Integrative Large-Scale Bayesian Learning: From Factor Analysis to Graphical Models

Alejandra Avalos-Pacheco, PhD
@AleAviP
alejandra.avalos@tuwien.ac.at

WU Research Seminar Series

May 5^{th}, 2023

Research motivation

- New technologies enable the gathering of large datasets.
- Two main statistical challenges:
  - Volume: High dim. data $\implies$ hard to handle and interpret.
  - Variety: Heterogeneous data $\implies$ systematic biases.

Integrate heterogeneous high-dimensional data

- New insights into the data
- Robust estimates, reducing the biases
- Gains in statistical power
- Make accurate decisions sooner

- Interested mainly in Bayesian modelling
  1. factor regression for data integration;
  2. multi-study graphical models;
  3. external data integration in trials.

Outline

Outline

Bayesian factor regression

Open Access

2022

Heterogeneous Large Datasets Integration Using Bayesian Factor Regression

Alejandra Avalos-Pacheco, David Rossell, Richard S. Savage

Author Affiliations -

Alejandra Avalos-Pacheco,[1,2] David Rossell,[3] Richard S. Savage[2]

[1]Harvard-MIT Center for Regulatory Science, Harvard Medical School, 210 Longwood Av, Boston, MA, USA

[2]Dept. of Statistics, University of Warwick, Coventry CV4 7AL, UK

[3]Dept. of Business and Economics, Universitat Pompeu Fabra, Carrer de Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain

Bayesian Anal. 17(1): 33-66 (2022). DOI: 10.1214/20-BA1240

Abstract

Two key challenges in modern statistical applications are the large amount of information recorded per individual, and that such data are often not collected all at once but in batches. These batch effects can be complex, causing distortions in both mean and variance. We propose a novel sparse latent factor regression model to integrate such heterogeneous data. The model provides a tool for data exploration

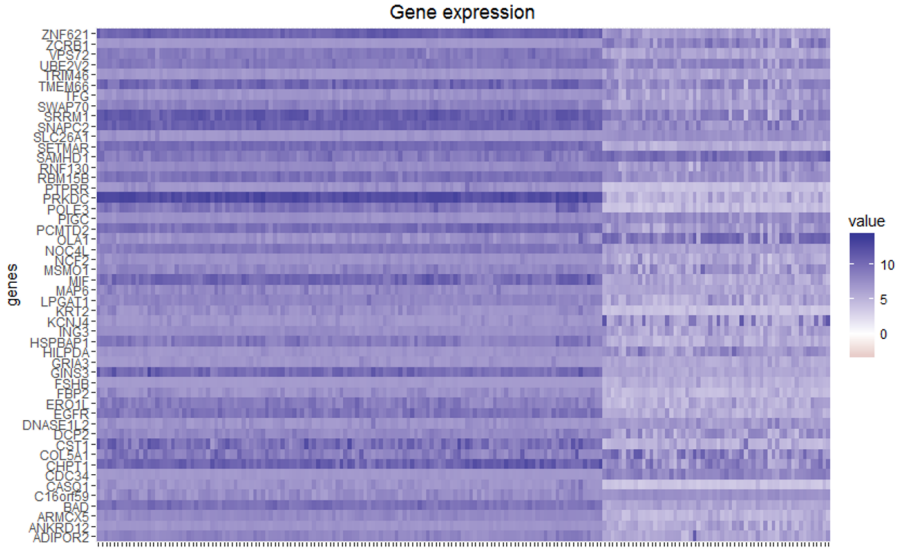
Bayesian Anal.
Vol.17 • No. 1 • 2022

Bayesian Analysis

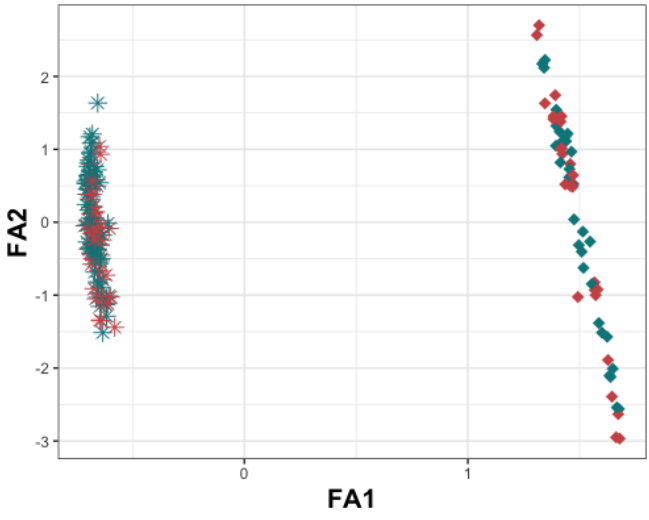
International Society for Bayesian Analysis

ISBA

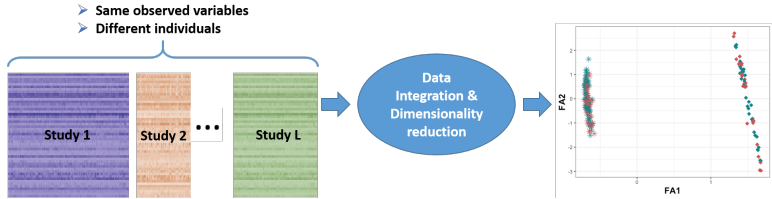
Gene expression: Ovarian cancer



Gene expression: Ovarian cancer



Factor regression and data integration



Challenge: **Batch effects** (non-biological experimental variation)

