Chapter 14
Markov Switching Tensor Regressions

Roberto Casarin, Radu Craiu and Qing Wang

Abstract A new flexible tensor-on-tenor regression model that accounts for latent regime changes is proposed. The coefficients are driven by a common hidden Markov process that addresses structural breaks to enhance the model flexibility and preserve parsimony. A new soft PARAFAC hierarchical prior is introduced to achieve dimensionality reduction while preserving the structural information of the covariate tensor. The proposed prior includes a new multi-way shrinking effect to address over-parametrization issues while preserving interpretability and model tractability. An efficient MCMC algorithm is introduced based on random scan Gibbs and back-fitting strategy. The model framework's effectiveness is illustrated using financial and commodity market volatility data. The proposed model exhibits superior performance compared to the current benchmark, Lasso regression.

14.1 Introduction

As data grow in volume and complexity, it is increasingly common to record them as high-dimensional arrays or tensors. Such structures appear in many applications and fields, such as neuroimaging [8, 11], biostatistics, financial networks [2], or even more generally, in time series [3]. People are often interested in characterizing the relationship between a tensor predictor and a scalar outcome [8] or tensor outcome [15]. Tensor regression has been studied extensively in a linear model framework. Nevertheless, a common challenge within the framework of regression is model

Roberto Casarin
Ca' Foscari University of Venice, Italy, e-mail: r.casarin@unive.it

Radu Craiu
University of Toronto, Canada, e-mail: radu.craiu@utoronto.ca

Qing Wang (✉)
Ca' Foscari University of Venice, Italy, e-mail: qing.wang@unive.it

G. Aneiros et al. (eds.), *New Trends in Functional Statistics and Related Fields*,
Contributions to Statistics, https://doi.org/10.1007/978-3-031-92383-8_14

misspecification. One of the sources of model misspecification is the presence of dynamic regimes, which naturally call for models with time-varying parameters.

In this paper, we assume a Hidden Markov chain dynamics for the regression coefficients because it allows for model parsimony while preserving a high level of flexibility compared to other time-varying parameter models. The model extends the soft tensor linear regression of [8, 14, 15] to an HMM or MS framework to accommodate structural breaks. In our multi-equation setting, the latent process is common to several tensor regression equations involving different response variables and, possibly, different sets of covariates. Using a common latent process facilitates the integration of information about the latent process from multiple outcomes and robustifies the estimation accounting for structural breaks. Regarding inference, we consider a Bayesian inference procedure combined with an efficient Gibbs sampler, which reduces computational costs and improves scalability [5].

14.2 A Markov-switching Tensor Regression Model

In our Markov Switching Tensor Regression Model (MSTR), we assume a system of N equations with time-varying parameters

$$\mathbf{y}_t = \boldsymbol{\mu}(s_t) + X_t \times_{1:M} \boldsymbol{B}(s_t) + \Sigma^{1/2}(s_t)\boldsymbol{\varepsilon}_t, \tag{14.1}$$

$t = 1, \ldots, T$, where $\times_{m:n}$ denotes the tensor contract product along the modes from m to n [10], $\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})'$ is the collection of response variables across equations, X_t is a $p_1 \times \ldots \times p_{M-1}$ covariate tensor, $\boldsymbol{B}(s_t) = (B_1(s_t), \ldots, B_N(s_t))$ is a M-modes coefficient tensor of size $p_1 \times \ldots \times p_{M-1} \times N$ and $B_\ell(s_t)$ is a $p_1 \times \ldots \times p_{M-1}$ coefficient tensor, $\boldsymbol{\mu}(s_t) = (\mu_1(s_t), \ldots, \mu_N(s_t))'$ an intercept vector, $\Sigma^{1/2}(s_t)$ denotes the Cholesky's decomposition of the positive definite $N \times N$ covariance matrix $\Sigma(s_t)$, $\boldsymbol{\varepsilon}_t \sim \mathcal{N}_N(0, I_N)$, i.i.d. for $t = 1, \ldots, T$.

