Detecting Change Points in Bayesian VAR: Consistency Results

Stefano Peluso *[1], Siddhartha Chib[2], and Antonietta Mira[3]

[1]Università degli Studi di Milano-Bicocca and Università della Svizzera italiana
[2]Washinghton University in St. Louis
[3]Università della Svizzera italiana and Università degli Studi dell'Insubria

August 31, 2024

Abstract

Vector autoregressive processes (VAR) model contemporaneous and lagged dependencies in multivariate time series. Due to changes in the model environment, abrupt changes in model parameters may occur, which requires the identification of unknown regimes. We propose a Bayesian method for inferring change-points in VAR models. We show that the posterior distribution of the change-point location asymptotically concentrates on the true change-point when we have a single change-point and a conjugate prior is assigned to the regime parameters. This result is extended to non-conjugate priors, under specific conditions on prior and data, and to multiple change-points, under controlled overlaps between estimated and true regimes. Simulated studies confirm the ability to recover the changes, and an application to macroeconomic US data shows the utility of the proposed method compared to established alternatives.

Keywords: Bayesian statistics, Change-points, Posterior ratio consistency

*Corresponding author: stefano.peluso@unimib.it

1 Introduction

Vector Autoregressive Processes (VAR) are an invaluable tool for modeling temporal and cross-sectional interdependencies in multivariate systems, with numerous applications in economics (Bernanke et al. [6], Huber et al. [24]), finance (Billio et al. [7]), genetics (Basu et al. [5]) and neuroscience (Seth et al. [40]). Theoretical properties of the VAR model have been investigated in a frequentist and Bayesian framework by, respectively, Lütkepohl [30] and Bańbura et al. [3] in the low-dimensional case, and by Basu and Michailidis [4] and Ghosh et al. [21, 22] in the high-dimensional case.

Given the complexities of modern data, several extensions of the basic VAR model have emerged. For example, Melnyk and Banerjee [33] introduced norm-driven sparsity structures in autoregressive matrices, while Schweinberger et al. [39] integrated temporal dependence with spatial structures. Lin and Michailidis [29] imposed a multi-block dependency structure among the coordinates of the VAR system, and Park et al. [35] and Wang et al. [45] explored low-dimensional factor and tensor representations of temporal dynamics. In some cases, stability and homogeneity have been replaced by partial stationarity [37, 17], heteroskedastic volatilities [14, 31], and model misspecifications [44]. The impact of abrupt change points on the parameters of the VAR model has been studied in the frequentist framework, with Bai et al. [1, 2] focusing on low-rank, high-dimensional cases, Hansen [23] addressing cointegrated time series, and Lee et al. [27] employing CUSUM-type tests [25].

To the best of our knowledge, Bayesian studies of VAR processes with change-points are limited, with the exception of Xuan and Murphy [47] and Knoblauch and Damoulas [26], where theoretical guarantees focus on correctly estimating regime parameters rather than identifying change-points. We aim to address this gap by proposing a Bayesian Gaussian VAR model with change-points. Our study investigates the theoretical properties of this model by analyzing the behavior of the marginal posterior distribution of change-point locations. We establish posterior ratio consistency, as defined by Cao et al. [8], Castelletti and Peluso [10], and Castelletti and Peluso [11], for a single change-point and conjugate regime parameter priors, under conditions where the competing model commits a finite error in the change-point location or where the error vanishes at a rate slower than the sample size. Additionally, a weaker form of posterior consistency is recovered when the change-point location error diminishes at a rate proportional to the sample size, under minimal conditions on the data proportions within the incorrect regimes. We extend these results to non-conjugate apriors, under specific conditions on the prior and the true generating process, and to multiple change-points, restricting the analysis to cases where estimated regimes are contaminated by at most one adjacent regime. Simulated experiments validate our theoretical findings, and an application to macroeconomic data demonstrates the utility of our methodology compared to established alternatives.

The statistical model is developed in Section 2.1 (likelihood part) and 2.2 (prior distributions). The theoretical results are discussed in Section 3, for the conjugate regime parameter priors and one single change-point (Section 3.1), and for non-conjugate and mis-specified priors and multiple change-points (Section 3.2). The results from simulated and real data are discussed in Section 4, while conclusions and further research possibilities are proposed in Section 5.

2 Model development

2.1 Vector autoregressive process with regimes

Let $\boldsymbol{y}_t \in \mathbb{R}^q$ be a vector of observations measured at time t, for $t = 1, 2, \dots, n$. The whole time sequence $\{1, 2, \dots, n\}$ is partitioned by m change-points $t_{n1} < t_{n2} < \cdots < t_{nm}$ into $m+1$ adjacent regimes $\mathcal{N}_{ni} = \{t : t_{n,i-1} \leq t < t_{ni}\}$, for $i = 1, \dots, m+1$, of cardinalities $n_1, \dots, n_m$, where $t_{n0} := 1$ and $t_{n,m+1} := n+1$. Within the generic i-th regime, we assume a Vector Autoregressive (VAR, West and Harrison [46]) dynamics of order $K \geq 1$:

$$\boldsymbol{y}_t = \sum_{k=1}^{K} \boldsymbol{B}_k^{(i)} \boldsymbol{y}_{t-k} + \boldsymbol{\epsilon}_t, \qquad t \in \mathcal{N}_{ni} \tag{1}$$

where $\boldsymbol{B}_k^{(i)}$ are $q \times q$ lag matrices, and $\boldsymbol{\epsilon}_t | \boldsymbol{\Omega}^{(i)} \sim N_q\left(0, (\boldsymbol{\Omega}^{(i)})^{-1}\right)$ is a q-variate time-independent Gaussian error, with regime-specific precision matrix $\boldsymbol{\Omega}^{(i)}$. We stack all n_i observations in $\boldsymbol{Y}_{\mathcal{N}_{ni}} = (\boldsymbol{y}_{t_{n,i-1}}, \dots, \boldsymbol{y}_{t_{ni}-K})^\top$, a $n_i \times q$ matrix, and define $\boldsymbol{Y}_n = \left(\boldsymbol{Y}_{\mathcal{N}_{n1}}^\top, \boldsymbol{Y}_{\mathcal{N}_{n2}}^\top, \dots, \boldsymbol{Y}_{\mathcal{N}_{n,m+1}}^\top\right)^\top$ as the whole $n \times q$ data matrix. In addition, we collect all lagged observations of Equation (1) in $\boldsymbol{z}_t = (\boldsymbol{y}_{t-1}^\top, \dots, \boldsymbol{y}_{t-K}^\top)^\top$ and then all measurements of the i-th regime in $\boldsymbol{Z}_{\mathcal{N}_{ni}} = (\boldsymbol{z}_{t_{n,i-1}}, \dots, \boldsymbol{z}_{t_{ni}-1})^\top$, a $n_i \times Kq$ matrix. In this way we rewrite (1) for all measurements in the more concise matrix form

$$\boldsymbol{Y}_{\mathcal{N}_{ni}} = \boldsymbol{Z}_{\mathcal{N}_{ni}} \boldsymbol{B}^{(i)} + \boldsymbol{E}_{\mathcal{N}_{ni}},$$

where $\boldsymbol{B}^{(i)} = (\boldsymbol{B}_1^{(i)}, \ldots, \boldsymbol{B}_K^{(i)})^\top$ of dimension $Kq \times q$ stacks by row the K lag matrices, and $\boldsymbol{E}_{\mathcal{N}_{ni}} = (\boldsymbol{\epsilon}_{t_{n,i-1}}, \ldots, \boldsymbol{\epsilon}_{t_{ni}-1})^\top$ is the $n_i \times q$ Gaussian error matrix. Within this framework, we have a Matrix-Normal distribution (Dawid [16]) for the error matrix,

$$\boldsymbol{E}_{\mathcal{N}_{ni}} \,|\, \boldsymbol{\Omega}^{(i)}, \mathcal{N}_{ni} \sim N_{n_i \times q}\left(\boldsymbol{0}_{n_i \times q}, \boldsymbol{I}_{n_i}, (\boldsymbol{\Omega}^{(i)})^{-1}\right),$$

or, equivalently, $\mathrm{vec}(\boldsymbol{E}_{\mathcal{N}_{ni}}) \,|\, \boldsymbol{\Omega}^{(i)}, \mathcal{N}_{ni} \sim N_{n_i q}\left(\boldsymbol{0}_{n_i q}, (\boldsymbol{\Omega}^{(i)})^{-1} \otimes \boldsymbol{I}_{n_i}\right)$, a $n_i q$-variate Gaussian distribution. The conditional density associated with the i-th regime is then, for $i = 1, \ldots, m+1$ and $\mathcal{N}_{n0} := \emptyset$,

$$\begin{aligned} & f\left(\boldsymbol{Y}_{\mathcal{N}_{ni}} \,|\, \boldsymbol{B}^{(i)}, \boldsymbol{\Omega}^{(i)}, \boldsymbol{Y}_{\mathcal{N}_{n,i-1}}, \ldots, \boldsymbol{Y}_{\mathcal{N}_{n1}}\right) \\ & \quad = \frac{|\boldsymbol{\Omega}^{(i)}|^{n_i/2}}{(2\pi)^{n_i q/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Omega}^{(i)}\left(\left(\boldsymbol{B}^{(i)} - \hat{\boldsymbol{B}}_{\mathcal{N}_{ni}}\right)^\top \boldsymbol{Z}_{\mathcal{N}_{ni}}^\top \boldsymbol{Z}_{\mathcal{N}_{ni}} \left(\boldsymbol{B}^{(i)} - \hat{\boldsymbol{B}}_{\mathcal{N}_{ni}}\right)\right.\right.\right. \\ & \qquad\qquad \left.\left.\left. + \hat{\boldsymbol{E}}_{\mathcal{N}_{ni}}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_{ni}}\right)\right]\right\}, \end{aligned}$$

where $\hat{\boldsymbol{E}}_{\mathcal{N}_{ni}} = \boldsymbol{Y}_{\mathcal{N}_{ni}} - \boldsymbol{Z}_{\mathcal{N}_{ni}} \hat{\boldsymbol{B}}_{\mathcal{N}_{ni}}$ and $\hat{\boldsymbol{B}}_{\mathcal{N}_{ni}} = \left(\boldsymbol{Z}_{\mathcal{N}_{ni}}^\top \boldsymbol{Z}_{\mathcal{N}_{ni}}\right)^{-1} \boldsymbol{Z}_{\mathcal{N}_{ni}}^\top \boldsymbol{Y}_{\mathcal{N}_{ni}}$. A similar expression can be found, for instance in Consonni et al. [15] in the context of a graphical model with no change-points, and in Paci and Consonni [34] in a VAR context with no change-points. In the current context, however, the presence of change-points requires that the conditioning on $\boldsymbol{Y}_{\mathcal{N}_{n1}}, \ldots, \boldsymbol{Y}_{\mathcal{N}_{n,i-1}}$ be explicit as some elements of $\boldsymbol{Z}_{\mathcal{N}_{ni}}$ belong to previous regimes.

Finally, denote with $\boldsymbol{\theta}$ the vector containing all the distinct regime parameters, that is the $(m+1)\left(Kq^2 + q(q+1)/2\right)$-dimensional vector

$$\boldsymbol{\theta} = \left(\left(\boldsymbol{\theta}^{(1)}\right)^\top, \ldots, \left(\boldsymbol{\theta}^{(m+1)}\right)^\top\right)^\top = \left(\mathrm{vec}(\boldsymbol{B}^{(i)})^\top, \mathrm{vech}(\boldsymbol{\Omega}^{(i)})^\top; i = 1, \ldots, m+1\right)^\top,$$

with $\mathrm{vech}(\cdot)$ being the half-vectorization; then we can explicit the whole likelihood as $f(\boldsymbol{Y}_n | \boldsymbol{\theta}, \mathcal{N}_{n1}, \ldots, \mathcal{N}_{n,m+1}) = \prod_{i=1}^{m+1} f\left(\boldsymbol{Y}_{\mathcal{N}_{ni}} | \boldsymbol{\theta}^{(i)}, \boldsymbol{Y}_{\mathcal{N}_{n,i-1}}, \ldots, \boldsymbol{Y}_{\mathcal{N}_{n1}}\right)$. In the sequel, for ease of notation $\mathcal{N}_{ni}$ will be written as $\mathcal{N}_i$, therefore omitting the dependence from n.

