Journal of Data Science 23 (1), 31–37
January 2025
DOI: 10.6339/25-JDS1161D
Discussion

Discussion of "Power Priors for Leveraging Historical Data: Looking Back and Looking Forward"☆

Fang Chen[1,*]
[1]SAS Institute Inc., Cary, NC, USA

1 Introduction

First and foremost, I would like to thank the editor-in-chief for the invitation to comment on the power prior article written by Chen et al. (2025). Since its introduction 25 years ago, the power prior (Ibrahim and Chen, 2000) has become one of the most widely used techniques for constructing informative prior distributions that use existing data information. The ease with which the power prior translates data information to distributional information about the model parameters is one of the key advantages that have made it so successful. It therefore comes as no surprise that how to conduct analyses based on power priors and related feature enhancements is a frequently requested topic from our software users. It is safe to say that the specific needs and functionality of power priors and related applications directly influence the design, features, and development of many statistical software packages (Gong and Chen, 2023; Alt et al., 2024).

It is delightful to see this ongoing effort to improve and expand the scope of power prior, and the extension of the original formula to partial borrowing, borrowing-by-parts, partial borrowing-by-parts, and propensity-score-based power priors. Because I am in the field of software development, I will focus my remarks on the practical and computational aspects of power priors. Many of my discussion points are based on a data analysis example that uses a simulated data set. The authors provided this simulated data set, which is similar to the proprietary ADNI data set that they used in the paper. I will start by providing details of the analysis before moving on to the discussion.

2 Data Analysis

The simulated data set is similar to the ADNI data set, and it consists of a historical portion and a current portion. In Figure 1 the graph at left shows an overlay of data from the two data subsets. The response variable is `y`, and there are nine covariates (`x1 - x8` and `z`), four of which are continuous and five are discrete. The overlaid density plots show that the covariates are similar to each other, at least marginally.

The graph at right contrasts two posterior distributions from fitting the historical data set and the current data set independently, using noninformative priors. There are twelve parameters in the model, eleven of which are regression coefficients (β_0 to β_8 and β_z, including the two regressors for the categorical variable `x4`) and the variance parameter σ^2. The red dotted lines represent the posterior from the historical data set, and the blue dashed lines represent the posterior of the current data set. With the exception of β_5, most parameters have distributions that are quite different from each other.

☆Main article: https://doi.org/10.6339/24-JDS1161.
* Email: fangk.chen@sas.com.

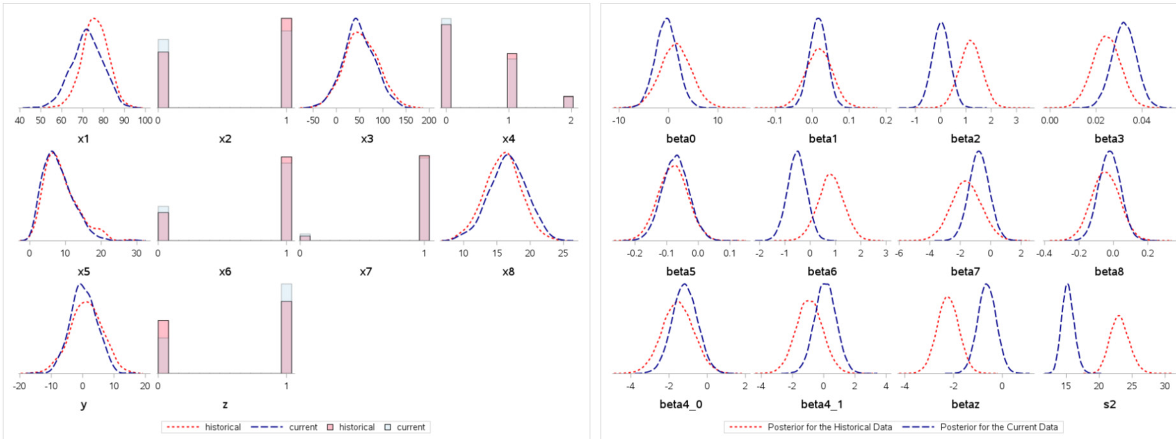

Figure 1: Density comparison of the covariates in the historical and current data sets (left graph); Posteriors of parameters from two independent models, historical data (red dotted) and current data (blue dashed), respectively (right graph).

I identified the optimal a_0^{opt} value by using deviance information criterion (DIC; Ibrahim et al., 2015). Figure 2 plots gridded a_0 values against corresponding DIC values. The minimum value occurs around 0.1, indicating less preference in borrowing from the historical data set. This could be attributed to the large differences between the fitted models, despite the similarity in the marginal distributions of the data.

