Bayesian Inference on the Order of Stationary Vector Autoregressions

Rachel L. Binks[*] Sarah E. Heaps[†] , Mariella Panagiotopoulou[‡],
Yujiang Wang[§] , and Darren J. Wilkinson[¶]

Abstract. Vector autoregressions (VARs) are a widely used tool for modelling multivariate time-series. It is common to assume a VAR is stationary; this can be enforced by imposing the stationarity condition which restricts the parameter space of the autoregressive coefficients to the stationary region. However, implementing this constraint is difficult due to the complex geometry of the stationary region. Fortunately, recent work has provided a solution for autoregressions of fixed order p based on a reparameterization in terms of a set of interpretable and unconstrained transformed partial autocorrelation matrices. In this work, focus is placed on the difficult problem of allowing p to be unknown, developing a prior and computational inference that takes full account of order uncertainty. Specifically, the multiplicative gamma process is used to build a prior which encourages increasing shrinkage of the partial autocorrelations with increasing lag. Identifying the lag beyond which the partial autocorrelations become equal to zero then determines p. Based on classic time-series theory, a principled choice of truncation criterion identifies whether a partial autocorrelation matrix is effectively zero. Posterior inference utilizes Hamiltonian Monte Carlo via Stan. The work is illustrated in a substantive application to neural activity data to investigate ultradian brain rhythms.

Keywords: Electroencephalography (EEG) data, Granger causality, increasing shrinkage prior, Stan, time-series decomposition, unconstrained reparameterization.

MSC2020 subject classifications: Primary 62F15, 62F30, 62M10; secondary 62P10.

1 Introduction

Vector autoregressive (VAR) processes are widely used to model multivariate time-series data in a variety of fields including neuroscience (Chiang et al., 2016), bioinformatics (Jiang et al., 2013; Hannaford et al., 2023), macroeconomics (Koop and Korobilis, 2010), and energy economics (Heaps et al., 2020). In an autoregression of order p, the random variable at time t is conditionally independent of its values at lags $p+1, p+2, \ldots$ given observations at the preceding p time points. Indeed the random variable at time t can be expressed as a noisy linear combination of these p values. The order of the autoregression

[*]School of Mathematics, Statistics and Physics, Newcastle University, rachel.binks@newcastle.ac.uk

[†]Department of Mathematical Sciences, Durham University, sarah.e.heaps@durham.ac.uk

[‡]School of Computing, Newcastle University, m.panagiotopoulou2@newcastle.ac.uk

[§]School of Computing, Newcastle University, yujiang.wang@newcastle.ac.uk

[¶]Department of Mathematical Sciences, Durham University, darren.j.wilkinson@durham.ac.uk

is therefore intrinsic to the characterization of the joint process and plays a vital role in forecasting. However, its value is typically not known *a priori.*

A common assumption when working with Gaussian time-series is that of stationarity, which posits that the means, variances and covariances of the process do not change over time. Since the overall level of many time-series exhibits periodic or systematic variation due to seasonality or time-trends, stationarity is often implausible as an assumption when modelling the raw data. However, stationary vector autoregressions frequently form the core building block of more sophisticated models, for example for differenced data in integrated models, for innovations from a time-varying mean in a time-series regression model or simply as components in state space models which are thought to be mean-reverting. From a practical perspective, enforcing stationarity prevents the predictive variance of the process from growing without bound into the future. This is often keenly motivated, for instance in applications where the goal is long-term forecasting or when modelling the dynamics of a linear system which is assumed to be in its equilibrium distribution. Moreover, stationarity admits various interpretations of the relationships between variables through the infinite-order moving average representation of the process, for example in Granger causality graphs or impulse response analysis.

Stationarity can be enforced by restricting the autoregressive coefficient matrices to lie within a constrained parameter space called the stationary region. However, the highly complex geometry of this region hampers the process of specifying a prior and subsequent computational inference. Fortunately, in recent work, Heaps (2023) solved the problem for vector autoregressions of fixed order by introducing an unconstrained and interpretable reparameterization of the stationary model. This is constructed by mapping the original model parameters to a set of partial autocorrelation matrices, which can be regarded as a vector analogue of the partial autocorrelation function of a univariate autoregression. A second transformation then scales the singular values of each of these partial autocorrelation matrices from $[0, 1)$ to the positive real line. The transformed partial autocorrelation matrices are interpretable and allow specification of a prior which is invariant with respect to the order of the components in the observation vector. Markov chain Monte Carlo (MCMC) methods for computational inference need only operate over a Euclidean space, making implementation routine. However, a clear limitation of this work is that inference is conditional on a fixed order of the process, with no account for the uncertainty in its value.