2.2 Change-point process and priors

In the previous subsection we assume knowledge of the regimes $\mathcal{N}_{ni}$, $i = 1, \ldots, m+1$, or equivalently of the change-points $t_{ni} = 1 + \lfloor n\tau_i \rfloor$, $i = 1, \ldots, m$, where τ_i denotes the proportion of data before the occurrence of the i-th change-point. Then the cardinality of the i-th regime can also be expressed as $n_i = \lfloor n\tau_i \rfloor - \lfloor n\tau_{i-1} \rfloor$, approximated by $n(\tau_i - \tau_{i-1})$ for samples large enough and $\tau_0 := 0$. The vector $(\tau_1, \ldots, \tau_m)$, together with the sample size n, is associated to a latent state process $(S_{n1}, \dots, S_{nn})$, indicator of the regime: $S_{nt} = i$ means that the time instant t belongs to the i-th regime, that is $t \in \mathcal{N}_{ni}$. Following Chib [13], the state variables can either retain the current state or move to the next higher state, with a prior probability of, respectively $p(S_{n,t+1} = i+1 | S_{nt} = i, \omega_i) = \omega_i$. Then, ω_i denotes the prior probability of a change from the i-th regime. Fixing $\omega_i \sim \text{Beta}(\alpha_i, \beta_i)$, the marginalization of ω_i results in the following prior on the change-point locations and on the number of change-points $m < n$:

$$\begin{aligned} p_n\left((\tau_1, \dots, \tau_m) = (\tau_1, \dots, \tau_m), M = m\right) &= p\left((T_{n1}, \dots, T_{nm}) = (t_{n1}, \dots, t_{nm}), M = m\right) \\ &= p(T_{n,m+1} > n \,|\, T_{nm} = t_{nm}) \prod_{i=1}^{m} p(T_{ni} = t_{ni} \,|\, T_{n,i-1} = t_{n,i-1}) \\ &= \prod_{i=1}^{m+1} \frac{B(\alpha_i + \mathbf{1}(i \leq m), \beta_i + n_i - 1)}{B(\alpha_i, \beta_i)}, \end{aligned} \tag{2}$$

where $B(\cdot, \cdot)$ denotes the Beta function. Given m change-points, a conjugate prior for the regime parameter $\boldsymbol{\theta}$ is

$$p(\boldsymbol{\theta}) = \prod_{i=1}^{m+1} p\left(\boldsymbol{B}^{(i)} \,|\, \boldsymbol{\Omega}^{(i)}\right) p\left(\boldsymbol{\Omega}^{(i)}\right)$$

where

$$\begin{aligned} p\left(\boldsymbol{B}^{(i)} \,|\, \boldsymbol{\Omega}^{(i)}\right) &= N_{Kq \times q}\left(\underline{\boldsymbol{B}}^{(i)}, (\boldsymbol{C}^{(i)})^{-1}, (\boldsymbol{\Omega}^{(i)})^{-1}\right), \\ p\left(\boldsymbol{\Omega}^{(i)}\right) &= \mathcal{W}_q(a^{(i)}, \boldsymbol{R}^{(i)}), \end{aligned}$$

a product of Matrix Normal Wishart prior distribution, where $\underline{\boldsymbol{B}}^{(i)}$ is the prior expected value of $\boldsymbol{B}^{(i)}$, $\boldsymbol{C}^{(i)}$ is the prior row precision matrix of dimension $Kq \times Kq$, $\boldsymbol{R}^{(i)}$ is a $q \times q$ positive semi-definite matrix, and $a^{(i)}$ is a scalar strictly greater than $q-1$, so that $\mathbb{E}\left(\boldsymbol{\Omega}^{(i)}\right) = a^{(i)}\left(\boldsymbol{R}^{(i)}\right)^{-1}$. The prior density is therefore

$$p\left(\boldsymbol{B}^{(i)}, \boldsymbol{\Omega}^{(i)}\right) = \frac{|\boldsymbol{\Omega}^{(i)}|^{\frac{a^{(i)}+q(K-1)-1}{2}}}{K(\boldsymbol{C}^{(i)}, \boldsymbol{R}^{(i)}, a^{(i)})} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Omega}^{(i)}\left[(\boldsymbol{B}^{(i)} - \underline{\boldsymbol{B}}^{(i)})^\top \boldsymbol{C}^{(i)} (\boldsymbol{B}^{(i)} - \underline{\boldsymbol{B}}^{(i)}) + \boldsymbol{R}^{(i)}\right]\right)\right\},$$

with

$$K(\boldsymbol{C}, \boldsymbol{R}, a) = \frac{(2\pi)^{\frac{Kq^2}{2}} 2^{\frac{aq}{2}} \Gamma_q(\frac{a}{2})}{|\boldsymbol{C}|^{\frac{q}{2}} |\boldsymbol{R}|^{\frac{a}{2}}}, \tag{3}$$

where $\Gamma_q\left(\frac{a^{(i)}}{2}\right) = \pi^{\frac{q(q-1)}{4}} \prod_{j=1}^{q} \Gamma\left(\frac{a^{(i)}}{2} + \frac{1-j}{2}\right)$ is the q-dimensional gamma function evaluated at $a^{(i)}/2$, and the prior normalizing constant is $\prod_{i=1}^{m+1} K(\boldsymbol{C}^{(i)}, \boldsymbol{R}^{(i)}, a^{(i)})$.

3 Theoretical results

3.1 Marginal likelihood and change-point detection

In this section we derive the marginal likelihood and the marginal posterior probability of the change-point location, and we show that the model is asymptotically able to recover the correct change-point in various settings. In particular, by establishing posterior ratio consistency, we show that the posterior mass will be concentrated on the correct model, for different kinds of errors committed by the alternative models.

Some algebraic manipulations show that the posterior distribution of the parameter vector $\boldsymbol{\theta}$ given the regimes and the number of change-points is

$$p(\boldsymbol{\theta} \mid \boldsymbol{Y}_n, m, \mathcal{N}_1, \ldots, \mathcal{N}_{m+1}) = \prod_{i=1}^{m+1} p\left(\boldsymbol{B}^{(i)}, \boldsymbol{\Omega}^{(i)} \mid \boldsymbol{Y}_{\mathcal{N}_i}, \boldsymbol{Y}_{\mathcal{N}_{i-1}}, \ldots, \boldsymbol{Y}_{\mathcal{N}_1}\right), \tag{4}$$

such that

$$\begin{aligned} p\left(\boldsymbol{B}^{(i)} \mid \boldsymbol{\Omega}^{(i)}, \boldsymbol{Y}_{\mathcal{N}_i}, \boldsymbol{Y}_{\mathcal{N}_{i-1}}, \ldots, \boldsymbol{Y}_{\mathcal{N}_1}\right) &= N_{Kq \times q}\left(\overline{\boldsymbol{B}}^{(i)}, (\boldsymbol{C}^{(i)} + \boldsymbol{Z}_{\mathcal{N}_i}^\top \boldsymbol{Z}_{\mathcal{N}_i})^{-1}, (\boldsymbol{\Omega}^{(i)})^{-1}\right), \\ p\left(\boldsymbol{\Omega}^{(i)} \mid \boldsymbol{Y}_{\mathcal{N}_i}, \boldsymbol{Y}_{\mathcal{N}_{i-1}}, \ldots, \boldsymbol{Y}_{\mathcal{N}_1}\right) &= \mathcal{W}_q\left(a^{(i)} + n_i, \boldsymbol{R}^{(i)} + \hat{\boldsymbol{E}}_{\mathcal{N}_i}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_i} + \boldsymbol{D}^{(i)}\right), \end{aligned} \tag{5}$$

where $\overline{\boldsymbol{B}}^{(i)} = (\boldsymbol{C}^{(i)} + \boldsymbol{Z}_{\mathcal{N}_i}^\top \boldsymbol{Z}_{\mathcal{N}_i})^{-1}(\boldsymbol{Z}_{\mathcal{N}_i}^\top \boldsymbol{Y}_{\mathcal{N}_i} + \boldsymbol{C}^{(i)} \underline{\boldsymbol{B}}^{(i)})$ is the posterior expectation matrix of $\boldsymbol{B}^{(i)}$, and $\boldsymbol{D}^{(i)} = (\underline{\boldsymbol{B}}^{(i)} - \hat{\boldsymbol{B}}_{\mathcal{N}_i})^\top \{(\boldsymbol{C}^{(i)})^{-1} + (\boldsymbol{Z}_{\mathcal{N}_i}^\top \boldsymbol{Z}_{\mathcal{N}_i})^{-1}\}^{-1} (\underline{\boldsymbol{B}}^{(i)} - \hat{\boldsymbol{B}}_{\mathcal{N}_i})$ is a measure of discrepancy between $\underline{\boldsymbol{B}}^{(i)}$ and $\hat{\boldsymbol{B}}_{\mathcal{N}_i}$ (prior and data). Define $\overline{\boldsymbol{B}}^{(i)} = (a^{(i)} + n_i)\left(\boldsymbol{R}^{(i)} + \hat{\boldsymbol{E}}_{\mathcal{N}_i}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_i} + \boldsymbol{D}^{(i)}\right)^{-1}$. Using prior and posterior densities in Equations 3 and 4, and the likelihood expression derived in Section 2.1, we can compute the marginal likelihood using the Chib [12] identity

$$\begin{aligned} m(\boldsymbol{Y}_n \mid m, \mathcal{N}_1, \ldots, \mathcal{N}_{m+1}) &= \frac{f\left(\boldsymbol{Y}_n \mid \boldsymbol{\theta}, m, \mathcal{N}_1, \ldots, \mathcal{N}_{m+1}\right) p(\boldsymbol{\theta} \mid m, \mathcal{N}_1, \ldots, \mathcal{N}_{m+1})}{p(\boldsymbol{\theta} \mid \boldsymbol{Y}_n, m, \mathcal{N}_1, \ldots, \mathcal{N}_{m+1})} \\ &= (2\pi)^{-nq/2} \prod_{i=1}^{m+1} \frac{K(\boldsymbol{C}^{(i)} + \boldsymbol{Z}_{\mathcal{N}_i}^\top \boldsymbol{Z}_{\mathcal{N}_i}, \boldsymbol{R}^{(i)} + \hat{\boldsymbol{E}}_{\mathcal{N}_i}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_i} + \boldsymbol{D}^{(i)}, a^{(i)} + n_i)}{K(\boldsymbol{C}^{(i)}, \boldsymbol{R}^{(i)}, a^{(i)})}, \end{aligned} \tag{6}$$

where the function $K(\cdot, \cdot, \cdot)$ is defined in (3). Using the one-to-one relationship between change-point locations and regimes, the posterior probability of the number and locations of change-points is therefore easily implied as proportional to the product of the above expression and the prior in (2).

In the next proposition, we show that, in case of a unique unknown change-point, the posterior probability will favor the correct one, when the alternative model commits a finite error γ in the estimated change-point location.

