $\beta_0 = \beta_{00} = 1$, $\beta_1 = 2$, $\beta_{10} = 1.75$, $n = 400$, and $n_0 = 200$. We also consider two additional scenarios: Scenario III, we set $\beta_0 = \beta_{00} = 1$, $\beta_1 = \beta_{10} = 2$ $n = 200$, and $n_0 = 400$; and Scenario IV, we set $\beta_0 = \beta_{00} = 1$, $\beta_1 = 2$, $\beta_{10} = 1.75$ $n = 400$, and $n_0 = 200$. The simulation results are given in Appendix B.

We generated 10,000 simulated datasets under each scenario. For each simulated dataset, we computed the posterior means, the posterior standard deviations, and the 95% HPD intervals of β_1 using the power prior (2.1) with 21 fixed a_0 values ranging from 0 to 1 with an increment of 0.05, and four estimated optimal a_0 values, namely, $a_{0,PLC}^{opt}$, $a_{0,MLC}^{opt}$, $a_{0,DIC}^{opt}$, and $a_{0,LPML}^{opt}$ given by (5.2), (5.4), (5.6), and (5.9), respectively, as well as using the normalized power prior (2.4). In all cases, an improper initial prior, $\pi_0(\beta_0, \beta_1) \propto 1$, was specified, and a uniform prior on (0,1) was taken for a_0 for the normalized power prior. Based on the 10,000 simulated datasets, we then calculated the average of the posterior means (Estimate), the average of the posterior standard deviations (SD), the coverage probability (CP) of the 95% HPD intervals, and the root of the mean square error (rMSE) for β_1.

The simulation results are shown in Table 1. The coverage probabilities and rMSE's are also plotted in Figure 1. From Table 1 and Figure 1, we see that under Scenario I, the highest CP is 0.9606, which is achieved at $a_0 = 0.5$. This empirical result is consistent with the theoretical result established in Theorem 4.1. The posterior estimates based on $a_{0,PLC}^{opt}$ were very similar to those under $a_0 = 0.10$ under both scenarios. The guide values $a_{0,MLC}^{opt}$, $a_{0,DIC}^{opt}$ and $a_{0,LPML}^{opt}$ led to similar posterior estimates under both scenarios. Compared to the SD's and rMSE's for fixed a_0 values, we see that $a_{0,DIC}^{opt}$, $a_{0,LPML}^{opt}$, and the random a_0 were equivalent to approximately borrowing 50% of the historical data under Scenario I and about 30%-40% of the historical data under Scenario II. In Scenario II, the random a_0 power had

the largest rMSE; the power prior with $a_{0,PLC}^{opt}$ had the smallest rMSE; and the power prior with guide values $a_{0,DIC}^{opt}$, and $a_{0,LPML}^{opt}$, as well as a random a_0 was over-borrowing the historical data, resulting in CPs around 90%, which were lower than 95% as expected. In general, the guide value $a_{0,PLC}^{opt}$ leads to less borrowing while the other three guide values and a random a_0 yield more borrowing.

6. Applications of the Power Prior

The power prior has been recently applied in many fields such as clinical trials, epidemiological studies, environmental health, genetics, health services research, etc.. In this section, we provide a snapshot overview on the use and implications of the power prior in various biomedical applications. The non-medical applications of the power prior are given in Appendix C.

6.1. Heritability Estimates in Human Genetics Research

In human genetics research, twin studies are often used as an initial process for testing a specific trait that is genetically influenced. One of the goals of these studies is to estimate the heritability in twin data. The heritability, denoted by h^2, is defined as twice the difference of the intraclass correlation coefficients of monozygous (MZ) and dizygous (DZ) twins. The twin data consist of n_1 MZ pairs of response variables, $\boldsymbol{y}_{1j}=(y_{1j1}, y_{ij2})'$ for j = 1,..., n_1; and n_2 DZ pairs of response variables, $\boldsymbol{y}_{2j}=(y_{2j1}, y_{2j2})'$ for j = 1,..., n_2. The assumed model is $\boldsymbol{y}_{ij} \sim N(X_{ij}\boldsymbol{\beta}, \boldsymbol{\Sigma_i})$, where $\Sigma_i=\sigma^2 \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}$ and X_{ij} is a matrix of covariates for i = 1, 2. The heritability of the trait is simply $h^2 = 2(\rho_1 - \rho_2)$. The published literature on heritability studies typically report sample sizes and intraclass correlations of MZ and DZ only. Therefore, the data from the k_{th} historical study can be written as $D_{0k}=(n_{01k}, n_{02k}, r_{01k}, r_{02k})$, where r_{01k} and r_{02k} are intraclass correlations of MZ and DZ, for k = 1,...,K_0. Write D_0 = (D_{01},..., D_{0K0}).

Often twin studies require great effort and much expense. Thus, the data are comprised of a small number of subjects. In typical analyses of human twin data in the literature, the standard errors of the estimates of the intraclass correlations are based on large sample theory, which may not be appropriate for small samples. To overcome these issues, Chen, Manatunga, and Williams [8] developed a new Bayesian scheme for analyzing human twin data. Specifically, they considered three types of prior distributions based on the complete data from previous studies (fully informative prior), summary statistics from the historical studies (semi-informative prior), and no historical information (non-informative prior), respectively. Their proposed semi-informative power prior based on D_0 takes the form

$$\pi\left(\boldsymbol{\beta},\sigma^2,\rho_1,\rho_2,\boldsymbol{a_0}|\mathbf{D_0}\right)\propto\prod_{k=1}^{K_0}\prod_{i=1}^{2}exp\left\{-\frac{a_{0k}}{2}\left(log\left[\left(n_{0ik}-\frac{3}{2}\right)^{-1}2\pi\left(1-\rho_i^2\right)^2\right]+\left(n_{0ik}-\frac{3}{2}\right)\left(z\left(r_{0ik}\right)-\eta_i\right)^2\right)\right\}$$
$$\times\frac{1}{\sigma^2}\prod_{k=1}^{K_0}a_{0k}^{\alpha_0}\,exp\left(-\lambda_0 a_{0k}\right), \tag{6.1}$$

where $\eta_i = z(\rho_i)$ and $z(r_{0ik})$, $k = 1,..., K_0$ are the Fisher's z transformations of ρ_i and r_{0ik}, respectively for $i = 1, 2$, $\boldsymbol{a}_0 = (a_{01}, \ldots, a_0K_0)$ and (α_0, λ_0) are pre-determined hyperparameters. From (6.1), we see that (i) the historica data are borrowed only through the ρ_i and (ii) no historical data are available for β and σ^2. In this sense, (6.1) can be viewed as a *partially borrowing power prior*. Using a simulation study, Chen, Manatunga, and Williams [8] empirically showed that the semi-informative prior is as informative as the fully informative prior if the purpose of the study is to estimate the intraclass correlations or heritability h^2 in twin studies.

6.2. Evaluating Water Quality in Environmental Science

In environmental statistics, one important issue is the evaluation of air or water quality standards. One objective of such studies is to estimate the water quality standards in water quality data. The data consist of n measurements of water quality (response variable), $\boldsymbol{y} = (y_1,..., y_n)$, and the assumed model is $y_i \sim N\left(\mu,\sigma^2\right)$ for $i = 1,..., n$. The historical data $D_0 = (n_0, \boldsymbol{y}_0)$ were collected in previous years and published by the US Environmental Protection Agency (USEPA).

One problem in evaluating water quality is that the current data are available over a short time period and consequently the sample size is inadequate to provide necessary precision in parameter estimation. To overcome this problem, Duan, Ye, and Smith [19] developed a novel Bayesian approach in analyzing water quality data. Specifically, they considered the power prior approach to incorporate historical data. Their proposed normalized power prior based on D_0 takes the form

$$\pi\left(\mu,\sigma^2,a_0|D_0\right)\propto\frac{a_0^{\frac{a_0n_0}{2}+k+\delta_0-2}(1-a_0)^{\lambda_0-1}}{\left(\frac{2\sigma^2}{n_0\hat{\sigma}_0^2}\right)^{\frac{a_0n_0}{2}+k}\Gamma\left(\frac{a_0n_0-3}{2}+k\right)}exp\left[-\frac{a_0n_0}{2\sigma^2}\left(\hat{\sigma}_0^2+\left(\mu-\bar{y}_0\right)^2\right]\right., \tag{6.2}$$

where $\hat{\sigma}_0^2 = 1/n_0\sum_{i=1}^{n_0}\left(y_{0i}-\bar{y}_0\right)^2$, k is a pre-specified constant, and δ_0 and λ_0 are known hyperparameters of the prior distribution of a_0. As shown in [19], the power prior approach in (6.2) improves the precision of the estimatesof the measurements of water quality over other approaches.

6.3. Application to Pediatric Quality of Care

In the pediatric quality of care clinics, investigators conducting new research often have access to data from previous studies. Therefore, it is scientifically reasonable and

advantageous to incorporate the information from previous studies in conducting a new study on pediatric quality of care.

