Markov Switching Multiple-equation Tensor Regressions

Roberto Casarin [1] Radu Craiu [2] Qing Wang [1]

[1]Ca' Foscari University of Venice [2]University of Toronto

Ca' Foscari University of Venice
Department of Economics

Introduction

Motivation

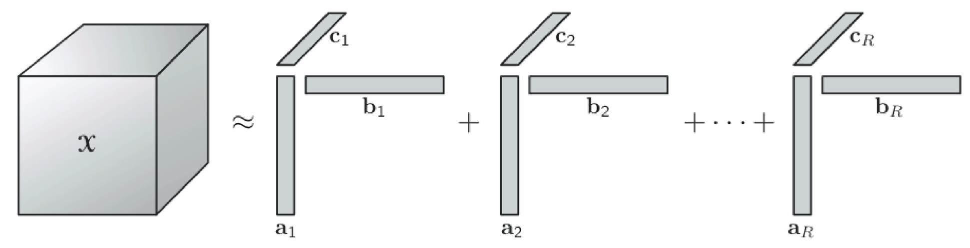
- Data arises naturally in high-dimensional array (tensor) structure in many applications, neuroimaging, spatial-temporal analysis, computer vision, financial networks, etc.
- Often people are interested in characterizing the relationship between a scalar outcome and tensor covariates (predictors).

Contributions:

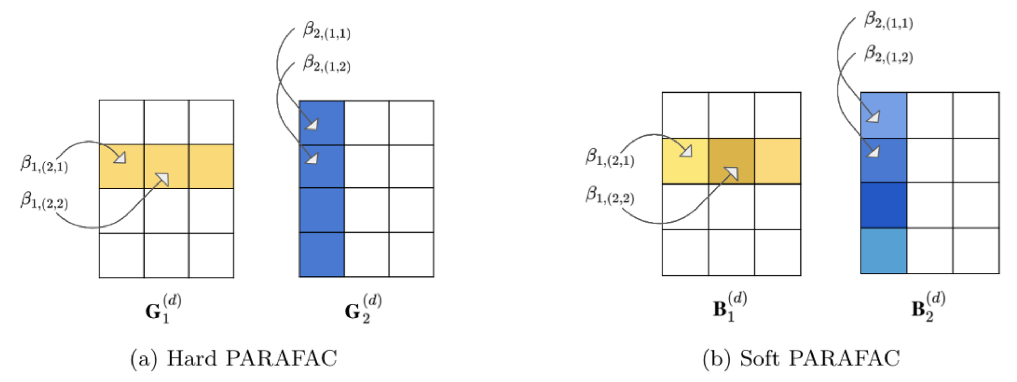
- Introduce a new flexible tensor model for multiple-equation regression that accounts for latent regime changes.
- Provide a suitable inference framework to deal with over-parametrization and overfitting.
- Propose an efficient MCMC algorithm for posterior approximation (Random Scan Gibbs Sampling and back-fitting strategy).

Dimensionality Reduction

We perform dimensionality reduction on the tensor coefficients using Soft PARAFAC (Papadogeorgou et al. 2021) decomposition to preserve structure information.
PARAFAC:



Hard Vs Soft PARAFAC:



