HHS Public Access

Author manuscript

J Am Stat Assoc. Author manuscript; available in PMC 2022 January 01.

Published in final edited form as:

J Am Stat Assoc. 2021 ; 116(535): 1521–1532. doi:10.1080/01621459.2020.1745813.

Bayesian Factor Analysis for Inference on Interactions

Federico Ferrari, **David B. Dunson**

Department of Statistical Science, Duke University

Abstract

This article is motivated by the problem of inference on interactions among chemical exposures impacting human health outcomes. Chemicals often co-occur in the environment or in synthetic mixtures and as a result exposure levels can be highly correlated. We propose a latent factor joint model, which includes shared factors in both the predictor and response components while assuming conditional independence. By including a quadratic regression in the latent variables in the response component, we induce flexible dimension reduction in characterizing main effects and interactions. We propose a Bayesian approach to inference under this Factor analysis for INteractions (FIN) framework. Through appropriate modifications of the factor modeling structure, FIN can accommodate higher order interactions. We evaluate the performance using a simulation study and data from the National Health and Nutrition Examination Survey (NHANES). Code is available on GitHub.

Keywords

Bayesian Modeling; Chemical Mixtures; Correlated Exposures; Quadratic regression; Statistical Interactions

1 Introduction

There is broad interest in incorporating interactions in linear regression. Extensions of linear regression to accommodate pairwise interactions are commonly referred to as quadratic regression. In moderate to high-dimensional settings, it becomes very challenging to implement quadratic regression since the number of parameters to be estimated is $2p + \binom{p}{2}$. Hence, classical methods such as least squares cannot be used and even common penalization and Bayesian methods can encounter computational hurdles. Reliable inferences on main effects and interactions is even more challenging when certain predictors are moderately to highly correlated.

A lot of effort has been focused on estimating pairwise interactions in moderate high-dimensional and ultra high-dimensional problems. We refer to the former when the number of covariates is between 20 and 100 and to the latter when $p > 100$. When $p = 100$, the number of parameters to be estimated is greater than 5000. When $p \in [20, 100]$, one-stage regularization methods like Bien et al. (2013) and Haris et al. (2016) can be successful. Some of these methods require a so-called heredity assumption (Chipman, 1996) to reduce dimensionality. Strong heredity means that the interaction between two variables is included in the model only if both main effects are. For weak heredity it suffices to have one main

effect in the model. Heredity reduces the number of models from $2^p + \binom{p}{2}$ to $\sum_{i=0}^{p} \binom{p}{i} 2^{\binom{i}{2}}$ or $\sum_{i=0}^{p} \binom{p}{i} 2^{pi - i(i+1)/2}$ for strong or weak heredity, respectively (Chipman, 1996). For ultra high-dimensional problems, two stage-approaches have been developed, see Hao et al. (2018) and Wang et al. (2019). However, these methods do not report uncertainties in model selection and parameter estimation, and rely on strong sparsity assumptions.

We are particularly motivated by studies of environmental health collecting data on mixtures of chemical exposures. These exposures can be moderately high-dimensional with high correlations within blocks of variables; for example, this can arise when an individual is exposed to a product having a mixture of chemicals and when chemical measurements consist of metabolites or breakdown products of a parent compound. There is a large public heath interest in studying E×E, E×G and G×G interactions, with E = environmental exposures and G = genetic factors. However, current methods for quadratic regression are not ideal in these applications due to the level of correlation in the predictors, the fact that strong sparsity assumptions are not appropriate, and the need for uncertainty quantification. Regarding the issue of sparsity, some exposures are breakdown products of the same compound, so it is unlikely that only one exposure has an effect on the outcome. Also, it is statistically challenging to tell apart highly correlated covariates with limited data. For this reason, it is appealing given the data structure to select blocks of correlated exposures together instead of arbitrarily selecting one chemical in a group.

To address these problems, one possibility is to use a Bayesian approach to inference in order to include prior information to reduce dimensionality while characterizing uncertainty through the posterior distribution. There is an immense literature on Bayesian methods for high-dimensional linear regression, including recent algorithms that can scale up to thousands of predictors (Bondell and Reich, 2012), (Rossell and Telesca, 2017), (Johndrow et al., 2017), (Nishimura and Suchard, 2018). In addition some articles have explicitly focused on quadratic regression and interaction detection (Zhang and Liu, 2007), (Cordell, 2009), (Mackay, 2014). Bayes variable selection and shrinkage approaches will tend to have problems when predictors are highly correlated; this has motivated a literature on Bayesian latent factor regression (Lucas et al., 2006), (Carvalho et al., 2008).

Latent factor regression incorporates shared latent variables in the predictor and response components. This provides dimensionality reduction in modeling of the covariance structure in the predictors and characterizing the impact of correlated groups of predictors on the response. Such approaches are closely related to principal components regression, but it tends to be easier to simultaneously incorporate shrinkage and uncertainty quantification within the Bayesian framework. In addition, within the Bayes latent factor regression paradigm, typical identifiability constraints such as orthogonality are not needed (see, for example Bhattacharya and Dunson (2011)). The main contribution of this article is to generalize Bayesian latent factor regression to accommodate interactions using an approach inspired by Wang et al. (2019). This is accomplished by including pairwise interactions in the latent variables in the response component. We refer to the resulting framework as Factor analysis for INteractions (FIN). There is a rich literature on quadratic and non-linear latent

variable modeling, largely in psychometrics (refer, for example, to Arminger and Muthén (1998)). However, to our knowledge, such approaches have not been used for inferences on interactions in regression problems.

In Section 2 we describe the proposed FIN framework, including extensions for higher order interactions. In Section 3 we provide theory on model misspecification and consistency. Section 4 contains a simulation study. Section 5 illustrates the methods on NHANES data. Code is available at https://github.com/fedfer/factor_interactions. Proofs of *Proposition 2* and *Proposition 3* are included in the Supplementary Material.

2 Model

2.1 Model and Properties

Let y_i denote a continuous health response for individual i, and $X_i = (x_{i1}, \cdots, x_{ip})^T$ denote a vector of exposure measurements. We propose a latent factor joint model, which includes shared factors in both the predictor and response components while assuming conditional independence. We include interactions among latent variables in the response component. We also assume that, given the latent variables, the explanatory variables and the response are continuous and normally distributed. We assume that the data have been normalized prior to the analysis so that we omit the intercept. The model is as follows:

$$\begin{aligned} y_i &= \eta_i^T \omega + \eta_i^T \Omega \eta_i + \epsilon_{y,i}, \quad \epsilon_{y,i} \sim N(0, \sigma^2), \\ X_i &= \Lambda \eta_i + \epsilon_i, \quad \epsilon_i \sim N_p(0, \Psi), \\ \eta_i &\sim N_k(0, I), \end{aligned} \tag{1}$$

where $\Psi = diag(\sigma_1^2, \cdots, \sigma_p^2)$. In a Bayesian fashion, we assume a prior for the parameters $\Theta = (\omega, \Omega, \Lambda, \Psi, \sigma^2)$ that will be specified in Section 2.2. Model (1) is equivalent to classical latent factor regression models; refer, for example, to West (2003), except for the $\eta_i^T \Omega \eta_i$ term. Here, Ω is a $k \times k$ symmetric matrix inducing a quadratic latent variable regression that characterizes interactions among the latent variables.

The above formulation can be shown to induce a quadratic regression of y on X. To build intuition consider the case in which $\sigma_j^2 = 0$ as done in West (2003) for the special case in which $\Omega = 0$. *The many-to-one map* $X_i = \Lambda \eta_i$ has multiple generalized inverses $\eta_i = \Lambda^T X_i + b$ such that $\Lambda b = 0$. If we substitute in the regression equation, we obtain

$$\begin{aligned} \mathbb{E}(y_i \mid X_i) &= (\Lambda^T X_i + b)^T \omega + (\Lambda^T X_i + b)^T \Omega (\Lambda^T X_i + b) = \\ &= X_i^T \Lambda \omega + X_i^T \Lambda \Omega \Lambda^T X_i + g(b) \end{aligned}$$

The following proposition gives a similar result in the non deterministic case:

Proposition 1. *Under model (1), the following are true:*

- **i.** $\mathbb{E}(y_i \mid X_i) = tr(\Omega V) + (\omega^T A) X_i + X_i^T (A^T \Omega A) X_i$,
- **ii.** $Cov(y_i, X_i) = \Lambda \omega$,

where $V = (\Lambda^T\Psi^{-1}\Lambda + I)^{-1}$ *and* $A = V\Lambda^T\Psi^{-1} = (\Lambda^T\Psi^{-1}\Lambda + I)^{-1}\Lambda^T\Psi^{-1}$.