- Cause distortions in both mean and variance
- Can lead to incorrect conclusions
- Limit our ability to see biological patterns of interest

Solution. A **novel sparse latent factor regression model** to integrate such heterogeneous data:

Factor analysis + factor regression + sparsity + batch effect correction

Factor Regression w/batch effect correction[1]

Stack all the studies in **one** data-set:

$$\underbrace{X}_{\mathbb{R}^{n\times p}} = \left(\underbrace{X^{(1)}}_{\mathbb{R}^{n_1\times p}}, \underbrace{X^{(2)}}_{\mathbb{R}^{n_2\times p}}, \ldots, \underbrace{X^{(l)}}_{\mathbb{R}^{n_l\times p}}\right), \quad n = n_1 + n_2 + \cdots + n_l$$

- Goals:
  1. Given $X \in \mathbb{R}^{n\times p}$ obtain $F \in \mathbb{R}^{n\times q}$, $q << p$
  2. Correct for systematic biases in mean and variance
- Model $\mathtt{x}_i = \boldsymbol{\Phi}\mathtt{f}_i + \boldsymbol{\beta}\mathtt{b}_i + \boldsymbol{\theta}\mathtt{v}_i + \mathtt{e}_i$
  - $\mathtt{f}_i \sim N(0, \mathbf{I})$: latent factors
  - $\boldsymbol{\Phi} \in \mathbb{R}^{p\times q}$: loading matrix
  - $\boldsymbol{\beta} \in \mathbb{R}^{p\times p_b}$: additive batch effects
  - $\mathtt{b}_i \in \{0,1\}^{p_b}$: batch indicators
  - $\boldsymbol{\theta} \in \mathbb{R}^{p\times p_v}$: regression coefficients
  - $\mathtt{v}_i \in \mathbb{R}^{p_v}$: observed covariates
  - $\mathtt{e}_{ij} \sim N(0, \mathcal{T}_{jl}{}^{-1})$, $\mathcal{T}_{jl}$: the j^{th} idiosyncratic precision element in batch l.

[1] Avalos-Pacheco A., Rossell D., Savage R. (2022) **Heterogeneous large datasets integration using Bayesian latent factor regression**, *Bayesian Analysis*, 17(1) pp. 33–66, doi.org/10.1214/20-ba1240.

Novel non-local spike-and-slab priors

- Challenge: How to choose the latent cardinality?
- Solution: induce **sparsity** and learn the underlying number of factors via a scalable **non-local prior** based formulation on the loadings.

Bayesian regularization and sparsity

Regularization as an optimization problem:

$$\widehat{\theta} \in \arg_{\theta}\min\{\underbrace{-\log p(Data \mid \theta)}_{loglikelihood} + \underbrace{\text{pen}(\theta)}_{\text{penality fun}}\}$$

Ex: linear regression: $\min_{\beta}\left[||Y - \beta X||^2 + \text{pen}(\beta)\right]$
Frequentist:

- L_1 penalty: Lasso (Tibshirani 1996)
- Non-convex penalties: SCAD (Fan and Li 2001)

Bayesian: penalty = -log (prior)

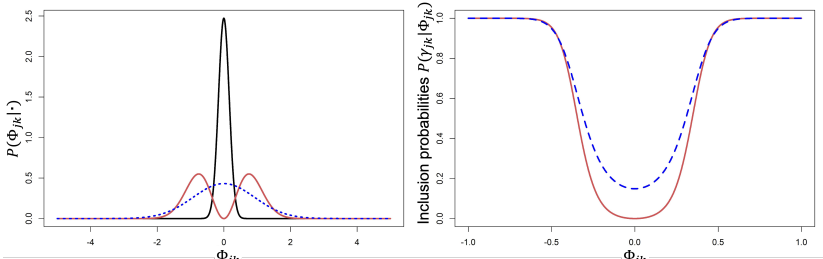
- Lasso: $\exp(-\lambda|\beta|) = \text{Laplace}(\beta;\lambda)$ prior
- $L_2 : \lambda\exp(-\beta^2) = N(\beta; 0, \lambda^{-1})$ prior

Novel non-local spike-and-slab priors

$$\mathbb{P}(\phi_{jk} \mid \gamma_{jk}, \lambda_0, \lambda_1) = (1-\gamma_{jk})\,\mathbb{P}(\phi_{jk} \mid \lambda_0) + \gamma_{jk}\,\mathbb{P}(\phi_{jk} \mid \lambda_1), \qquad \gamma_{jk} = \{0, 1\}$$

- ⋆ Normal-spike-and-slab
  [George & McCulloch (1993)]
- ⋆ Laplace-spike-and-slab
  [Ročková & George (2016)]

- ★ Normal-spike-and-MOM-slab
- ★ Laplace-spike-and-MOM-slab



Novel non-local spike-and-slab priors

Non-local priors [Johnson & Rossell (2010)]

An absolutely continuous measure with density $\mathbb{P}(\phi_{jk})$ is a non-local prior if $\lim_{\phi_{jk} \to 0} \mathbb{P}(\phi_{jk}) = 0$.