In the context of pediatric quality of care, Neelon and O'Mally [57] compared common specifications of the power prior and explored whether it is preferable to use fixed a_0 or random a_0, which was also discussed in Section 2.3. They empirically showed that the normalized power prior provides a measure of congruence between the current and historical data, so that the historical data were downweighted more substantially as the studies diverged. They suggested that in real world problems involving large datasets and models with several parameters, the normalized power prior may lead to considerably more downweighting than desired. Thus, they further recommended that it is perhaps more appropriate to assign a_0 a fixed value based on expert opinion about the relevance of the historical data for the current analysis. Neelon and O'Mally [57] then applied the power prior methods to a pair of studies designed to improve delivery of care in pediatric clinics.

6.4. Analysis of Randomized Therapeutic Trials

In randomized therapeutic trials (RTTs), historical data provide a valuable source of information for the motivation and design of later trials. One objective of these studies is to estimate the intervention effect. In the presence of previous studies, meta-analysis is a well-known approach for estimating the overall treatment effect. When one is interested in the effect of the study-specific subpopulation, however, the historical data based on meta-analysis would receive too much weight. As a solution to this problem, Charlotte et al. [46] established a new Bayesian method for analyzing data from randomized therapeutic trials. They evaluated the use of the power prior distribution, illustrated with data from a large randomized clinical trial on the effect of ST-wave analysis in intrapartum fetal monitoring.

Charlotte et al. [46] advocated the use of a power prior distribution with pre-specified fixed study weights based on differences in study characteristics. They further proposed obtaining a ranking of the historical studies via expert elicitation, based on relevance for the current study, and then specified study weights accordingly.

6.5. Benchmark Dose Estimation in Toxicology

The benchmark approach is a useful tool in toxicology. One of the aims of these studies is to estimate the benchmark dose in the toxicological experiment. The benchmark dose, denoted by BMD, is defined as the dose of an environmental toxicant that corresponds to a prescribed change in response compared to the background response level. The toxicological data consist of n binomial responses $\mathbf{y} = (y_1,...,y_n)$ of an adverse event at a specific dose level. The assumed model is $y_i \sim b(n_i, p_i)$, where n_i is the number of animals tested at dose level x_i and p_i is the probability that the animals give an adverse response. That is, $p_i = \frac{exp(\beta_1+\beta_2 x_i)}{1+exp(\beta_1+\beta_2 x_i)}$ for $i = 1,..., n$. The historical $D_0 = (n_0, \mathbf{y}_0, X_0)$ are available from a recent report by the USEPA.

The typical BMD analysis employed by the U.S. EPA ignores the possibility that other models might partially reflect the true dose-response relationship. An alternative approach is to estimate a Bayesian model averaged (BMA) BMD. Shao [62] compared three methods

for integrating historical data from a previous study, including a pooled data analysis, a hierarchical model, and the power prior approach, in risk assessment for BMA BMD estimation. He empirically showed that the power prior had little influence on current estimates when the historical and current data were incompatible.

In the context of the toxicology studies, Shao [37] discussed another power prior approach, which borrows "partially" from the common parameter shared in the models for the historical data and the current data. Using this power prior, which is called the *partial borrowing power prior* in this paper, Shao [37] empirically showed that the partially borrowing power prior successfully achieved the reduction of the uncertainty in the estimates of both the parameters of interest and the benchmark dose.

6.6. Hospital Anxiety and Depression Scale in Psychology

In psychometrics, item response theory (IRT) models have been commonly used to express the probability of an item response as a function of the item psychometric properties and the individual latent characteristics (generally-called abilities). In order to accurately estimate the item parameters and the individual abilities, a large number of respondents and many items in a test are needed. However, in practice, the administration of a test with a large number of items on a large number of subjects may not always be possible. Therefore, the use of collateral or historical information in model estimation assumes a particular importance. In Bayesian estimation of IRT models, non-informative priors may lead to unstable estimates and poor convergence of the Gibbs sampling algorithm. To overcome these problems, Matteucci and Veldkamp [56] introduced the power prior in Bayesian estimation of IRT models.

Using the data from the Hospital Anxiety and Depression Scale, Matteuchi and Veldkamp [56] demonstrated the efficiency of the power prior approach in terms of measurement precision with small samples. In addition, Matteuchi and Veldkamp [56] empirically showed that the power prior improves not only the precision of the ability estimates but also convergence of the Gibbs sampling algorithm.

6.7. Application to Non-inferiority Trials for Anti-infective Products

In the context of the design and analysis of non-inferiority (NI) trials for anti-infective products, Gamalo, Tiwari, and LaVange [47] developed a new methodological approach to determine NI margins that can utilize all relevant historical data through a novel power adjusted Bayesian meta-analysis. They also provided a Bayesian decision rule for the NI analysis that is based on a broader use of available prior information and a sample-size determination that is based on this Bayesian decision rule. They used the power prior as a means to discount historical data and then proposed a new prior, called the *order restricted power prior*, for combining historical data from different types of studies such as randomized double-blind studies, randomized open-label studies, observational studies, animal models of disease, and Pharmacokinetic-Pharmacodynamic (PK-PD) profiles.

For illustrative purposes, we consider that the historical data are from two types of studies. Let $\bar{\boldsymbol{Y}}^{H1} = \left(\bar{Y}_1^{H1}, \ldots, \bar{Y}_k^{H1}\right)$ denote the random variables corresponding to estimates of the treatment response from k historical studies that are not randomized control trials (RCTs) and also let $\bar{\boldsymbol{Y}}^{H2} = \left(\bar{Y}_{k+1}^{H2}, \ldots, \bar{Y}_m^{H2}\right)$ denote the random variables corresponding to estimates of the treatment response from $m - k$ historical studies that are RCTs. Write the combined historical data as $\bar{\boldsymbol{Y}}^{H} = \left(\left(\bar{\boldsymbol{Y}}^{H1}\right)', \left(\bar{\boldsymbol{Y}}^{H2}\right)'\right)$. Assume that $\bar{Y}_i^{H1} \sim N\left(\mu+\alpha_i, \sigma_i^2\right)$ for $i = 1,\ldots, k$ and $\bar{Y}_i^{H2} \sim N\left(\mu+\alpha_i, \sigma_i^2\right)$ for $i = k + 1,\ldots, m$, where μ is the population treatment response, α_i, $i = 1,\ldots, m$, are random effects corresponding to the study effets, and σ_i^2 $i = 1, \ldots, m$, are within-study variabilities. Let $\boldsymbol{\theta}=(\mu, \alpha_{\mathbf{1}}, \ldots, \alpha_{\mathbf{m}})'$ and $\boldsymbol{\sigma}^2=\left(\sigma_1^2, \ldots, \sigma_m^2\right)'$. The order restricted power prior proposed by [47] is given by

$$\pi\left(\boldsymbol{\theta}|\bar{\boldsymbol{Y}}^{\mathbf{H}}, \boldsymbol{\sigma}^{\mathbf{2}}, \mathbf{a_{01}}, \mathbf{a_{02}}\right) \propto \left[\prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}\sigma_i} exp\left\{-\frac{1}{2\sigma_i^2}\left(\bar{Y}_i^{H1} - \mu - \alpha_i\right)^2\right\}\right]^{a_{01}} \times \left[\prod_{j=k+1}^{n} \frac{1}{\sqrt{2\pi}\sigma_j} exp\left\{-\frac{1}{2\sigma_j^2}\left(\bar{Y}_j^{H2} - \mu - \alpha_j\right)^2\right\}\right]^{a_{02}} \pi_0(\boldsymbol{\theta}), \quad 0<a_{01}<a_{02}<1, \tag{6.3}$$

where $\pi_0(\boldsymbol{\theta})$ is the initial prior for $\boldsymbol{\theta}$. The joint order restricted power prior therefore takes the form $\pi\left(\boldsymbol{\theta}, \mathbf{a_{01}}, \mathbf{a_{02}}|\bar{\boldsymbol{Y}}^{\mathbf{H}}, \boldsymbol{\sigma}\right) \propto \pi\left(\boldsymbol{\theta}|\bar{\boldsymbol{Y}}^{\mathbf{H}}, \boldsymbol{\sigma}^{\mathbf{2}}, \mathbf{a_{01}}, \mathbf{a_{02}}\right) \pi_0(a_{01}, a_{02})$, where $\pi_0(a_{01}, a_{02})$ is the initial prior for (a_{01}, a_{02}). The order constraints on the power parameters in (6.3) serve as a means to downweight the influence of the historical data, which are not RCTs. Gamalo, Tiwari, and LaVange [47] specified a Dirichlet distribution as the initial prior for the transformed variables $(u, v, 1 - u - v)$, where $u = a_{01}$ and $v = a_{02} - a_{01}$, a Dirichlet process prior for $(a_1,\ldots, a_m)'$, and an improper uniform initial prior for μ.