This shows that the induced regression of y on X from model (1) is indeed a quadratic regression. Let us define the induced main effects as $\beta_X = A^T\omega$ and the matrix containing the first order interactions as $\Omega_X = A^T\Omega A$. Notice that we could define Ω as a diagonal matrix and we would still estimate pairwise interactions between the regressors, further details are given in Sections 2.3 and 2.4.

In epidemiology studies, it is of interest to include interactions between chemical exposures and demographic covariates. The covariates are often binary variables, like *race* or *sex*, or continuous variables that are non-normally distributed, like *age*. Hence, we do not want to assume a latent normal structure for the covariates. Letting $Z_i = (z_{i1}, \cdots, z_{iq})^T$ be a vector of covariates, we modify model (1) to include a main effect for Z_i and an interaction term between Z_i and the latent factor η_i:

$$\begin{aligned} y_i &= \eta_i^T\omega + \eta_i^T\Omega\eta_i + Z_i^T\alpha + \eta_i^T\Delta Z_i + \epsilon_{y,i}, \quad \epsilon_{y,i} \sim N(0, \sigma^2), \\ X_i &= \Lambda\eta_i + \epsilon_i, \quad \epsilon_i \sim N_p(0, \Psi), \\ \eta_i &\sim N_k(0, I), \end{aligned} \tag{2}$$

where Δ is a $k \times q$ matrix of interaction coefficients between the latent variables and the covariates, and $\alpha = (\alpha_1, \cdots, \alpha_q)$ are main effects for the covariates. Following *Proposition 1* we have that

$$\mathbb{E}(\eta_i^T\Delta Z_i \mid X_i, Z_i) = \mathbb{E}(\eta_i^T \mid X_i)\Delta Z_i = X_i^T(A^T\Delta)Z_i,$$

where $(A^T\Delta)$ is a $p \times q$ matrix of pairwise interactions between exposures and covariates. In the sequel, we focus our development on model (1) for ease in exposition, but all of the details can be easily modified to pertain to model (2).

2.2 Priors and MCMC Algorithm

In this section we define the priors for $(\omega, \Omega, \Lambda, \Psi, \sigma^2)$, briefly describe the computational challenges given by model (1) and summarize our Markov Chain Monte Carlo sampler in Algorithm 1. We choose an Inverse-Gamma distribution with parameters $(\frac{1}{2}, \frac{1}{2})$ for σ^2 and σ_j^2 for $j = 1, \cdots, p$. The elements of ω and Ω are given independent Gaussian priors. For $\Lambda = \{\lambda_{i,j}\}$, a typical choice to attain identifiability requires $\lambda_{i,j} = 0$ for $j > i$ and $\lambda_{j,j} > 0$ for $j = 1, \cdots, k$ (Geweke and Zhou, 1996). However, some Bayesian applications, like covariance estimation (Bhattacharya and Dunson, 2011), do not require identifiability of Λ. The same holds for inference on induced main effects and interactions for model (1). Notice that model (1) is invariant to rotations:

$$\begin{aligned} y_i &= \eta_i^T PP^T\omega + \eta_i^T PP^T\Omega PP^T\eta_i + \epsilon_{y,i}, \quad \epsilon_{y,i} \sim N(0, \sigma^2), \\ X_i &= \Lambda PP^T\eta_i + \epsilon_i, \quad \epsilon_i \sim N_p(0, \Psi), \end{aligned}$$

where P is a $k \times k$ orthogonal matrix $P(PP^T = I)$. However, the induced main effects satisfy

$$\beta_X = \Psi^{-1}\Lambda P(P^T\Lambda^T\Psi^{-1}\Lambda P + P^T P)^{-1}P^T\omega = \Psi^{-1}\Lambda(\Lambda^T\Psi^{-1}\Lambda + I)^{-1}\omega\,.$$

The same holds for induced interactions, showing that we do not need to impose identifiability constraints on Λ. We choose the Dirichlet-Laplace (DL) prior of Bhattacharya et al. (2015) row-wise, corresponding to

$$\begin{aligned} \lambda_{j,h} \mid \phi_{jh}, \tau_j &\sim DE(\phi_{jh}\tau_j) \qquad h = 1, \cdots, k \\ \phi_j &\sim Dir(a, \cdots, a) \qquad \tau_j \sim Gamma(ka, 1/2), \end{aligned}$$

where $j = 1, \cdots, p$, $\boldsymbol{\phi}_j = (\boldsymbol{\phi}_{j1}, \cdots, \boldsymbol{\phi}_{jk})$, DE refers to the zero mean double-exponential or Laplace distribution, and k is an upper bound on the number of factors, as the prior allows effective deletion of redundant factor loadings through row-wise shrinkage. The DL prior provides flexible shrinkage on the factor loadings matrix, generalizing the Bayesian Lasso (Park and Casella, 2008) to have a carefully chosen hierarchical structure on exposure-specific ($\boldsymbol{\tau}_j$) and local ($\boldsymbol{\phi}_{jh}$) scales. This induces a prior with concentration at zero, to strongly shrink small signals, and heavy-tails, to avoid over-shrinking large signals. The DL prior induces near sparsity row-wise in the matrix Λ, as it is reasonable to assume that each variable loads on few factors.

In Section 2.4, we describe how the above prior specification induces an appealing shrinkage prior on the main effects and interactions, and discuss hyperparameter choice. In practice, we recommend the rule of thumb that chooses k such that $\frac{\sum_{j=1}^{k} v_j}{\sum_{j=1}^{p} v_j} > 0.9$, where v_j is the j^{th} largest singular value of the correlation matrix of X. *Proposition 2* in Section 3 provides theoretical justification for this criterion. As an alternative to row-wise shrinkage, we could have instead used column-wise shrinkage as advocated in Bhattacharya et al. (2015) and Legramanti et al. (2019). Although such approaches can be effective in choosing the number of factors, we found in our simulations that they can lead to over-shrinkage of the estimated main effects and interactions.

The inclusion of pairwise interactions among the factors in the regression of the outcome y_i rules out using a simple data augmentation Gibbs sampler, as in West (2003), Bhattacharya and Dunson (2011). The log full conditional distribution for η_i is:

$$\begin{aligned} &-\frac{1}{2}\left[\eta_i^T\left(\frac{\omega\omega^T}{\sigma_y^2} + \Lambda^T\Psi^{-1}\Lambda + I - 2\frac{\Omega Y_i}{\sigma_y^2}\right)\eta_i - 2\eta_i^T\left(\Lambda^T\Psi^{-1}X_i + \frac{\omega Y_i}{\sigma_y^2}\right)\right] - \\ &-\frac{1}{2}\left[\frac{2\eta_i^T\omega\eta_i^T\Omega\eta_i}{\sigma_y^2} + \frac{(\eta_i^T\Omega\eta_i)^2}{\sigma_y^2}\right] + C, \end{aligned}$$

where C is a normalizing constant. We update the factors η_i using the Metropolis-Adjusted Langevin Algorithm (MALA) (Grenander and Miller, 1994), (Roberts et al., 1996). Sampling the factors is the main computational bottleneck of our approach since we have to

update n vectors, each of dimension k. The overall MCMC algorithm and the MALA step are summarized in Algorithm 1.

2.3 Higher Order Interactions

FIN can be generalized to allow for higher order interactions. In particular, we can obtain estimates for the interaction coefficients up to the Q^{th} order with the following model:

$$\mathbb{E}(y_i \mid \eta_i) = \sum_{h=1}^{k} \omega_h^{(1)} \eta_{ih} + \sum_{h=1}^{k} \omega_h^{(2)} \eta_{ih}^2 + \cdots + \sum_{h=1}^{k} \omega_h^{(Q)} \eta_{ih}^Q, \tag{3}$$

which is a polynomial regression in the latent variables. We do not include interactions between the factors, so that the number of parameters to be estimated is Qk. When $Q = 2$, this model is equivalent to Ω being a diagonal matrix. Recall that $\eta_i | X_i \sim N(AX, V)$, where A and V are defined in *Proposition 1*. Since we do not include interactions among the factors, let us just focus on the marginal distribution of the j^{th} factor, i.e $\eta_{ih} \mid X_i \sim N(\mu_h, \sigma_h^2)$ where $\mu_h = \sum_{j=1}^{p} a_{hj} X_{ij}$ and $\sigma_h^2 = V_{hh}$. Below we provide an expression for $E(\sum_{q=1}^{Q} \omega_h^{(q)} \eta_{ih}^q \mid X)$, which can be calculated using non-central moments of a Normal distribution, see Winkelbauer (2012) for a reference.