$$\mathbb{P}(\phi_{jk} \mid \lambda_1) = \frac{\phi_{jk}^2}{\lambda_1} N(\phi_{jk}; 0, \lambda_1)$$

Advantages of sparse loadings:

1. Facilitates interpretation of factors as linear combinations of a smaller set of variables.
2. The sparsity assumption has significant potential gains for estimation accuracy.
3. Allows one to work with large q and let the data learn how many factors are needed.
4. Non-local priors have strongest frequentist properties for model selection

Problem: Computational challenges of Non-local priors that can render them impractical

Solution:

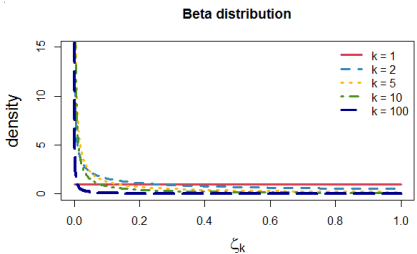
1. Novel scalable EM algorithms
2. Closed-form updates
3. **R** package publicly available at `https://github.com/AleAviP/BFR.BE`

Hyper prior on the latent indicators

Indian buffet process (IBP) prior Griffiths and Ghahramani (2005)

$$\Gamma = \{\gamma_{jk}\}^{P,\infty}$$

$$\gamma_{jk}|\zeta_k \sim \text{Bernoulli}(\zeta_k)$$
$$\zeta_k|\alpha, 1 \sim \text{Beta}(\alpha/k, 1)$$

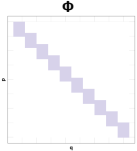


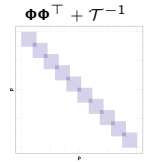
Priors

- Idiosyncratic precisions: $\tau_{jl} \mid \eta, \xi \sim$ Gamma$(\eta/2, \eta\xi/2)$
- Regression parameters: $(\theta_j, \beta_j) \sim N(0, \psi\mathbf{I})$

Simulation

Synthetic data **without batch effects**, $n = 100$, $q^* = 10$, $p = 1,000$ or $1,500$ parameters, truly sparse loadings $\mathbf{\Phi}^*$.

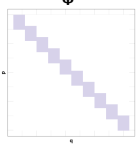


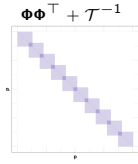


	$p = 1,000$						$p = 1,500$					
Model	$\hat{q}$	$\lvert\hat{\mathbf{\Phi}}\rvert_0$	$\lVert\mathbb{E}[x] - \hat{\mathbb{E}}[x]\rVert_F$	$\lVert\mathrm{Cov}[x_i] - \widehat{\mathrm{Cov}}[x_i]\rVert_F$	it	time	$\hat{q}$	$\lvert\hat{\mathbf{\Phi}}\rvert_0$	$\lVert\mathbb{E}[x] - \hat{\mathbb{E}}[x]\rVert_F$	$\lVert\mathrm{Cov}[x_i] - \widehat{\mathrm{Cov}}[x_i]\rVert_F$	it	time
	$q = 100$											
Flat	100.0	100000.0	209.5	185.7	3.0	5.1	100.0	100000.0	259.2	280.2	3.0	7.1
Normal-SS	31.0	1228.6	109.0	144.6	4.3	9.9	56.4	1568.2	181.3	231.9	4.0	13.8
MOM-SS	**9.7**	856.8	79.4	**143.3**	4.8	**10.8**	**9.2**	745.4	105.0	**245.6**	4.0	**17.3**
FastBFA	83.6	1389.9	198.1	141.9	12.0	1.7	87.2	1763.9	208.2	211.3	6.5	1.0
LASSO-BIC	10.0	4787.3	54.1	271.4	NA	19.6	10.0	7976.6	66.1	409.3	NA	31.6

Simulation

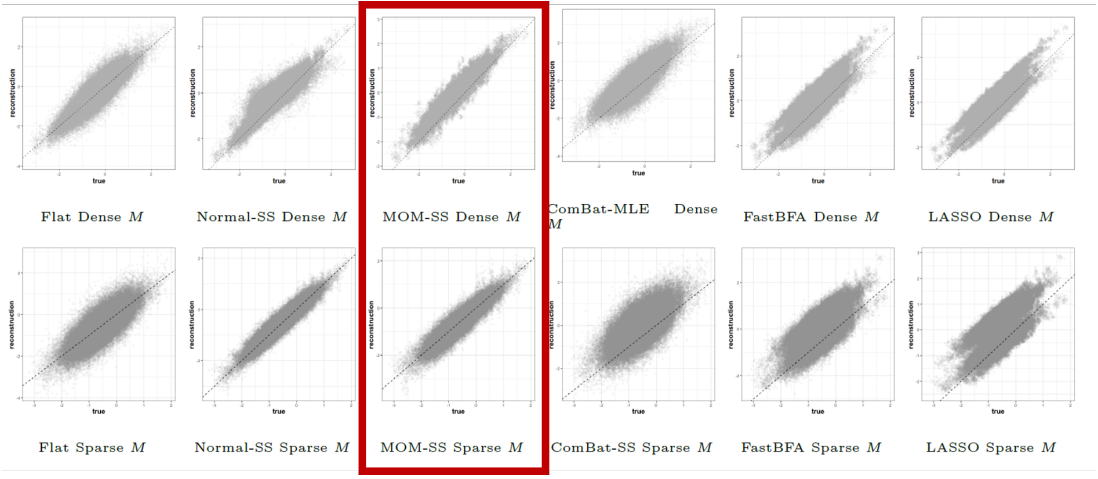
Synthetic data **with batch effects** for $n = 200$, $q^* = 10$, $p = 250$ or 500 parameters, truly sparse loadings $\mathbf{\Phi}^*$.