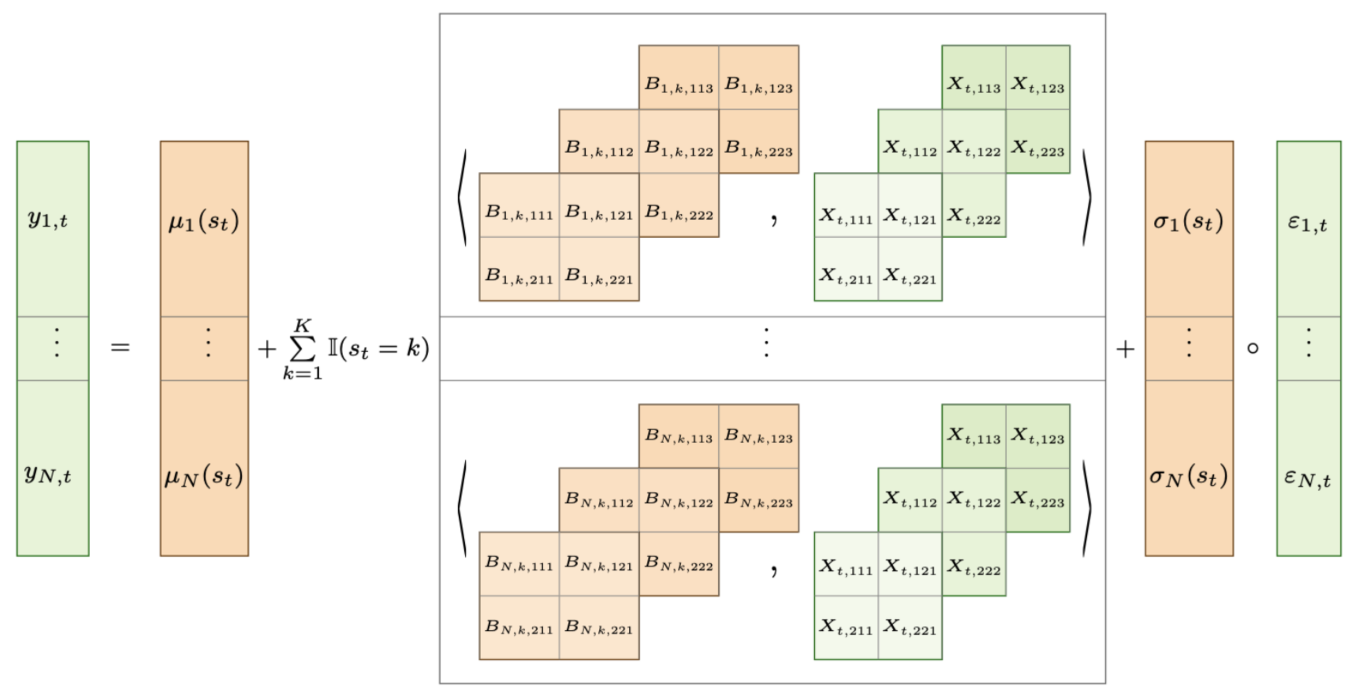
where

$$\beta^{(d)}_{m,j_m} \sim \mathcal{N}_{q_m}\left(\gamma^{(d)}_{m,j_m}, \tau\sigma^2_m\zeta^{(d)} I_{q_m}\right)$$

Soft PARAFAC enables a higher-rank approximation using a low-rank decomposition for the tensor coefficients.

The Model

A Markov-Switching Multiple-equation Tensor Regression Model:



where $t = 1, \ldots, T$, $X_t, B_\ell(s_t)$ are $p_1 \times p_2$ matrices, $\langle \cdot, \cdot \rangle$ denotes inner product. The latent process is a K-state Markov chain process and the parametrization used is

$$\mu_\ell(s_t) = \sum_{k=1}^{K} \mu_{\ell k}\mathbb{I}(s_t = k), \quad B_\ell(s_t) = \sum_{k=1}^{K} B_{\ell k}\mathbb{I}(s_t = k), \quad \sigma_\ell(s_t) = \sum_{k=1}^{K} \sigma_{\ell k}\mathbb{I}(s_t = k)$$

Assume the following decomposition:

$$B_{\ell k} = \sum_{d=1}^{D} B^{(d)}_{\ell,k,1} * B^{(d)}_{\ell,k,2}$$

where $*$ is the Hadamard product, $B^{(d)}_{\ell,k,m}, m = 1, 2$ are the multiplicative factors. D is the number of components used to decompose the tensor.