$$\mathbb{E}\left(\sum_{q=1}^{Q} \omega_j^{(q)} \eta_j^q \mid X\right) = \sum_{f=1}^{\left\lfloor \frac{Q+1}{2} \right\rfloor} \sum_{q=f}^{\left\lfloor \frac{Q+1}{2} \right\rfloor} \omega_h^{(2q-1)} \sigma_h^{2q-2f} b_{qf}^o \sum_{k_+ = 2f-1} \binom{2f-1}{k_1 \cdots k_p} \prod_{j=1}^{p} (a_{hj} X_j)^{k_j} + \sum_{f=0}^{\left\lfloor \frac{Q+1}{2} \right\rfloor} \sum_{q=f \vee 1}^{\left\lfloor \frac{Q+1}{2} \right\rfloor} \omega_h^{(2q)} \sigma_h^{2q-2f} b_{qf}^e \sum_{k_+ = 2f} \binom{2f}{k_1 \cdots k_p} \prod_{j=1}^{p} (a_{hj} X_j)^{k_j},$$

where $b_{qf}^o = \frac{(2q-1)!}{(2f-1)!(q-f)!2^{q-f}}$, $b_{qf}^e = \frac{(2q)!}{(2f)!(q-f)!2^{q-f}}$ and $k_+ = \sum_{j=1}^{p} k_h$. We just need to sum up over the index h in (3) and we can read out the expressions for the intercept,

Algorithm 1 MCMC algorithm for sampling the parameters of model (1)

Step 1 Sample $\eta_i,\ i = 1, \cdots, n$ via Metropolis-Hastings using as a proposal distribution a $N(\eta_i + \frac{1}{2}\nabla_{\eta_i} log(\pi(\eta_i \mid -)), \epsilon I_k)$.

Step 2 Sample the main effects coefficients ω from a multivariate normal distribution:

$$\pi(\omega \mid -) \sim N\left((\frac{\eta^T\eta}{\sigma^2} + I_n / 100)^{-1}\eta(y - diag(\eta\Omega\eta)) / \sigma^2, (\frac{\eta^T\eta}{\sigma^2} + I_n / 100)^{-1}\right)$$

where η is the matrix with rows equal to η_i.

Step 3 Sample upper triangular part of Ω, namely Ω^{U}, from a multivariate normal distribution:

$$\pi(\Omega^{\mathrm{U}} \mid -) \sim N\left((\frac{\eta^{*T}\eta^*}{\sigma^2} + \frac{p(p+1)}{2})^{-1}\eta^*(y - \eta\omega)) / \sigma^2, (\frac{\eta^{*T}\eta^*}{\sigma^2} + I\frac{p(p+1)}{2} / 100)^{-1}\right)$$

where η^* is a matrix containing the pairwise interactions of among the columns of η. Then set $\Omega = \frac{\Omega + \Omega^T}{2}$

Step 4 Sample σ^{-2} from a Gamma distribution:

$$\pi(\sigma^{-2} \mid -) \sim Gamma\left(\frac{1+n}{2}, \frac{1}{2} + \frac{1}{2}(y - \eta\omega - diag(\eta\Omega\eta^T))^T(y - \eta\omega - diag(\eta\Omega\eta^T)\right)$$

Step 5 D λ_j the rows of Λ, for $j = 1, \cdots, p$. Sample p conditionally independent posteriors:

$$\pi(\lambda_j \mid -) \sim N\left((D_j^{-1} + \frac{\eta^T\eta}{\sigma_j^2})^{-1}\eta^T\sigma_j^{-2}X^{(j)}, (D_j^{-1} + \frac{\eta^T\eta}{\sigma_j^2})^{-1}\right)$$

where $X^{(j)}$ is the j^{th} column of the matrix X, $D_j = diag(\tau_j^2\psi_{j1}\phi_{j1}^2, \cdots, \tau_j^2\psi_{jk}\phi_{jk}^2)$.

Step 6 Sample $\widehat{\psi_{jh}}$ for $j = 1, \cdots, p$ and $h = 1, \cdots, k$ from independent Inverse Gaussin distribution:

$$\pi(\psi_{jh}) \sim InvGauss\left(\tau_j\phi_{jh} / \mid \lambda_{jh} \mid, 1\right) \text{ and set } \psi_{jh} = 1/\widehat{\psi_{jh}}$$

Step 7 Sample τ_j for $j = 1, \cdots, p$ from independent Generalized Inverse Gaussian distributions:

$$\pi(\tau_j \mid -) \sim GInvGauss\left(1 - k, 1, 2\sum_{h=1}^{k}\frac{\mid \lambda jh \mid}{\phi_{jh}}\right)$$

Step 8 In order to update ϕ_{jh} for $j = 1, \cdots, p$ and $h = 1 \cdots, k$, sample T_{jh} from independent Generalized Inverse Gaussian distributions:

$$\pi(T_{jh} \mid -) \sim GInvGauss\left(a - 1, 1, 2 \mid \lambda_{jh} \mid\right)$$

Then set $\phi_{jh} = \frac{T_{jh}}{\sum_{h=1}^{k} T_{jh}}$

Step 9 Sample σ_j^{-2} for $j = 1, \cdots, p$ from conditionally independent gamma distributions

$$\pi(\sigma_j^{-2} \mid -) \sim Gamma\left(\frac{1+n}{2}, \frac{1}{2} + \frac{1}{2}\sum_{i=1}^{n}(X_{ij} - \lambda_j^T\eta_i)\right)$$

main effects and interactions up to the Q^{th} order. In particular, we have that the intercept is equal to $\sum_{h=1}^{k}\sum_{q=1}^{\left\lceil\frac{Q-1}{2}\right\rceil}\omega_h^{(2q)}V_{hh}^{2q}b_{q0}^e$. When $Q = 2$ this reduces to $\sum_{h=1}^{k}\omega_h^{(2)}V_{hh}^2 = tr(\Omega V)$, where $\Omega = diag(\omega_1^{(2)}, \cdots, \omega_k^{(2)})$. The expression for the main effects coefficients on X_j is

$\sum_{h=1}^{k}\sum_{q=1}^{\left\lfloor\frac{Q+1}{2}\right\rfloor}\omega_h^{(2q-1)}V_{hh}^{2q-1}b_{q1}^{o}a_{hj}$. When $Q=2$ this becomes $\sum_{h=1}^{k}\omega_h^{(1)}a_{hj}$, hence $\beta_X = A^T\omega$. Similarly the expression for the interaction between X_j and X_l is equal to $\sum_{h=1}^{k}\sum_{q=1}^{\left\lfloor\frac{Q-1}{2}\right\rfloor}2\omega_h^{(2q)}V_{hh}^{2q}b_{q1}^{e}a_{hj}a_{hl}$ and when $Q=2$ we have $\sum_{h=1}^{k}2\omega_h^{(2)}a_{hj}a_{hl}$ which is equal to $2[A^T\Omega A]_{(j,l)}$.

In general, if we are interested in the q^{th} order interactions, we can find the expression on the top summation for $f=\frac{q+1}{2}$ when q is odd and on the bottom summation for $f=\frac{q}{2}$ when q is even. Finally notice that with Qk parameters we manage to estimate $\sum_{q=0}^{Q}\binom{p}{q}$ parameters thanks to the low dimensional factor structure in the covariates.

2.4 Induced Priors

In this section, we show the behavior of the induced priors on the main effects and pairwise interaction coefficients under model (1) using simulated examples, and we show the induced grouping of coefficients when we have prior information on the covariance structure of X. We endow ω with a normal prior having zero mean and covariance equal to Ξ, where Ξ is a diagonal matrix. Then, conditional on Λ and Ψ, the induced prior on β_X is also Normal with mean 0 and covariance equal to $A^T A$. Recall from *Proposition 1* that the induced main effect coefficients are equal to $\beta_X^T=\omega^T(\Lambda^T\Psi^{-1}\Lambda)^{-1}\Lambda^T\Psi^{-1}$. This expression is equivalent to West (2003) and we can similarly characterize the limiting case of $\Psi\rightarrow 0$, i.e. when the factors explain all of the variability in the matrix of regressors X_i. Let $\Psi=sI$ and $s\rightarrow 0$, together with enforcing Λ to be orthogonal, we have that $\beta_X=\Lambda\omega$. It follows that β_X has the *generalised singular g-prior* (or *gsg*-prior) distribution defined by West (2003), whose density is proportional to $exp(-\frac{1}{2}\beta_X^T\Lambda^T\Xi^{-1}\Lambda\beta)$.