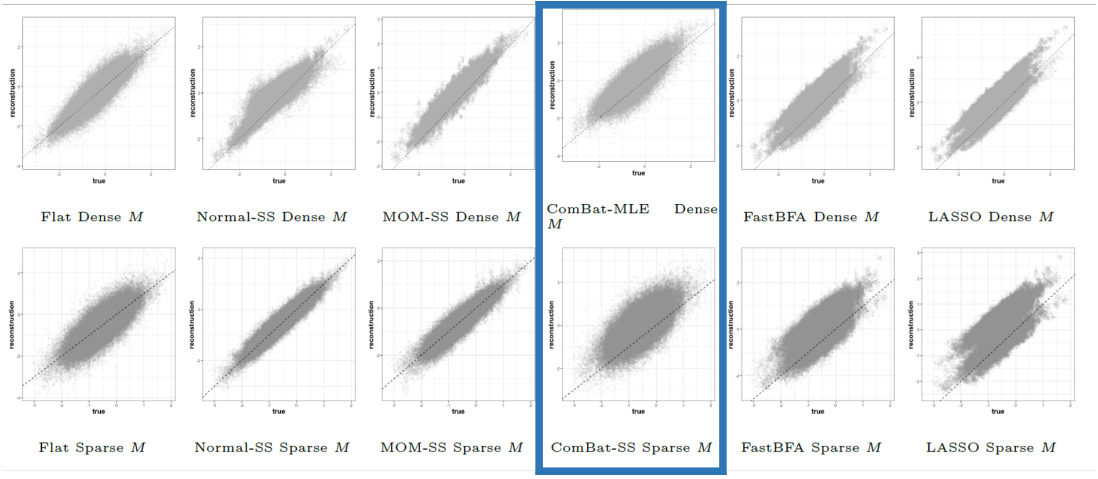


	$p = 250$						$p = 500$					
Model	$\hat{q}$	$\lvert\hat{\mathbf{\Phi}}\rvert_0$	$\lvert\lvert\mathbb{E}[X] - \hat{\mathbb{E}}[X]\rvert\rvert_F$	$\lvert\lvert F\mathbf{\Phi}^\top - \mathbb{E}[F \mid \Delta, X]\hat{\mathbf{\Phi}}^\top\rvert\rvert_F$	it	time	$\hat{q}$	$\lvert\hat{\mathbf{\Phi}}\rvert_0$	$\lvert\lvert\mathbb{E}[X] - \hat{\mathbb{E}}[X]\rvert\rvert_F$	$\lvert\lvert F\mathbf{\Phi}^\top - \mathbb{E}[F \mid \Delta, X]\hat{\mathbf{\Phi}}^\top\rvert\rvert_F$	it	time
	$q = 100$											
Flat	100.0	25000.0	96.8	100.6	4.4	1.3	100.0	25000.0	147.8	152.5	4.0	3.6
Normal-SS	10.0	765.8	45.7	54.8	5.0	1.7	10.6	1146.3	60.0	72.6	5.0	4.0
MOM-SS	**10.0**	740.4	63.8	**72.4**	6.0	**1.5**	**10.0**	1158.7	85.7	**108.3**	5.4	**3.6**
ComBat-MLE	100.0	25000.0	169.0	182.9	8.7	0.0	100.0	25000.0	232.7	252.4	4.9	0.0
FastBFA	10.0	337.0	51.9	168.3	12.7	0.3	11.3	681.8	75.8	247.9	11.9	0.4
LASSO-BIC	10.3	1374.0	39.6	178.9	NA	3.5	10.3	2613.9	49.8	252.1	NA	8.8