Gamalo, Tiwari, and LaVange [47] illustrated their proposed method through three case studies, including determination of the effect of antibacterial drugs in reducing all-cause mortality in hospital-acquired or ventilator-associated bacterial pneumonia (HABP/VABP) patients, estimation of the NI margin for trials in HABP/VABP drug development, and sample size determination for the treatment of HABP/VABP. They empirically showed that the approach of incorporating prior information in the sample size calculations for NI trials can result in significant reductions in sample size.

7. A Case Study in Cancer Clinical Trials

Chen et al. [11, 29] and Ibrahim et al. [15, 63] demonstrated the use of the power prior in survival analysis settings in the context of melanoma cancer clinical trials. Interferon (IFN) was used in two previous Eastern Cooperative Oncology Group (ECOG) phase III

melanoma clinical trials, E1684 and E1690. The first trial, E1684, was a two arm clinical trial comparing high-dose interferon (IFN) to Observation (OBS) (Kirkwood et al. [64]). There were a total of $n_0 = 286$ patients enrolled in the study. The treatment effect favoring IFN that was seen in E1684 with respect to both relapse-free survival (RFS) and overall survival (OS) was larger than expected and was accompanied by substantial side effects due to the high-dose regimen. As a result, ECOG began a second trial (E1690) in 1991 to attempt to confirm the results of E1684 and to study the benefit of IFN given at a lower dose. The ECOG trial E1690 was a three arm phase III clinical trial, and had treatment arms of high dose interferon, low dose interferon, and observation (Kirkwood et al. [65]). This study had $n = 427$ patients on the high dose interferon arm and observation arm combined. The two datasets were quite similar with respect to the distributions of several prognostic factors, including Breslow depth, number of nodes, performance status, site of primary, and stage of disease. Prognostic factor analyses were conducted to examine the significance of time trend covariates and institutional effects for each study alone, as well as for the combined studies, and these factors were highly non-significant.

Chen, Harrington, and Ibrahim [29] and Ibrahim, Chen, and Chu [63] considered an analysis of the E1690 data, using E1684 as the historical data, which was incorporated via the power prior. They presented a power prior using the treatment covariate alone based on a piecewise exponential model, parametric cure rate model, and semiparametric cure rate model. Here, we present an analysis based on RFS and four covariates, which are treatment, age, gender, and the interaction of treatment and gender. In our analysis, we used $n_0 = 285$ with deletion of one observation due to missing age and gender. We use the cure rate model of [11] to carry out the Bayesian analysis for these data. The cure rate model has been a key component in the design of adjuvant melanoma ECOG trials, and this model was used to design E1690, E1694, and the E1697 adjuvant melanoma trials.

For the i^{th} patient, $i = 1,...,n$, let y_i denote the observed RFS time or censoring time and let v_i be the censoring indicator variable taking a value of 1 if y_i is an RFS time and 0 if it is a censoring time. Also, let trt_i denote the treatment indictor such that $\text{trt}_i = 1$ if the i^{th} patient received IFN and $\text{trt}_i = 0$ if the i^{th} patient received OBS. As the gender indicator, gender_i is 1 if the i^{th} patient is male and 0 if the i^{th} patient is female. The likelihood function for the E1690 data (denoted by D with $n = 427$) is given by

$L(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{D}) = \prod_{i=1}^{n} \left\{ exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) f_0(y_i|\boldsymbol{\lambda}) \right\}^{\nu_i} exp\left\{ -exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) F_0(y_i|\boldsymbol{\lambda}) \right\}$, where

$\boldsymbol{\beta} = (\beta_0, \cdots, \beta_4)'$, $\boldsymbol{x}_i'\boldsymbol{\beta} = \beta_0 + \mathbf{trt_i}\beta_1 + \mathbf{age_i}\beta_2 + \mathbf{gender_i}\beta_3 + (\mathbf{trt_i} \times \mathbf{gender_i})\beta_4$, and $F_0(y|\boldsymbol{\lambda})$ is the cumulative distribution function and $f_0(y|\boldsymbol{\lambda})$ is the corresponding density function. We further assume a piecewise exponential model for $F_0(y|\boldsymbol{\lambda})$, which is given by

$F_0(y|\boldsymbol{\lambda}) = 1 - exp\left\{ -\lambda_j (y - s_{j-1}) - \sum_{g=1}^{j-1} \lambda_g (s_g - s_{g-1}) \right\}$, where

$s_{j-1} \leq y < s_j$, $s_0 = 0 < s_1 < \cdots < s_{J-1} < s_J = \infty$, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)'$. Let D_0 denote the historical E1684 data and the corresponding likelihood function $L(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{D_0})$ is defined in a similar way as the one for the current data. Then, the power prior in (2.1) reduces to $\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{D_0}, \mathbf{a_0}) \propto [L(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{D_0})]^{a_0} \pi_0(\boldsymbol{\beta}, \boldsymbol{\lambda})$, where $0 \leq a_0 \leq 1$, $\pi_0(\boldsymbol{\beta}, \boldsymbol{\lambda})$ is the initial prior,

and $L(\boldsymbol{\beta}, \boldsymbol{\lambda}|\mathbf{D_0})$ is the likelihood function based on the historical data. We specify a joint noninformative uniform initial prior for $(\boldsymbol{\beta}, \boldsymbol{\lambda})$, i.e., $\pi_0(\boldsymbol{\beta}, \boldsymbol{\lambda}) \propto 1$.

Figure 2 shows the plots of DIC and LPML versus a_0 for various values of J. We see from Figure 2 that DIC (LPML) is roughly a convex (concave) function of a_0 for a given value of J and the entire DIC (LPML) curve for $J = 5$ is below (above) those corresponding to $J = 2$, $J = 10$, and $J = 15$. These results show that $J = 5$ is the best choice according to both the DIC and LPML criteria. In terms of a_0, the respective best DIC values for $J = 2$, 5, 10, and 15 are 1040.626, 1038.072, 1044.742, and 1045.578, attained at $a_0 = 0.2$, 0.5, 0.6, and 0.7 while the respective best LPML values for $J = 2$, 5, 10, and 15 are −520.324, −519.06, −522.403, and −522.902, attained at $a_0 = 0.3$, 0.5, 0.6, and 0.7. These results empirically show that when the model becomes more complex (i.e., J becomes larger), the optimal a_0 becomes larger. This is quite interesting as these results essentially imply that the historical and current data become more comparable under more complex models.

The posterior estimates of the regression parameters, including posterior means, posterior standard deviations (SDs), and 95% HPD intervals, for $a_0 = 0$, 0.5, and 1 under $J = 5$ are given in Table 2. From Table 2, we see that (i) the posterior SDs decrease as a_0 increases for all parameters, implying that the incorporation of historical data improves the precision of the posterior estimates; (ii) the 95% HPD intervals for β_2, β_3, β_4, and $\beta_1 + \beta_4$ include 0 for all a_0's; and (iii) the 95% HPD intervals for β_0 and β_1 include 0 when $a_0 = 0$ but do not include 0 when $a_0 = 0.5$ and 1. Note that based on our notation and model setup, β_1 quabtufues the treatment effect for female patients while $\beta_1 + \beta_4$ captures the treatment effect for male patients. We also see from Table 2 that the posterior means and 95% HPD intervals are −0.311 and (−0.620, 0.012), −0.340 and (−0.616, −0.078), and −0.352 and (−0.593, −0.126) for $a_0 = 0$, 0.5, and 1, respectively, for β_1; and −0.022 and (−0.447, 0.416), −0.108 and (−0.461, 0.259), and −0.147 and (−0.442, 0.180) for $a_0 = 0$, 0.5, and 1, respectively, for $\beta_1 + \beta_4$. We also computed the posterior estimates of the hazard ratios (HRs) of the treatment effect for females ($\exp(\beta_1)$) and for males ($\exp(\beta_1 + \beta_4)$). The estimated HRs of the treatment effect and corresponding 95% HPD intervals are 0.732 and (0.538, 1.012), 0.712 and (0.540, 0.925), and 0.703 and (0.553, 0.882) for $a_0 = 0$, 0.5, and 1, respectively, for females; and 0.978 and (0.640, 1.516), 0.898 and (0.631, 1.295), and 0.864 and (0.643, 1.197) for $a_0 = 0$, 0.5, and 1, respectively, for males. The above results indicate that the treatment effect favoring IFN with respect to RFS can be seen only for female patients but not for male patients.

8. The Power Prior in Bayesian Design

Many examples of the use of the power prior in experimental design settings can be found in Spiegelhalter, Abrams, and Myles (2004). The power prior has also been used more recently in clinical trial design settings in [21, 24, 38, 39]. Specifically, Chen et al. [21] developed a new and general method to determine Bayesian sample size using historical data for a non-inferiority trial. Ibrahim et al. [24] and Chen et al. [38, 39] adapted this method using the partial borrowing power prior in Bayesian meta-experimental design as well as Bayesian design for superiority trials with recurrent events data.