Now, consider the extension presented in the previous section, where we include powers of the factors in the regression of y_i. In Figure 1, we show the induced marginal priors for main effects, pairwise interactions and 3^{rd} order interactions when $p=20$ and $k=5$, 10 when ω and Ω are given $\mathcal{N}(0,1)$ priors element-wise. Increasing (or decreasing) the variance of the priors on ω and Ω will directly increase (or decrease) the variance of the induced main effects and pairwise interactions, as β_X and Ω_X are linear functions of ω and Ω respectively. For a fixed k, there is increasing shrinkage towards zero with higher orders of interaction. However, we avoid assuming exact sparsity corresponding to zero values of the coefficients, a standard assumption of other methods. Although most of the mass is concentrated around zero, the distributions have heavy tails. We can indeed notice that the form of the priors resembles a mixture of two normal distributions with different variances, and that we place a higher mixture weight on the normal distribution concentrated around zero as we increase the order of interactions. This is because higher order interactions contain products of the elements of A, previously defined in *Proposition 1*, and the elements of A are affected by the DL prior shrinkage, since A is a function of Λ. Also, notice that the priors have higher variance as we increase the number of latent factors k.

In environmental epidemiology, it is common to have prior knowledge of groups of exposures that are highly correlated and it is natural to include such information in the specification of Λ. One possibility is to impose a block sparsity structure in which each group of chemicals is restricted to load on the same factor. Then, cross group dependence is allowed including additional factors and endowing the factor loadings with a DL prior. Suppose that the variables in X can be divided in l groups: $S_1, S_2, \ldots S_l$ of dimensions $p_1, p_2, \ldots, p_l$, where $l < k$ and $p = \sum_{r=1}^{l} p_r$. Letting $\Lambda = [\Lambda^B \Lambda']$, where Λ^B is $p \times l$, we can assign a block sparsity structure to Λ^B:

$$\lambda^B_{p1+1,1} = \ldots = \lambda^B_{p,1} = 0$$
$$\lambda^B_{1,2} = \ldots = \lambda^B_{p1,2} = \lambda^B_{p1+p2+1,2} = \ldots = \lambda^B_{p,2} = 0$$
$$\ldots$$
$$\lambda^B_{1,l} = \ldots = \lambda^B_{p-p_l,l} = 0$$

In the Supplementary Material we show the effect of the block sparsity structure on the a priori induced groupings of main effects and interactions when $l = k$, so that $\Lambda = \Lambda^B$.

3 Properties of the Model

In this section we prove that the posterior distribution of $\Theta = (\omega, \Omega, \sigma^2, \Lambda, \Psi)$ is weakly consistent for a broad set of models. Let $KL(\Theta_0, \Theta)$ denote the Kullback-Leibler divergence between $p(X, y|\Theta_0)$ and $p(X, y|\Theta)$, where

$$p(X, y \mid \Theta_0) = \int p(X \mid \Lambda_0, \Psi_0, \eta) p(y \mid \omega_0, \Omega_0, \sigma_0^2, \eta) p(\eta) d\eta .$$

We will assume that $p(X, y|\Theta_0)$ represents the true data-generating model. This assumption is not as restrictive as it may initially seem. The model is flexible enough to always characterize and model quadratic regression in the response component, while accurately approximating any covariance structure in the predictor component. In fact it always holds that:

$$\mathbb{E}(y_i \mid X_i) = \beta_0 X_i + X_i \Omega_0 X_i,$$
$$X_i \sim N(0, \Lambda_0 \Lambda_0^T + \Psi_0),$$

where β_0 and Ω_0 are functions of Θ_0 as in *Proposition 1*, and the true number of factors is k_0. When $k_0 = p$, we can write any covariance matrix as $\Lambda_0 \Lambda_0^T + \Psi_0$. We take an "overfitted" factor modeling approach, related to Bhattacharya and Dunson (2011), Rousseau and Mengersen (2011), and choose k to correspond to an upper bound on the number of factors. In practice, we recommend the rule of thumb that chooses k such that $\frac{\sum_{j=1}^{k} v_j}{\sum_{j=1}^{p} v_j} > 0.9$, where v_j is the j^{th} largest singular value of the correlation matrix of X. We have found this choice to have good performance in a wide variety of simulation cases. However, there is nonetheless a potential concern that k may be less than k_0 in some cases. *Proposition 2* quantifies the

distance in terms of Kullback-Leibler divergence between the true data generating model and the likelihood under model miss-specification as n approaches infinity.

Proposition 2. *Fix* Λ_0, $\Psi_0 = s_0 I_p$, k_0, *and assume that* $k < k_0$. *As n increases the posterior distribution of* Λ *and* $\Psi = sI_p$ *concentrates around* Λ^* *and* Ψ^*, *satisfying:*

$$KL((\Lambda_0, \Psi_0); (\Lambda^*, \Psi^*)) \leq \sum_{j=k+1}^{k_0} \frac{v_j}{s_0},$$

where v_j *is the* j^{th} *largest singular value of* $\Lambda_0 \Lambda_0^T$.

Unsurprisingly, the bound of *Proposition 2* resembles the Eckart-Young theorem for low-rank approximation based on the Singular Value Decomposition of a matrix. The Eckart-Young theorem states that the rank k approximation $\widehat{\Omega}$ of a matrix Ω minimizing the Frobenius norm is such that $\|\widehat{\Omega} - \Omega\|_F = \sqrt{\sum_{j=k+1}^{p} v_j^2}$. In a similar fashion as Principal Component Analysis and Factor Analysis, we can inspect the singular values of the correlation matrix of the regressors in order to choose the number of factors to include in the model, and thanks to *Proposition 2* we know how far the posterior distribution will concentrate from the truth.

The next proposition provides a sufficient condition in order to achieve posterior consistency when $k \geq k_0$. Notice that we achieve posterior consistency on the induced main effects and pairwise interactions.

Proposition 3. *Fix* $\Theta_0 = (\omega_0, \Omega_0, \sigma_0^2, \Lambda_0, \Psi_0, k_0)$. *Whenever* $k \geq k_0$, *for any* $\delta > 0$ *there exists an* $\epsilon > 0$ *such that:*

$$\{\Theta : d_\infty(\Theta_0, \Theta) < \delta\} \subset \{\Theta : KL(\Theta_0, \Theta) < \epsilon\}$$

where d_∞ *is the sup-norm.*

One can easily define a prior on Θ such that it places positive probability in any small neighborhood of Θ_0, according to d_∞. The prior defined in Section 2.2 satisfies this condition. Consequently, the posterior distribution of Θ is weakly consistent due to Schwartz (1965). The proofs of *Proposition 2* and *Proposition 3* can be found in the Supplementary Material.

4 Simulation Experiments

In this section we compare the performance of our FIN method with four other approaches: PIE (Wang et al., 2019), RAMP (Hao et al., 2018), Family (Haris et al., 2016) and HierNet (Bien et al., 2013). These methods are designed for inference on interactions in moderate to high dimensional settings. We generate 25 and 50 covariates in three ways:

$$X_i \sim N_p(0, \Lambda\Lambda^T + I_p), \quad \lambda_{i,j} \sim N(0,1), \quad \text{(factor)}$$
$$X_i \sim N_p(0, W), \quad [W]_{i,j} = 0.8^{|i-j|}, \quad \text{(linear)}$$
$$X_i \sim N_p(0, I_p). \quad \text{(indepedent)}$$

In the factor scenario we set the true number of factors equal to 7 for $p = 25$ and equal to 17 when $p = 50$. FIN achieved similar performance when we chose a smaller number of latent factors. The average absolute correlation in the covariates is between 0.25 and 0.3 for the factor and linear scenarios when $p = 25$. These two simulation scenarios are the most similar to the environmental epidemiology data analysis in Section 5. The complexity gains of FIN with respect to a Bayesian linear models with interactions is analyzed in the Supplementary Material.

For each scenario, we generate the continuous outcome according to a linear regression with pairwise interactions:

$$y_i = X_i^T \beta_0 + X_i^T \Omega_0 X_i + \epsilon_i,$$

where half of the main effects are different from zero and $\epsilon_i \sim N(0, 1)$ for $i = 1, \cdots, 500$. We distinguish between a sparse matrix of pairwise interactions Ω_0, with only 5% non-zero interactions, or dense, where 20% of the elements are different from zero.