Extending this work, we focus on development of a prior and associated procedures for computational inference which allow the order of stationary vector autoregressions to be learned. Although the general problem is hitherto unaddressed in the literature, it has been widely studied in the special case of univariate stationary autoregressions. In an approach which is similar in spirit to ours, Barnett et al. (1996) enforce stationarity by reparameterizing the univariate model in terms of its partial autocorrelations, each with support restricted to $(-1, 1)$. A univariate stationary autoregression of order p has a non-zero partial autocorrelation at lag p and zero partial autocorrelations at all higher lags. Spike-and-slab priors, with an atom of probability at zero, are then chosen for each partial autocorrelation in a model with a large (maximum) value for p, so

that identifying the lag of the final non-zero autocorrelation identifies the order of the process. Vermaak et al. (2004) use the same reparameterization of the model to enforce stationarity but frame the problem of order determination as a model selection problem and use reversible jump Markov chain Monte Carlo to learn the order of the process. In Huerta and West (1999), stationarity is enforced through a different reparameterization of the univariate autoregression in terms of the reciprocal roots of its characteristic equation. Under this parameterization, the process is stationary if and only if the reciprocal roots have moduli less than 1 and the order of the process is determined by the number of reciprocal roots with non-zero modulus. Priors are assigned to the real and complex reciprocal roots with atoms of probability at moduli 0 in each case, thereby allowing inference on the model order. In the Bayesian framework, the only generalization of these ideas to stationary *vector* autoregressions appears in Huerta and Prado (2006) who extend Huerta and West (1999) by considering a multivariate generalization of the characteristic equation. However, because this generalization is only available when the autoregressive coefficient matrices are diagonal, the approach is limited to the class of diagonal vector autoregressive processes. Other recent work which addresses the problem of order determination in vector autoregressions includes Zhang et al. (2021) and Fan et al. (2022) but their focus is on classes of rank-reduced models and stationarity is not enforced.

In this paper, we provide the first methodology for quantifying uncertainty in model order for the full class of stationary vector autoregressions. To this end, we enforce stationarity using the transformed partial autocorrelation parameterization of Heaps (2023), exploiting a number of its properties to build a shrinkage prior for an overparameterized model. In particular, under this parameterization, the model of order k is nested within the model of order $k+1$. We can therefore fit an overparameterized model with purposefully more lags than are required and construct a prior which increasingly shrinks the transformed partial autocorrelation matrices at higher lags towards zero. By identifying the lag beyond which the partial autocorrelations become essentially equal to zero, we learn about the order of the process. The interpretability of the reparameterized model allows classical theory on the sampling distribution of the partial autocorrelation function to inform specification of the shrinkage prior and subsequent decision-making about whether a partial autocorrelation matrix is effectively zero. This is based on a truncation criterion for the max norm of the matrix which is completely general and could be applied with any prior for an overparameterized stationary model. We also provide code for implementation of computational inference via Stan thereby facilitating straightforward use by a variety of practitioners across the spectrum of fields which rely on vector autoregressions for modelling and forecasting applications.

The remainder of the paper is structured as follows. In Section 2 we discuss the core reparameterization of stationary vector autoregressive models in terms of transformed partial autocorrelation matrices. In Section 3 we discuss the prior distribution assigned to the unknowns in our reparameterized model. Section 4 considers posterior inference and the use of a truncation criterion to determine model order. In Section 5 we apply our model and inferential procedures in a set of simulation experiments before considering a substantive application to neural activity data in Section 6. Finally Section 7 provides some concluding remarks.

2   Stationary vector autoregressions

2.1   Vector autoregressions

Without loss of generality, suppose that the m-variate process $\{\boldsymbol{y}_t\}$ can be modelled as a zero-mean vector autoregression of order p, denoted $\text{VAR}_m(p)$,

$$\boldsymbol{y}_t = \phi_1 \boldsymbol{y}_{t-1} + \ldots + \phi_p \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t, \tag{1}$$

where the errors $\boldsymbol{\epsilon}_t$ form a sequence of uncorrelated, zero-mean multivariate normal random vectors, $\boldsymbol{\epsilon}_t \sim \text{N}_m(\boldsymbol{0}, \Sigma)$. The continuous model parameters therefore comprise the autoregressive coefficient matrices $\phi_i \in M_{m \times m}(\mathbb{R})$, $i = 1, \ldots, p$, which are denoted collectively as $\Phi \in M_{m \times m}(\mathbb{R})^p$, and the error variance matrix $\Sigma \in \mathcal{S}_m^+$, where $M_{m \times n}(V)$ and $\mathcal{S}_m^+$ denote the space of $m \times n$ matrices with entries in V and the space of $m \times m$ symmetric, positive definite matrices, respectively. Defining B as the backshift operator, such that $B^k \boldsymbol{y}_t = \boldsymbol{y}_{t-k}$, it is common to express (1) as