To put the methodology in a specific context, Chen et al. [21] considered designing a clinical trial to evaluate the performance of a new generation of drug-eluting stents (DES) ("test device") with a non-inferiority comparison to the first generation of DES ("control device"). The trial had two arms: test device and control device. The primary endpoint is the 12-month Target Lesion Failure (TLF) (binary) composite endpoint, which is an ischemia-driven revascularization of the target lesion (TLR), myocardial infarction (MI) (Q-wave and non-Q-wave) related to the target vessel, or (cardiac) death related to the target vessel. Historical data were available from two previous trials on the first generation of DES. The first trial conducted in 2002 evaluated the safety and effectiveness of the slow release paclitaxel-eluting stent for treatment of de novo coronary artery lesions. The second trial conducted in 2004 expanded on the first trial, studied more complex de novo lesions, and involved multiple overlapping stents and smaller and larger diameter stents. The historical data based on lesion size matched criteria are subsets of the data published in Stone et al. [66, 67]. The numbers of failures, the numbers of patients, and the percentages of the 12-Month TLF were 44, 535, and 8.2% for historical trial 1 and 33, 304, and 10.9% for historical trial 2.

Let (y_t, n_t) and (y_c, n_c) be the data corresponding to the test device and the control device, respectively. Assume that the ratio of the two sample sizes, $r=\frac{n_c}{n_t}$, is fixed and typically small. Thus, $n_t=\frac{n}{1+r}$ and $n_c=\frac{rn}{1+r}$ where $n = n_t + n_c$ is the total sample size, The goal of the trial is to show that the test device is non-inferior to the control device. We assume that y_t and y_c independently follow binomial distributions $b(p_t, n_t)$ and $b(p_c, n_c)$, respectively. Then, the joint distribution of $y^{(n)} = (y_t, y_c)'$ is given by

$f\left(\boldsymbol{y}^{(n)}|\boldsymbol{\theta}\right) \propto p_t^{y_t}(1-p_t)^{n_t-y_t}p_c^{y_c}(1-p_c)^{n_c-y_c}$, where $\boldsymbol{\theta}=(p_t,p_c)$. The design parameter is the difference between p_t and p_c, namely, $p_t - p_c$. The hypotheses for non-inferiority testing are H_0: $p_t - P_c \geq \delta$ versus H_1: $p_t - P_c <$, where δ is a prespecified non-inferiority margin. The trial is successful if H_1 is accepted. Let $\Theta_0=\{\boldsymbol{\theta}=(\mathbf{p_t},\mathbf{p_c}):\mathbf{p_t}-\mathbf{p_c}\geq\delta\}$ and $\Theta_1=\{\boldsymbol{\theta}=(\mathbf{p_t},\mathbf{p_c}):\mathbf{p_t}-\mathbf{p_c}<\delta\}$. Following Chen et al. [21], the key design quantity is defined as

$$\beta_s^{(n)}=E_s\left[1\left\{P\left(p_t-p_c<\delta|\boldsymbol{y}^{(n)},\pi^{(f)}\right)\geq\gamma\right\}\right], \quad (8.1)$$

where the indicator function $1\{A\}$ is 1 if A is true and 0 otherwise, $\gamma > 0$ is a prespecified Bayesian credible level, the probability $P\left(p_t-p_c<\delta|\boldsymbol{y}^{(n)},\pi^{(f)}\right)$ is computed with respect to the posterior distribution of $\boldsymbol{\theta}$ given the data $\boldsymbol{y}^{(n)}$ and the feting prior $\pi^{(f)}(\boldsymbol{\theta})$, and the expectation E_s is taken with respect to the marginal distribution of $\boldsymbol{y}^{(n)}$ under the sampling prior $\pi^{(s)}(\boldsymbol{\theta})$. Let $\overline{\Theta}_0$ and $\overline{\Theta}_1$ denote the closures of Θ_0 and Θ_1. Let $\pi_0^{(s)}(\boldsymbol{\theta})$ denote a sampling prior with support $\Theta_B=\overline{\Theta}_0\cap\overline{\Theta}_1$. Also let $\pi_1^{(s)}(\boldsymbol{\theta})$ denote a sampling prior with support $\Theta_1^*\subset\Theta_1$. Then, $\beta_{s0}^{(n)}$ and $\beta_{s1}^{(n)}$ given in (8.1) corresponding to $\pi^{(s)}=\pi_0^{(s)}$ and $\pi^{(s)}=\pi_1^{(s)}$ are the Bayesian type I error and power, respectively.

Let $\boldsymbol{y}_{c0} = (y_{c01}, y_{c02})' = (44,33)'$ denote the historical data for the control medical device. The partial borrowing power prior with fixed $a_0 = (a_{01}, a_{02})'$ is given by

$$\pi\left(p_t, p_c|\boldsymbol{y}_{c0}, \boldsymbol{a}_0\right) \propto \pi_0\left(p_t\right) p_c^{a_{01}y_{c01}+a_{02}y_{c02}}\left(1-p_c\right)^{a_{01}\left(n_{c01}-y_{c01}\right)+a_{02}\left(n_{c02}-y_{c02}\right)}\pi_0\left(p_c\right), \quad (8.2)$$

where $0 \leq a_{01}, a_{02} \leq 1$, and $\pi_0(p_t)$ and $\pi_0(p_c)$ are initial priors. Assuming $\pi_0\left(p_c\right) \propto p_c^{-1}(1-p_c)^{-1}$, the posterior distribution of p_c is given by

$$\pi\left(p_c|y_c, \boldsymbol{y}_{c0}, \boldsymbol{a}_0\right) \propto p_c^{y_c+a_{01}y_{c01}+a_{02}y_{c02}-1}(1-p_c)^{\left(n_c-y_c\right)+a_{01}\left(n_{c01}-y_{c01}\right)+a_{02}\left(n_{c02}-y_{c02}\right)-1}, \quad (8.3)$$

and the normalized power prior for (p_t, p_c, a_0) given multiple historical datasets $\boldsymbol{y}_{c0}$ is of the form

$$\pi\left(p_t, p_c, \boldsymbol{a}_0|\boldsymbol{y}_{c0}\right) \propto \pi_0\left(p_t\right) \frac{1}{C\left(\boldsymbol{a}_0\right)} p_c^{a_{01}y_{c01}+a_{02}y_{c02}-1}(1-p_c)^{a_{01}\left(n_{c01}-y_{c01}\right)+a_{02}\left(n_{c02}-y_{c02}\right)-1}\prod_{k=1}^{2} a_{0k}^{b_{01}-1}\left(1-a_{0k}\right)^{b_{02}-1}\Bigg], \quad (8.4)$$

where $C\left(\boldsymbol{a}_0\right) = B\left(a_{01}y_{c01}+a_{02}y_{c02}, a_{01}\left(n_{c01}-y_{c01}\right)+a_{02}\left(n_{c02}-y_{c02}\right)\right)$, $B(\cdot,\cdot)$ denotes the complete beta function, and $b_{01} > 0$ and $b_{02} > 0$ are prespecified hyperparameters. Note that in (8.4), we assume that the a_{0k}'s are independent and distributed as a_{0k} ~ beta(b_{01}, b_{02}) for k = 1,2.

We consider (8.2) or (8.4) after integrating out a_0 as the fitting prior $\pi^{(f)}\left(\boldsymbol{\theta}|\boldsymbol{y}_{\mathbf{c0}}\right)$. For the sampling prior, $\pi_\ell^{(s)}\left(\boldsymbol{\theta}\right)$, l = 0,1, we take

$$\pi_\ell^{(s)}\left(\boldsymbol{\theta}\right) = \begin{cases} \Delta_{\{p_t=p_c^*+\delta, p_c=p_c^*\}} & \text{for} \quad \ell=0, \\ \Delta_{\{p_t=p_c^*, p_c=p_c^*\}} & \text{for} \quad \ell=1, \end{cases} \quad (8.5)$$

where $0<p_c^*<1$ is the design value of the 12 month TLF for the future data and $\Delta_{\{A\}}$ denotes the point mass at the event A, that is, $P(A) = 1$.

As discussed in [21], we set the margin to be $\delta = 4.1\%$, took an improper beta(0,0) initial prior for $\pi_0(p_c)$, and specified $b_{01} = b_{02} = 1$ for the initial priors of the a_{0k}'s in (8.4). In the sampling prior, we assumed a point mass prior at $p_c^*=9.2\%$ for $\pi^{(s)}(p_c)$, where 9.2% was the pooled proportion for the two historical control datasets. We first computed the powers and the type I errors for various sample sizes based on the Bayesian procedure without the incorporation of historical data ($\boldsymbol{a}_0 = \mathbf{0}$) as well as with power priors for random and fixed $\boldsymbol{a}_0$. Table 3 shows the results. From Table 3, we see that (i) without incorporation of historical data, a total sample size of 1480 is required in order to achieve 80% power; (ii) With incorporation of the historical data, a sample size of (n_t, n_c) = (810, 270) achieves 80% power; and (iii) the power prior with random $\boldsymbol{a}_0$ borrows approximately 30% of the historical data. Thus, the Bayesian sample size determination (SSD) procedure with incorporation of historical data leads to a reduction in the sample size.