Proposition 3.1. *Given a true proportion of data τ_{01} in the first regime, and $\gamma \in (-1,1)$ such that $|\gamma| < min\{\tau_{01}, 1-\tau_{01}\}$, we have, with $\bar{P}$-probability 1, as $n \to \infty$, that $\frac{p(\tau_1=\tau_{01}+\gamma, M=1|\mathbf{Y}_n)}{p(\tau_1=\tau_{01}, M=1|\mathbf{Y}_n)} \to 0$.*

Proof. For the true regimes $\mathcal{N}_{01}$ and $\mathcal{N}_{02}$, we have that $n_{01} = \lfloor n\tau_{01} \rfloor$ and $n_{02} = n - \lfloor n\tau_{01} \rfloor$. Then we can write

$$
\begin{aligned}
p(\tau_1 = \tau_{01}, M=1|\boldsymbol{Y}_n) &= p(T_{n1} = 1 + n_{01}, M = 1|\boldsymbol{Y}_n) \\
&\propto \frac{B(\alpha_1+1, \beta_1+n_{0,1}-1)B(\alpha_2, \beta_2+n-n_{0,1}-1)}{\left(\left|\boldsymbol{C}^{(1)} + \boldsymbol{Z}_{\mathcal{N}_{01}}^\top \boldsymbol{Z}_{\mathcal{N}_{01}}\right| \left|\boldsymbol{C}^{(2)} + \boldsymbol{Z}_{\mathcal{N}_{02}}^\top \boldsymbol{Z}_{\mathcal{N}_{02}}\right|\right)^{q/2}} \\
&\cdot \frac{\Gamma_q\left(\frac{a+n_{0,1}}{2}\right)\Gamma_q\left(\frac{a+n-n_{0,1}}{2}\right)}{\left|\boldsymbol{R}^{(1)} + \hat{\boldsymbol{E}}_{\mathcal{N}_{01}}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_{01}} + \boldsymbol{D}^{(1)}\right|^{\frac{a+n_{0,1}}{2}} \left|\boldsymbol{R}^{(2)} + \hat{\boldsymbol{E}}_{\mathcal{N}_{02}}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_{02}} + \boldsymbol{D}^{(2)}\right|^{\frac{a+n-n_{0,1}}{2}}}
\end{aligned}
$$

We now notice that $\left|\boldsymbol{C}^{(i)} + \boldsymbol{Z}_{\mathcal{N}_i^0}^\top \boldsymbol{Z}_{\mathcal{N}_i^0}\right| \sim n_{0,i}^q \left|\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(i)}\right|^{-1}$, and similarly, $\left|\boldsymbol{R}^{(i)} + \hat{\boldsymbol{E}}_{\mathcal{N}_i^0}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_i^0} + \boldsymbol{D}^{(i)}\right| \sim n_{0,i}^q \left|\boldsymbol{\Omega}_0^{(i)}\right|^{-1}$, for $i = 1, 2$. For $\delta_n \sim n\gamma$,

$$
\begin{aligned}
p(\tau_1 = \tau_{01}+\gamma, M=1|\boldsymbol{Y}_n) &\sim p(T_{n1} = 1 + n_{01} + \delta_n, M = 1|\boldsymbol{Y}_n) \\
&\propto \frac{B(\alpha_1+1, \beta_1+n_{0,1}+\delta_n-1)B(\alpha_2, \beta_2+n_{0,2}-\delta_n-1)}{\left(\left|\boldsymbol{C}^{(1)} + \boldsymbol{Z}_{\mathcal{N}_1}^\top \boldsymbol{Z}_{\mathcal{N}_1}\right| \left|\boldsymbol{C}^{(2)} + \boldsymbol{Z}_{\mathcal{N}_2}^\top \boldsymbol{Z}_{\mathcal{N}_2}\right|\right)^{q/2}} \\
&\cdot \frac{\Gamma_q\left(\frac{a+n_{0,1}+\delta_n}{2}\right)\Gamma_q\left(\frac{a+n_{0,2}-\delta_n}{2}\right)}{\left|\boldsymbol{R}^{(1)} + \hat{\boldsymbol{E}}_{\mathcal{N}_1}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_1} + \boldsymbol{D}^{(1)}\right|^{(a+n_{0,1}+\delta_n)/2} \left|\boldsymbol{R}^{(2)} + \hat{\boldsymbol{E}}_{\mathcal{N}_2}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_2} + \boldsymbol{D}^{(2)}\right|^{(a+n_{0,2}-\delta_n)/2}}
\end{aligned}
$$

Assuming $\gamma \in (0, 1-\tau_{01})$, we have that $n_1 > n\tau_{01}+1$ for n sufficiently large. Then we have $\left|\boldsymbol{C}^{(2)} + \boldsymbol{Z}_{\mathcal{N}_2}^\top \boldsymbol{Z}_{\mathcal{N}_2}\right| \sim (n_{0,2}-\delta_n)^q \left|\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right|^{-1}$, and, for $\boldsymbol{W}_n := \left(\boldsymbol{I}_q + \frac{\delta_n}{n_{0,1}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}$,

$$
\left|\boldsymbol{C}^{(1)} + \boldsymbol{Z}_{\mathcal{N}_1}^\top \boldsymbol{Z}_{\mathcal{N}_1}\right| \quad \sim \quad \left|n_{0,1}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1} + \delta_n\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\right| = n_{0,1}^q \left|\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right|^{-1} / \left|\boldsymbol{W}_n\right|.
$$

Similarly to above, $\left|\boldsymbol{R}^{(2)} + \hat{\boldsymbol{E}}_{\mathcal{N}_2}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_2} + \boldsymbol{D}^{(2)}\right| \sim (n_{0,2}-\delta_n)^q \left|\boldsymbol{\Xi}_0^{(2)}\right|^{-1}$, whilst the equivalent derivation in the first regime $\mathcal{N}_1$ is a little more involved, since $\hat{\boldsymbol{B}}_{\mathcal{N}_1}$ may be not consistent, being $\hat{\boldsymbol{B}}_{\mathcal{N}_1} \sim \boldsymbol{W}_n \boldsymbol{B}_{0,1} + (\boldsymbol{I}_q - \boldsymbol{W}_n)\boldsymbol{B}_{0,2}$. In this case it can be shown that

$$
\begin{aligned}
\left|\boldsymbol{R}^{(1)} + \hat{\boldsymbol{E}}_{\mathcal{N}_1}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_1} + \boldsymbol{D}^{(1)}\right| &\sim n_{0,1}^q \left|\boldsymbol{\Omega}_0^{(1)}\right|^{-1} \cdot \left|\boldsymbol{I}_q + \frac{\delta_n}{n_{0,1}}\boldsymbol{\Omega}_0^{(1)}\left(\boldsymbol{\Omega}_0^{(2)}\right)^{-1}\right. \\
&+\boldsymbol{\Omega}_0^{(1)}\left(\boldsymbol{B}_{0,1} - \boldsymbol{B}_{0,2}\right)^\top \left(\boldsymbol{I}_q - \boldsymbol{W}_n\right)^\top \left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}\left(\boldsymbol{I}_q - \boldsymbol{W}_n\right)\left(\boldsymbol{B}_{0,1} - \boldsymbol{B}_{0,2}\right) \\
&\left.+\frac{\delta_n}{n_{0,1}}\boldsymbol{\Omega}_0^{(1)}\left(\boldsymbol{B}_{0,1} - \boldsymbol{B}_{0,2}\right)^\top \boldsymbol{W}_n^\top \left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\boldsymbol{W}_n\left(\boldsymbol{B}_{0,1} - \boldsymbol{B}_{0,2}\right)\right| \\
&=: n_{0,1}^q \left|\boldsymbol{\Omega}_0^{(1)}\right|^{-1} \cdot \left|\boldsymbol{I}_q + \boldsymbol{\Omega}_0^{(1)}\boldsymbol{\Xi}_0^{(1)}\right|.
\end{aligned}
$$

where $\boldsymbol{\Xi}_0^{(1)}$ is clearly defined. Incidentally, we stress at this point that, if $\delta_n/n_{0,1} \to 0$, then $\hat{\boldsymbol{B}}_{\mathcal{N}_1}$ is consistent, $\boldsymbol{W}_n \sim \boldsymbol{I}_q$, and the above expression reduces to $n_{0,1}^q \left|\boldsymbol{\Omega}_0^{(1)}\right|^{-1}$. Also, using the approximation $\Gamma(x+\alpha) \sim \Gamma(x)x^\alpha$, for $\alpha \in \mathbb{R}$ and $x \to \infty$, we can write, for some constant C, that $\Gamma_q\left(\frac{a+n_i}{2}\right) \sim C\Gamma(n_i/2)^q n_i^\beta$, where $\beta := \frac{q}{2}(a-(q-1)/2)$. Using the same approximation, we note that $B(\alpha, \beta+x) \sim \Gamma(\alpha)x^{-\alpha}$, for $\alpha, \beta \in \mathbb{R}$ and $x \to \infty$. Putting the pieces together, we have that, for $\delta_n \to +\infty$,

$$
\begin{aligned}
\frac{p\left(T_1 = t_n^* + \delta_n, M=1|\boldsymbol{Y}_n\right)}{p\left(T_1 = t_n^*, M=1|\boldsymbol{Y}_n\right)} &\propto \frac{B(\alpha_1+1, \beta_1+n_{0,1}+\delta_n-1)}{B(\alpha_1+1, \beta_1+n_{0,1}-1)} \frac{B(\alpha_2, \beta_2+n_{0,2}-\delta_n-1)}{B(\alpha_2, \beta_2+n_{0,2}-1)} \\
&\cdot \frac{\Gamma_q\left(\frac{a+n_{0,1}+\delta_n}{2}\right)}{\Gamma_q\left(\frac{a+n_{0,1}}{2}\right)} \frac{\Gamma_q\left(\frac{a+n_{0,2}-\delta_n}{2}\right)}{\Gamma_q\left(\frac{a+n_{0,2}}{2}\right)} \prod_{i=1,2} \left(\frac{\left|\boldsymbol{C}^{(i)} + \boldsymbol{Z}_{\mathcal{N}_i^0}^\top \boldsymbol{Z}_{\mathcal{N}_i^0}\right|}{\left|\boldsymbol{C}^{(i)} + \boldsymbol{Z}_{\mathcal{N}_i}^\top \boldsymbol{Z}_{\mathcal{N}_i}\right|}\right)^{q/2} \\
&\cdot \frac{\left|\boldsymbol{R}^{(1)} + \hat{\boldsymbol{E}}_{\mathcal{N}_{01}}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_{01}} + \boldsymbol{D}^{(1)}\right|^{(a+n_{0,1})/2}}{\left|\boldsymbol{R}^{(1)} + \hat{\boldsymbol{E}}_{\mathcal{N}_1}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_1} + \boldsymbol{D}^{(1)}\right|^{(a+n_{0,1}+\delta_n)/2}} \frac{\left|\boldsymbol{R}^{(2)} + \hat{\boldsymbol{E}}_{\mathcal{N}_{02}}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_{02}} + \boldsymbol{D}^{(2)}\right|^{(a+n_{0,2})/2}}{\left|\boldsymbol{R}^{(2)} + \hat{\boldsymbol{E}}_{\mathcal{N}_2}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_2} + \boldsymbol{D}^{(2)}\right|^{(a+n_{0,2}-\delta_n)/2}} \\
&\sim C\left(1+\frac{\delta_n}{n_{0,1}}\right)^{q\frac{n_{0,1}+\delta_n}{2}} \left(\frac{\left|\boldsymbol{\Omega}_0^{(1)}\right|}{\left|\boldsymbol{\Omega}_0^{(2)}\right|}\right)^{\delta_n/2} \left|\boldsymbol{I}_q + \boldsymbol{\Omega}_0^{(1)}\boldsymbol{\Xi}_0^{(1)}\right|^{-\frac{n_{0,1}+\delta_n}{2}}
\end{aligned}
$$

Since $\delta_n \sim \gamma n_{0,1}$, we have that the above posterior ratio is asymptotically equivalent to

$$
\begin{aligned}
& C\left|\frac{\tau_{01}}{\tau_{01}+\gamma}\left[\boldsymbol{\Omega}_0^{(2)}\left(\boldsymbol{\Omega}_0^{(1)}\right)^{-1}\right]^{\gamma/(\tau_{01}+\gamma)}\left(\boldsymbol{I}_q + \boldsymbol{\Omega}_0^{(1)}\boldsymbol{\Xi}_0^{(1)}\right)\right|^{-n(\tau_{01}+\gamma)/2} \\
&= C\left|\boldsymbol{\Omega}_0^{(2)}\left(\boldsymbol{\Omega}_0^{(1)}\right)^{-1}\right|^{-n\gamma/2}\left(\left(\frac{\tau_{01}}{\tau_{01}+\gamma}\right)^q \prod_{j=1}^q\left(1+eig_j\left(\boldsymbol{\Omega}_0^{(1)}\boldsymbol{\Xi}_0^{(1)}\right)\right)\right)^{-n(\tau_{01}+\gamma)/2} \qquad (7)
\end{aligned}
$$

First treat the simplified case where $\boldsymbol{B}_0^{(1)} = \boldsymbol{B}_0^{(2)}$. In this case, $\boldsymbol{\Xi}_0^{(1)}$ reduces to $\frac{\gamma}{\tau_{01}}\left(\boldsymbol{\Omega}_0^{(2)}\right)^{-1}$ and then

$$
\frac{p\left(T_1 = t_n^* + \delta_n, M=1|\boldsymbol{Y}_n\right)}{p\left(T_1 = t_n^*, M=1|\boldsymbol{Y}_n\right)} \propto C\prod_{j=1}^q\left(\frac{\tau_{01}+\gamma\lambda_j}{\tau_{01}+\gamma}\lambda_j^{-\gamma/(\tau_{01}+\gamma)}\right)^{-n(\tau_{01}+\gamma)/2},
$$

where λ_j is the j-th eigenvalue of $\boldsymbol{\Omega}_0^{(1)}\left(\boldsymbol{\Omega}_0^{(2)}\right)^{-1}$. If $\lambda_j = 1$ for $j = 1, \ldots, q-1$ and $\lambda_q = \epsilon \neq 1$, then the posterior ratio is proportional to $\epsilon^{n\gamma/2}\left(\frac{\tau_{01}+\gamma\epsilon}{\tau_{01}+\gamma}\right)^{-n(\tau_{01}+\gamma)/2} \to 0$, and the same is true for any eigenvalues, if at least one of them is different from one.