How to interpret a_0 is another question that users frequently ask. It is relevant in the variations of the power priors that are presented in the paper and thus warrants treatment here. If the data are i.i.d., it is reasonable to interpret the discounting parameter as the percentage of information that is contained in the historical data. For example, setting a_0 to 0.4 is roughly equivalent to borrowing 40% of the information from the historical data set. This can

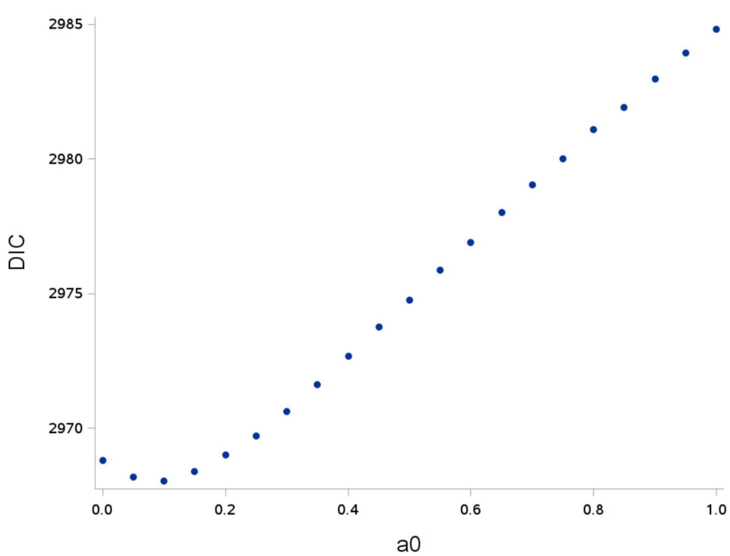

Figure 2: DIC values for different a_0 values.

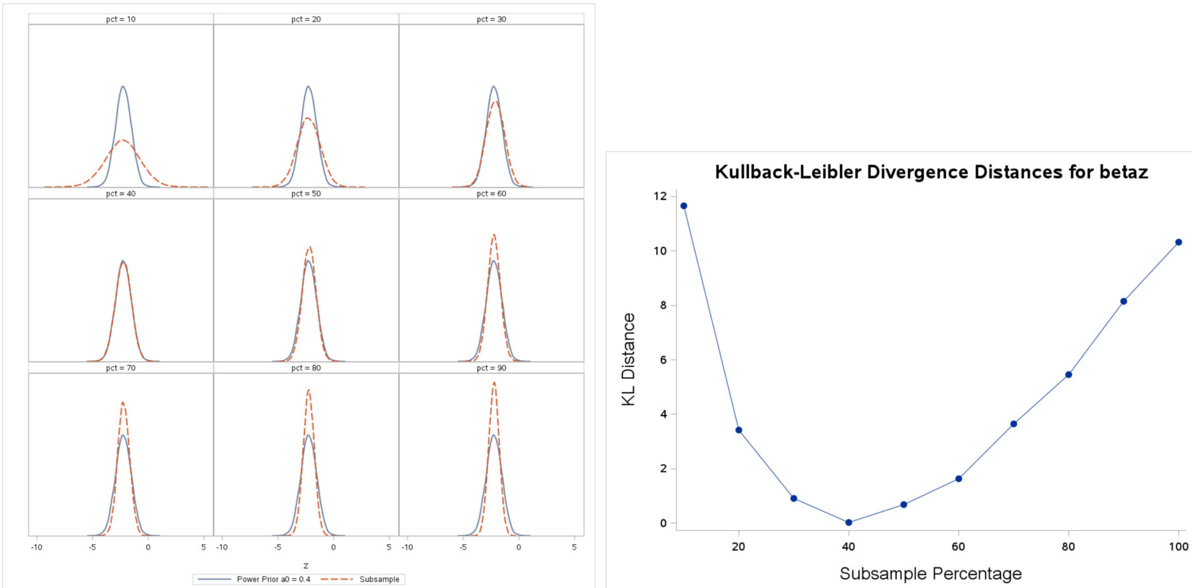

Figure 3: Left graph: density comparison of the power prior for β_z ($a_0 = 0.4$, blue solid) and the distributions of β_z using subsampled historical data of different proportions (red dashed, 200 repeats); Right graph: Kullback-Leibler divergence distance between the two densities in each panel of the left graph. The x-axis represents the subsample percentage used in the simulation.

also be illustrated by a simple simulation: subsample the historical data set according to different percentages ($10\%, 20\%, \ldots, 90\%$), fit regression model on reduced-size data sets, compare subsample densities (averaged over 200 repeats) with that of the power prior that uses a desired a_0, and find the closest match.