To carry out a sensitivity analysis of the Bayesian SSD, we consider n = 1200 with n_t = 900 and n_c = 300, three different values of p_c^*, namely 8.0%, 9.2%, and 10.0%, and various Bayesian credible levels for γ. The powers and type I errors for the normalized power priors with various initial priors for a_{0k}'s as well as the power prior with fixed $a_0 = (0.3, 0.3)'$ are

shown in Table 4. From this table, we see that the type I errors were not controlled at 5% for both the normalized power prior and the power prior with fixed $a_0 = (0.3, 0.3)'$ when $\gamma = 0.95\%$ and $p_c^* = 8.0\%$, because of the fact that the historical data and the current data from the control device are not compatible. Also, we see that the type I errors corresponding to the normalized power prior is quite sensitive to the specification of the initial prior beta(b_{01}, b_{02}) for a_{0k} in (8.4). However, when γ increases or when the historical control data are downweighted, the type I error decreases. In particular, when $\gamma = 0.96$ and 0.97, the type I errors were 0.049 and 0.035 for the power prior with fixed $a_0 = (0.3, 0.3)'$. In addition, if a point mass sampling prior at $p_c = 8.0\%$ is assumed, the type I errors under the normalized power prior were 0.041 when $(b_{01}, b_{02}) = (1,1)$ and $\gamma = 0.97$ and 0.047 when $(b_{01}, b_{02}) = (1,10)$ and $\gamma = 0.96$.

In the binomial setting, closed-form expressions for the penalized likelihood-type criterion (PLC), the marginal likelihood criterion (MLC), and the deviance information criterion (DIC) in Section 5 are available. Let $\eta_1(\boldsymbol{a}_0) = y_c + a_{01}y_{c01} + a_{02}y_{c02}$ and $\eta_2(\boldsymbol{a}_0) = (n_c - y_c) + a_{01}(n_{c01} - y_{c01}) + a_{02}(n_{c02} - y_{c02})$. Using (8.3), assuming an improper beta(0,0) initial prior for p_c, and ignoring the binomial coefficient $\binom{n_c}{y_c}$, we have

$$m^*(\boldsymbol{a}_0) = B(\eta_1(\boldsymbol{a}_0), \eta_2(\boldsymbol{a}_0)). \quad (8.6)$$

The guide values of $\boldsymbol{a}_0$ based on PLC and MLC are given as

$$\boldsymbol{a}_{0,PLC}^{opt} = \underset{\mathbf{0}<\boldsymbol{a}_0<1}{arg\,min} G(\boldsymbol{a}_0) \quad \text{and} \quad \boldsymbol{a}_{0,MLC}^{opt} = \underset{\mathbf{0}<\boldsymbol{a}_0<1}{arg\,min} [-2\log\{m(\boldsymbol{a}_0)\}], \quad (8.7)$$

$G(\boldsymbol{a}_0) = -2\log[m^*(\boldsymbol{a}_0)] + \sum_{k=1}^{2} \frac{\log(n_{c0k})}{a_{0k}}$, $m(\boldsymbol{a}_0) = m^*(\boldsymbol{a}_0)/C(\boldsymbol{a}_0)$, and $m^*(\boldsymbol{a}_0)$ and $C(\boldsymbol{a}_0)$ are defined in (8.6) and (8.4), respectively. Ignoring $\binom{n_c}{y_c}$, we take the deviance function as $Dev(p_c) = -2[y_c \log p_c + (n_c - y_c)\log(1 - p_c)]$. Using (8.3), the posterior mean of p_c is $\bar{p}_c(\boldsymbol{a}_0) = \eta_1(\boldsymbol{a}_0)/[\eta_1(\boldsymbol{a}_0) + \eta_2(\boldsymbol{a}_0)] = (y_c + a_{01}y_{c01} + a_{02}y_{c02})/(n_c + a_{01}n_{c01} + a_{02}n_{c02})$. It can be shown that the posterior mean of Dev(p_c) is given by

$$\begin{aligned} E[Dev(p_c)|y_c, \boldsymbol{y}_{c0}, \boldsymbol{a}_0] &= -\frac{2}{B(\eta_1(\boldsymbol{a}_0), \eta_2(\boldsymbol{a}_0))} \int_0^t [y_c \log(p_c) + (n_c - y_c)\log(1 - p_c)] p_c^{\eta_1(\boldsymbol{a}_0)-1}(1 - p_c)^{\eta_2(\boldsymbol{a}_0)-1} dp_c \\ &= -2[y_c\psi(\eta_1(\boldsymbol{a}_0)) + (n_c - y_c)\psi(\eta_2(\boldsymbol{a}_0)) - n_c\psi(\eta_1(\boldsymbol{a}_0) + \eta_2(\boldsymbol{a}_0))], \end{aligned} \quad (8.8)$$

where $\psi(\eta) = \frac{d}{d\eta}\log\Gamma(\eta)$ is the digamma function. Then, the guide value of $\boldsymbol{a}_0$ based on DIC is given by

$$\boldsymbol{a}_{0,DLC}^{opt} = \underset{\mathbf{0}<\boldsymbol{a}_0<1}{arg\,min} \left\{ 2E[Dev(p_c)|y_c, \boldsymbol{y}_{c0}, \boldsymbol{a}_0] + 2\left[y_c \log \bar{p}_c(\boldsymbol{a}_0) + (n_c - y_c)\log\left(1 - \bar{p}_c(\boldsymbol{a}_0)\right)\right]\right\}, \quad (8.9)$$

where $E\left[Dev\left(p_c\right)|y_c, \boldsymbol{y}_{c0}, \boldsymbol{a}_0\right]$ is given by (8.8). The powers and type I errors under these three guide values are also given in Table 4. From this table, we see that (i) the power prior with $\boldsymbol{a}_{0,PLC}^{opt}$ leads to a slightly lower power but a better controlled type I error; (ii) the powers and type I errors under the power priors with $\boldsymbol{a}_{0,MLC}^{opt}$ and $\boldsymbol{a}_{0,DIC}^{opt}$ are similar; and (iii) both $\boldsymbol{a}_{0,MLC}^{opt}$ and $\boldsymbol{a}_{0,DIC}^{opt}$ require higher Bayesian credible levels in order to control the type I errors and maintain good powers at the same time.

9. Concluding Remarks

We have provided a comprehensive review of the power prior and its applications in this article. As seen in earlier sections, the power prior has been used in a wider variety of contexts and disciplines, ranging from experimental design to data analysis. We also demonstrated a large number of attractive theoretical properties of the power for inference, such as its asymptotic normality, conjugacy, log-concavity for Gibbs sampling, its connection to hierarchical models, its semi-automatic nature for variable subset selection, and its important role in the design of clinical trials. The power prior continues to be highly used today, especially in clinical trials contexts, in both design and analysis settings. It is becoming a standard approach to informative prior elicitation. It is also one of the recommended priors by the US Food and Drug Administration (FDA) for analyzing and designing medical device trials.

The power prior has also been implemented in software packages such as SAS Proc MCMC and WinBUGS. Some comments are in order here for the different variations of the power prior. First, we note that the normalized power prior (2.4) is computationally difficult to work with especially in regression settings, and thus the joint power prior (2.3) is more preferred for this purpose. Second, the partial borrowing power prior is flexible and different than (2.3) in that it allows borrowing information from historical data on a subset of the model parameters that is common to both the historical dataset and the current study. The partial borrowing power prior is most useful in clinical trials setting, for example, where one only has historical data on the control arm, and no historical data on the treatment arm for the current study. Third, the partial discounting power prior, which is different than the partial borrowing power prior, is most useful in settings with latent variables or random effects, where one wishes to discount the likelihood function of the historical data but not discount the distribution of the latent variables or the random effects. Fourth, the order restricted power prior, which is another variation of the power prior, is quite attractive in combining historical data from different types of previous trials with ordered importance. The partial borrowing power prior, the partial discounting power prior, and the order restricted power prior all show great promise for future use in both the design and analysis of clinical trials. All of these three variations of the power prior are promising topics of future research.

Regarding fixed or random a_0 in design or analysis, our experience shows that taking a_0 fixed and doing several sensitivity analyses for different values of a_0 is much computationally feasible and easier to interpret than taking a_0 random. In addition, a fixed a_0 may also be more convenient to elicit via expert opinion. The a_0 random case is

computationally very difficult and it often gives answers similar to the a_0 fixed case, so its advantages appear to be minimal. As a result, using fixed a_0 may be more preferred and desirable in many applications. A related issue is the specification of a useful guide value for a_0. In Section 5, we have derived the closed-form expressions of the penalized likelihood-type criterion, marginal likelihood criterion, DIC, and LPML criterion for the linear model with known sampling-level variances. When the sampling-level variances are unknown, the closed-form expressions of these criteria except for the LPML criterion can still be derived. Thus, for the LPML criterion, a sampling-based Monte Carlo method needs to be used in order to obtain a guide value of a_0. Although the approaches based on the penalized likelihood-type, marginal likelihood, DIC, and LPML criteria are attractive, their properties have not been fully investigated and much more work needs to be done in finding optimal guide values for a_0. This is also a topic of future research.