In the general case without imposition of $\boldsymbol{B}_0^{(1)} = \boldsymbol{B}_0^{(2)}$, using the Woodbury matrix identity $\boldsymbol{I}_q - \boldsymbol{W}_n = \left(\boldsymbol{I}_q + \frac{\tau_{01}}{\gamma}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}$ we can write $\boldsymbol{\Xi}_0^{(1)}$ as follows

$$
\frac{\gamma}{\tau_{01}}\left(\boldsymbol{\Omega}_0^{(2)}\right)^{-1} + (\boldsymbol{B}_{0,1} - \boldsymbol{B}_{0,2})^\top\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)} + \frac{\tau_{01}}{\gamma}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}(\boldsymbol{B}_{0,1} - \boldsymbol{B}_{0,2}).
$$

Therefore, from Weyl's inequalities we have

$$eig_j\left(\boldsymbol{\Omega}_0^{(1)}\boldsymbol{\Xi}_0^{(1)}\right) \geq eig_j\left(\frac{\gamma}{\tau_{01}}\boldsymbol{\Omega}_0^{(1)}\left(\boldsymbol{\Omega}_0^{(2)}\right)^{-1}\right) = \lambda_j, \tag{8}$$

so that in the general case

$$\frac{p\left(T_1 = t_n^* + \delta_n, M = 1|\boldsymbol{Y}_n\right)}{p\left(T_1 = t_n^*, M = 1|\boldsymbol{Y}_n\right)} \leq C\prod_{j=1}^{q}\left(\frac{\tau_{01}+\gamma\lambda_j}{\tau_{01}+\gamma}\lambda_j^{-\tau_{01}/(\tau_{01}+\gamma)}\right)^{-n(\tau_{01}+\gamma)/2} \to 0.$$

Finally, in a similar way, it can be shown that, when $\gamma \in (-\tau_{01}, 0)$, $\delta_n \sim \gamma n \to -\infty$, and the posterior ratio is proportional to

$$C\left(1-\frac{\delta_n}{n_{0,2}}\right)^{q\frac{n_{0,2}-\delta_n}{2}}\left(\frac{\left|\boldsymbol{\Omega}_0^{(2)}\right|}{\left|\boldsymbol{\Omega}_0^{(1)}\right|}\right)^{-\delta_n/2}\left|\boldsymbol{I}_q+\boldsymbol{\Omega}_0^{(2)}\boldsymbol{\Xi}_0^{(2)}\right|^{-\frac{n_{0,2}-\delta_n}{2}}$$

$$= C\left|\frac{1-\tau_{01}}{1-\tau_{01}+|\gamma|}\left[\boldsymbol{\Omega}_0^{(1)}\left(\boldsymbol{\Omega}_0^{(2)}\right)^{-1}\right]^{|\gamma|/(1-\tau_{01}+|\gamma|)}\left(\boldsymbol{I}_q+\boldsymbol{\Omega}_0^{(2)}\boldsymbol{\Xi}_0^{(2)}\right)\right|^{-n(1-\tau_{01}+|\gamma|)/2},$$

where

$$\boldsymbol{\Xi}_0^{(2)} = \frac{|\gamma|}{1-\tau_{01}}\left(\boldsymbol{\Omega}_0^{(2)}\right)^{-1} + \left(\boldsymbol{B}_{0,2}-\boldsymbol{B}_{0,1}\right)^{\top}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}+\frac{1-\tau_{01}}{|\gamma|}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}\left(\boldsymbol{B}_{0,2}-\boldsymbol{B}_{0,1}\right),$$

so that also in this case posterior ratio consistency holds following the same line of reasoning given above for positive γ, with γ, τ_{01},$\boldsymbol{\Omega}_0^{(1)}$, $\boldsymbol{\Omega}_0^{(2)}$ and $\boldsymbol{\Xi}_0^{(1)}$ now respectively replaced by $|\gamma|$, $1-\tau_{01}$,$\boldsymbol{\Omega}_0^{(2)}$, $\boldsymbol{\Omega}_0^{(1)}$ and $\boldsymbol{\Xi}_0^{(2)}$.

In the next proposition, we show that it is possible to obtain posterior ratio consistency even when the estimation error in the regime proportion vanishes at a rate lower than the sample size.

Proposition 3.2. *Given a true proportion of data τ_{01} in the first regime, and for any $\delta_n \in \mathbb{R}\setminus\{0\}$ such that $|\delta_n| \to \infty$ and $|\delta_n|/n \to 0$, we have, with $\bar{P}$-probability 1, as $n \to \infty$, that $\frac{p(\tau_1=\tau_{01}+\delta_n/n, M=1|\boldsymbol{Y}_n)}{p(\tau_1=\tau_{01}, M=1|\boldsymbol{Y}_n)} \to 0$.*

Proof. First restricting to the case of $\delta_n \to +\infty$ and $\boldsymbol{B}_0^{(1)} = \boldsymbol{B}_0^{(2)}$, from the proof of Proposition 3.1, we know that the posterior ratio is proportional to

$$C\left|\boldsymbol{\Omega}_0^{(2)}\left(\boldsymbol{\Omega}_0^{(1)}\right)^{-1}\right|^{-\delta_n/2}\left[\left(\frac{n_{01}}{n_{01}+\delta_n}\right)^q\prod_{j=1}^{q}\left(1+\frac{\delta_n}{n_{01}}\lambda_j\right)\right]^{-(n_{01}+\delta_n)/2},$$

where λ_j is the j-th eigenvalue of $\boldsymbol{\Omega}_0^{(1)}\left(\boldsymbol{\Omega}_0^{(2)}\right)^{-1}$. The above expression is asymptotically equivalent to

$$C\left|\boldsymbol{\Omega}_0^{(2)}\left(\boldsymbol{\Omega}_0^{(1)}\right)^{-1}\right|^{-\delta_n/2}\prod_{j=1}^{q}\exp\left[-\frac{\delta_n}{2}\left(\lambda_j-1\right)\right].$$

When $\lambda_j = 1$ for $j = 1, \ldots, q-1$ and $\lambda_q = \epsilon$, note that the log posterior ratio becomes proportional to $\frac{\delta_n}{2}(\log \epsilon - \epsilon + 1) \to -\infty$, proving posterior ratio consistency, and the same is true for any eigenvalues of $\mathbf{\Omega}_0^{(1)}$ and $\mathbf{\Omega}_0^{(2)}$, if at least one of them is different in the two matrices. The extension of the proof to the case with $\boldsymbol{B}_0^{(1)} \neq \boldsymbol{B}_0^{(2)}$ and with $\delta_n \to -\infty$ follows the same structure of the proof of Proposition 3.1.

Next, we establish a posterior consistency result, specifically marginal posterior consistency, which is weaker than posterior ratio consistency. This result holds when the estimation error in the proportion of data attributed to the regimes decreases at a rate of n^{-1}, as studied in Shimizu [42].

Proposition 3.3. *Let τ_{01} be the true proportion of data in the first regime and $B_d^c(\tau) = (\tau - d, \tau + d)^c$, for some $d > 0$. With $\bar{P}$ probability 1, we have $\int_{B^c_{\epsilon/n}(\tau_{01})} p(\tau_1 = \tau, M = 1|\boldsymbol{Y}_n)d\tau \to 0$ for all $\epsilon > 0$ sufficiently large.*

Proof. Following the proof of Proposition 3.1, we can express the marginal probability of the change-point as

$$
\begin{aligned}
p\left(\tau_1 = \tau_{01} + \frac{\epsilon}{n}, M = 1|\boldsymbol{Y}_n\right) &\propto \frac{\left(\frac{n_{01}+\delta_n}{n_{01}}\right)^{\frac{q}{2}(n_{01}+\delta_n)} (2/e)^{qn/2}}{\left|\left(\mathbf{\Omega}_0^{(1)}\right)^{-1} + \mathbf{\Xi}_0^{(1)}\right|^{(n_{01}+\delta_n)/2} \left|\mathbf{\Omega}_0^{(2)}\right|^{-(n-n_{01}-\delta_n)/2}} \\
&\propto \frac{e^{\epsilon q/2}(2/e)^{qn/2}}{\left|\left(\mathbf{\Omega}_0^{(1)}\right)^{-1} + \mathbf{\Xi}_0^{(1)}\right|^{n\tau_{01}/2} \left|\mathbf{\Omega}_0^{(2)}\right|^{-n(1-\tau_{01})/2}},
\end{aligned}
$$

and similarly

$$
p\left(\tau_1 = \tau_{01} - \frac{\epsilon}{n}, M = 1|\boldsymbol{Y}_n\right) \propto \frac{e^{\epsilon q/2}(2/e)^{qn/2}}{\left|\left(\mathbf{\Omega}_0^{(2)}\right)^{-1} + \mathbf{\Xi}_0^{(2)}\right|^{n(1-\tau_{01})/2} \left|\mathbf{\Omega}_0^{(1)}\right|^{-n\tau_{01}/2}}.
$$

The marginal posterior density outside a ball around the true model can be bounded above by

$$
\begin{aligned}
\int_{B^c_{\epsilon/n}(\tau_0)} p(\tau_1 = \tau, M = 1|\boldsymbol{Y}_n)d\tau &= \int_0^{\tau_0 - \epsilon/n} p(\tau_1 = \tau, M = 1|\boldsymbol{Y}_n)d\tau \\
&\quad + \int_{\tau_0+\epsilon/n}^1 p(\tau_1 = \tau, M = 1|\boldsymbol{Y}_n)d\tau \\
&\leq p\left(\tau_1 = \tau_0 - \frac{\epsilon}{n}, M = 1|\boldsymbol{Y}_n\right)\left(\tau_{01} - \frac{\epsilon}{n}\right) \\
&\quad + p\left(\tau_1 = \tau_0 + \frac{\epsilon}{n}, M = 1|\boldsymbol{Y}_n\right)\left(1 - \tau_{01} - \frac{\epsilon}{n}\right).
\end{aligned}
$$

The second term in the expression above converges to zero if

$$q(1-\ln 2)+\tau_{01}\ln\left|\left(\mathbf{\Omega}_0^{(1)}\right)^{-1}+\mathbf{\Xi}_0^{(1)}\right|+(1-\tau_{01})\ln\left|\mathbf{\Omega}_0^{(2)}\right|>0.$$

When $\boldsymbol{B}_0^{(1)}=\boldsymbol{B}_0^{(2)}$, it is sufficient to have

$$q(1-\ln 2)+q\tau_{01}\ln\left(\frac{\epsilon}{\tau_{01}}+\min_j eig_j\left(\mathbf{\Omega}_0^{(2)}\left(\mathbf{\Omega}_0^{(1)}\right)^{-1}\right)\right)+\ln\left|\mathbf{\Omega}_0^{(2)}\right|>0,$$

that is

$$\frac{\epsilon}{\tau_{01}}>(e/2)^{1/\tau_{01}}\left|\mathbf{\Omega}_0^{(2)}\right|^{-1/(q\tau_{01})}-\min_j eig_j\left(\mathbf{\Omega}_0^{(2)}\left(\mathbf{\Omega}_0^{(1)}\right)^{-1}\right).$$