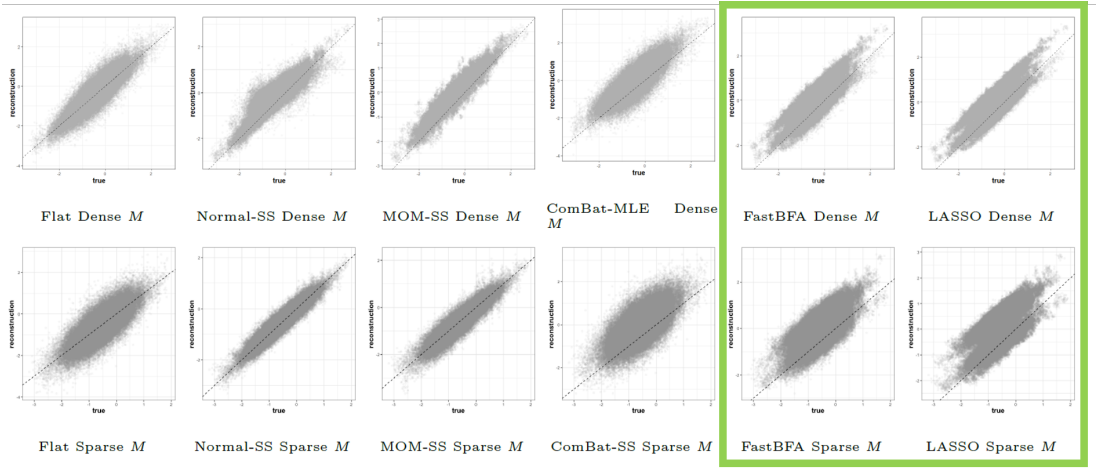
Simulation studies



Simulation studies



Simulation studies



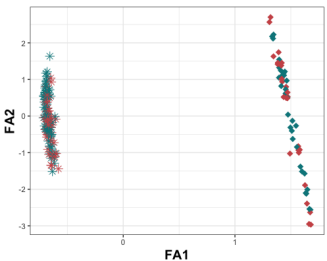
Ovarian cancer datasets

Ovarian cancer: `curatedOvarianData 1.16.0`, $p = 1,007$ genes

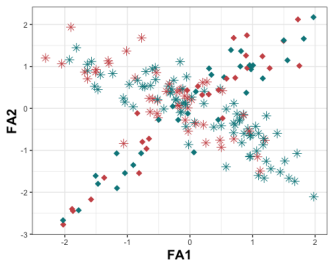
1. Illumina Human microRNA array `E.MTAB.386`, $n_1 = 129$ patients
2. `GSE30161`, $n_2 = 58$ patients

Age at initial pathologic diagnosis has been used as covariate.

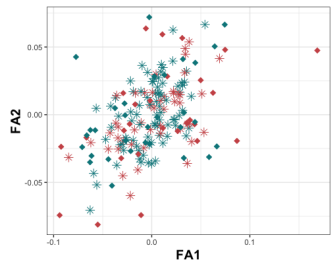
Ov, No correction



Ov, ComBat-MLE



Ov, MOM-SS

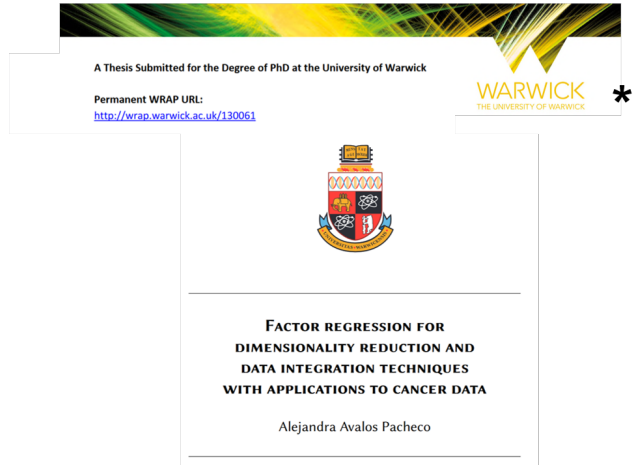


Survival analysis

Survival analysis for ovarian ($p = 1,007$ genes) data sets.

	Ovarian		
	$\hat{q}$	$\lvert\hat{\mathbf{\Phi}}\rvert_0$	CI
Flat	100.0	100700.0	0.634
Normal-SS	7.8	7854.6	0.568
MOM-SS	**4.0**	4028.0	**0.588**
ComBat-MLE 90%	101.0	101707.0	0.589
ComBat-MLE 70%	41.0	41287.0	0.588
ComBat-FastBFA	100.0	100700.0	0.527

Bayesian factor regression


A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:
http://wrap.warwick.ac.uk/130061

WARWICK
THE UNIVERSITY OF WARWICK

*

FACTOR REGRESSION FOR
DIMENSIONALITY REDUCTION AND
DATA INTEGRATION TECHNIQUES
WITH APPLICATIONS TO CANCER DATA

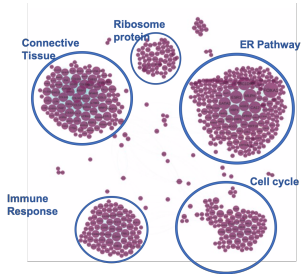
Alejandra Avalos Pacheco

* Winner of the Savage Award 2019, category Applied Methodology, awarded by the International Society of Bayesian Statistics (ISBA)

Outline

Multi-study factor analysis[2,3]

- Model: $\mathtt{x}_{is} = \boldsymbol{\Phi}\mathtt{f}_{is} + \boldsymbol{\Lambda}_s\mathtt{l}_{is} + \mathtt{e}_{is}$
  $i = 1, \ldots, n_s, s = 1, \ldots, S$
  - $\boldsymbol{\Phi} \in \mathbb{R}^{p \times q}$: common factor loadings
  - $\boldsymbol{\Lambda}_s \in \mathbb{R}^{p \times q_s}$: specific factor loadings
  - $\mathtt{f}_{is} \in \mathbb{R}^q$: common factors
  - $\mathtt{l}_{is}$: batch indicators
  - $\mathtt{e}_{is} \sim N(0, \boldsymbol{\mathcal{T}}_s^{-1})$, $\boldsymbol{\mathcal{T}}_s = \text{diag}\{\boldsymbol{\mathcal{T}}_{1s}, \ldots, \boldsymbol{\mathcal{T}}_{ps}\}$.



$$\Sigma_s = \underbrace{\boldsymbol{\Phi}\boldsymbol{\Phi}^\top}_{\Sigma_{\boldsymbol{\Phi}}} + \underbrace{\boldsymbol{\Lambda}_s\boldsymbol{\Lambda}_s^\top}_{\Sigma_{\boldsymbol{\Lambda}_s}} + \boldsymbol{\mathcal{T}}_s^{-1}$$

- $\Sigma_{\boldsymbol{\Phi}}$: crucial for identifying reproducible biological pathways shared by different cancer studies that traditional factor analysis approaches may miss due to batch effects[2,3].