Acknowledgments

We would like to thank the Editor, the Associate Editor, and the two anonymous reviewers for their very helpful comments and suggestions, which have led to an improved version of the paper. Dr. J. G. Ibrahim's and Dr. M.-H. Chen research was partially supported by NIH grant #GM 70335.

References

1. Ibrahim JG, Chen MH. Power prior distributions for regression models. Statistical Science. 2000; 15:46–60.
2. Berry DA. Bayesian Methods in Phase III Trials. Drug Information Journal. 1991; 25:345–368.
3. Eddy, DM.; Hasselblad, V.; Schachter, R. Meta-analysis by the Confidence Profile Method: The Statistical Synthesis of Evidence. Academic Press; New York: 1992.
4. Berry DA, Hardwick J. Using Historical Controls in Clinical Trials: Application to ECMO. Statistical Decision Theory and Related Topics. 1993; 5:141–156.
5. Lin, Z. Statistical Methods for Combining Historical Controls with Clinical Trials Data. Duke University; 1993. Unpublished Doctorial Dissertation
6. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian Approaches to Randomised Trials (with discussion). Journal of Royal Statistical Society, Series A. 1994; 157:357–416.
7. Berry, DA.; Stangl, DK. Bayesian Biostatistics. Marcel Dekker; New York: 1996.
8. Chen MH, Manatunga AK, Williams CJ. Heritability Estimates from Human Twin Data by Incorporating Historical Prior Information. Biometrics. 1998; 54:1348–1362. [PubMed: 9883538]
9. Ibrahim JG, Chen MH. Prior Distributions and Bayesian Computation for Proportional Hazards Models. Sankhya, Series B. 1998; 60:48–64.
10. Ibrahim JG, Ryan LM, Chen MH. Use of Historical Controls to Adjust for Covariates in Trend Tests for Binary Data. Journal of the American Statistical Association. 1998; 93:1282–1293.
11. Chen MH, Ibrahim JG, Sinha D. A New Bayesian Model for Survival Data with A Surviving Fraction. Journal of the American Statistical Association. 1999; 94:909–919.
12. Chen MH, Ibrahim JG, Yiannoutsos C. Prior Elicitation, Variable Selection and Bayesian Computation for Logistic Regression Models. Journal of the Royal Statistical Society, Series B. 1999; 61:223–242.
13. Ibrahim JG, Chen MH, MacEachern SN. Bayesian Variable Selection for Proportional Hazards Models. Canadian Journal of Statistics. 1999; 27:701–717.
14. Chen MH, Ibrahim JG, Shao QM. Power Prior Distributions for Generalized Linear Models. Journal of Statistical Planning and Inference. 2000; 84:121–137.
15. Ibrahim JG, Chen MH, Sinha D. On Optimality Properties of The Power Prior. Journal of the American Statistical Association. 2003; 98:204–213.

16. Chen MH, Ibrahim JG. The Relationship Between the Power Prior and Hierarchical Models. Bayesian Analysis. 2006; 1:551–574.
17. De Santis F. Using Historical Data for Bayesian Sample Size Determination. Journal of the Royal Statistical Society, Series A. 2007; 170:95–113.
18. De Santis F. Power Priors and Their Use in Clinical Trials. American Statistical Association. 2006; 60:122–129.
19. Duan Y, Ye K, Smith EP. Evaluating Water Quality Using Power Priors to Incorporate Historical Information. Environmetrics. 2006; 17:95–106.
20. Neelon B, O'Malley AJ, Margolis PA. Bayesian Analysis Using Historical Data With application to Pediatric Quality of Care. ASA Proceedings of the Bayesian Statistical Sciences Section. 2008:2960–2967.
21. Chen MH, Ibrahim JG, Lam P, Yu A, Zhang Y. Bayesian Design of Non-inferiority Trials for Medical Devices Using Historical Data. Biometrics. 2011; 67:1163–1170. [PubMed: 21361889]
22. Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials. Biometrics. 2011; 67:1047–1056. [PubMed: 21361892]
23. Hobbs BP, Sargent DJ, Carlin BP. Commensurate Priors for Incorporating Historical Information in Clinical Trials Using General and Generalized Linear Models. Bayesian Analysis; 2012; 7:639–674.
24. Ibrahim JG, Chen MH, Xia A, Liu T. Bayesian Meta-experimental Design: Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes. Biometrics. 2012; 68:578–586. [PubMed: 21955084]
25. Spiegelhalter, DJ.; Abrams, KR.; Myles, JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. John Wiley & Sons; Chichester: 2004.
26. Berry, SM.; Carlin, PR.; Lee, JJ.; Müller, P. Bayesian Adaptive Methods for Clinical Trials. John Wiley & Sons; Boca Raton: 2012.
27. Zellner A. Optimal Information Processing and Bayes's Theorem (with discussion). The American Statistician. 1988; 42:278–284.
28. Zellner A. Information Processing and Bayesian Analysis. Journal of Econometrics. 2002; 107:41–50.
29. Chen MH, Harrington DP, Ibrahim JG. Bayesian Cure Rate Models for Malignant Melanoma: A Case Study of Eastern Cooperative Oncology Group Trial E1690. Applied Statistics. 2002; 51:135–150.
30. Duan, Y. A Modified Bayesian Power Prior with Applications in Water Quality Evaluation. Virginia Polytechnic Institute and State University; 2005. Unpublished Doctoral Dissertation
31. Chen MH, Ibrahim JG, Shao QM, Weiss RE. Prior Elicitation for Model Selection and Estimation in Generalized Linear Mixed Models. Journal of Statistical Planning and Inference. 2003; 111:57–76.
32. Chen MH, Dey DK, Shao QM. Bayesian Analysis of Binary Data Using Skewed Logit Models. Calcutta Statistical Association Bulletin. 2001; 51:11–30.
33. Chen MH, Dey DK. A Unified Bayesian Analysis for Correlated Ordinal Data Models. Brazilian Journal of Probability and Statistics. 2000; 14:87–111.
34. Chen MH. Bayesian Analysis of Correlated Mixed Categorical Data by Incorporating Historical Prior Information. The Special Issue of Communications in Statistics: Theory and Methods. 1998; 27:1341–1361.
35. Chen MH, Dey DK. Bayesian Modeling of Correlated Binary Responses via Scale Mixture of Multivariate Normal Link Functions. Sankhyā, Series A. 1998; 60:322–343.
36. Chen MH, Dey DK. Variable Selection for Multivariate Logistic Regression Models. Journal of Statistical Planning and Inference. 2003; 111:37–55.
37. Shao, K. Bayesian Model Averaging for Toxicity Study Design and Benchmark Dose Estimation. Carnegie Mellon University; 2011. Unpublished Doctorial Dissertation