The plot at the left of Figure 3 displays such a visual comparison. The blue line is the marginal prior distribution of β_z with $a_0 = 0.4$. This curve is the same in each of the panels. The red lines represent the marginal distributions of β_z using bootstrap data of different proportions, from 10% to 90%. Visually, it is clear that the 40% bootstrap data provide the closest and, in fact, an almost identical fit to the power prior distribution of β_z with $a_0 = 0.4$. The Kullback-Leibler divergence distances between the two distributions are plotted in the graph at right in Figure 3, leading to the same conclusion. Comparisons using other parameters are identical and not repeated here.

3 Partial and Borrowing-by-Parts Power Priors

The partial-borrowing power prior is useful, and I envision this to be an area of development with much practical potential. In complex models that involve clustered data, it could be unrealistic to assume that cluster-specific random variables (e.g., random-effects parameters) from two data sets arise from the same distributions and are exchangeable. The partial-borrowing power prior offers a natural way to limit the impact of nuisance parameters or latent variables through integration.

Computationally speaking, the partial power prior is easy to implement and can be implemented for many different types of models without requiring analytical derivations. One can

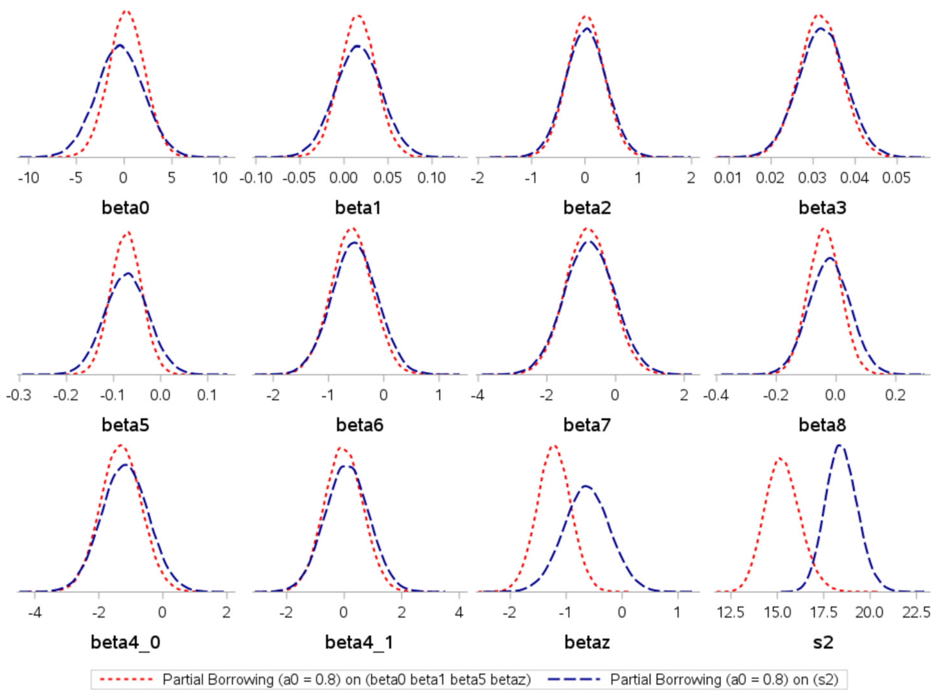

Figure 4: Comparison of two partial Power Prior analyses. The red dotted lines are posterior distributions using pPP with $a_0 = 0.8$ on (β_0, β_1, β_5, and β_z); the blue dashed lines are posterior distributions using pPP with the same a_0 value on only the variance parameter.

fit a power prior with a fixed a_0 by using only the historical data set, keep MCMC samples of the desired parameters, use parametric distributions to approximate the marginal samples (e.g., multivariate normal to β, and inverse gamma or inverse Wishart for the scaled variance or covariance parameters), and use the approximated distributions for the current-stage analysis. For large samples, the partial borrowing power prior can be approximated accurately.

I see a few practical issues and would like to seek the authors' input on how to address them. The first issue relates to interpretation. A partial power prior with a fixed a_0 value is unlikely to be equivalent to a power prior (i.e., to borrow the same amount of information from the historical data) with the same a_0 value. If we can make a link between the value of a power prior a_0 and the percentage of information used from the historical data set, what can we say about the partial power prior? How can we tell if using an a_0 value in a partial power prior is borrowing too much, or too little, if an intuitive understanding of the a_0 is not necessarily accurate? Do the authors see merit in using simulation methods to establish such a link between the a_0 and the percentage of information from the historical data for the purposes of interpretation?