$$\boldsymbol{\epsilon}_t = (I_m - \phi_1 B - \ldots - \phi_p B^p) \boldsymbol{y}_t = \phi(B) \boldsymbol{y}_t,$$

in which I_m is the $m \times m$ identity matrix and $\phi(u) = (I_m - \phi_1 u - \ldots - \phi_p u^p)$, $u \in \mathbb{C}$, is referred to as the characteristic polynomial. A vector autoregression is stable if and only if all the roots of $\det\{\phi(u)\} = 0$ lie outside the unit circle. Since all stable processes are stationary, and unstable stationary processes are not generally of interest, this is often referred to as the stationarity condition for Φ and the subset of $M_{m \times m}(\mathbb{R})^p$ over which the condition is satisfied is referred to as the stationary region, denoted $\mathcal{C}_{p,m}$.

2.2   Reparameterization over the stationary region

As illustrated in Section S1 of the Supplementary Materials (Binks et al., 2024), the stationary region $\mathcal{C}_{p,m}$ has a complex geometry, especially when p or m is large. With no standard distributions over $\mathcal{C}_{p,m}$, this complicates the process of specifying a prior that conveys meaningful information, for example, concerning the relative sizes of the autocorrelations at different lags. Moreover, it is difficult to design an efficient MCMC sampler which targets a distribution with support constrained to $\mathcal{C}_{p,m}$. In recent work, Heaps (2023) proposes a solution which addresses both issues, reparameterizing the model over the stationary region in terms of a set of interpretable, unconstrained parameters. The reparameterization involves two bijective mappings. First, the original model parameters $(\Sigma, \Phi) \in \mathcal{S}_m^+ \times \mathcal{C}_{p,m}$ are mapped to a new parameter set $\{\Sigma, (P_1, \ldots, P_p)\} \in \mathcal{S}_m^+ \times \mathcal{V}^p$ in which $\mathcal{V}$ denotes the subset of matrices in $M_{m \times m}(\mathbb{R})$ whose singular values are less than one. The matrix P_{s+1} is referred to as the $(s+1)$-th partial autocorrelation matrix. It is defined as the conditional cross-covariance matrix between $\boldsymbol{y}_{t+1}$ and $\boldsymbol{y}_{t-s}$ given $\boldsymbol{y}_t, \ldots, \boldsymbol{y}_{t-s+1}$ which has been standardized through

$$P_{s+1} = \Sigma_s^{-1/2} \text{Cov}(\boldsymbol{y}_{t+1}, \boldsymbol{y}_{t-s} | \boldsymbol{y}_t, \ldots, \boldsymbol{y}_{t-s+1}) \Sigma_s^{* -1/2},$$

$s = 0, \ldots, p-1$, in which Σ_s and Σ_s^* are the conditional variances

$$\Sigma_s = \text{Var}(\boldsymbol{y}_{t+1} | \boldsymbol{y}_t, \ldots, \boldsymbol{y}_{t-s+1}) \quad \text{and} \quad \Sigma_s^* = \text{Var}(\boldsymbol{y}_{t-s} | \boldsymbol{y}_{t-s+1}, \ldots, \boldsymbol{y}_t)$$

and $\Sigma^{1/2}$ denotes the symmetric matrix-square-root. Full details of the mapping and its inverse, which proceed by recursion, are described in Heaps (2023). The steps are also outlined in Section S2 of the Supplementary Materials (Binks et al., 2024), with a numerical example for the simple case where $m = 2$ and $p = 2$.

A second transformation then maps each partial autocorrelation matrix $P \in \mathcal{V}$ to an unconstrained square matrix $A \in M_{m \times m}(\mathbb{R})$ through

$$A = (I_m - PP^{\mathrm{T}})^{-1/2}P.$$

Denoting the singular value decomposition of P by $P = U\mathrm{diag}(r_1, \ldots, r_m)V^{\mathrm{T}}$ in which the singular values satisfy $1 > r_1 \geq r_2 \geq \cdots \geq r_m \geq 0$, the corresponding factorization of A is given by $A = U\mathrm{diag}(\tilde{r}_1, \ldots, \tilde{r}_m)V^{\mathrm{T}}$ where $\tilde{r}_i = r_i/(1 - r_i^2)^{1/2} \geq 0$, $i = 1, \ldots, m$. Therefore the second transformation can be regarded as an orientation-preserving mapping which simply scales the singular values of P from $[0, 1)$ to the positive real line.