When $\boldsymbol{B}_0^{(1)}\neq\boldsymbol{B}_0^{(2)}$, by Weyl's inequalities we have

$$\left|\left(\mathbf{\Omega}_0^{(1)}\right)^{-1}+\mathbf{\Xi}_0^{(1)}\right|\geq\left|\left(\mathbf{\Omega}_0^{(1)}\right)^{-1}+\frac{\epsilon}{\tau_{01}}\left(\mathbf{\Omega}_0^{(2)}\right)^{-1}\right|,$$

and therefore the lower bound above for $\frac{\epsilon}{\tau_{01}}$ is still sufficient. A similar line of reasoning can show that the first term in the marginal posterior density converges to zero if

$$\frac{\epsilon}{1-\tau_{01}}>(e/2)^{1/(1-\tau_{01})}\left|\mathbf{\Omega}_0^{(1)}\right|^{-1/(q(1-\tau_{01}))}-\min_j eig_j\left(\mathbf{\Omega}_0^{(1)}\left(\mathbf{\Omega}_0^{(2)}\right)^{-1}\right).$$

□

The above proposition indicates that models assuming a change-point located at a distance of ϵ/n from the true τ_0 will be assigned a vanishing marginal posterior probability. Our result extends Theorem 1 of Shimizu [42] in several directions: it accounts for temporal dependence, multiple time series, and moves from a Normal-gamma conjugate prior to a Matrix Normal-Wishart prior. Intuitively, when the error in the change-point of the alternative model decreases rapidly, identifying the correct model becomes more challenging. Furthermore, it is crucial that the error term ϵ in ϵ/n is *sufficiently large*, meeting a lower bound specified in the proof. This bound depends on the matrices $\mathbf{\Omega}_0^{(1)}$ and $\mathbf{\Omega}_0^{(2)}$ and the proportions τ_{01} and $1-\tau_{01}$ of the data in the two regimes.

3.2 Extension to non-conjugacy and to multiple changes

In this section, we extend previous results to accommodate any choice of prior distribution, including non-conjugate priors. In such cases, the posterior density associated with specific change-point locations is no longer available in closed form. Building on the work of Garel and Hallin [20], we examine the Local Asymptotic Normality (LAN) of the model within regimes, leading to a Gaussian asymptotic approximation of the posterior distribution of the regime parameters. We also demonstrate the asymptotic equivalence of the marginal posterior distribution of the change-point location to that of the conjugate case, under specified conditions, regardless of the specific prior used. This holds within a model that assumes the correct location for the change-point. For models with incorrect regimes, we extend our marginal posterior consistency results to the case of misspecified models, following the approach of Shalizi [41].

Proposition 3.4. *Given a true proportion of data τ_{01} in the first regime, and γ such that $|\gamma| < min\{\tau_{01}, 1-\tau_{01}\}$, for a prior $p(\boldsymbol{\theta}, \tau_1) > 0$ for all $(\boldsymbol{\theta}, \tau_1) \in \Theta \times [0,1]$, and if the roots of $\left|\boldsymbol{I} - \sum_{j=1}^{K} \boldsymbol{B}_{0j}^{(i)} z^j\right| = 0$ are such that $|z| < 1$ for $i = 1, \dots, m+1$, we have, with $\bar{P}$-probability 1, as $n \to \infty$, that* $\frac{p(\tau_1=\tau_0+\gamma, M=1|\boldsymbol{Y}_n)}{p(T_1=t_n^*, M=1|\boldsymbol{Y}_n)} \to 0$.

Proof. The model is Local Asymptotic Normal (LAN), since conditions A1, A2 and C1-C5 in Garel and Hallin [20] are verified for each regime. In particular, condition A1 is the stated assumption on the determinantal equation, and condition A2 is trivially satified since the moving average part of the model is absent; conditions C1, C2, C3 and C5 are regularity conditions on the likelihood (continuity, existence of partial derivatives, finiteness and continuity of the expected Fisher information matrix), all satisfied because of the assumed normality; finally, condition C4 is on the initial value of the stochastic process, which we always satisfy since we assume it as known. Furthermore, note that $\boldsymbol{B}_k^{(i)} \in \mathbb{R}^{q \times q}$ and

$$\boldsymbol{\Omega}_k^{(i)} \in \bigcup_{n \in \mathbb{N}} \left\{ \boldsymbol{\Omega} \in \mathbb{R}^{q \times q} : \ \boldsymbol{\Omega} = \boldsymbol{\Omega}^\top, \ \mathrm{tr}\,(\boldsymbol{\Omega}) = n, \ \boldsymbol{x}^\top \boldsymbol{\Omega} \boldsymbol{x} \geq 0 \right\},$$

for all $k = 1, \dots, K$ and $i = 1, \dots, m+1$. Therefore the parameter space is σ-compact, that is it is a countable union of compact subsets. The LAN property, continuous prior on $\boldsymbol{\theta}$ strictly positive around the true $\boldsymbol{\theta}_0$ and σ-compactness imply a Berstein-von Mises-type theorem [43, Lemma 10.6] on the regime parameter posterior distribution, under the correctly specified change-point model. In other words, for any prior respecting the stated conditions, the posterior distribution can be asymptotically well approximated by a Gaussian distribution that only depends on the likelihood. Then the marginal likelihood, in light of Equation (5) of the main text, will be asymptotically equivalent to the one computed in the conjugate case.

On the other hand, when the change-point is misspecified, we need to verify the validity of assumptions 1-7 in Shalizi [41], for posterior distribution consistency in misspecified models. Assumption 1 of Shalizi [41] is merely a measurability assumption which is valid. Assumptions 2 and 3 are on the dynamics of the likelihood ratio. In particular, it is required that the Kullback-Leibler divergence rate

$$h(\boldsymbol{\theta}, \tau_1) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\ln \frac{f\left(\boldsymbol{Y}_n \mid \boldsymbol{\theta}_0, m_0, \mathcal{N}_1^0, \ldots, \mathcal{N}_{m+1}^0\right)}{f\left(\boldsymbol{Y}_n|\boldsymbol{\theta}, m, \mathcal{N}_1, \ldots, \mathcal{N}_{m+1}\right)}\right]$$

exists (Assumption 2) and that

$$\lim_{n\to\infty} \frac{1}{n} \ln \frac{f\left(\boldsymbol{Y}_n|\boldsymbol{\theta}_0, m_0, \mathcal{N}_1^0, \ldots, \mathcal{N}_{m+1}^0\right)}{f\left(\boldsymbol{Y}_n|\boldsymbol{\theta}, m, \mathcal{N}_1, \ldots, \mathcal{N}_{m+1}\right)} = h(\boldsymbol{\theta}, \tau_1),$$

(equipartition property, Assumption 3). This holds in our case, since it can be shown that, with $\gamma > 0$,

$$\begin{aligned} 2h(\boldsymbol{\theta}, \tau_1) \sim{} & \frac{n_{01}}{n} \ln \frac{\left|\boldsymbol{\Omega}_0^{(1)}\right|}{\left|\boldsymbol{\Omega}^{(1)}\right|} + \frac{n_{02}}{n} \ln \frac{\left|\boldsymbol{\Omega}_0^{(2)}\right|}{\left|\boldsymbol{\Omega}^{(2)}\right|} + \gamma \ln \frac{\left|\boldsymbol{\Omega}^{(1)}\right|}{\left|\boldsymbol{\Omega}^{(2)}\right|} + \gamma \mathrm{tr}\left(\left(\boldsymbol{\Omega}_0^{(2)}\right)^{-1} - \left(\boldsymbol{\Omega}_0^{(1)}\right)^{-1}\right) \\ & + \frac{n_{01}}{n} \mathrm{tr}\left(\boldsymbol{\Omega}_0^{(1)} \left(\boldsymbol{B}_{0,1} - \boldsymbol{B}_{0,2}\right)^\top \left(\boldsymbol{I}_q - \boldsymbol{W}_n\right)^\top \left(\boldsymbol{\Omega}_0^{(1)}\right)^{-1} \boldsymbol{W}_n^{-1} \left(\boldsymbol{I}_q - \boldsymbol{W}_n\right)\left(\boldsymbol{B}_{0,1} - \boldsymbol{B}_{0,2}\right)\right), \end{aligned}$$

which is almost surely finite and continuous in n, so that Assumption 4 and 7 of Shalizi [41] are also respectively satisfied. An analogous expression can be found with $\gamma < 0$. The remaining Assumptions 5 and 6 of Shalizi [41] are related to the prior choice, and they are always satisfied, since priors are defined to have a support that include the true value of $\boldsymbol{\theta}$ and τ_1. Therefore, we can use Theorem 3 in Shalizi [41], for which we know that $p((\boldsymbol{\theta}, \tau_1) \in A|\boldsymbol{Y}_n) \to 0$ on any set A with $\min_{(\boldsymbol{\theta},\tau_1)\in A} h(\boldsymbol{\theta}, \tau_1) > \min_{(\boldsymbol{\theta},\tau_1)\in\Theta\times[0,1]} h(\boldsymbol{\theta}, \tau_1)$. Our result then follows for $A = \Theta \times \{\tau_0 + \gamma\}$, $\gamma \neq 0$ for all n. □

Remark 3.1. *It is of interest to see* $h(\boldsymbol{\theta}, \tau_1)$ *in the proof of Proposition 3.4 as a generalization, to a possibly misspecifed change-point, of the usual Kullback-Leibler divergence of a VAR model* $\ln\left(|\boldsymbol{\Omega}_0| \,/\, |\boldsymbol{\Omega}|\right)$ *(Feutrill and Roughan 19), to which we reduce when* $n_{01} = n$ *(no change-point) and* $\delta_n = 0$ *(no mispecification) at every* n*.*

We also extend the results of Section 3.1 to a generic number m of change-points, under the restriction that in the alternative models the estimated regimes include a limited number of observations from wrong regimes.

Proposition 3.5. *Given true proportions of data* τ_{0k} *in the* k*-th regime, for* $k = 1, \ldots, m$*, and* $\gamma_k \in \mathbb{R}$ *different from zero for some* k*, if* $\mathcal{N}_k \cap \mathcal{N}_{0,k-1} \cap \mathcal{N}_{0,k+1} = \emptyset$ *for all regimes, we have, with* $\bar{P}$*-probability 1, as* $n \to \infty$*, that* $\frac{p(\tau_1=\tau_{01}+\gamma_1,\ldots,\tau_m=\tau_{0m}+\gamma_m, M=m|\boldsymbol{Y}_n)}{p(\tau_1=\tau_{01},\ldots,\tau_m=\tau_{0m}, M=m|\boldsymbol{Y}_n)} \to 0.$

Proof. For a generic number m of change-points, fix $\delta_{ni} \sim n\gamma_i$, $i = 1, \ldots, m+1$, and $\delta_{n0} = \delta_{n,m+1} = 0$. Then

$$\begin{aligned}
&p(\tau_1 = \tau_{01} + \gamma_1, \ldots, \tau_m = \tau_{0m} + \gamma_m, M = m|\boldsymbol{Y}_n) \\
&= p(T_{n1} = 1 + n_{01} + \delta_{n1} - \delta_{n0}, \ldots, T_{n,m+1} = 1 + n_{0,m+1} + \delta_{n,m+1} - \delta_{nm}, M = m|\boldsymbol{Y}_n) \\
&\propto \prod_{i=1}^{m+1} \frac{B(\alpha_i + 1, \beta_i + n_{0i} + \delta_{ni} - \delta_{n,i-1} - 1)}{\left(\left|\boldsymbol{C}^{(i)} + \boldsymbol{Z}_{\mathcal{N}_i}^\top \boldsymbol{Z}_{\mathcal{N}_i}\right|\right)^{q/2}} \frac{\Gamma_q\left(\frac{a + n_{0i} + \delta_{ni} - \delta_{n,i-1}}{2}\right)}{\left|\boldsymbol{R}^{(i)} + \hat{\boldsymbol{E}}_{\mathcal{N}_i}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_i} + \boldsymbol{D}^{(i)}\right|^{(a + n_{0i} + \delta_{ni} - \delta_{n,i-1})/2}},
\end{aligned}$$

where γ_i denotes the error between the evaluated and true i-th change-point, chosen appropriately to respect $0 < \tau_1 < \tau_2, \ldots, \tau_m < T$.