[2] De Vito et. al. (2019) Biometrics
[3] De Vito et. al. (2021) The Annals of Applied Statistics

Multiplicative gamma process shrinkage prior[4]

- common factor loadings

$$\phi_{jk} \mid \omega_{jk}\gamma_k \sim N(0, \omega_{jk}^{-1}\gamma_k^{-1}),$$

$$\omega_{jk} \sim \Gamma(\tfrac{\eta}{2}, \tfrac{\eta}{2}), \quad \gamma_k = \textstyle\prod_l \delta_l, \quad \delta_1 \sim \Gamma(a_1, 1) \quad \delta_l \sim \Gamma(a_2, 1)$$

$$l = 1, \ldots, \infty$$

- study-specific factor loadings

$$\lambda_{jk_s} \mid \omega_{jk_s}^{s}\gamma_{k_s}^{s} \sim N(0, \omega_{jk_s}^{s-1}\gamma_{k_s}^{s-1})$$

$$l = 1, \ldots, \infty$$

[4] Battacharya and Dunson (2011), Biometrika

BMSFA: Shrinkage prior

- Infinitely many factors:

$$\boldsymbol{\Phi} \in \mathbb{R}^{p \times \infty} \text{ and } \boldsymbol{\Lambda}_s \in \mathbb{R}^{p \times \infty}$$

- Degree of shrinkage increases across the column index
- Use of the prior on a parameter-expanded loading matrix
- Inference via Gibbs sampler

Challenges:

1. MCMC implementations of existing BMSFA methods are
   - computationally expensive
   - not scalable in high dimensional settings
2. Current MSFA methods do not take into account covariate effects
   - batch effects
   - pre-treatment patient characteristics

Multi-study factor regression (MSFR)[7,8]

- Challenge: Keeping track of the observed variables
- Solution: Obtain a covariance structure that models the study/batch specific covariances in addition to the common component, keeping track of the observed variables, such as the demographic information
- Model:

$$\mathtt{x}_{is} = \mathbf{\Phi}\mathtt{f}_{is} + \mathbf{\Lambda}_s\mathtt{l}_{is} + \beta\mathtt{b}_{is} + \mathtt{e}_{is}$$

  $i = 1, \ldots, n_s, s = 1, \ldots, S$
- **R**-package available at: `https://github.com/rdevito/MSFR` (Authors: De Vito R. and Avalos-Pacheco A.)

[7] De Vito R.,Avalos-Pacheco A., *Multi-study Factor Regression for Heterogeneous Data: An Application in Nutritional Epidemiology* arXiv:2304.13077 [submited]

[8] Avalos-Pacheco A., Jewson J., Rossell D., De Vito R., *Sparse Bayesian Factor Models for Single and Multi-Study Data* [working project]

Simulation studies

	$\hat{q}$	$\hat{q}_s$	$\hat{\beta}$	$\hat{\Phi}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\lambda}_5$	$\hat{\lambda}_6$	$\hat{\Sigma}_1$	$\hat{\Sigma}_2$	$\hat{\Sigma}_3$	$\hat{\Sigma}_4$	$\hat{\Sigma}_5$	$\hat{\Sigma}_6$
Scenario 1: $q = 3, q_s = 1, S = 2, p_b = 2, p = 20, n_s = \{500, 500, 500\}$																
MSFR	3.00	1.00	0.978	0.996	0.959	0.931					0.974	0.987				
MSFA&LR	3.00	1.00	0.993	0.973	0.922	0.875					0.923	0.942				
MSFA	4.95	1.04	NA	0.915	0.900	0.844					0.889	0.904				
FR	3.57	NA	0.991	0.956	NA	NA					0.912	0.888				
Scenario 2: $q = 4, q_s = 1, S = 6, p_b = 7, p = 42, n_s = \{1257, 1444, 2126, 4940, 2314, 897\}$																
MSFR	4.00	1.00	0.952	1.000	0.983	0.982	0.990	0.995	0.991	0.981	0.999	0.999	0.999	0.999	0.999	0.999
MSFA&LR	1.14	4.13	0.999	0.602	0.017	0.036	0.033	0.002	0.006	0.011	0.841	0.848	0.858	0.866	0.832	0.847
MSFA	8.22	2.51	NA	0.953	0.927	0.908	0.939	0.982	0.973	0.927	0.965	0.965	0.964	0.966	0.966	0.965
FR	5.00	NA	0.998	0.971	NA	NA	NA	NA	NA	NA	0.860	0.874	0.868	0.872	0.876	0.892
Scenario 3: $q = 4, q_s = 1, S = 6, p_b = 9p = 42, n_s = \{1257, 1444, 2126, 4940, 2314, 897\}$																
MSFR	3.99	1.00	0.940	0.999	0.970	0.975	0.982	0.986	0.981	0.971	0.999	0.999	0.999	0.999	0.999	0.999
MSFA&LR	1.36	4.17	0.998	0.617	0.033	0.074	0.050	0.022	0.026	0.030	0.852	0.856	0.864	0.869	0.847	0.851
MSFA	8.06	2.28	NA	0.923	0.925	0.903	0.936	0.982	0.976	0.920	0.929	0.930	0.929	0.929	0.931	0.928
FR	5.00	NA	0.998	0.971	NA	NA	NA	NA	NA	NA	0.859	0.874	0.867	0.871	0.876	0.892