The Hierarchical Priors

We propose a new multi-way shrinking prior (Guhaniyogi et al. 2017) to address over-parametrization and control shrinkage effects at different levels:

$$\gamma^{(d)}_m \sim \mathcal{N}_{p_m}(\mathbf{0}, \tau\zeta^{(d)} W^{(d)}_m)$$

$$w^{(d)}_{m,j_m} \sim \mathcal{E}xp((\lambda^{(d)}_m)^2/2)$$

$$\lambda^{(d)}_m \sim \mathcal{G}a(a_\lambda, b_\lambda)$$

$$\kappa^2_m \sim \mathcal{G}a(a_\kappa, b_\kappa)$$

$$\tau \sim \mathcal{G}a(a_\tau, b_\tau)$$

$$(\zeta^{(1)}, \ldots, \zeta^{(D)}) \sim \mathcal{D}ir(\alpha/D, \ldots, \alpha/D)$$

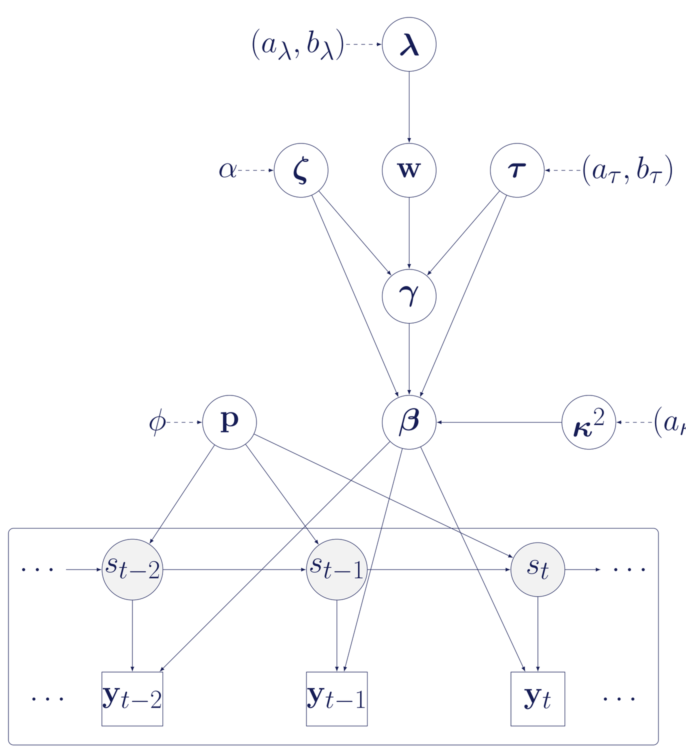


Figure 1. DAG of the Bayesian Markov-switching Matrix Regression model.

MCMC Algorithm

The joint posterior we are interested in is $p(\boldsymbol{\theta}, \boldsymbol{s} \mid \boldsymbol{y}, \boldsymbol{X})$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$ is the collection of the state-specific parameters $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k, \boldsymbol{\zeta}_k, \tau_k, \boldsymbol{\lambda}_k, \boldsymbol{w}_k, \boldsymbol{\sigma}_k, \boldsymbol{\mu}_k)$ and $\mathbf{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$, $\mathbf{X} = (X_1, \ldots, X_T)$, $\boldsymbol{s} = (s_1, \ldots, s_T)$ are the collection of response variables, covariates and state variables, respectively.

The joint posterior is not tractable, we use the full conditionals of the parameters to approximate it.

We propose a MCMC procedure based on Gibbs sampling to sample the unknowns from 3 blocks.

For the hidden states, we apply a Forward Filtering Backward Sampling (FFBS) strategy.

We perform Random-Partial-Scan Gibbs to randomly select a subset of components to update for each iteration to improve the efficiency of the Gibbs Sampler after the first 10 iterations.