With $m = 2$ change points, and restricting to alternative models where each regime is contamimated by data from at most one adjacent regime, we have to distinguish three scenarios: (a) $\gamma_1 > 0, \gamma_2 > 0$, (b) $\gamma_1 > 0, \gamma_2 < 0$ and (c) $\gamma_1 < 0, \gamma_2 < 0$. In the scenarios with $\gamma_1 > 0$, observations belonging to $\mathcal{N}_{02}$ are wrongly allocated to $\mathcal{N}_1$, and then we have $\left|\boldsymbol{C}^{(1)} + \boldsymbol{Z}_{\mathcal{N}_1}^\top \boldsymbol{Z}_{\mathcal{N}_1}\right| \sim \left|n_{0,1}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1} + \delta_{n1}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\right|$; if otherwise $\gamma_1 < 0$, a portion of the observations in the correct regime $\mathcal{N}_{01}$ will be allocated in $\mathcal{N}_1$, and $\left|\boldsymbol{C}^{(1)} + \boldsymbol{Z}_{\mathcal{N}_1}^\top \boldsymbol{Z}_{\mathcal{N}_1}\right| \sim (n_{0,1} + \delta_{n1})^q \left|\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right|^{-1}$. Similarly, in those scenarios with $\gamma_2 < 0$, $\left|\boldsymbol{C}^{(3)} + \boldsymbol{Z}_{\mathcal{N}_3}^\top \boldsymbol{Z}_{\mathcal{N}_3}\right| \sim \left|n_{0,3}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}\right)^{-1} - \delta_{n2}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\right|$, whilst if $\gamma_2 > 0$ we have $\left|\boldsymbol{C}^{(3)} + \boldsymbol{Z}_{\mathcal{N}_3}^\top \boldsymbol{Z}_{\mathcal{N}_3}\right| \sim (n_{0,3} - \delta_{n2})^q \left|\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}\right|^{-1}$. For the second regime, in scenario (a) $\mathcal{N}_2$ is made of a portion of data in $\mathcal{N}_{02}$ and in $\mathcal{N}_3^0$, so that $\left|\boldsymbol{C}^{(2)} + \boldsymbol{Z}_{\mathcal{N}_2}^\top \boldsymbol{Z}_{\mathcal{N}_2}\right| \sim \left|(n_{0,2} - \delta_{n1})\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1} + \delta_{n2}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}\right)^{-1}\right|$; in scenario (b) $\mathcal{N}_2$ is made of a portion of data in $\mathcal{N}_{02}$, so that $\left|\boldsymbol{C}^{(2)} + \boldsymbol{Z}_{\mathcal{N}_2}^\top \boldsymbol{Z}_{\mathcal{N}_2}\right| \sim (n_{0,2} + \delta_{n2} - \delta_{n1})^q \left|\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right|^{-1}$; finally, in scenario (c) $\mathcal{N}_2$ is made of a portion of data in $\mathcal{N}_{02}$ and in $\mathcal{N}_{01}$, so that $\left|\boldsymbol{C}^{(2)} + \boldsymbol{Z}_{\mathcal{N}_2}^\top \boldsymbol{Z}_{\mathcal{N}_2}\right| \sim \left|(n_{0,2} + \delta_{n2})\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1} - \delta_{n1}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}\right|$.

Similar considerations can be developed for on $\left|\boldsymbol{R}^{(i)} + \hat{\boldsymbol{E}}_{\mathcal{N}_i}^\top \hat{\boldsymbol{E}}_{\mathcal{N}_i} + \boldsymbol{D}^{(i)}\right|$, starting from the different asymptotic expressions for $\hat{\boldsymbol{B}}_{\mathcal{N}_i} = \left(\boldsymbol{Z}_{\mathcal{N}_i}^\top \boldsymbol{Z}_{\mathcal{N}_i}\right)^{-1} \boldsymbol{Z}_{\mathcal{N}_i}^\top \boldsymbol{Y}_{\mathcal{N}_i}$ in the three scenarios. In scenario (a) described above $\hat{\boldsymbol{B}}_{\mathcal{N}_3}$ is consistent, whilst

$$\hat{\boldsymbol{B}}_{\mathcal{N}_1} \sim \left(\boldsymbol{I}_q + \frac{\delta_{n1}}{n_{0,1}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1} \boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1} \boldsymbol{B}_0^{(1)} + \left(\boldsymbol{I}_q + \frac{n_{0,1}}{\delta_{n1}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1} \boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1} \boldsymbol{B}_0^{(2)}$$

and

$$\begin{aligned}
\hat{\boldsymbol{B}}_{\mathcal{N}_2} \sim &\left(\boldsymbol{I}_q + \frac{\delta_{n2}}{n_{0,2} - \delta_{n1}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}\right)^{-1} \boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1} \boldsymbol{B}_0^{(2)} \\
&+ \left(\boldsymbol{I}_q + \frac{n_{0,2} - \delta_{n1}}{\delta_{n2}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1} \boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}\right)^{-1} \boldsymbol{B}_0^{(3)}.
\end{aligned}$$

Overall, in this scenario the posterior ratio is bounded above by

$$C\prod_{j=1}^{q}\left(\frac{\tau_{01}+\gamma_1\lambda_{12j}}{\tau_{01}+\gamma_1}\lambda_{12j}^{-\gamma_1/(\tau_{01}+\gamma_1)}\right)^{-n(\tau_{01}+\gamma_1)/2}$$
$$\cdot\prod_{j=1}^{q}\left(\frac{\tau_{02}-\tau_{01}-\gamma_1+\gamma_2\lambda_{23j}}{\tau_{02}-\tau_{01}-\gamma_1+\gamma_2}\lambda_{23j}^{-\gamma_2/(\tau_{02}-\tau_{01}-\gamma_1+\gamma_2)}\right)^{-n(\tau_{01}-\tau_{01}-\gamma_1+\gamma_2)/2},$$

where λ_{ikj} is the j-th eigenvalue of $\boldsymbol{\Omega}_0^{(i)}\left(\boldsymbol{\Omega}_0^{(k)}\right)^{-1}$, and an equivalent of Proposition 3.1 can be recovered. Note that in the above expression the first term is related to the contamination of the first regime by observations in the second regime, whilst the second term stems from the contamination of the second regime from observations in the third regime. In scenario (b) $\hat{\boldsymbol{B}}_{\mathcal{N}_1}$ behaves as in scenario (a),

$$\hat{\boldsymbol{B}}_{\mathcal{N}_3}\sim\left(\boldsymbol{I}_q-\frac{\delta_{n2}}{n_{0,3}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}\right)^{-1}\boldsymbol{B}_0^{(3)}+\left(\boldsymbol{I}_q-\frac{n_{0,3}}{\delta_{n2}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\boldsymbol{B}_0^{(2)}$$

and $\hat{\boldsymbol{B}}_{\mathcal{N}_2}$ is consistent, so that overall the posterior ratio is asymptotically bounded above by

$$C\prod_{j=1}^{q}\left(\frac{\tau_{02}-\tau_{01}+\gamma_2-\gamma_1\lambda_{21j}}{\tau_{02}-\tau_{01}+\gamma_2-\gamma_1}\lambda_{21j}^{\gamma_1/(\tau_{02}-\tau_{01}+\gamma_2-\gamma_1)}\right)^{-n(\tau_{02}-\tau_{01}+\gamma_2-\gamma_1)/2}$$
$$\cdot\prod_{j=1}^{q}\left(\frac{1-\tau_{02}-\gamma_2\lambda_{32j}}{1-\tau_{02}-\gamma_2}\lambda_{32j}^{\gamma_2/(1-\tau_{02}-\gamma_2)}\right)^{-n(1-\tau_{02}-\gamma_2)/2},$$

for which it is again feasible to recover a result as in Proposition 3.1. Finally in scenario (c) $\hat{\boldsymbol{B}}_{\mathcal{N}_1}$ is consistent, $\hat{\boldsymbol{B}}_{\mathcal{N}_3}$ behaves as in scenario (b), also

$$\hat{\boldsymbol{B}}_{\mathcal{N}_2}\sim\left(\boldsymbol{I}_q-\frac{\delta_{n1}}{n_{0,2}-\delta_{n2}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\boldsymbol{B}_0^{(2)}+$$
$$\left(\boldsymbol{I}_q-\frac{n_{0,2}-\delta_{n2}}{\delta_{n1}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}\boldsymbol{B}_0^{(1)},$$

and posterior ratio consistency as in Proposition 3.1 is still available, since the ratio is in the limit bounded above by

$$C\prod_{j=1}^{q}\left(\frac{\tau_{01}+\gamma_1\lambda_{12j}}{\tau_{01}+\gamma_1}\lambda_{12j}^{-\gamma_1/(\tau_{01}+\gamma_1)}\right)^{-n(\tau_{01}+\gamma_1)/2}$$
$$\cdot\prod_{j=1}^{q}\left(\frac{1-\tau_{02}-\gamma_2\lambda_{32j}}{1-\tau_{02}-\gamma_2}\lambda_{32j}^{\gamma_2/(1-\tau_{02}-\gamma_2)}\right)^{-n(1-\tau_{02}-\gamma_2)/2}.$$

For a general m number of change-points, we can follow the same development above. If regimes $k_1, k_2, \ldots, k_I$ are contaminated, respectively, by observations in regimes $k_1', k_2', \ldots, k_I'$, then we can still derive posterior ratio consistency, and the posterior ratio between the alternative model and the true one is bounded above by

$$C\prod_{i=1}^{I}\prod_{j=1}^{q}\left(\frac{|\mathcal{N}_{k_i}\cap\mathcal{N}_{0,k_i}|/n+|\gamma_k|\lambda_{k_ik_i'j}}{\tau_k-\tau_{k-1}}\lambda_{k_ik_i'j}^{-|\gamma_k|/(\tau_k-\tau_{k-1})}\right)^{-n(\tau_k-\tau_{k-1})/2}\to 0.$$

□

Remark 3.2. *The assumption $\mathcal{N}_k\cap\mathcal{N}_{0,k-1}\cap\mathcal{N}_{0,k+1}=\emptyset$ in Proposition 3.5 is due to technical reasons, and we believe it is not too limiting, since it restricts the comparison to alternative models where the estimated change points are not too wrong, in the sense that their corresponding estimated regimes can wrongly incorporate observations from at most one adjacent other regime. For instance, for the case with $m=2$ change-points, this means excluding from the comparison the scenario with $\gamma_1<0$ and $\gamma_2>0$, where $\mathcal{N}_2$ is made of all data in $\mathcal{N}_{02}$ and a portion of data in $\mathcal{N}_{01}$ and in $\mathcal{N}_{03}$. In this scenario, with $\delta_{n1}=n\gamma_1$ and $\delta_{n2}=n\gamma_2$ we have $\left|\boldsymbol{C}^{(1)}+\boldsymbol{Z}_{\mathcal{N}_1}^\top\boldsymbol{Z}_{\mathcal{N}_1}\right|\sim\left|(n_{0,1}+\delta_{n1})\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}\right|$, $\left|\boldsymbol{C}^{(3)}+\boldsymbol{Z}_{\mathcal{N}_3}^\top\boldsymbol{Z}_{\mathcal{N}_3}\right|\sim\left|(n_{0,3}-\delta_{n3})\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}\right)^{-1}\right|$,*

$$\left|\boldsymbol{C}^{(2)}+\boldsymbol{Z}_{\mathcal{N}_2}^\top\boldsymbol{Z}_{\mathcal{N}_2}\right|\sim\left|n_{0,2}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}+\delta_{n2}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}\right)^{-1}-\delta_{n1}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}\right|,$$

$\hat{\boldsymbol{B}}_{\mathcal{N}_1}$ and $\hat{\boldsymbol{B}}_{\mathcal{N}_3}$ are consistent, and

$$\begin{aligned}\hat{\boldsymbol{B}}_{\mathcal{N}_2} &\sim \left(\boldsymbol{I}_q+\frac{\delta_{n2}}{n_{0,2}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}-\frac{\delta_{n1}}{n_{0,2}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\boldsymbol{B}_0^{(2)}\\ &+\left(\boldsymbol{I}_q+\frac{n_{0,2}}{\delta_{n2}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}-\frac{\delta_{n1}}{\delta_{n2}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(3)}\right)^{-1}\boldsymbol{B}_0^{(3)}\\ &+\left(\boldsymbol{I}_q-\frac{n_{0,2}}{\delta_{n1}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}-\frac{\delta_{n2}}{\delta_{n1}}\left(\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(2)}\right)^{-1}\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right)^{-1}\boldsymbol{B}_0^{(1)}.\end{aligned}$$

We conjecture that posterior ratio consistency can still be recovered, for a form of a posterior ratio that generalizes those derived in the proofs of the previous propositions.