The second issue is how to better understand the impact of the partial power prior on the non-borrowing parameters. As indicated by the authors in Section 6 of the paper, partial borrowing on different parameters has an uneven impact on other parameter estimates. For example, in the NTP analysis, pPP on β_0 not only impacts the mean estimate of β_0 but also impact that of the β_1 parameter; the reverse—that pPP on β_1 barely changes the mean estimates of either parameters—is not true. This uneven impact is also empirically observed in this data analysis.

Figure 4 shows the overlay of two analyses: the red dotted lines are posterior densities that use the partial power prior with $a_0 = 0.8$ on a subset of the regression coefficients (β_0, β_1,

β_5, and β_z); the blue dashed lines are posterior densities that use the partial power prior with the same a_0 value but only on the variance parameter. In comparison, we can see pronounced differences between these five parameters. This is expected, because we borrowed heavily from the historical data set on these parameters and they should be different. Weak and equal impact on some parameters (β_2, β_3, β_6, and β_7) indicates relatively weak borrowing in these dimensions. This is also expected, because we used non-informative priors on these parameters and would expect similar, if not equal, effect. But we can clearly see the impact of the partial power priors on β_8 and the two β_4 parameters, indicating that partial borrowing could influence the analysis in ways that are not obvious. This could be the result of the integration step that carries some information from other parameters to the current-stage analysis. However, using two different ways to integrate leads to different impacts on the non-borrowing parameters can be perplexing, because they appear to bring different amounts of information to the current stage analysis. Is there a way to better understand how different partial borrowing could retain or strip information on other parameters and to evaluate this potentially hidden impact on the analysis?

The borrowing-by-parts power prior is the answer to quest to separate information. That is, if we want to separate the informativeness and noninformativeness in a power prior, such as by borrowing on a subset of parameters and not borrowing on others, we can assign positive values of a_0 (informative) on a subset of parameters and 0 (noninformative) on the other. Is this equivalent to fitting a power prior or a partial power prior on a submodel by removing all the non-borrowing parameters and keeping only the parameters of interest?

The authors illustrated how to construct a borrowing-by-parts power prior in a binary model and a regression model, but it feels as if this could be a technically challenging prior to construct in a general setting. How can we borrow more information on a subset of regression parameters while discounting others? A general question I want to pose to the authors is this: when should one consider using borrowing-by-parts power priors or the partial version of the prior, taking into account the computational burdens they could entail?

4 Propensity-Score-Based Power Priors

The power prior based on propensity score matching offers an automatic way of selecting a_0, and it is a computationally convenient prior to construct, with many out-of-the-box software packages, such as PROC PSMATCH in SAS (SAS Institute Inc, 2025) and the R-package MatchIt (Ho et al., 2011), providing capability for fitting propensity score models. I implemented the inverse probability of treatment weighting based power prior (iptwPP) on the simulated data set and got the following results, which I'd like to see the authors comment on.

The graph at left in Figure 5 shows a density plot of the $a_{0i} = e(x_{0i}) \cdot w(x_{0i})$ values (defined in Section 4.4 of Chen et al., 2025) used in the iptwPP, where i is the observation index for the historical data set. The values of a_{0i} range between 0 and 1. Most of the iptwPP values are close to 0, which leads to a small amount of borrowing on an observational level. The mean of the iptwPP a_{0i} values is 0.07, with a standard deviation of 0.1.

The iptwPP-based posterior distributions are displayed in the graph at right in Figure 5, using the dotted red lines. Interestingly, and also surprisingly, the posterior distributions look very similar to that of the power prior using $a_0 = 0.1$, the optimal a_0 based on DIC.

Although this is based on only one analysis and therefore could be coincidental, the fact that two different approaches to finding weight parameters lead to the same results is an interesting result in itself and perhaps worth exploring further. The ADNI data analysis example in Section 6