Simulation Results

Markov-switching Tensor Regression

Simulation settings

- 2 sets of true coefficients are used to represent 2 different regimes, both i.i.d covariates and AR(1) covariates are used in the simulation.
- Matrix predictor with dimensions 14 x 14
- Regime specific intercepts: $\mu_1 = \mu_2 = 0$
  Regime specific variances: $\sigma_1^2 = 2, \sigma_2^2 = 0.1$.
- Number of observations: 800
- Gibbs iterations: 3000

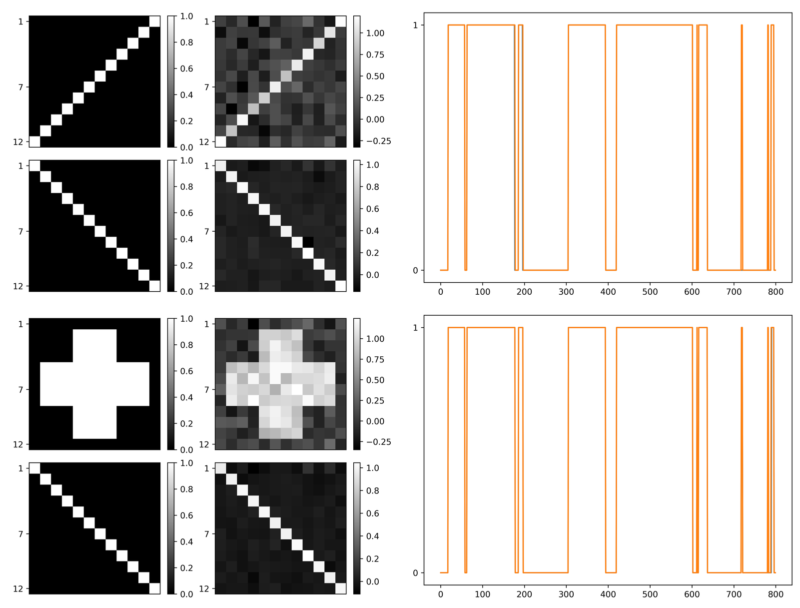


Figure 2. Markov-switching model with Diagonal and Anti-diagonal coefficients (first row) and with Cross and Diagonal coefficients (second row).

Empirical Applications

Oil prices on S&P 500 and disaggraggated markets
We examine the impact of oil price volatility on the stock market returns (S&P 500) at an aggregate level and on the financial sector, energy sector and other sectors of S&P 500 at the disaggregate level.