For each value of p we have six simulation scenarios: factor, linear or independent combined with sparse or dense pairwise interactions. We generate the non-zero main effects and interaction coefficients from a Uniform distribution in the interval $(-1, -0.5)\cup(0.5, 1)$ such that the regression equation follows the strong heredity constraint. Strong heredity allows an interaction between two variables to be included in the model only if the main effects are. This is done to favor RAMP, Family and HierNet, which assume the heredity condition. We repeat the simulations 50 times and evaluate the performance on a test dataset of 500 units computing predictive mean square error, mean square error for main effects, Frobenious norm (FR) for interaction effects, and percentage of true positives (TP) and true negatives (TN) for main effects and interactions. The percentage of TP and TN main effects is defined as follows:

$$\text{TP(main effects)} = \frac{1}{p}\sum_{j=1}^{p} \mathbb{1}(\hat{\beta}_j \neq 0, \beta_{0j} \neq 0, sign(\hat{\beta}_j) = sign(\beta_{0j}))$$
$$\text{TN(main effects)} = \frac{1}{2}\sum_{j=1}^{p} \mathbb{1}(\hat{\beta}_j = 0, \beta_{0j} = 0),$$

where $\hat{\beta}_j$ is the estimated main effect for feature j and β_{0j} is the true coefficient. FIN is the only method reporting uncertainty quantification and we set $\hat{\beta}_j = 0$ whenever zero is included in the 95% credible interval. We equivalently define the percentage of true positive and true negative interactions.

The MCMC algorithm was run for 5000 iterations with a burn-in of 4000. We observed good mixing. In particular, the Effective Sample Size (ESS) was always greater than 900 across our simulations, both for main effects and interactions. We set the hyperparameter a of the Dirichlet-Laplace prior equal to 1/2. We obtained similar results for a in the interval $[1/p, p]$. The results are summarized in Table 1-2 for $p = 25$ and in Table 1-2 of the Supplementary Material for $p = 50$. Across all the simulations, we chose k such that $\frac{\sum_{j=1}^{k} v_j}{\sum_{j=1}^{p} v_j} > 0.9$.

In the factor scenario, FIN outperforms the other methods in predictive performance and estimation of main effects and interactions, whereas the rate recovery of true main effects and interactions is comparable to HierNet and PIE with sparse Ω_0 and outperforms the other methods when Ω_0 is dense. The latter scenario is the most challenging with respect to selection of main effects and interactions. Most of the other methods either select or shrink to zero all the effects. In the linear scenario, FIN also shows the best performance together with PIE and Hiernet. Despite the model misspecification with independent covariates, FIN has a comparable predictive performance with respect to the other methods, which do not take into account correlation structure in the covariates. The 95% predictive intervals provided by FIN contained the true value of y_i on average approximately 95% of the time in the factor scenario, 89% for the linear scenario, and 79% for the independent scenario. The average bias in the posterior predictive mean is negligible in each simulation scenario.

The optimization method performed by HierNet (Bien et al., 2013) tends to favor interactions only in presence of large component main effects, and in doing so overshrinks interactions estimates, especially in the *dense scenario*. Penalized regression techniques PIE (Wang et al., 2019) and RAMP (Hao et al., 2018) tend to over-shrink coefficient estimates and select too few predictors, particularly in the dense scenario. On the other hand, FAMILY (Haris et al., 2016) performs a relaxed version of the penalized algorithm by refitting an unpenalized least squares model, which results in a high false positive rate of main effects. We also considered different signal-to-noise ratios with $\epsilon_i \sim N(0, \frac{1}{4})$ and $\epsilon_i \sim N(0, 4)$. The results are very similar to the results we have presented; hence, we omit them.

5 Environmental Epidemiology Application

The goal of our analysis is to assess the effect of ten phthalate metabolites, four perfluoroalkyl (pfas) metabolites and fourteen metals on body mass index (BMI). Phthalates are mainly used as plasticizers and can be found in toys, detergents, food packaging, and soaps. They have previously been associated with increased BMI (Hatch et al., 2008) and waist circumference (WC) (Stahlhut et al., 2007). There is a growing health concern for the association of phthalates (Kim and Park, 2014), (Zhang et al., 2014) and pfas metabolites (Braun, 2017) with childhood obesity. Metals have already been associated with an increase in waist circumference and BMI, see Padilla et al. (2010) and Shao et al. (2017), using data from the National Health and Nutrition Examination Survey (NHANES).

We also consider data from NHANES, using data from the years 2015 and 2016. We select a subsample of 7602 individuals for which the measurement of BMI is not missing, though

FIN can easily accommodate missing outcomes through adding an imputation step to the MCMC algorithm. Figure 3 contains a plot of the correlation between exposures. Several pairwise correlations are missing, as for example between pfas and most metals, because some chemicals are only measured within subsamples of the data. The average absolute correlation between the 28 exposures is around 0.28, similarly to the *factor* and *linear* simulation scenarios presented in Section 4. We also include in the analysis cholesterol, creatinine, race, sex, education and age. We apply the $\log_{10}$ transformation to the chemicals, cholesterol and creatinine. Histograms of the chemical measurements can be found in Figure 1 of the Supplementary Material. We also apply the $\log_{10}$ transformation to BMI in order to make its distribution closer to normality, which is the assumed marginal distribution in our model. The log-transformation is commonly applied in environmental epidemiology in order to reduce the influence of outliers and has been employed in several studies using NHANES data (Nagelkerke et al., 2006), (Lynch et al., 2010), (Buman et al., 2013). We leave these transformations implicit for the remainder of the section.

We assume a latent normal structure for the chemicals, which are included in the matrix X, and use the other variables as covariates, which are included in the matrix Z. We estimate a quadratic regression according to model (2). We specify independent Gaussian priors for elements of α and Δ. Algorithm 1 can be easily adapted for model (2). The matrix X has 60% missing data and Figure 2 of the Supplementary material contains a plot of the missingness pattern. Since we are modeling the chemical measurements, we can simply add a sampling step to the MCMC algorithm to sample the missing data according to (2). Similarly, 0.4% of chemicals have been recorded under the limit of detection (LOD). In order to be coherent with our model we can sample these observations as:

$$X_{ij} \mid X_{ij} \in [-\infty, \log_{10}(\text{LOD}_j)] \sim TN(\eta_i^T \lambda_j, \sigma_j^2, -\infty, \log_{10}(\text{LOD}_j))$$

where LOD_j is the limit of detection for exposure j and $TN(\mu, \sigma^2, a, b)$ is a truncated normal distribution with mean μ, variance σ^2 and support in $[a, b]$. We imputed the missing data using MICE (White et al., 2011) to compute the correlation matrix of chemicals. We noticed from the Eigendecomposition of the correlation matrix that the first 13 eigenvectors explain more than 90% of the total variability; hence we set the number of factors equal to 13.

Figure 2 on the right shows the posterior mean of the matrix of factor loadings Λ, before and after applying the MatchAlign algorithm of Poworoznek and Dunson (2019), which resolves rotational ambiguity and column label switching for the posterior samples of Λ. The matrix of factor loadings reflects the correlation structure of the chemicals. We can distinguish three families of chemicals: metals collected from urine, pfas and phathalates. The pfas chemicals load mostly on the 1^{st} factor, the metals from urine on the 8^{th} factor together with the phthalates, which is expected since there is high correlation between the two groups of chemicals. Finally, a group of highly correlated phthalates loads on the 13^{th} factor.

We also estimated a regression with pairwise interactions using the methods PIE, RAMP, Family and HierNet introduced in Section 4. These methods do not directly deal with

missing data, so we imputed the missing data using MICE (White et al., 2011). Figure 3 shows the estimated main effects of the chemicals. The signs of the coefficients are generally consistent across different methods.

Figure 4 shows the posterior mean of the matrix of chemical interactions and of the matrix $A^T\Delta$ of pairwise interactions between exposures and covariates. As expected, we estimate a "dense" matrix of interactions. This is due to exposures being breakdown products of the same compound and high correlation between chemicals belonging to the same family. For example the correlation between the pfas metabolites is equal to 0.7, with only 1977 observations containing complete measurements. Interactions between highly correlated pfas metabolites have been observed in animal studies (Wolf et al., 2014), (Ding et al., 2013). Linear (Henn et al., 2011), (Lin et al., 2013) and nonlinear interactions (Valeri et al., 2017) between metals have been associated with neurodevelopment. Interactions between phathalates and other chemicals have been related to human semen quality (Hauser et al., 2005). Finally, we estimate several interactions between chemicals and age, cholesterol and creatinine, which are usually expected in environmental epidemiology applications (Barr et al., 2004). The code for reproducing the analysis is available at https://github.com/fedfer/factor_interactions.

6 Discussion

We proposed a novel method that exploits the correlation structure of the predictors and allows us to estimate interaction effects in high dimensional settings, assuming a latent factor model. Using simulated examples, we showed that our method has a similar performance to state-of-the-art methods for interaction estimation when dealing with independent covariates and outperforms the competitors when there is moderate to high correlation among the predictors. We provided a characterization of uncertainty with a Bayesian approach to inference. Our FIN approach is particularly motivated by epidemiology studies with correlated exposures, as illustrated using data from NHANES.