The latent process $\{s_t, t = 1, \ldots, T\}$ is a K-state homogeneous Markov chain with transition probability $\mathbb{P}(s_t = j | s_{t-1} = i) = p_{ij}$, $i, j = 1, \ldots, K$ and the tensor regression parametrization used is

$$\mu_\ell(s_t) = \sum_{k=1}^{K} \mu_{\ell k} \mathbb{I}(s_t = k), B_\ell(s_t) = \sum_{k=1}^{K} B_{\ell k} \mathbb{I}(s_t = k), \Sigma(s_t) = \sum_{k=1}^{K} \Sigma_k \mathbb{I}(s_t = k).$$

See [7] and [4] for alternative coefficient parameterisations.

Since, in many applications, the number of covariates in Eq. 14.1 is large, a dimensionality reduction strategy is needed. In this paper, we follow [14] and [2, 3] and consider a low-rank representation combined with a hierarchical prior distribution. The hierarchical prior allows for shrinking effects in the coefficient matrices $B_{\ell k}$, $k = 1, \ldots, K$, and the low-rank representation induces further shrinking effects along different modes. To simplify exposition, we assume $M = 3$, and the number of elements in each mode is p_1, p_2 and N. In our PARAFAC representation, the

state-specific coefficient matrix is written as follows:

$$B_{\ell k} = \sum_{d=1}^{D} B_{\ell,1,k}^{(d)} \circ B_{\ell,2,k}^{(d)}, \tag{14.2}$$

where $\circ$ is the element-by-element Hadamard product and $B_{\ell,m,k}^{(d)}$ $m = 1, 2$ are multiplicative factors as in [14]. The soft PARAFAC prior distribution includes three stages. At the first stage, we assume an inverse gamma prior distribution $\mathcal{IG}(a_\sigma, b_\sigma)$ for $\sigma_{\ell,k}^2$ and a matrix-variate normal distribution for the coefficient tensor:

$$B_{\ell,m,k}^{(d)} \sim \mathcal{MN}_{p_1,p_2}\left(G_{\ell,m,k}^{(d)}, \tau_{\ell,k}\kappa_{\ell,m,k}^2\zeta_{\ell,k}^{(d)} I_{p_1}, I_{p_2}\right), \tag{14.3}$$

where $\mathcal{MN}_{p_1,p_2}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ denotes the matrix-variate normal distribution with $p_1 \times p_2$ mean matrix $\mathbf{M}$, $p_1 \times p_1$ row covariance matrix $\mathbf{U}$ and $p_2 \times p_2$ column covariance matrix $\mathbf{V}$. See [9]. The location matrix $G_{\ell,m,k}^{(d)}$ is parametrized as follows:

$$G_{\ell,m,k}^{(d)} = \begin{cases} \boldsymbol{\gamma}_{\ell,1,k}^{(d)} \otimes \boldsymbol{\iota}_{p_2}, & \text{if } m = 1, \\ \boldsymbol{\iota}_{p_1} \otimes \boldsymbol{\gamma}_{\ell,2,k}^{(d)}, & \text{if } m = 2, \end{cases}$$

where $\otimes$ denotes the *outer product*, $\boldsymbol{\iota}_n = (1, \ldots, 1)'$ is the n-dimensional unit vector, $\boldsymbol{\gamma}_{\ell,1,k}^{(d)}$ and $\boldsymbol{\gamma}_{\ell,2,k}^{(d)}$ are the PARAFAC margins, which are vectors of sizes p_1 and p_2, respectively. In the conditional mean of the factors $B_{\ell,k}^{(d)}$, we have

$$\mathbb{E}\left(B_{\ell,k} \mid \boldsymbol{\gamma}_{\ell,1,k}^{(d)}, \boldsymbol{\gamma}_{\ell,2,k}^{(d)}\right) = \sum_{d=1}^{D} \mathbb{E}\left(B_{\ell,1,k}^{(d)}\right) \circ \mathbb{E}\left(B_{\ell,2,k}^{(d)}\right) = \sum_{d=1}^{D} \left(G_{\ell,1,k}^{(d)} \circ G_{\ell,2,k}^{(d)}\right)$$
$$= \sum_{d=1}^{D} (\boldsymbol{\gamma}_{\ell,1,k}^{(d)} \circ \boldsymbol{\iota}_{p_1}) \otimes (\boldsymbol{\gamma}_{\ell,2,k}^{(d)} \circ \boldsymbol{\iota}_{p_2}) = \sum_{d=1}^{D} \boldsymbol{\gamma}_{\ell,1,k}^{(d)} \otimes \boldsymbol{\gamma}_{\ell,2,k}^{(d)}.$$