We emphasize that the above result is a direct extension of Proposition 3.1 from the case of a single change-point ($m=1$) to an arbitrary number of change-points. Additionally, a corresponding result that extends Proposition 3.2 to multiple change-points is also available under the same conditions.

4 Simulations and Data Analysis

We conducted two simulated experiments. In the first experiment, we computed the Bayes factor (BF) between the true model (in the denominator) and an alternative model that estimates a change point with an error δ_n, to assess our ability to detect the correct change-point location. Figure 1 displays the log BF for $q = 2$, $n = 150{,}000$, with the change point located at $t_{1n} = 90{,}000$. This scenario involves a change in the $\mathbf{B}$ matrix, from

$$\mathbf{B}^{(1)} = \begin{pmatrix} 0.6 & 0.2 \\ 0 & 0.1 \end{pmatrix} \quad \text{to} \quad \mathbf{B}^{(2)} = \begin{pmatrix} 0.6 & 0.1 \\ 0.1 & 0.5 \end{pmatrix},$$

with the precision matrix remaining unchanged at $\mathbf{\Omega}^{(1)} = \mathbf{\Omega}^{(2)} = \mathbf{I}_2$. Similar results can be obtained with different parameter values, and the locations of the change points are more easily identified when the $\mathbf{B}$ matrices in the two regimes are more distinct. On the left side of the figure, we observe a symmetric decay of the BF as we move away from the true change-point location, consistently identifying the correct model as the most likely. The right side of the figure zooms in on the area around the change-point of the two time series. In the second experiment, shown in Figure 2, we modified the previous setting to include a change in the precision matrix, from $\mathbf{\Omega}^{(1)} = \mathbf{I}_2$ to $\mathbf{\Omega}^{(2)} = 0.7\mathbf{I}_2$. This adjustment made the model selection problem easier due to the increase in the difference between the two regimes, resulting in a BF that is lower than in the previous scenario for any level of δ_n. Additionally, when the regimes have different precision matrices, we observe an asymmetric BF decay for positive and negative errors in the change-point location. As expected, it is easier to distinguish models that mistakenly include observations from the high-variability regime in the low-variability regime.

In the second experiment, we compare alternative models across various true change points. Table 1 presents the results for a single potential change point, showing the true proportion of data in the first regime (τ_{01}, where $\tau_{01} = 1$ indicates that there are no change points). The proportion of times that each model is selected, based on 100 randomly generated samples of size $n = 150,000$, clearly identifies the correct model. Models with larger errors are never selected.

Then, in Table 2, we extend the experiments to two potential change points that occur at different values of τ_{01} and τ_{02}. In more detail, $\tau_{01} = 1$ again denotes absence of regimes, $\tau_{01} + \tau_{02} = 1$ denotes the identification of one change point, $\tau_{01} + \tau_{02} < 1$ of two change points.

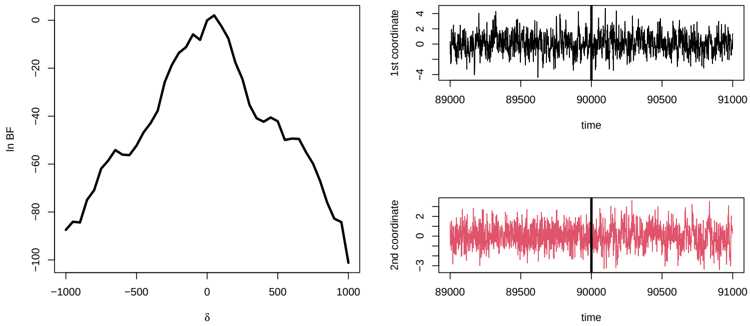

Figure 1: Left: true log Bayes Factor as a function of δ_n, the change-point estimation error. Right: first (top) and second (bottom) coordinate of the bivariate time series, around the change-point (vertical line) in the matrix $\boldsymbol{B}$.

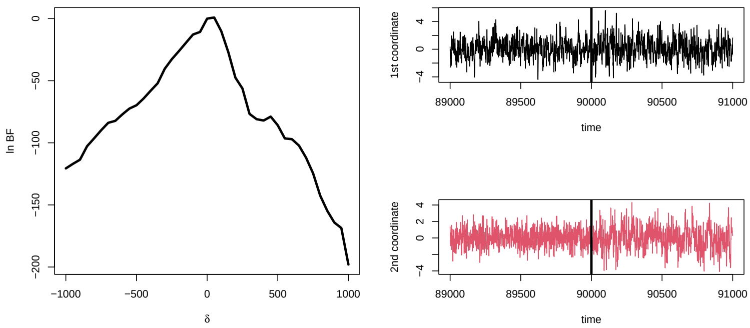

Figure 2: Left: true log BF (line) as a function of δ, the change-point estimation error. Right: first (top) and second (bottom) coordinate of the bivariate time series, around the change-point (vertical line) in both the matrices $\boldsymbol{B}$ and $\boldsymbol{\Omega}$.

	$100\delta_1/T$										
τ_{01}	-0.17	-0.13	-0.1	-0.07	-0.03	0	0.03	0.07	0.1	0.13	0.17
1.00	0	0	0	0.00	0.00	**1.00**	0.00	0.00	0	0	0
0.95	0	0	0	0.00	0.01	**0.97**	0.02	0.00	0	0	0
0.90	0	0	0	0.00	0.01	**0.93**	0.06	0.00	0	0	0
0.85	0	0	0	0.00	0.01	**0.98**	0.01	0.00	0	0	0
0.80	0	0	0	0.00	0.01	**0.97**	0.02	0.00	0	0	0
0.75	0	0	0	0.00	0.00	**0.99**	0.01	0.00	0	0	0
0.70	0	0	0	0.00	0.03	**0.93**	0.04	0.00	0	0	0
0.65	0	0	0	0.00	0.01	**0.94**	0.05	0.00	0	0	0
0.60	0	0	0	0.00	0.01	**0.94**	0.05	0.00	0	0	0
0.55	0	0	0	0.00	0.02	**0.92**	0.05	0.01	0	0	0
0.50	0	0	0	0.00	0.06	**0.90**	0.03	0.01	0	0	0
0.45	0	0	0	0.00	0.03	**0.93**	0.04	0.00	0	0	0
0.40	0	0	0	0.00	0.01	**0.98**	0.01	0.00	0	0	0
0.35	0	0	0	0.00	0.03	**0.95**	0.02	0.00	0	0	0
0.30	0	0	0	0.01	0.04	**0.93**	0.02	0.00	0	0	0
0.25	0	0	0	0.01	0.03	**0.93**	0.03	0.00	0	0	0
0.20	0	0	0	0.00	0.03	**0.94**	0.03	0.00	0	0	0
0.15	0	0	0	0.00	0.04	**0.91**	0.04	0.01	0	0	0
0.10	0	0	0	0.00	0.05	**0.93**	0.02	0.00	0	0	0
0.05	0	0	0	0.00	0.02	**0.94**	0.04	0.00	0	0	0

Table 1: Simulated results with zero or one change points. The true data proportion in the first regime is denoted by the row, and the evaluated model commits a mistake denoted by the column. Cells contain proportion of times, over 100 simulated datasets of size 150000, the column model is selected, with best models in boldface.

				(τ_1, τ_2)			
(τ_{01}, τ_{02})	$(1, 0)$	$(0.75, 0.25)$	$(0.5, 0.5)$	$(0.5, 0.25)$	$(0.25, 0.75)$	$(0.25, 0.5)$	$(0.25, 0.25)$
$(1, 0)$	**-701549.1**	-701678.2	-701679.2	-701806.7	-701678.4	-701806.9	-701806.8
$(0.75, 0.25)$	-720171.3	**-714971.6**	-717947.8	-715100.4	-719421.4	-715100.2	-718075.3
$(0.5, 0.5)$	-734191.4	-732803.8	**-728379.2**	-728480.6	-731959.7	-731456.1	-728507.2
$(0.5, 0.25)$	-720198.2	-719445.0	-717967.3	**-715127.4**	-719436.7	-718097.4	-718095.0
$(0.25, 0.75)$	-745498.6	-745336.5	-744736.2	-744837.7	**-741778.8**	-741881.3	-741881.2
$(0.25, 0.5)$	-734179.2	-731947.2	-734291.0	-731437.1	-731949.1	**-728492.1**	-731447.1
$(0.25, 0.25)$	-720201.2	-719445.7	-717974.2	-718101.7	-719446.0	-718104.5	**-715127.0**

Table 2: Simulated results with zero, one or two change points. The true data proportions in the first and second regimes are denoted by the row, and the data proportions in the regimes in the evaluated model are denoted by the column. Cells contain average, over 100 simulated datasets of size 150000, log marginal likelihoods, with best models in boldface.

Finally, we apply our approach to three key monthly US macroeconomic variables ($q = 3$), which have been extensively used in studies such as McCracken and Ng [32], Carriero et al. [9], and Dufays et al. [18]. These data are publicly available through the Federal Reserve Bank of St. Louis and can also be accessed at `http://qed.econ.queensu.ca/jae/datasets/dufays001/` [18]. The variables include the civilian unemployment rate (UNRATE), the federal funds rate (FEDFUNDS), and the spread between 10-year Treasury bills and federal funds rates (T10YFFM), covering the period from February 1959 to May 2017 ($n = 700$).

We set $a = q + 1$ and assume $\boldsymbol{R}^{(1)} = \boldsymbol{R}^{(2)} = \boldsymbol{I}_q/100$, centering the precision matrices in both regimes on high variances with significant prior uncertainty. Furthermore, we fix $\underline{\boldsymbol{B}}^{(1)} = \underline{\boldsymbol{B}}^{(2)} = \mathbf{0}$ and $\boldsymbol{C}^{(1)} = \boldsymbol{C}^{(2)} = \boldsymbol{I}_q/100$, reflecting the a priori belief in the absence of lagged dependencies among the series, albeit with high uncertainty. Evaluating the marginal likelihoods as shown in Table 2, but using a finer grid for change points with a step size of 0.1, we identified two change points (indicated by vertical bars in Figure 3): one in July 1976 and another in March 1988. These changes define an intermediate regime characterized by unusual spikes in FEDFUNDS and T10YFFM. Although we explored the possibility of additional change points, they resulted in lower marginal likelihoods.

For comparison, we also implemented two additional methods for detecting change points in VAR models that are available in the literature: the threshold block segmentation scheme (TBSS) by Bai et al. [1] and Safikhani and Shojaie [38], and the FASTCPD method by Zhang and Dawn [48] and Li and Zhang [28]. The results of these methods are shown in the lower part of Figure 3. TBSS and FASTCPD identified seven and four change points, respectively, but the three methods did not converge on the same set of changes. The two changes identified by our approach are similar, though not identical, to two of the change points detected by TBSS and FASTCPD. When we evaluated the marginal likelihood at the change points identified by the alternative methods, the log Bayes factors were -117.27 and -179.2, favoring our selected model.