NHANES data are obtained using a complex sampling design, that includes oversampling of certain population subgroups, and contains sampling weights for each observation that are inversely proportional to the probability of begin sampled. We did not employ sampling weights in our analysis because our goal was to study the association between exposures and BMI rather than providing population estimates. One possibility to include the sampling weights in our method is to jointly model the outcome and the survey weights (Si et al., 2015), without assuming that the population distribution of strata is known.

Our MCMC algorithm can be efficiently employed for n and p in the order of thousands and hundreds respectively, which allows us to estimate around 5000 interactions when $p = 100$. However, it is necessary to speed up the computations in order to apply our method to bigger p, which is common with genomics data. The computational bottleneck is the Metropolis Hastings step described in Section 2.2. One possibility is to include the heredity constraint (Chipman, 1996) while estimating the factors.

In order to allow departures from linearity and Gaussianity, it is of interest to model the regression on the health outcome as a non-linear function of latent factors. Non parametric latent models have desirable properties in term of convergence rates (Zhou et al., 2017) and large support for density estimation (Kundu and Dunson, 2014). Verma and Engelhardt (2018) developed a dimension reduction approach with latent variables for single cell RNA-seq data building on Gaussian process latent variable models (GP-LVM). Although attractive from a modeling perspective, a major challenge is efficient posterior computation. Another promising direction to decrease modeling assumptions is to rely on a copula factor model related to Murray et al. (2013).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grant 1R01ES028804-01 of the National Institute of Environmental Health Sciences of the United States Institutes of Health. The authors would like to thank Evan Poworoznek, Antonio Canale, Michele Caprio, Amy Herring, Elena Colicino and Emanuele Aliverti for helpful comments.

Appendix

Proof of Proposition 1. (i) Let us drop the i index for notation simplicity and always assume that we are conditioning on all the parameters. The posterior distribution of η is Normal with covariance $V = (\Lambda^T\Psi^{-1}\Lambda + I)^{-1}$ and mean AX where $A = V\Lambda^T\Psi^{-1} = (\Lambda^T\Psi^{-1}\Lambda + I)^{-1}\Lambda^T\Psi^{-1}$. This follows from a simple application of Bayes Theorem. Now:

$$\mathbb{E}(y \mid X) = \mathbb{E}(\mathbb{E}(y \mid \eta) \mid X) = \mathbb{E}(\eta^T\omega + \eta^T\Omega\eta \mid X) = \\ = \omega^T\mathbb{E}(\eta \mid X) + \mathbb{E}(\eta^T\Omega\eta \mid X)$$

Recall that the expectation of a quadratic form $\eta^T\Omega\eta$ of a random vector η with mean μ and covariance matrix Σ is equal to $tr(\Omega\Sigma) + \mu^T\Omega\mu^T$.

$$\mathbb{E}(y \mid X) = \omega^T AX + tr(\Omega V_\eta) + (AX)^T\Omega(AX) = \\ = tr(\Omega V) + (\omega^T A)X + X^T(A^T\Omega A)X$$

(ii) Recall that $\eta \sim N(0, I)$, $y = \eta^T\omega + \eta^T\Omega\eta_i + \epsilon_y$ and $X = \Lambda\eta + \epsilon$, from simple algebra it follows that

$$Cov(y, X) = \omega^T Cov(\eta, \eta)\Lambda^T + Cov(\eta^T\Omega\eta, \Lambda\eta)$$

From the prior specification $Cov(\eta, \eta) = I$, hence let us focus on the term $Cov(\eta^T\Omega\eta, \Lambda\eta)$ and show that it is equal to 0_p:

$$Cov(\eta^T\Omega\eta, \Lambda\eta) = Cov(\sum_{j=1}^{p}\sum_{l=1}^{p}\omega_{j,l}\eta_j\eta_l, \Lambda\eta) =$$
$$= \sum_{j=1}^{p}\sum_{l=1}^{p}\omega_{j,l}Cov(\eta_j\eta_l, \begin{pmatrix}\lambda_{1,1}\eta_1 + \ldots + \lambda_{1,k}\eta_k \\ \ldots \\ \lambda_{p,1}\eta_1 + \ldots + \lambda_{p,k}\eta_k\end{pmatrix}) =$$
$$= \sum_{j=1}^{p}\sum_{l=1}^{p}\omega_{j,l}Cov(\eta_j\eta_l, \begin{pmatrix}\lambda_{1,j}\eta_j + \lambda_{1,l}\eta_l \\ \ldots \\ \lambda_{p,j}\eta_j + \lambda_{p,l}\eta_l\end{pmatrix}) =$$
$$= \sum_{j=1}^{p}\sum_{l=1}^{p}\omega_{j,l}\left[Cov(\eta_j\eta_l, \begin{pmatrix}\lambda_{1,j}\eta_j \\ \ldots \\ \lambda_{p,j}\eta_j\end{pmatrix}) + Cov(\eta_j\eta_l, \begin{pmatrix}\lambda_{1,l}\eta_l \\ \ldots \\ \lambda_{p,l}\eta_l\end{pmatrix})\right]$$

Now $Cov(\eta_j\eta_l, \eta_j) = E(\eta_j^2\eta_l) = 0$. In fact when $j \neq l$, we have that $E(\eta_j^2\eta_l) = E(\eta_j^2)E(\eta_l) = 0$ and when $j = l$, $E(\eta_j^3) = 0$ since $\eta_j \sim N(0, 1)$.

References

Arminger G and Muthén BO (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. Psychometrika 63(3), 271–300.

Barr DB, Wilder LC, Caudill SP, Gonzalez AJ, Needham LL, and Pirkle JL (2004). Urinary creatinine concentrations in the us population: implications for urinary biologic monitoring measurements. Environmental Health Perspectives 113(2), 192–200.

Bhattacharya A and Dunson DB (2011). Sparse Bayesian infinite factor models. Biometrika, 291–306. [PubMed: 23049129]

Bhattacharya AK, Pati D, Pillai NS, and Dunson DB (2015). Dirichlet-Laplace priors for optimal shrinkage. Journal of the American Statistical Association 110 512, 1479–1490. [PubMed: 27019543]

Bien J, Taylor J, and Tibshirani R (2013). A lasso for hierarchical interactions. Annals of Statistics 41 (3), 1111. [PubMed: 26257447]

Bondell HD and Reich BJ (2012). Consistent high-dimensional Bayesian variable selection via penalized credible regions. Journal of the American Statistical Association 107 (500), 1610–1624. [PubMed: 23482517]

Braun JM (2017). Early-life exposure to edcs: role in childhood obesity and neurodevelopment. Nature Reviews Endocrinology 13(3), 161.

Buman MP, Winkler EA, Kurka JM, Hekler EB, Baldwin CM, Owen N, Ainsworth BE, Healy GN, and Gardiner PA (2013). Reallocating time to sleep, sedentary behaviors, or active behaviors: associations with cardiovascular disease risk biomarkers, nhanes 2005–2006. American Journal of Epidemiology 179(3), 323–334. [PubMed: 24318278]

Carvalho CM, Chang J, Lucas JE, Nevins JR, Wang Q, and West M (2008). High-dimensional sparse factor modeling: Applications in gene expression genomics. Journal of the American Statistical Association 103(484), 1438–1456. PMID: 21218139. [PubMed: 21218139]

Chipman H (1996). Bayesian variable selection with related predictors. Canadian Journal of Statistics 24 (1), 17–36.

Cordell HJ (2009). Detecting gene–gene interactions that underlie human diseases. Nature Reviews Genetics 10(6), 392.

Ding G, Zhang J, Chen Y, Wang L, Wang M, Xiong D, and Sun Y (2013). Combined effects of pfos and pfoa on zebrafish (danio rerio) embryos. Archives of Environmental Contamination and Toxicology 64 (4), 668–675. [PubMed: 23479250]

Geweke J and Zhou G (1996). Measuring the pricing error of the arbitrage pricing theory. The review of financial studies 9 (2), 557–587.

Grenander U and Miller MI (1994). Representations of knowledge in complex systems. Journal of the Royal Statistical Society: Series B (Methodological) 56(4), 549–581.

Hao N, Feng Y, and Zhang HH (2018). Model selection for high-dimensional quadratic regression via regularization. Journal of the American Statistical Association 113(522), 615–625.