In the second stage, we assume that the margins from the PARAFAC decomposition follow a multivariate normal distribution:

$$\boldsymbol{\gamma}_{\ell,m,k}^{(d)} \sim \mathcal{N}_{p_m}(\mathbf{0}, \tau_{\ell,k}\zeta_{\ell,k}^{(d)} W_{\ell,m,k}^{(d)}), \tag{14.4}$$

and assume the distributions are centred around the null vector and have random scales to allow for shrinkage at different levels. At the third stage, we borrow from [13] and specify the priors for scale parameters to induce shrinkage across PARAFAC components and fibers:

$$\tau_{\ell,k} \sim \mathcal{G}a(a_\tau, b_\tau), \quad \kappa_{\ell,m,k}^2 \sim \mathcal{G}a(a_\kappa, b_\kappa), \quad w_{\ell,m,j_m,k}^{(d)} \sim \mathcal{E}xp((\lambda_{\ell,m,k}^{(d)})^2/2),$$
$$\lambda_{\ell,m,k}^{(d)} \sim \mathcal{G}a(a_\lambda, b_\lambda), \quad \left(\zeta_{\ell,k}^{(1)}, \ldots, \zeta_{\ell,k}^{(D)}\right) \sim \mathcal{D}ir(\alpha/D, \ldots, \alpha/D),$$

where $\mathcal{G}a(a,b)$, $\mathcal{E}xp(\lambda)$ and $\mathcal{D}ir(\nu_1,\ldots,\nu_D)$ denote the Gamma, Exponential and Dirichlet distributions, respectively. Compared to [14] our prior assumes the global scale $\tau_{\ell,k}$ contributes not only to the variance of $\boldsymbol{\gamma}^{(d)}_{\ell,m,k}$ but also to that of one of the tensor coefficients $B^{(d)}_{\ell,m,k}$. This allows for stronger shrinkage effects and a full factorization of the prior variance, as detailed below. The matrices $W^{(d)}_{\ell,m,k} = \text{diag}(w^{(d)}_{\ell,1,1,k},\ldots,w^{(d)}_{\ell,m,j_m,k},\ldots,w^{(d)}_{\ell,M,p_M,k})$, where $j_m = 1,\ldots,p_m$ denotes the j_mth element along mode m, are the row-specific parameters that shrink the individual elements of the margins. Together with the prior on $\lambda^{(d)}_{\ell,m,k}$, they lead to an adaptive LASSO-type penalty on $\boldsymbol{\gamma}^{(d)}_{\ell,m,k}$ ([1]). The parameter $\zeta^{(d)}_{\ell,k}$ is component-specific and allows a subset of the D components to contribute substantially to the PARAFAC approximation while leaving the values of other components close to zero. The transition probabilities $(p_{1k},\ldots,p_{Kk})$ are assumed to follow a Dirichlet distribution:

$$(p_{1k},\ldots,p_{Kk}) \sim \mathcal{D}ir(\nu_1,\ldots,\nu_K). \tag{14.5}$$

Dimensionality reduction is achieved by noticing that the number of equation- and state-specific coefficients needed to be estimated reduces from $p_1 \times p_2$ to $D(p_1+p_2)$ by applying a low-rank approximation. The choice of rank D for the soft PARAFAC decomposition of the tensor coefficient can lead to significant changes in computational costs, with a higher value of D triggering drastic increases in computational time. However, the increase in D doesn't necessarily guarantee a vast boost in inferential performance. Intuitively, the soft PARAFAC can expand away from the low-rank hard PARAFAC structure and achieve a higher-rank representation of the tensor coefficient. See [5] for further discussion.