5 Conclusions and further directions

We have proposed a Bayesian methodology for detecting change-points in Vector Autoregressive Processes, with a focus on achieving posterior consistency in identifying change-point locations, even when alternative models assume incorrect locations with finite or

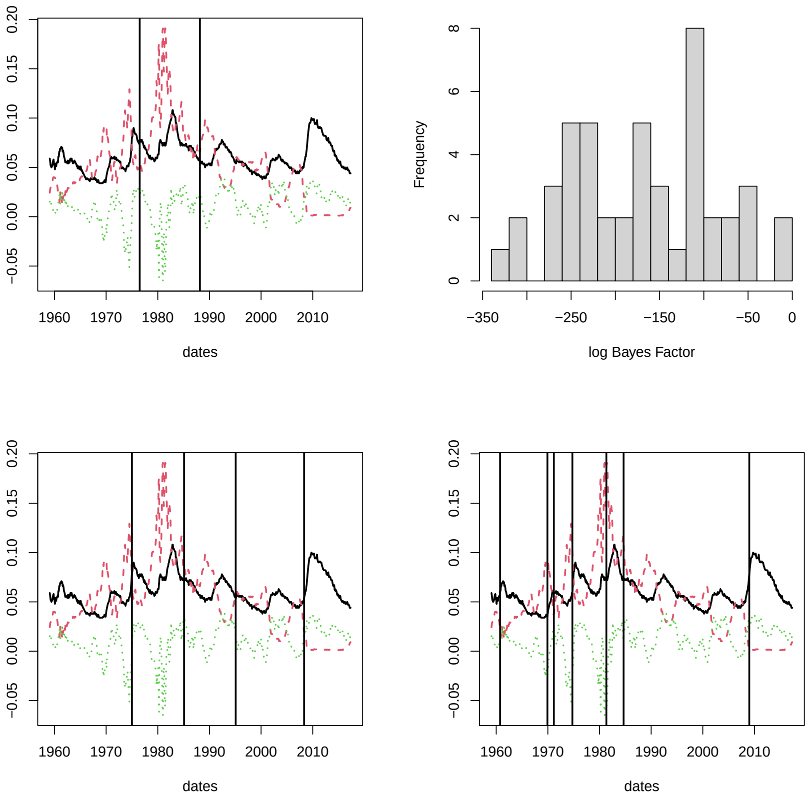

Figure 3: Top left: the three macroeconomic monthly series UNRATE, FEDFUNDS and T10YFFM, from February 1959 to May 2020, with change-points as vertical bars, estimated by our method. Top right: log Bayes Factor between alternative change point configurations and the best selected model. Bottom left: macroeconomic series change-points, estimated by TBSS. Bottom right: macroeconomic series change-points, estimated by FASTCPD.

vanishing errors. Our approach extends conjugate analysis to non-conjugate and misspecified priors, as well as to scenarios involving multiple change-points, under specific conditions related to contamination among regimes. The theoretical results are validated through simulations, and we demonstrate the practical utility of our model on a multivariate macroeconomic series, comparing its performance with established alternative methods.

Settings with an increasing number of time series q are also of scientific interest. Notable works in this direction include Basu and Michailidis [4] and Ghosh et al. [21, 22], although, to our knowledge, no change-point analysis has been explored in the literature. A generalization of our approach to the cases where q increases with n is currently under investigation. We conjecture that similar proof structures can be applied, provided that appropriate sparsity constraints are introduced in both contemporaneous and lagged dependencies. These constraints would involve imposing conditional independences in the precision matrices $\boldsymbol{\Omega}^{(i)}$ and the lag matrices $\boldsymbol{B}^{(i)}$. In this high-dimensional context, some of the key results we have relied on, such as the asymptotic equivalence between $\left|\boldsymbol{C}^{(1)} + \boldsymbol{Z}_{\mathcal{N}_{01}}^\top \boldsymbol{Z}_{\mathcal{N}_{01}}\right|$ and $n_{01}^q \left|\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}\right|^{-1}$, are no longer applicable. However, they can be replaced by similar expressions that involve sparsity-driven submatrices of $\boldsymbol{\Omega}_{0,\boldsymbol{Y}}^{(1)}$. For a similar approach to bridge low- and high-dimensional settings within the graphical modeling framework, see Castelletti and Peluso [11].

Finally, it is possible to extend the proposed methodology to multivariate change points, where multiple change point processes are linked to different subsets of the regime parameters, as in Peluso et al. [36]. This allows for distinguishing the timing of regime changes in the lag matrix from those in the covariance matrix. While such an extension is likely to introduce additional notational complexity, it is expected to improve finite-sample performance in identifying change-points, due to the increased number of observations available within each parameter-specific regime.

Acknowledgments

SP acknowledges support from the SNSF project CRSK-2_190656 and, jointly with AM, from the SNSF project 200021_200557.

References

[1] P. Bai, A. Safikhani, and G. Michailidis. Multiple change points detection in low rank and sparse high dimensional vector autoregressive models. *IEEE Transactions on Signal Processing*, 68:3074–3089, 2020.

[2] P. Bai, A. Safikhani, and G. Michailidis. Multiple change point detection in reduced rank high dimensional vector autoregressive models. *Journal of the American Statistical Association*, pages 1–17, 2023.

[3] M. Bańbura, D. Giannone, and L. Reichlin. Large bayesian vector auto regressions. *Journal of applied Econometrics*, 25(1):71–92, 2010.

[4] S. Basu and G. Michailidis. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 2015.

[5] S. Basu, A. Shojaie, and G. Michailidis. Network granger causality with inherent grouping structure. *The Journal of Machine Learning Research*, 16(1):417–453, 2015.

[6] B. S. Bernanke, J. Boivin, and P. Eliasz. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1):387–422, 2005.

[7] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics*, 104(3):535–559, 2012.

[8] X. Cao, K. Khare, and M. Ghosh. Posterior graph selection and estimation consistency for high-dimensional bayesian dag models. *The Annals of Statistics*, 47(1): 319–348, 2019.

[9] A. Carriero, T. E. Clark, and M. Marcellino. Large bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, 212(1): 137–154, 2019.

[10] F. Castelletti and S. Peluso. Network structure learning under uncertain interventions. *Journal of the American Statistical Association*, pages 1–12, 2022.

[11] F. Castelletti and S. Peluso. Bayesian learning of network structures from interventional experimental data. *Biometrika*, page asad032, 2023.

[12] S. Chib. Marginal likelihood from the gibbs output. *Journal of the american statistical association*, 90(432):1313–1321, 1995.

[13] S. Chib. Estimation and comparison of multiple change-point models. *Journal of econometrics*, 86(2):221–241, 1998.

[14] T. E. Clark. Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, 29(3):327–341, 2011.

[15] G. Consonni, L. La Rocca, and S. Peluso. Objective bayes covariate-adjusted sparse graphical model selection. *Scandinavian Journal of Statistics*, 44(3):741–764, 2017.

[16] A. P. Dawid. Some matrix-variate distribution theory: notational considerations and a bayesian application. *Biometrika*, 68(1):265–274, 1981.

[17] X. Ding, Z. Qiu, and X. Chen. Sparse transition matrix estimation for high-dimensional and locally stationary vector autoregressive models. *Electronic Journal of Statistics*, 11(2):3871–3902, 2017.

[18] A. Dufays, Z. Li, J. V. Rombouts, and Y. Song. Sparse change-point var models. *Journal of Applied Econometrics*, 36(6):703–727, 2021.

[19] A. Feutrill and M. Roughan. A review of shannon and differential entropy rate estimation. *Entropy*, 23(8):1046, 2021.

[20] B. Garel and M. Hallin. Local asymptotic normality of multivariate arma processes with a linear trend. *Annals of the Institute of Statistical Mathematics*, 47:551–579, 1995.

[21] S. Ghosh, K. Khare, and G. Michailidis. High-dimensional posterior consistency in bayesian vector autoregressive models. *Journal of the American Statistical Association*, 2018.

[22] S. Ghosh, K. Khare, and G. Michailidis. Strong selection consistency of bayesian vector autoregressive models based on a pseudo-likelihood approach. *The Annals of Statistics*, 49(3):1267–1299, 2021.

[23] P. R. Hansen. Structural changes in the cointegrated vector autoregressive model. *Journal of econometrics*, 114(2):261–295, 2003.

[24] F. Huber, T. Krisztin, and M. Pfarrhofer. A bayesian panel vector autoregression to analyze the impact of climate shocks on high-income economies. *The Annals of Applied Statistics*, 17(2):1543–1573, 2023.

[25] C. Inclan and G. C. Tiao. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89(427): 913–923, 1994.

[26] J. Knoblauch and T. Damoulas. Spatio-temporal bayesian on-line changepoint detection with model selection. In *International Conference on Machine Learning*, pages 2718–2727. PMLR, 2018.

[27] S. Lee, S. Lee, and K. Maekawa. Change point test for structural vector autoregressive model via independent component analysis. *Journal of Statistical Computation and Simulation*, 93(5):687–707, 2023.

[28] X. Li and X. Zhang. fastcpd: Fast change point detection in r. *arXiv preprint arXiv:2404.05933*, 2024.

[29] J. Lin and G. Michailidis. Regularized estimation and testing for high-dimensional multi-block vector-autoregressive models. *Journal of Machine Learning Research*, 18, 2017.

[30] H. Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.

[31] H. Lütkepohl and A. Netsunajev. Structural vector autoregressions with heteroskedasticity: A review of different volatility models. *Econometrics and statistics*, 1:2–18, 2017.

[32] M. W. McCracken and S. Ng. Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589, 2016.

[33] I. Melnyk and A. Banerjee. Estimating structured vector autoregressive models. In *International Conference on Machine Learning*, pages 830–839. PMLR, 2016.

[34] L. Paci and G. Consonni. Structural learning of contemporaneous dependencies in graphical var models. *Computational Statistics & Data Analysis*, 144:106880, 2020.

[35] B. U. Park, E. Mammen, W. Härdle, and S. Borak. Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association*, 104(485):284–298, 2009.

[36] S. Peluso, S. Chib, and A. Mira. Semiparametric multivariate and multiple change-point modeling. *Bayesian Analysis*, 14(3):727–751, 2019.

[37] C. E. Quintos. Fully modified vector autoregressive inference in partially nonstationary models. *Journal of the American Statistical Association*, 93(442):783–795, 1998.

[38] A. Safikhani and A. Shojaie. Joint structural break detection and parameter estimation in high-dimensional nonstationary var models. *Journal of the American Statistical Association*, 117(537):251–264, 2022.

[39] M. Schweinberger, S. Babkin, and K. B. Ensor. High-dimensional multivariate time series with additional structure. *Journal of Computational and Graphical Statistics*, 26(3):610–622, 2017.

[40] A. K. Seth, A. B. Barrett, and L. Barnett. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8):3293–3297, 2015.

[41] C. R. Shalizi. Dynamics of Bayesian updating with dependent data and misspecified models. *Electronic Journal of Statistics*, 3(none):1039 – 1074, 2009. doi: 10.1214/09-EJS485. URL `https://doi.org/10.1214/09-EJS485`.

[42] K. Shimizu. Asymptotic properties of bayesian inference in linear regression with a structural break. *Journal of Econometrics*, 2022.

[43] A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[44] D. Wang and R. S. Tsay. Rate-optimal robust estimation of high-dimensional vector autoregressive models. *The Annals of Statistics*, 51(2):846–877, 2023.

[45] D. Wang, Y. Zheng, H. Lian, and G. Li. High-dimensional vector autoregressive time series modeling via tensor decomposition. *Journal of the American Statistical Association*, 117(539):1338–1356, 2022.

[46] M. West and J. Harrison. *Bayesian forecasting and dynamic models*. Springer Science & Business Media, 2006.

[47] X. Xuan and K. Murphy. Modeling changing dependency structure in multivariate time series. In *Proceedings of the 24th international conference on Machine learning*, pages 1055–1062, 2007.

[48] X. Zhang and T. Dawn. Sequential gradient descent and quasi-newton's method for change-point analysis. In *International Conference on Artificial Intelligence and Statistics*, pages 1129–1143. PMLR, 2023.