Haris A, Witten D, and Simon N (2016). Convex modeling of interactions with strong heredity. Journal of Computational and Graphical Statistics 25 (4), 981–1004. [PubMed: 28316461]

Hatch EE, Nelson JW, Qureshi MM, Weinberg J, Moore LL, Singer M, and Webster TF (2008). Association of urinary phthalate metabolite concentrations with body mass index and waist circumference: a cross-sectional study of nhanes data, 1999–2002. Environmental Health 7(1), 27. [PubMed: 18522739]

Hauser R, Williams P, Altshul L, and Calafat AM (2005). Evidence of interaction between polychlorinated biphenyls and phthalates in relation to human sperm motility. Environmental Health Perspectives 113 (4), 425–430. [PubMed: 15811833]

Henn BC, Schnaas L, Ettinger AS, Schwartz J, Lamadrid-Figueroa H, Hernández-Avila M, Amarasiriwardena C, Hu H, Bellinger DC, Wright RO, et al. (2011). Associations of early childhood manganese and lead coexposure with neurodevelopment. Environmental Health Perspectives 120(1), 126–131. [PubMed: 21885384]

Johndrow JE, Orenstein P, and Bhattacharya A (2017). Bayes shrinkage at gwas scale: Convergence and approximation theory of a scalable mcmc algorithm for the horseshoe prior. arXiv preprint arXiv:1705.00841.

Kim SH and Park MJ (2014). Phthalate exposure and childhood obesity. Annals of Pediatric Endocrinology & Metabolism 19(2), 69. [PubMed: 25077088]

Kundu S and Dunson DB (2014). Latent factor models for density estimation. Biometrika 101 (3), 641–654.

Legramanti S, Durante D, and Dunson DB (2019). Bayesian cumulative shrinkage for infinite factorizations. arXiv preprint arXiv:1902.04349.

Lin C-C, Chen Y-C, Su F-C, Lin C-M, Liao H-F, Hwang Y-H, Hsieh W-S, Jeng S-F, Su Y-N, and Chen P-C (2013). In utero exposure to environmental lead and manganese and neurodevelopment at 2 years of age. Environmental Research 123, 52–57. [PubMed: 23578827]

Lucas J, Carvalho C, Wang Q, Bild A, Nevins JR, and West M (2006). Sparse statistical modelling in gene expression genomics. Bayesian inference for gene expression and proteomics 1, 0–1.

Lynch BM, Dunstan DW, Healy GN, Winkler E, Eakin E, and Owen N (2010). Objectively measured physical activity and sedentary time of breast cancer survivors, and associations with adiposity: findings from nhanes (2003–2006). Cancer Causes & Control 21 (2), 283–288. [PubMed: 19882359]

Mackay TF (2014). Epistasis and quantitative traits: using model organisms to study gene–gene interactions. Nature Reviews Genetics 15(1), 22.

Murray JS, Dunson DB, Carin L, and Lucas JE (2013). Bayesian gaussian copula factor models for mixed data. Journal of the American Statistical Association 108 (502), 656–665. [PubMed: 23990691]

Nagelkerke NJ, Bernsen RM, Sgaier SK, and Jha P (2006). Body mass index, sexual behaviour, and sexually transmitted infections: an analysis using the nhanes 1999–2000 data. BMC Public Health 6 (1), 199. [PubMed: 16884541]

Nishimura A and Suchard MA (2018). Prior-preconditioned conjugate gradient for accelerated Gibbs sampling in “large n & large p” sparse Bayesian logistic regression models. arXiv preprint arXiv:1810.12437.

Padilla MA, Elobeid M, Ruden DM, and Allison DB (2010). An examination of the association of selected toxic metals with total and central obesity indices: Nhanes 99-02. International Journal of Environmental Research and Public Health 7(9), 3332–3347. [PubMed: 20948927]

Park T and Casella G (2008). The Bayesian lasso. Journal of the American Statistical Association 103(482), 681–686.

Poworoznek E and Dunson DB (2019). Efficiently resolving column switching in Bayesian matrix sampling with clustering. https://poworoznek.github.io/website/overview.html.

Roberts GO, Tweedie RL, et al. (1996). Exponential convergence of langevin distributions and their discrete approximations. Bernoulli 2(4), 341–363.

Rossell D and Telesca D (2017). Nonlocal priors for high-dimensional estimation. Journal of the American Statistical Association 112(517), 254–265. [PubMed: 29881129]

Rousseau J and Mengersen K (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73(5), 689–710.

Schwartz L (1965). On Bayes procedures. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 4 (1), 10–26.

Shao W, Liu Q, He X, Liu H, Gu A, and Jiang Z (2017). Association between level of urinary trace heavy metals and obesity among children aged 6–19 years: Nhanes 1999–2011. Environmental Science and Pollution Research 24 (12), 11573–11581. [PubMed: 28321702]

Si Y, Pillai NS, Gelman A, et al. (2015). Bayesian nonparametric weighted sampling inference. Bayesian Analysis 10 (3), 605–625.

Stahlhut RW, van Wijngaarden E, Dye TD, Cook S, and Swan SH (2007). Concentrations of urinary phthalate metabolites are associated with increased waist circumference and insulin resistance in adult us males. Environmental Health Perspectives 115 (6), 876–882. [PubMed: 17589594]

Valeri L, Mazumdar MM, Bobb JF, Claus Henn B, Rodrigues E, Sharif OI, Kile ML, Quamruzzaman Q, Afroz S, Golam M, et al. (2017). The joint effect of prenatal exposure to metal mixtures on neurodevelopmental outcomes at 20–40 months of age: evidence from rural bangladesh. Environmental Health Perspectives 125 (6), 067015. [PubMed: 28669934]

Verma A and Engelhardt B (2018). A robust nonlinear low-dimensional manifold for single cell RNA-seq data. bioRxiv, 443044.

Wang C, Jiang B, and Zhu L (2019). Penalized interaction estimation for ultrahigh dimensional quadratic regression. arXiv preprint arXiv:1901.07147.

West M (2003). Bayesian factor regression models in the “large p, small n” paradigm. In Bayesian Statistics, pp. 723–732. Oxford University Press.

White IR, Royston P, and Wood AM (2011). Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine 30(4), 377–399. [PubMed: 21225900]

Winkelbauer A (2012). Moments and absolute moments of the normal distribution. arXiv preprint arXiv:1209.4340.

Wolf CJ, Rider CV, Lau C, and Abbott BD (2014). Evaluating the additivity of perfluoroalkyl acids in binary combinations on peroxisome proliferator-activated receptor-α activation. Toxicology 316, 43–54. [PubMed: 24374136]

Zhang Y and Liu JS (2007). Bayesian inference of epistatic interactions in case-control studies. Nature Genetics 39(9), 1167. [PubMed: 17721534]

Zhang Y, Meng X, Chen L, Li D, Zhao L, Zhao Y, Li L, and Shi H (2014). Age and sex-specific relationships between phthalate exposures and obesity in chinese children at puberty. PloS One 9(8), e104852. [PubMed: 25121758]

Zhou S, Pati D, Bhattacharya A, and Dunson D (2017). Adaptive posterior convergence rates in non-linear latent variable models. arXiv preprint arXiv:1701.07572.

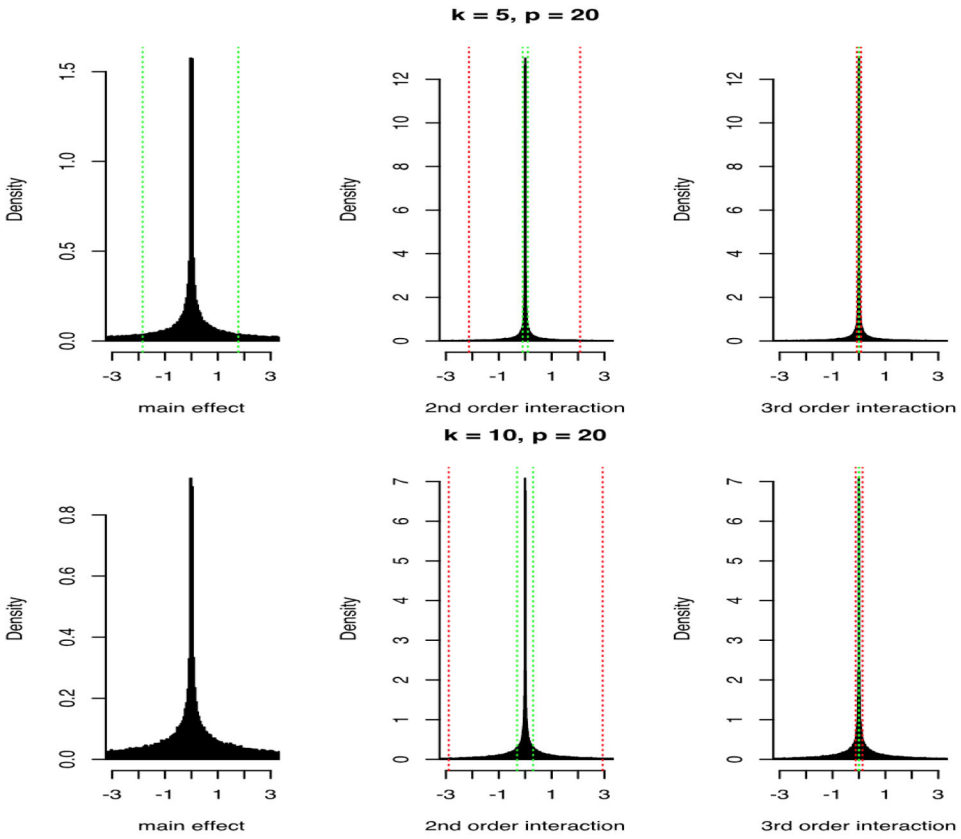

Figure 1:

Induced priors on main effects, pairwise interactions and 3^{rd} order interactions for $p = 20$ and $k = 5, 10$. The green lines corresponds to 0.25 and 0.75 quartiles and the red lines to the 0.05 and 0.95.