38. Chen MH, Ibrahim JG, Xia HA, Liu T, Henessey V. Bayesian Sequential Meta-analysis Design in Evaluating Cardiovascular Risk in a New Antidiabetic Drug Development Program. Statistics in Medicine. 2014; 33:1600–1618. [PubMed: 24343859]
39. Chen MH, Ibrahim JG, Zeng D, Hu K, Jia C. Bayesian Design of Superiority Clinical Trials for Recurrent Events Data with Applications to Bleeding and Transfusion Events in Myelodyplastic Syndrome. Biometrics. 2014; 70:1003–1013. [PubMed: 25041037]
40. Ibrahim JG, Chen MH, Ryan LM. Bayesian Variable Selection for Time Series Count Data. Statistica Sinica. 2000; 10:971–987.
41. Chen MH, Shao QM. Monte Carlo Methods on Bayesian Analysis of Constrained Parameter Problems. Biometrika. 1998; 85:73–87.
42. Liang F. A Double Metropolis-Hastings Sampler for Spatial Models with Intractable Normalizing Constants. Journal of Statistical Computation and Simulation. 2010; 80:1007–1022.
43. A Monte Carlo Metropolis-Hastings Algorithm for Sampling from Distributions with Intractable Normalizing Constants. Neural computation. 2013; 25:2199–2234. [PubMed: 23607562]
44. Liang F, Jin IH, Song Q, Liu JS. An Adaptive Exchange Algorithm for Sampling from Distributions with Intractable Normalizing Constants. Journal of the American Statistical Association. 2015 in press.
45. Duan Y, Smith EP, Ye K. Using Power Priors to Improve the Binomial Test of Water Quality. Journal of Agricultural, Biological, and Environmental Statistics. 2006; 11:151–168.
46. Charlotte R, Irene K, Kristel JMJ, Krel GMM, Herbert JAH. Incorporation of Historical Data in the Analysis of Randomized Therapeutic Trials. Contemporary Clinical Trials. 2011; 32:848–855. [PubMed: 21729767]
47. Gamalo MA, Tiwari RC, LaVange LM. Bayesian Approach to the Design and Analysis of Non-inferiority Trials for Anti-infective Products. Pharmaceutical Statistics. 2014; 13:25–40. [PubMed: 23913880]
48. Zellner, A. Bayesian Analysis in Econometrics and Statistics. Edward Elgar; Cheltenham: 1997.
49. Chen CF. On Asymptotic Normality of Limiting Density Functions with Bayesian Implications. Journal of the Royal Statistical Society, Series B. 1985; 97:540–546.
50. Ibrahim JG, Laud PW. A Predictive Approach to the Analysis of Designed Experiments. Journal of the American Statistical Association. 1994; 89:309–319. 1994.
51. Laud PW, Ibrahim JG. Predictive Model Selection. Journal of the Royal Statistical Society, Series B. 1995; 57:247–262.
52. Bedrick EJ, Christensen R, Johnson W. A new perspective on priors for generalized linear models EJ Bedrick, R Christensen, W Johnson. Journal of the American Statistical Association. 1996; 91:1450–1460.
53. Chen MH, Ibrahim JG. Conjugate Priors for Generalized Linear Models. Statistica Sinica. 2003; 13:461–476.
54. Ibrahim JG. On Properties of Predictive Priors in Linear Models. The American Statistician. 1997; 51:333–337.
55. Laud PW, Ibrahim JG. Predictive Specification of Prior Model Probabilities in Variable Selection. Biometrika. 1996; 83:267–274.
56. Matteucci M, Veldkamp BP. Bayesian Estimation of IRT Models with Power Priors. Advances in Latent Variables. In press.
57. Neelon B, O'Malley AJ. Bayesian Analysis Using Power Priors with Application to Pediatric Quality of Care. Journal of Biometrics and Biostatistics. 2010; 1:103.
58. Spiegelhalter DJ, Best NG, Carlin BP, Linde AVD. Bayesian Measures of Model Complexity and Fit. Journal of Royal Statistical Society, Series B. 2002; 64:583–639.
59. Geisser, S. Predictive Inference: An Introduction. Chapman Hall; New York: 1993.
60. Gelfand AE, Dey DK, Chang H. Model Determinating Using Predictive Distributions with Implementation Via Sampling-based Methods (with Discussion). Bayesian Statistics. 1992; 4:147–167.
61. Ibrahim JG, Chen MH, Sinha D. Bayesian Semi-parametric Models for Survival Data with A Cure Fraction. Biometrics. 2001; 57:383–388. [PubMed: 11414560]

62. Shao K. A Comparison of Three Methods for Integrating Historical Information for Bayesian Model Averaged Benchmark Dose Estimation. Environmental Toxicology and Pharmacology. 2012; 34:288–296. [PubMed: 22647377]

63. Ibrahim JG, Chen MH, Chu H. Bayesian Methods in Clinical Trials: A Bayesian Analysis of ECOG Trials E1684 and E1690. BMC Medical Methodology. 2012; 12:1–12.

64. Kirkwood JM, Strawderman MH, Ernstoff MS, Smith TJ, Borden EC, Blum RH. Interferon alfa-2b Adjuvant Therapy of Highrisk Resected Cutaneous Melanoma: The Eastern Cooperative Oncology Group Trial EST 1684. Journal of Clinical Oncology. 1996; 14:7–17. [PubMed: 8558223]

65. Kirkwood JM, Ibrahim JG, Sondak VK, Richards J, Flaherty LE, Ernstoff MS, Smith TJ, Rao U, Steele M, Blum RH. The Role of High- and Low-dose Interferon Alfa-2b in High-risk Melanoma: First Analysis of Intergroup Trial E1690/S9111/C9190. Journal of Clinical Oncology. 2000; 18:2444–2458. [PubMed: 10856105]

66. Stone GW, Ellis SG, Cannon L, Mann JT, Greenberg JD, O'Shaughnessy CD, DeMaio S, Hall P, Popma JJ, Koglin J, Russell ME, for the TAXUS V Investigators. Comparison of a Polymer-based Paclitaxel-Eluting Stent with a Bare Metal Stent in Patients with Complex Coronary Artery Disease: A randomized controlled trial. Journal of the American Medical Association. 2005; 294:1215–1223. [PubMed: 16160130]

67. Stone GW, Ellis SG, Cox DA, Hermiller J, O'Shaughnessy CD, Mann JT, Turco M, Caputo R, Bergin P, Greenberg JD, Popma JJ, Russell ME, for the TAXUS-IV Investigators. A Polymer-based, Paclitaxel-Eluting Stent in Patients with Coronary Artery Disease. The New England Journal of Medicine. 2004; 350:221–231. [PubMed: 14724301]

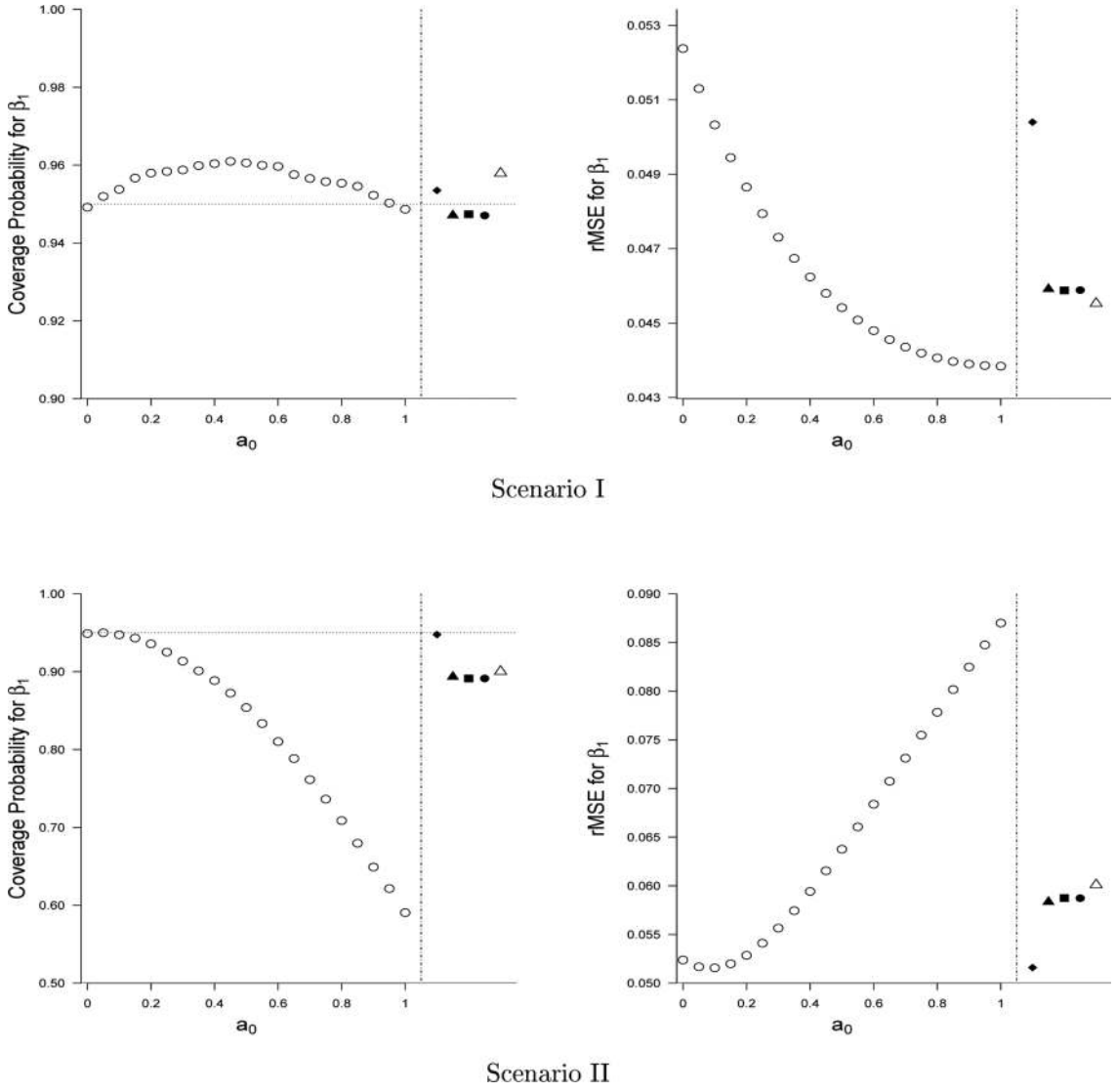
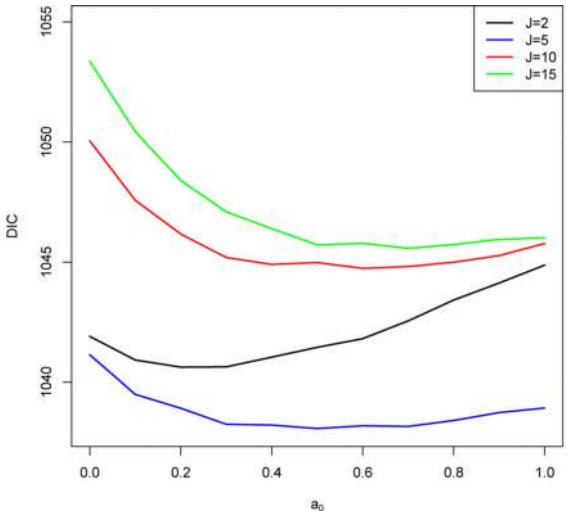


Figure 1.