14.3 Posterior Approximation

In this section, we assume tensor-valued covariates and denote with $B^{(d)}_{\ell,m,\tilde{j}_m,k}$ the j_mth slice of tensor $B^{(d)}_{\ell,m,k}$ along the mode m, where $\tilde{j}_m = (:,\cdots,:,j_m,:,\cdots,:)$ and with $B^{(d)}_{\ell,m,\tilde{j}_m,k}$ the $p_1 \times \cdots \times p_{m-1} \times p_{m+1} \times \cdots \times p_M$ tensor with $M-1$ modes. In the case of $M = 2$, introduced in the previous section, $B^{(d)}_{\ell,m,k}$ is a matrix. The slice $B^{(d)}_{\ell,m,\tilde{j}_m,k}$ is the j_1th row of $B^{(d)}_{\ell,m,k}$ when $m = 1$ with $\tilde{j}_1 = (j_1, j)$, $j = 1,\ldots,p_2$ or the j_2th column of $B^{(d)}_{\ell,m,k}$ when $m = 2$ with $\tilde{j}_2 = (i, j_2)$, $i = 1,\ldots,p_1$, respectively. Thus, choosing $j_1 = i$ and $j_2 = j$ and the jth and ith elements of $\tilde{j}_1$ and $\tilde{j}_2$, respectively one get the coefficient $B_{\ell,k,ij} = \sum_{d=1}^{D} B^{(d)}_{\ell,1,(i,j),k} B^{(d)}_{\ell,2,(i,j),k}$. For $\ell = 1,\ldots,N$, $k = 1,\ldots,K$, $m = 1,\ldots,M$, $d = 1,\ldots,D$ and $j_m = 1,\ldots,p_m$ define the $q_m \times 1$ vector $\boldsymbol{\beta}^{(d)}_{\ell,m,j_m,k} = \text{vec}(B^{(d)}_{\ell,m,\tilde{j}_m,k})$, with $q_m = \prod_{l \neq m} p_l$, obtained by stacking vertically all 1-mode fibers of the tensor following a lexicographic order of the indexes. We further define the collections $\boldsymbol{\beta}_k = (\boldsymbol{\beta}_{1k},\ldots,\boldsymbol{\beta}_{Nk})$ and $\boldsymbol{\gamma}_k = (\boldsymbol{\gamma}_{1k},\ldots,$

$\boldsymbol{\gamma}_{Nk}$), with $\boldsymbol{\beta}_{\ell,k} = (\boldsymbol{\beta}^{(1)}_{\ell,1,1,k}, \ldots, \boldsymbol{\beta}^{(d)}_{\ell,m,j_m,k}, \ldots, \boldsymbol{\beta}^{(D)}_{\ell,M,p_M,k})'$ and $\boldsymbol{\gamma}_{\ell,k} = (\gamma^{(1)}_{\ell,1,1,k}, \ldots, \gamma^{(d)}_{\ell,m,j_m,k}, \ldots, \gamma^{(D)}_{\ell,M,p_M,k})'$, where $\gamma^{(d)}_{\ell,m,j_m,k}$ is the j_mth entry, $j_m = 1, \ldots, p_m$, of the PARAFAC marginal vector $\boldsymbol{\gamma}^{(d)}_{\ell,m,k}$ defined in Sect. 14.2.