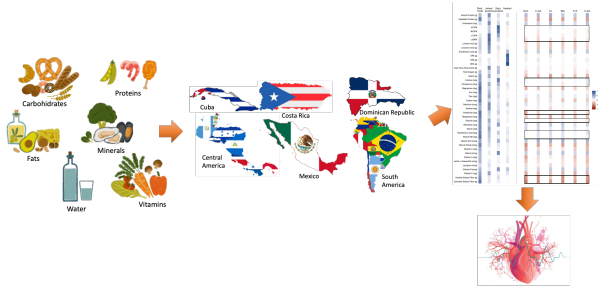
Dietary patterns in the HCHS/SOL data

Dietary patterns

- Synthesize multiple related dietary components in one or more combined variables.
- Drawback: Their limited reproducibility across subpopulations
- Solution: Use MSFR to study the **shared** and **ethnic background site-specific** dietary patterns in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), while controlling for **covariates** or confounders that do not affect the dietary patterns like: education level and alcohol intake.

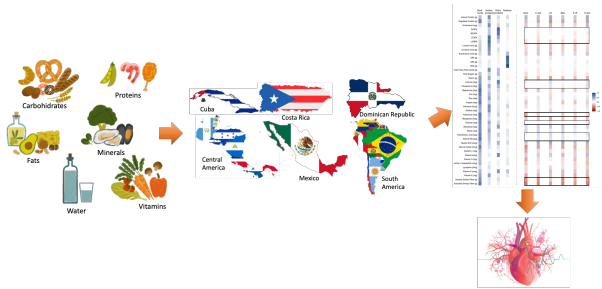
The HCHS/SOL data

- We study data from the HCHS/SOL study:
  - 12978 enrolled adults aged 18–74 years
  - from six Hispanic/Latino ethnic backgrounds $S = 6$ (Cuban, Dominican Republic, Mexican, Puerto Rican, Central and South American) at baseline (2008-2011), $n_s = \{1257, 1444, 2126, 4940, 2314, 897\}$.
  - data from $p = 42$ nutrients
  - non-nutritional data in our model: BMI, Gender, education levels (3), alcohol intake (3), and number of cigarette packs (3), obtaining $p_b = 9$ levels.

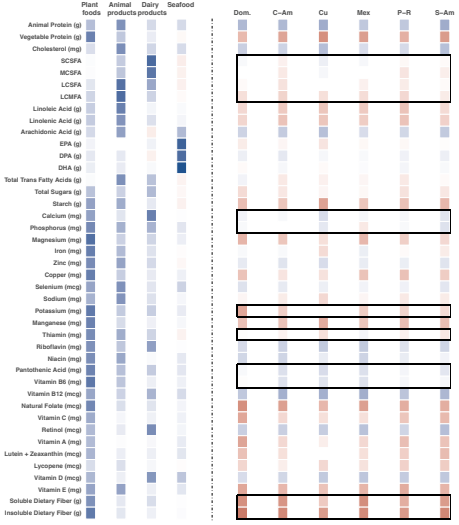


Aims

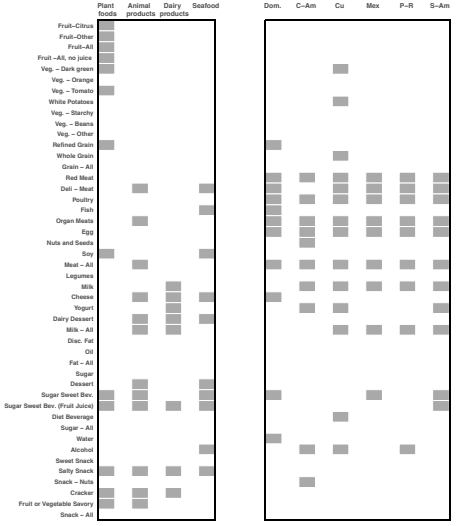
1. analyze the obtained factors loadings of the data
2. study the loading association with different food groups
3. validate our obtained dietary patterns, associating each factor with the alternative eating index (AHEI-2010):
   - a measure of overall diet quality
   - related to cardiometabolic disease risk (Liese et al., 2015)
   - the higher the AHEI-2010 index, the lower the chronic disease risk.