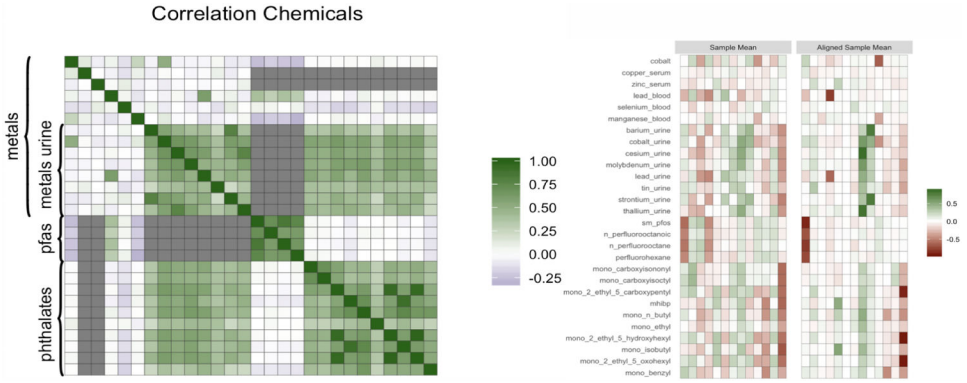

Figure 2:

On the left, correlation between the exposures, the colour grey indicates missing pairwise correlation. On the right, posterior mean of the matrix Λ of factor loadings before and after applying the MatchAlign algorithm.

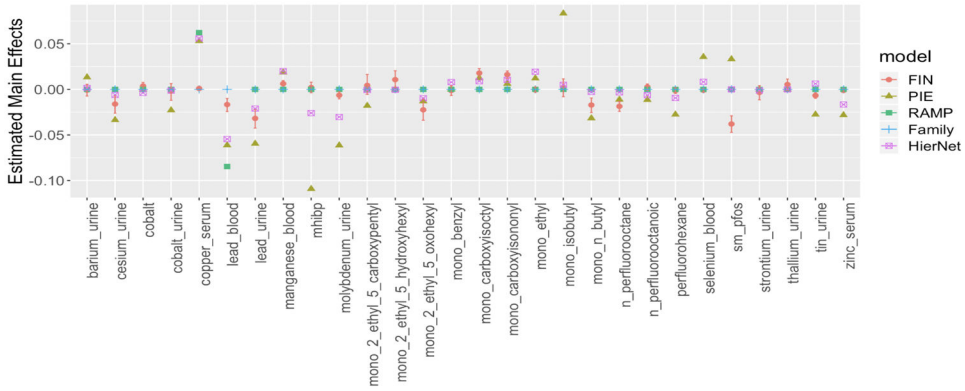

Figure 3:

Estimated main effects using FIN with 95% credible intervals and estimated coefficients using RAMP, hierNet, Family and PIE.

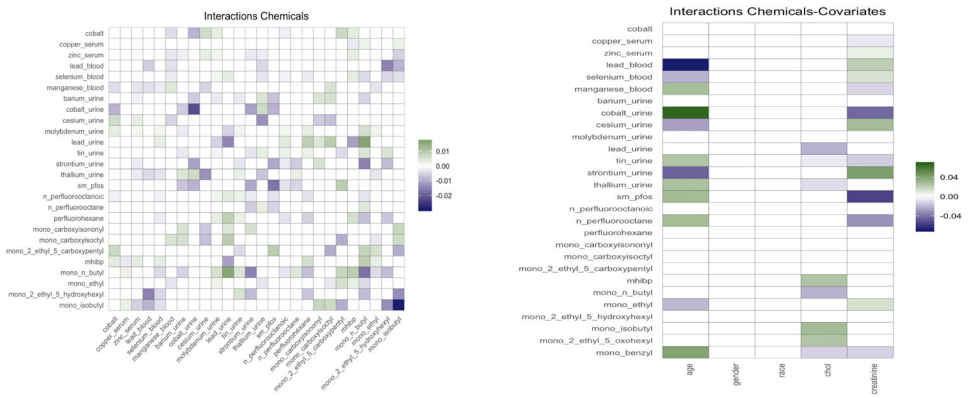

Figure 4:

On the left, posterior mean of the matrix of chemicals interactions. On the right, posterior mean of the matrix $A^T\Delta$ of pairwise interactions between exposures and covariates. The white boxes indicates that the 99% credible interval contains zero.

Table 1:

Results from simulation study with $p = 25$ and *dense* Ω_0 in the three scenarios: factor, linear and independent for $n = 500$. We computed test error, Frobenious norm, MSE for main effects, percentage of true positives and true negatives for main effects and interactions for Hiernet, Family, PIE, RAMP and FIN model with $a = 0.5$ across 50 simulations. Test error, FR, and main MSE are presented as ratios compared to the best performing model.

		HierNet	FAMILY	PIE	RAMP	FIN
factor	test error	1.974	16.689	7.067	64.717	1
	FR	1.361	1.013	1.418	1.620	1
	main MSE	1.167	1.062	1.807	4.225	1
	TP main	0.920	0.988	0.155	0.270	0.753
	TN main	0.067	0.007	0.921	0.773	0.475
	TP int	0.151	0.807	0.105	0.037	0.699
	TN int	0.889	0.233	0.929	0.962	0.387
linear	test error	1	2.662	1.688	6.309	1.565
	FR	1	1.049	1.075	1.289	1.016
	main MSE	2.421	1	1.766	4.259	1.263
	TP main	1	0.996	0.177	0.301	0.572
	TN main	0.002	0.005	0.904	0.805	0.718
	TP int	0.532	0.818	0.280	0.028	0.635
	TN int	0.849	0.278	0.887	0.968	0.570
independent	test error	1	6.150	2.759	10.729	7.128
	FR	1.175	1.548	1	2.042	1.654
	main MSE	1	1.529	1.756	2.446	2.031
	TP main	1	1	0.241	0.074	0.302
	TN main	0	0.002	0.930	0.985	0.888
	TP int	0.989	0.952	0.641	0.005	0.412
	TN int	0.937	0.414	0.908	1.000	0.914

Table 2:

Results from simulation study with $p = 25$ and *sparse* Ω_0 in the three scenarios: factor, linear and independent for $n = 500$. We computed test error, Frobenious norm, MSE for main effects, percentage of true positives and true negatives for main effects and interactions for Hiernet, Family, PIE, RAMP and FIN model with $a = 0.5$ across 50 simulations. Test error, FR, and main MSE are presented as ratios compared to the best performing model.

		HierNet	FAMILY	PIE	RAMP	FIN
factor	test error	1.284	5.274	1.206	4.225	1
	FR	1.189	1.259	1	2.157	1.284
	main MSE	3.430	1.560	1	1.590	1.312
	TP main	0.667	0.823	0.698	0.583	0.812
	TN main	0.445	0.259	0.863	0.834	0.716
	TP int	0.514	0.839	0.562	0.031	0.448
	TN int	0.959	0.580	0.974	0.965	0.941
linear	test error	1.199	5.271	1	5.060	1.486
	FR	3.889	6.859	1	7.916	5.370
	main MSE	1	3.563	1.387	1.392	1.726
	TP main	1	0.845	0.857	0.952	0.976
	TN main	0.484	0.272	0.845	0.815	0.807
	TP int	0.970	0.887	0.964	0.077	0.917
	TN int	0.970	0.645	0.987	0.975	0.894
independent	test error	1.425	9.685	1	12.746	3.438
	FR	12.956	18.036	1	21.604	9.607
	main MSE	1	6.082	3.056	4.326	3.055
	TP main	1	0.830	0.860	0.630	0.900
	TN main	0.418	0.585	0.847	0.915	0.898
	TP int	1	0.852	1	0.071	0.921
	TN int	0.993	0.868	0.990	0.995	0.957