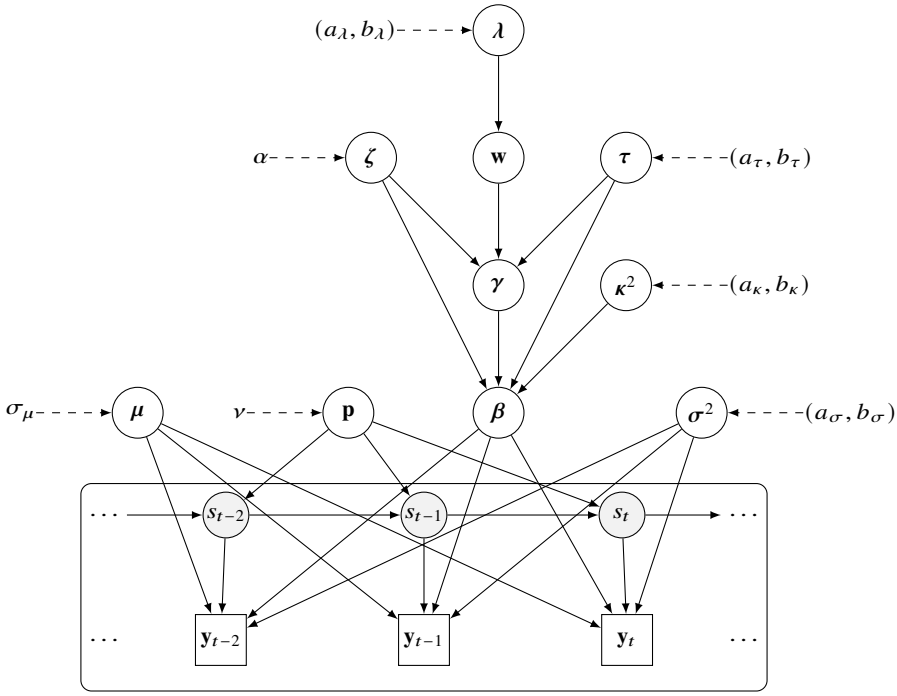

Fig. 14.1: Directed Acyclic Graph of the Bayesian Markov-switching Matrix Regression model. It exhibits the hierarchical structure of the observations $\mathbf{y}_t$ (boxes), the latent state variables s_t (grey circles), the parameters $\beta^{(d)}_{\ell,m,j_m,k}$, $\mu_{\ell,k}$, p_{ik} and $\sigma^2_{\ell,k}$, the hyper-parameters of the first stage $\gamma^{(d)}_{\ell,m,j_m,k}$ and $\kappa^2_{\ell,m,k}$, the second stage $\tau_{\ell,k}$, $\zeta^{(d)}_{\ell,m,k}$ and $w^{(d)}_{\ell,m,j_m,k}$ and the third stage $\lambda^{(d)}_{\ell,m,k}$ (white circles). The directed arrows show the conditional independence structure of the model

We summarize our Bayesian model in the Directed Acyclic Graph (DAG) representation of Fig. 14.1 where $\mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_K)$ is the collection of transition probabilities, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_K)$, $\boldsymbol{\sigma}^2 = (\boldsymbol{\sigma}^2_1, \ldots, \boldsymbol{\sigma}^2_K)$, and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$ denote the collections across equations and states of the regression coefficients, the PARFAC factors, the error scale parameters and intercepts, respectively, where $\boldsymbol{\mu}_k = (\mu_{1,k}, \ldots, \mu_{N,k})'$, $\boldsymbol{\sigma}^2_k = (\sigma^2_{1,k}, \ldots, \sigma^2_{N,k})'$ and $\mathbf{p}_k = (p_{1k}, \ldots, p_{Kk})'$. See [5] for the derivation and further details on the sampling methods.

14.4 Empirical Application

To illustrate our MSTR model with $K = 3$ regimes and rank $D = 2$ (MSTR(3,2) in the following), we study the relationship between the daily volatility index of the US market, also known as VIX and the crude oil ETF oil volatility index (OVX) with several other financial indicators. This is motivated by the fact that VIX has been recognized as the key measure of the market's expectations and sentiments, and predicting VIX is crucial for developing investment strategies.

A vector of the log-VIX averages at different window sizes of $\{1, 5, 10, 22, 66\}$ days is usually employed to study the long-range dependence in VIX data [6]. This choice mirrors daily, weekly, bi-weekly, monthly and quarterly components and returns the family of heterogeneous autoregressive (HAR) processes. The model specification is shown in (14.6) and (14.7).