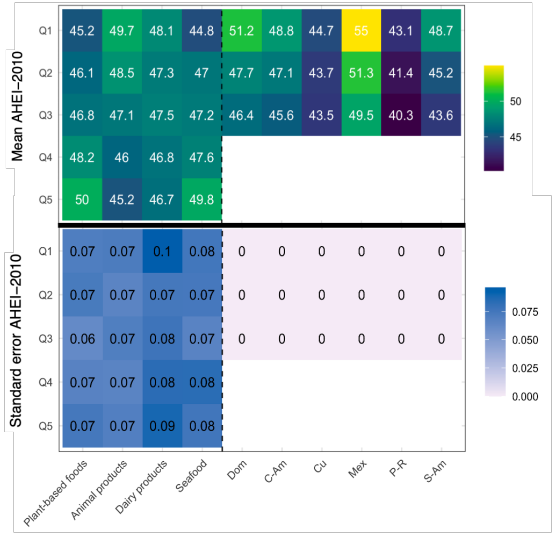
Heat-maps of Φ and λ_s



Φ and λ_s association with food groups



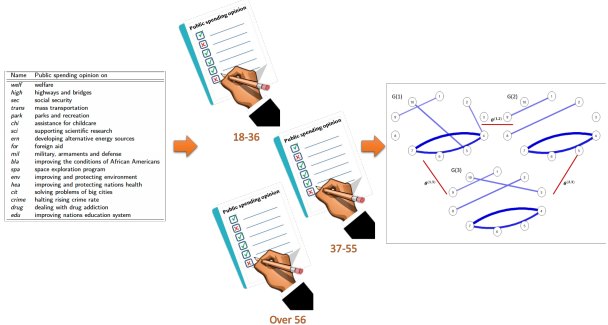
AHEI-2010 index for HCHS/SOL data



Outline

Multiple Ising Graphs

- Challenge: Our models can only deal with continuous outcome data.
- Solution: A model to study the heterogeneity induced in a set of **binary variables** by external factors.



Multiple Ising Graphs[9]

- Multiple graphical models for Gaussian random variables have been widely studied
  - Guo et al. (2011), Danaher et al. (2014), Peterson et al. (2015), Ha et al. (2021)
  - they can capture the heterogeneity of the data involved in more realistic settings
- There have been only few proposals for multinomial sampling models
  - Hojsgaard (2003), Corander (2003), Nyman et al. (2014, 2016).
- Multiple Ising models:
  - More general, allow context-specific independences to vary not only with respect to adjacent vertices.
  - Model the heterogeneity induced in a set of binary variables by external factors.
  - Factors may influence the joint dependence relationships represented by a set of graphs across different groups.

[9] Lazzerini A., Avalos-Pacheco A., Lupparelli M., Stingo F. C. **Bayesian Inference of Multiple Ising Models for Heterogeneous Public Opinion Survey Networks** [submitted].

Conclusions

- Methods to **integrate** large datasets from **multiple studies** with **continuous** and **binary** outputs.
- **Link** and borrow strength across related sub-populations, via **efficient** EM, ECM and VI algorithms.
- The non-local prior based formulations induce sparsity and learn the number of factors.
- The MRF prior on the binary indicator of edge inclusion
  - encourage the selection of the same edges in related graphs
  - can learn which sup-populations are similar, and which ones are not.
- We show the usefulness of our approaches in:
  1. cancer genomics
  2. nutriotional epidemiology
  3. public opinion studies in the US
- **R**-code publicly available

References

Factor Models

1. Avalos-Pacheco A., De Vtio R. **Integrative Factor Models for Heterogeneous Biomedical Data**. CLADAG conference proceeding, [submitted].
2. Avalos-Pacheco A., Rossell D., Savage R. (2022) **Heterogeneous large datasets integration using Bayesian latent factor regression**, *Bayesian Analysis*, 7(1) pp. 33–66, doi.org/10.1214/20-ba1240.
3. Avalos-Pacheco A., (2019) **Factor regression for dimensionality reduction and data integration techniques with applications to cancer data**, . *PhD thesis, University of Warwick*, doi.org/10.1214/20-ba1240.

Multi-study Factor Models

4. Hansen B., Avalos-Pacheco A., Russo M., De Vito R., **A Variational Bayes Approach to Factor Analysis**, Bayesian Statistics, New Generations New Approaches. BAYSM 2022, Springer Proceedings in Mathematics and Statistics, [in press].
5. Hansen B., Avalos-Pacheco A., Russo M., De Vito R., **Fast Variational Inference for Bayesian Factor Analysis in Single and Multi-Study Settings** [in preparation & available up to request].
6. De Vito R., Avalos-Pacheco A., **Multi-study Factor Regression Model: An Application in Nutritional Epidemiology**, arXiv:2304.13077, [submitted].
7. Avalos-Pacheco A., Jewson J., Rossell D., De Vito R., **Sparse Bayesian Factor Models for Single and Multi-Study Data** [working project].

Graphical Models

8. Lazzerini A., Avalos-Pacheco A., Lupparelli M., Stingo F. C. **Bayesian Inference of Multiple Ising Models for Heterogeneous Data** [submitted].
9. Avalos-Pacheco A., Lazzerini A., Lupparelli M., Stingo F. C. **Profile Undirected Graphical Models** [working project].

R Packages

10. BFR, Avalos-Pacheco A.: `https://github.com/AleAviP/BFR.BE`
11. MSFR, De Vito R., Avalos-Pacheco A.: `https://github.com/rdevito/MSFR`

R code

12. MIG: Lazzerini A. `https://github.com/kinglaz90/phd`