Plots of coverage probability (left panel) and rMSE (right panel) of β_1, where ○ indicates the results based on fixed values of a_0 (evenly-spaced from 0 to 1), and ◆, ▲, ■, ●, and Δ display the results based on $a_{0,PLC}^{opt}$, $a_{0,MLC}^{opt}$, $a_{0,DIC}^{opt}$, and $a_{0,LPML}^{opt}$, and the normalized power prior.



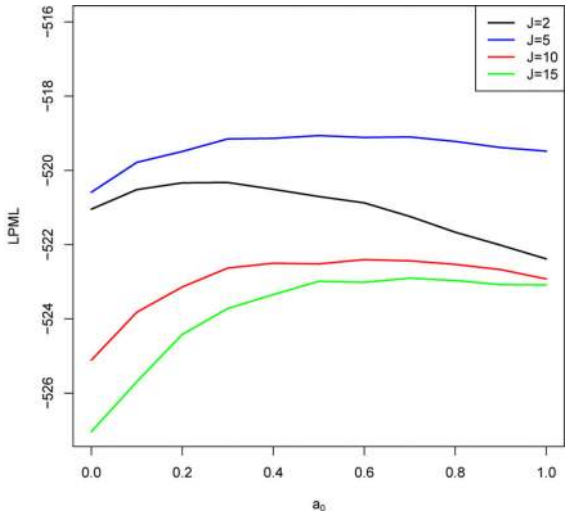


Figure 2.
DIC and LPML plots for E1684 and E1690.

Table 1

Simulation Results for the Posterior Estimates of β_1

a_0	Scenario I				Scenario II			
	Estimate	**SD**	**CP**	**rMSE**	**Estimate**	**SD**	**CP**	**rMSE**
0.00	1.9989	0.0521	0.9492	0.0523	1.9989	0.0521	0.9492	0.0523
0.10	1.9989	0.0510	0.9538	0.0503	1.9887	0.0510	0.9476	0.0515
0.20	1.9990	0.0499	0.9580	0.0486	1.9793	0.0499	0.9359	0.0528
0.30	1.9990	0.0489	0.9588	0.0473	1.9707	0.0489	0.9137	0.0556
0.40	1.9990	0.0480	0.9604	0.0462	1.9627	0.0480	0.8887	0.0594
0.50	1.9990	0.0472	0.9606	0.0454	1.9552	0.0472	0.8540	0.0637
0.60	1.9991	0.0463	0.9597	0.0448	1.9483	0.0463	0.8103	0.0683
0.70	1.9991	0.0456	0.9566	0.0443	1.9419	0.0456	0.7615	0.0731
0.80	1.9991	0.0448	0.9554	0.0440	1.9358	0.0448	0.7089	0.0778
0.90	1.9991	0.0441	0.9523	0.0439	1.9302	0.0441	0.6492	0.0824
1.00	1.9991	0.0435	0.9487	0.0438	1.9249	0.0435	0.5906	0.0870
$a_{0,\mathrm{PLC}}^{\mathrm{opt}}$	1.9989	0.0510	0.9535	0.0503	1.9891	0.0510	0.9476	0.0515
$a_{0,\mathrm{MLC}}^{\mathrm{opt}}$	1.9992	0.0448	0.9471	0.0459	1.9794	0.0492	0.8935	0.0583
$a_{0,\mathrm{DIC}}^{\mathrm{opt}}$	1.9992	0.0448	0.9474	0.0458	1.9783	0.0491	0.8913	0.0587
$a_{0,\mathrm{LPML}}^{\mathrm{opt}}$	1.9992	0.0448	0.9471	0.0458	1.9784	0.0491	0.8915	0.0587
Random	1.9991	0.0472	0.9579	0.0455	1.9687	0.0512	0.9003	0.0601

Table 2

Posterior Estimates for E1690 using E1684 as the historical data (J=5)

a_0	Parameter	Posterior Mean	Posterior SD	95% HPD Interval
0.0	β_0	0.204	0.127	(−0.043, 0.451)
	β_1	−0.311	0.163	(−0.620, 0.012)
	β_2	0.119	0.067	(−0.008, 0.252)
	β_3	−0.290	0.195	(−0.675, 0.091)
	β_4	0.290	0.274	(−0.263, 0.805)
	$\beta_1 + \beta_4$	−0.022	0.223	(−0.447, 0.416)
0.5	β_0	0.260	0.101	(0.056, 0.450)
	β_1	−0.340	0.138	(−0.616, −0.078)
	β_2	0.094	0.055	(−0.010, 0.204)
	β_3	−0.219	0.161	(−0.536, 0.091)
	β_4	0.233	0.230	(−0.209, 0.682)
	$\beta_1 + \beta_4$	−0.108	0.183	(−0.461, 0.259)
1.0	β_0	0.286	0.088	(0.110, 0.454)
	β_1	−0.352	0.120	(−0.593, −0.126)
	β_2	0.086	0.048	(−0.013, 0.175)
	β_3	−0.182	0.137	(−0.461, 0.078)
	β_4	0.205	0.197	(−0.179, 0.599)
	$\beta_1 + \beta_4$	−0.147	0.159	(−0.442, 0.180)

Table 3

Powers and Type I Errors for 12-Month TLF with $p_c^*=9.2\%$

Total Sample Size		1000	1080	1200	1280	1480
n_t		750	810	900	960	1110
n_c		250	270	300	320	370
No Borrowing $\boldsymbol{a}_0 = (0, 0)$	Power	0.648	0.676	0.718	0.738	0.800
	Type I Error	0.049	0.048	0.048	0.050	0.044
Random $\boldsymbol{a}_0$ using (8.4)	Power	0.843	0.878	0.897	0.902	0.914
	Type I Error	0.038	0.031	0.029	0.036	0.039
Fixed $\boldsymbol{a}_0 = (0.3, 0.3)$	Power	0.840	0.856	0.884	0.892	0.923
	Type I Error	0.030	0.027	0.028	0.030	0.032

Table 4

Powers and Type I Errors under Three p_c^*'s and Various γ's for 12-Month TLF with $(n_t, n_c) = (900, 300)$

Fitting Prior	γ	$p_c^* = 8.0\ \%$		$p_c^* = 9.2\ \%$		$p_c^* = 10.0\ \%$	
		Power	Type I Error	Power	Type I Error	Power	Type I Error
Power Prior with $a_{0k} \sim$ beta(b_{01}, b_{02}) in (8.4)							
$(b_{01}, b_{02}) = (1, 1)$	0.95	0.945	0.070	0.882	0.039	0.799	0.034
$(b_{01}, b_{02}) = (1, 5)$	0.95	0.916	0.061	0.832	0.033	0.760	0.026
$(b_{01}, b_{02}) = (1, 10)$	0.95	0.868	0.053	0.791	0.038	0.728	0.032
$(b_{01}, b_{02}) = (1, 1)$	0.96	0.935	0.055	0.880	0.022	0.765	0.026
	0.97	0.917	0.041	0.848	0.015	0.719	0.009
$(b_{01}, b_{02}) = (1, 5)$	0.96	0.899	0.047	0.803	0.027	0.722	0.021
Power Prior with fixed $\boldsymbol{a}_0 = (a_{01}, a_{02})'$							
$\boldsymbol{a}_0 = (0.3, 0.3)$	0.95	0.965	0.065	0.884	0.028	0.788	0.018
	0.96	0.953	0.049	0.856	0.021	0.750	0.013
	0.97	0.940	0.035	0.820	0.015	0.703	0.008
$a_0 = a_{0,\mathrm{PLC}}^{\mathrm{opt}}$ in (8.7)	0.95	0.918	0.055	0.829	0.035	0.755	0.027
	0.96	0.898	0.045	0.798	0.026	0.722	0.021
	0.97	0.871	0.033	0.758	0.018	0.673	0.014
$a_0 = a_{0,\mathrm{MLC}}^{\mathrm{opt}}$ in (8.7)	0.97	0.900	0.072	0.851	0.046	0.780	0.028
	0.98	0.877	0.053	0.811	0.030	0.734	0.019
$a_0 = a_{0,\mathrm{DIC}}^{\mathrm{opt}}$ in (8.9)	0.97	0.912	0.071	0.854	0.041	0.788	0.027
	0.98	0.889	0.051	0.815	0.028	0.738	0.018