$$\mathrm{VIX}_t = \mu_1(s_t) + \left\langle B_1(s_t), \begin{pmatrix} \mathrm{SP}_{t-1} & \dots & \mathrm{SP}_{t-h} & \dots & \mathrm{SP}_{t-44} \\ \mathrm{ER}_{t-1} & \dots & \mathrm{ER}_{t-h} & \dots & \mathrm{ER}_{t-44} \\ \mathrm{Oil}_{t-1} & \dots & \mathrm{Oil}_{t-h} & \dots & \mathrm{Oil}_{t-44} \\ \mathrm{OVX}_{t-1} & \dots & \mathrm{OVX}_{t-h} & \dots & \mathrm{OVX}_{t-44} \end{pmatrix} \right\rangle + \sigma_1(s_t)\epsilon_{1t}, \quad (14.6)$$

$$\mathrm{OVX}_t = \mu_2(s_t) + \left\langle B_2(s_t), \begin{pmatrix} \mathrm{SP}_{t-1} & \dots & \mathrm{SP}_{t-h} & \dots & \mathrm{SP}_{t-44} \\ \mathrm{ER}_{t-1} & \dots & \mathrm{ER}_{t-h} & \dots & \mathrm{ER}_{t-44} \\ \mathrm{Oil}_{t-1} & \dots & \mathrm{Oil}_{t-h} & \dots & \mathrm{Oil}_{t-44} \\ \mathrm{VIX}_{t-1} & \dots & \mathrm{VIX}_{t-h} & \dots & \mathrm{VIX}_{t-44} \end{pmatrix} \right\rangle + \sigma_2(s_t)\epsilon_{2t}. \quad (14.7)$$

Left plots in **Fig.** 14.2 show that the MSTR(3,2) in-sample fitting is better than the one of the ordinary least squares and linear LASSO. The stepwise lines in the right plots show that by taking advantage of the Hidden Markov process, the MSTR$(3, 2)$ model detects three oil volatility regimes: low volatility (regime 1), moderately high volatility (regime 2) and high volatility (regime 3). The regime identification follows from the prior identifying restriction $\mu_{11} < \mu_{12} < \mu_{13}$.

Fig. 14.3 provides the coefficient estimates of the Markov-switching Tensor Regression model MSTR$(3, 2)$. In each plot the pairs of coefficients $(B_\ell(1), B_\ell(2))$ and $(B_\ell(1), B_\ell(3))$ are depicted (dots) for the VIX equation ($\ell = 1$, left plots) and OVX equation ($\ell = 2$, right plots).

The regime separation can be further described by inspecting the estimated effects of h-day log-return of oil prices (Oil_{t-h}, $h = 1, \dots, 44$) and S&P 500 (SP_{t-h} $h = 1, \dots, 44$) on VIX (blue dots, left) and OVX (red dots, right). The dots in the plots correspond to the values of parameters in different pairs of volatility regimes. The 70% HPD regions (grey ellipses) provide evidence of coefficient heterogeneity across regimes (asymmetric effects), equations (market asymmetry) and lags (long-term effects). The asymmetric effects in the coefficients are more substantial across regime 1 and 3 than regime 1 and 2. Moreover, there is evidence of a more balanced impact of the h-day S&P 500 log-returns on both markets (VIX and OVX) compared to the impacts of oil prices. There is also evidence of non-negligible long-term effects of oil prices on the stock market (dark blue dots in the left-bottom plot of Panel a).