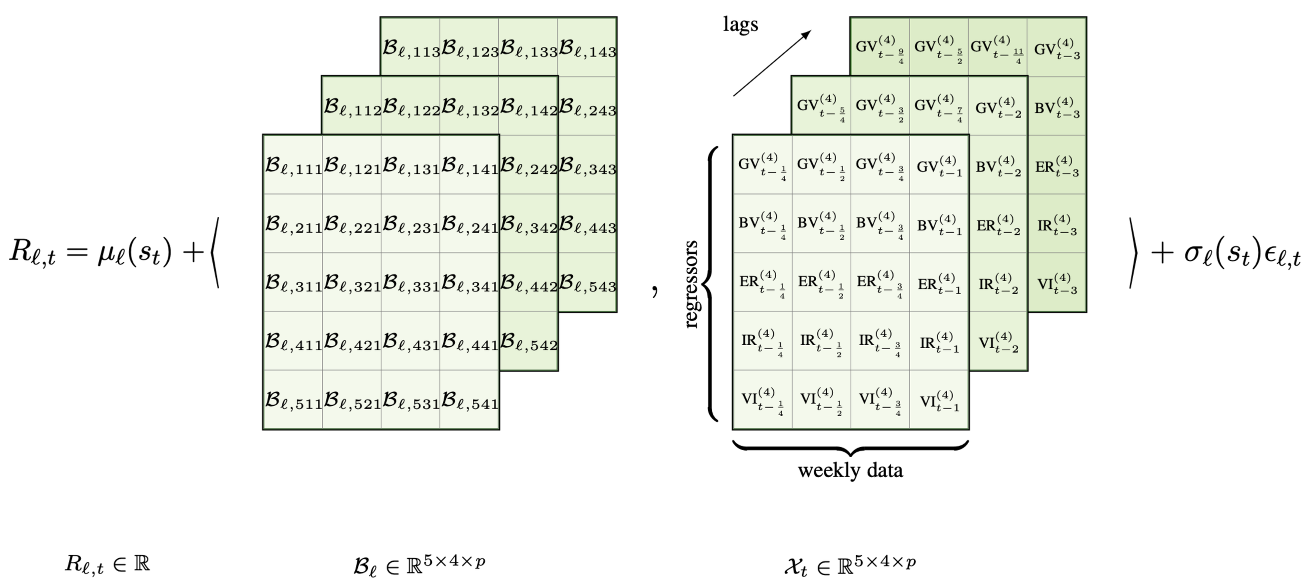


Figure 3. Graphic Representation of Tensor Regression for Macro Application

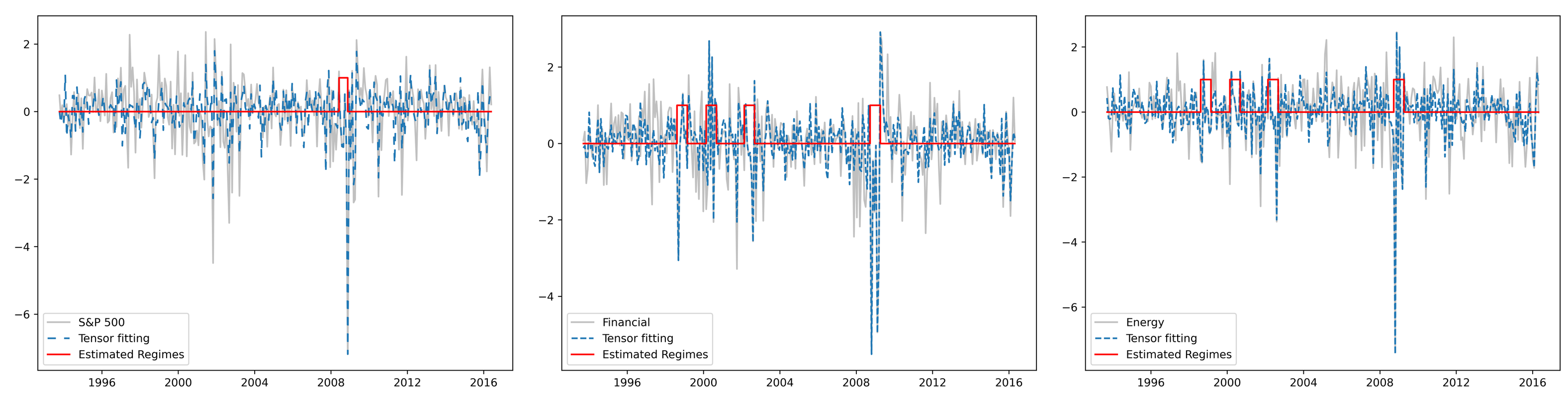


Figure 4. Tensor Regression with Markov Switching (blue dashed line) and estimated hidden states (red solid line). True data is shown in solid silver line

Conclusions

- A new Markov-switching tensor regression model is proposed where a hidden Markov chain process allows for structural changes in the parameters of the regression model.
- A low-rank representation of the coefficient tensor and hierarchical prior distribution are proposed to introduce shrinkage effects to overcome overparametrization.
- An efficient MCMC sampler is proposed based on back-fitting and random scan strategies.
- The tensor regression model is readily to be used with tensor covariates with order 2 or 3.

References

Papadogeorgou, G., Z. Zhang, and D. B. Dunson (2021). "Soft Tensor Regression.". In: J. Mach. Learn. Res. 22, pp. 219–1.

Guhaniyogi, R., S. Qamar, and D. B. Dunson (2017). "Bayesian tensor regression". In: The Journal of Machine Learning Research 18.1, pp. 2733–2763.

qing.wang@unive.it