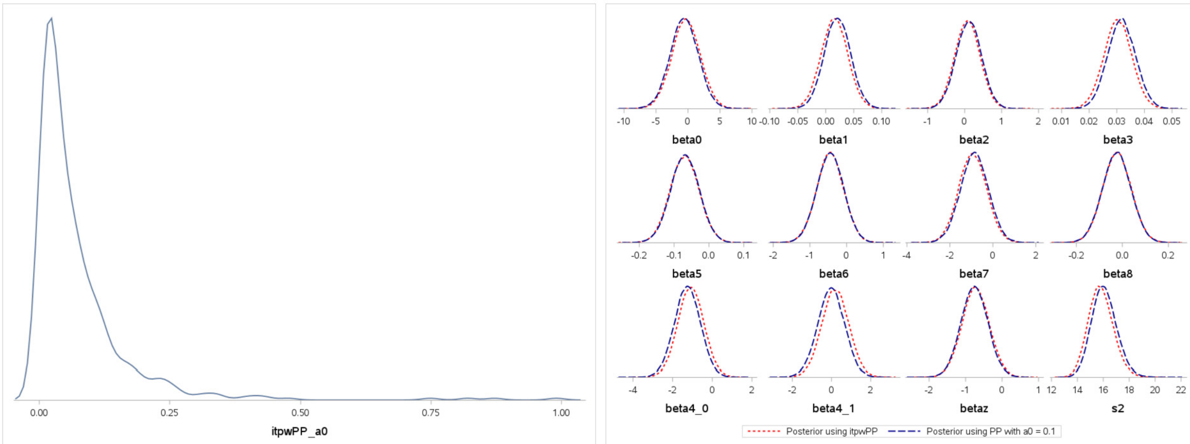

Figure 5: Density of iptwPP a_0 values (left); Posterior density comparison between the iptwPP-based approach and the PP approach with $a_0^{\text{opt}} = 0.1$ (right).

includes the iptwPP approach but is shown without much detail, and I wonder if the authors had a similar experience in using the iptwPP. Both approaches are procedure-oriented and data/model-driven, highly valued characteristics in the applied world, because they circumvent the need for the practitioner to defend a specific choice of a_0 values. If the iptwPP can lead to results that have optimal properties in model fitting, it can further underscore the usefulness of the approach. On the other hand, this agreement provides interpretability to the deviance-based search approach and gives us more confidence in making a claim that an optimal value of a_0 indicates average similarity between the two data sets.

There are some technical concerns that I hope the authors would be able to address as well. How important is the validity of the propensity model in constructing the PS-based power prior? How sensitive is the iptwPP to propensity score model misspecification? In addition, the PS-based method requires a technical adjustment: the stabilized weight discounting a_{0i} values are not guaranteed to be between 0 and 1 and need to be transformed. In this analysis, adjustment was done using

$$\frac{a_{0i} - \min(\mathbf{a_0})}{\max(\mathbf{a_0}) - \min(\mathbf{a_0})}$$

where $\mathbf{a_0}$ was the vector of all weight-adjusted a_{0i} values, $\min(\mathbf{a_0})$ was set to 0, and $\max(\mathbf{a_0})$ was set to slightly higher than the maximum computed a_{0i} value. Which value of maximum is chosen can dramatically change the a_{0i} values, and by extension the analysis. What is the right value to use? It could depend on the true max of stabilized weight discounting values, not just the observed max. That consideration can pose another layer of difficulty.

To end the discussion, I want to congratulate the authors for continuing to improve the power priors and for expanding our understanding and knowledge in this area. I thank the editor-in-chief and the authors again for inviting me to participate in this conversation on the most recent innovations in this wonderful Bayesian tool.

Acknowledgement

The author would like to express gratitude to Jun Yan for providing valuable suggestions that improved the quality of the manuscript and thank Ed Huddleston for editing the paper.

References

Alt EM, Chen X, Carvalho LM, Ibrahim JG (2024). *hdbayes: Bayesian Analysis of Generalized Linear Models with Historical Data*. R package version 0.1.1. https://CRAN.R-project.org/package=hdbayes

Chen MH, Guan Z, Lin M, Sun M (2025). Power priors for leveraging historical data: Looking back and looking forward. *Journal of Data Science*, 23(1): 1–30. https://doi.org/10.6339/24-JDS1161

Gong Y, Chen F (2023). Bayesian models with the power prior using PROC BGLIMM. *SAS Statistics Research and Applications Paper 2023-03*. SAS Institute Inc., Cary, NC. https://support.sas.com/content/dam/SAS/support/en/technical-papers/bayesian-models-with-the-power-prior-using-proc-bglimm.pdf

Ho DE, Imai K, King G, Stuart EA (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8): 1–28. https://doi.org/10.18637/jss.v042.i08

Ibrahim JG, Chen MH (2000). Power prior distributions for regression models. *Statistical Science*, 15(1): 46–60. https://doi.org/10.1214/ss/1009212673

Ibrahim JG, Chen MH, Gwon Y, Chen F (2015). The power prior: Theory and applications. *Statistical Science*, 34(28): 3724–3749.

SAS Institute Inc (2025). *SAS/STAT User's Guide*. SAS Institute Inc., Cary, NC. Revised January 2025. https://documentation.sas.com/doc/en/pgmsascdc/v_059/statug/titlepage.htm