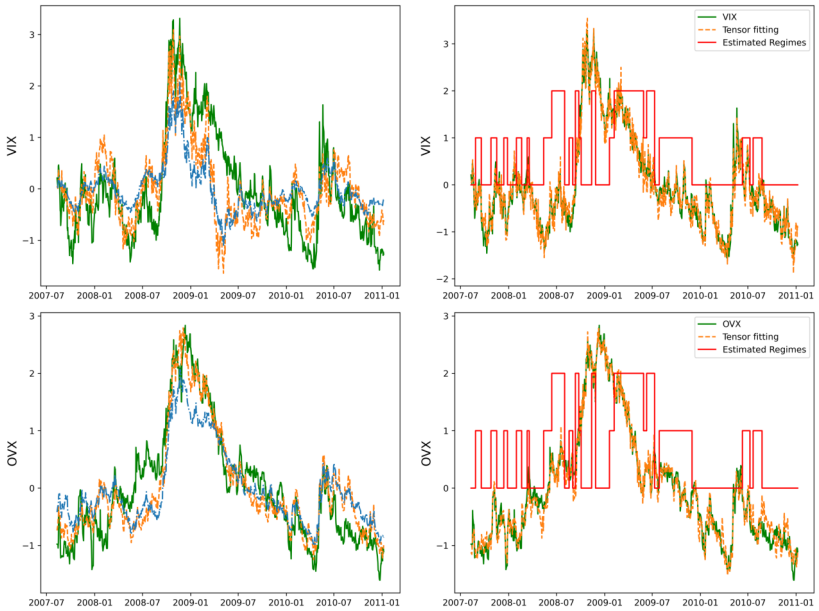

Fig. 14.2: In-sample fitting versus the actual data. Left: Least Squares (orange dashed) and LASSO (blue dashed) fitting. Right: Markov-Switching Tensor Regression model MSTR$(3,2)$ fitting (orange dashed) and estimated hidden states (red solid). For all plots, the green solid line represents the actual data for the VIX and VOX.

14.5 Conclusion

This paper proposes a Markov Switching Tensor-on-Tensor Regression Model (MSTR) for high dimensional data where a common hidden Markov chain process drives dependencies between equations and allows for regime changes and time-varying coefficients. A low-rank representation of the tensor coefficient is used to achieve dimensionality reduction. A hierarchical prior distribution introduces further shrinkage effects in the regression coefficients. An efficient MCMC sampler based on Random Partial Scan Gibbs and a back-fitting strategy is introduced. We illustrate our MSTR with a real-world application to oil and stock market volatility data. Our Bayesian MSTR model outperforms competing models and can capture structural changes in the parameters by identifying distinct volatility regimes. The MSTR can also capture the heterogeneity and asymmetric effects in the coefficient across the equations.

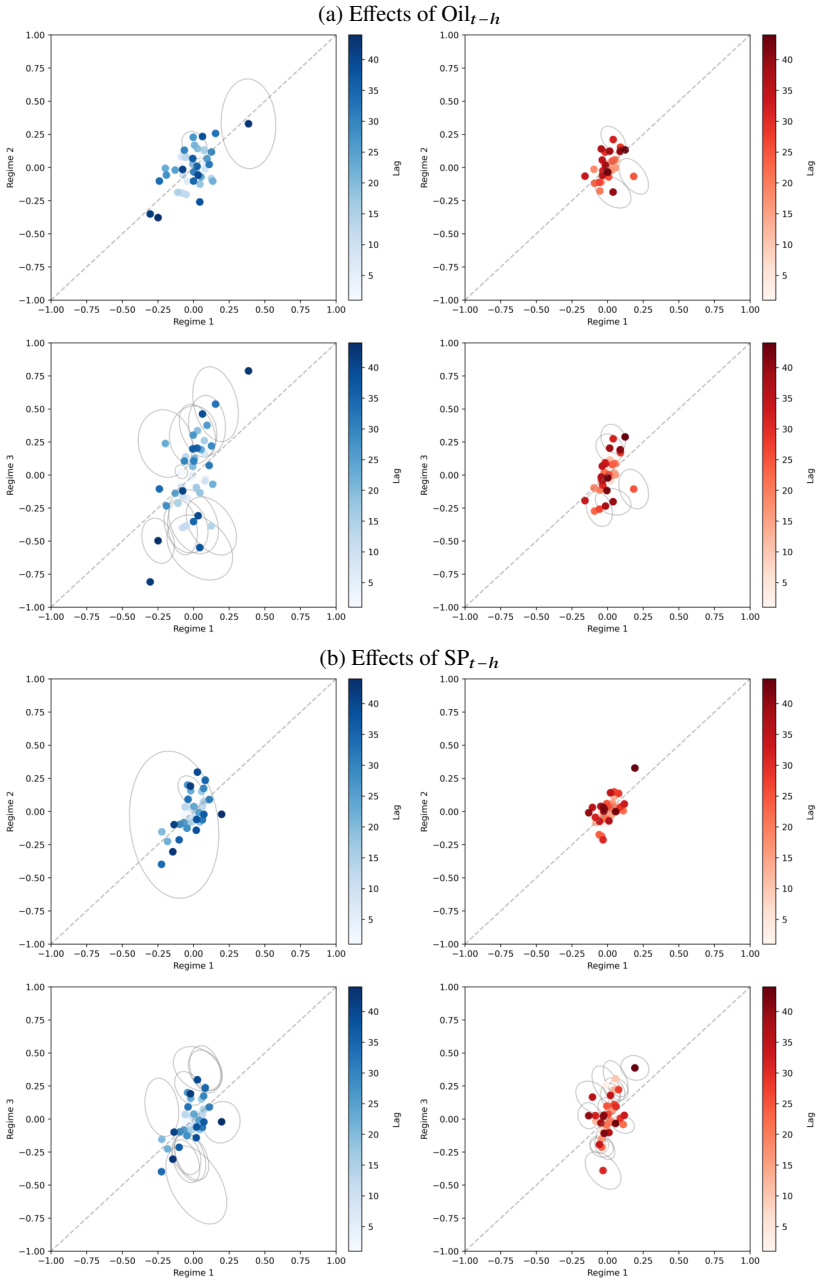

Fig. 14.3: Markov-switching Tensor Regression model MSTR$(3,2)$ coefficient estimates. Effects of h-day Oil (panel a) and S&P 500 (panel b) log-returns on VIX (left column) and OVX (right column) for $h \in \{1,\ldots,44\}$. In each plot, lighter and darker colours represent smaller and larger h, respectively, and grey ellipses exhibit significant asymmetric effects across regimes (70% Highest Posterior Density regions).

References

1. Armagan, A., Dunson, D.B., Lee, J.: Generalized double Pareto shrinkage. Statistica Sinica **23**(1), 119 (2013)
2. Billio, M., Casarin, R., Iacopini, M.: Bayesian Markov-switching tensor regression for time-varying networks. J. Am. Stat. Assoc. **119**(545), 109–121 (2024)
3. Billio, M., Casarin, R., Iacopini, M., Kaufmann, S.: Bayesian dynamic tensor regression. J. Bus. Econ. Stat. **41**(2), 429–439 (2023)
4. Casarin, R., Sartore, D., Tronzano, M.: A Bayesian Markov-switching correlation model for contagion analysis on exchange rate markets. J. Bus. Econ. Stat. **36**(1), 101–114 (2018)
5. Casarin, R., Craiu, R.V., Wang, Q.: Markov Swithching Multiple-equation Tensor Regression. J. Multivar. Anal., (2025)
6. Fernandes, M., Medeiros, M.C., Scharth, M.: Modeling and predicting the CBOE market volatility index. J. Bank. Finance **40**, 1–10 (2014)
7. Frühwirth-Schnatter, S.: Finite Mixture and Markov Switching Models. Springer (2006)
8. Guha, S., Rodriguez, A.: Bayesian regression with undirected network predictors with an application to brain connectome data. J. Am. Stat. Assoc. **116**(534), 581–593 (2021)
9. Gupta, A.K., Nagar, D.K.: Matrix Variate Distributions. Chapman and Hall/CRC, Boca Raton (2018)
10. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**, 455–500 (2009)
11. Spencer, D., Guhaniyogi, R., Shinohara, R., Prado, R.: Bayesian tensor regression using the Tucker decomposition for sparse spatial modeling. arXiv:2203.04733 (2022).
12. Wang, K., Xu, Y.: Bayesian tensor-on-tensor regression with efficient computation. Stat. Interface **17**, 199 (2024)
13. Guhaniyogi, R., Qamar, S., Dunson, D.B.: Bayesian tensor regression. J. Mach. Learn. Res. **18**(79), 1–31 (2017)
14. Papadogeorgou, G., Zhang, Z., Dunson, D.B.: Soft tensor regression. J. Mach. Learn. Res. **22**, 1–53 (2021)
15. Wang, K., Xu, Y.: Bayesian tensor-on-tensor regression with efficient computation. Stat. Interface **17**, 199 (2024)