3   Prior distribution

3.1   Shrinkage prior for transformed partial autocorrelations

The relationship between the singular value decompositions of P and A, described in the previous section, has two important implications which we exploit in building our prior. First, the spectral norms of P and A, $r_1 = \|P\|_2$ and $\tilde{r}_1 = \|A\|_2$, are clearly related through the monotonic mapping: $\tilde{r}_1 = r_1/(1 - r_1^2)^{1/2}$. The relative sizes of the unconstrained parameters A_s across lags $s = 1, \ldots, p$ therefore relate directly to the relative sizes of the partial autocorrelation matrices P_s across lags. Second, $P = 0_m$ if and only if $A = 0_m$ in which 0_m denotes the $m \times m$ matrix of zeros. It follows from the definition of the partial autocorrelation matrices that for $k < p$, $P_k \neq 0_m$ and $P_{k+s} = 0_m$ for $s = 1, \ldots, p - k$ if and only if $\phi_k \neq 0_m$ and $\phi_{k+s} = 0_m$ for $s = 1, \ldots, p - k$. The order of a $\mathrm{VAR}_m(p)$ process is therefore $k < p$ if and only if $A_k \neq 0_m$ and $A_{k+s} = 0_m$ for $s = 1, \ldots, p - k$. Under the unconstrained parameterization, it follows that the model of order $k < p$ is nested within the model of order $k + 1$.

A common Bayesian approach to quantifying uncertainty on the dimension of nested models is to fit an overparameterized model with purposefully more components than are required. By using a shrinkage prior, components that are shrunk enough to be deemed negligible in the likelihood can then be discarded. Consequently, inference on both the continuous model parameters and the model dimension are available from a single within-model MCMC sampler, without recourse to transdimensional MCMC; see, for instance, Rousseau and Mengersen (2011) and Bhattacharya and Dunson (2011) in the context of mixture and factor models, respectively. We can therefore borrow ideas from this literature by adopting a shrinkage prior for $A_1, \ldots, A_{p_{\max}}$ with a large value for $p_{\max}$. By identifying the lag beyond which the A_s can be taken as zero matrices, we therefore learn about the order p of the process. Moreover, we can convey the very reasonable idea that the partial autocorrelations at higher lags are likely to be smaller than those at lower lags by choosing a shrinkage prior for the A_s, $s = 1, \ldots, p_{\max}$, whose degree of shrinkage increases with the lag s.

A popular increasing shrinkage prior is the multiplicative gamma process (MGP) (Bhattacharya and Dunson, 2011) originally developed as a structured sequence of global-local shrinkage priors for the loadings matrix in infinite factor models. Denoting the (i,j)th element in A_s by $a_{s,ij}$ we adopt a prior of this form by choosing

$$a_{s,ij}|\lambda_{s,ij}, \tau_s \sim \mathrm{N}(0, \lambda_{s,ij}^{-1}\tau_s^{-1}),$$

independently for $i,j = 1, \ldots, m$, $s = 1, \ldots, p_{\max}$, where the local precision parameters at lag s are assigned the prior

$$\lambda_{s,ij} \sim \mathrm{Gam}(a/2, a/2),$$

independently for $i,j = 1, \ldots, m$, $s = 1, \ldots, p_{\max}$, and the global precision parameter at lag s is constructed as

$$\tau_s = \prod_{k=1}^{s} \delta_k, \quad \delta_1 \sim \mathrm{Gam}(a_1, 1), \quad \delta_k \sim \mathrm{Gam}(a_2, 1), \quad k \geq 2,$$

in which the δ_k are independent. The global precisions τ_s are therefore a cumulative product of gamma random variables whose prior expectation $E(\tau_s)$ increases with s when $a_2 > 1$. Guidelines on the choice of hyperparameter a_1 and a_2 can be found in Durante (2017) who presents a numerical method for checking that the global variances $\theta_s = 1/\tau_s$ are stochastically decreasing in s near zero, that is, $\Pr\{\theta_s \in (0,\theta]\}$ is non-decreasing in s for any θ in a small neighbourhood of zero.

The multiplicative gamma process prior does not place any mass at zero and so none of the A_s, and hence P_s, matrices are shrunk exactly to zero. We define the *effective order* p^* of the model as the value of $s \leq p_{\max}$ such that P_s fails a criterion for truncation to zero when $s = p^*$ but passes for $s = p^* + 1, \ldots, p_{\max}$. Applying the truncation criterion to the standardized P_s matrices, rather than the unconstrained A_s matrices, allows classical theory from univariate time-series analysis to inform our judgement in a manner which is robust with respect to the scale of the data as well as its length and dimension m. Further details on the truncation criterion are provided in Section 4.2. It is worth emphasising that although our prior encourages increasing shrinkage of the P_s matrices towards zero as the lag increases, this ordering does not have to be replicated in the posterior if it is not supported by the data. Further, although by definition P_{p^*} is not truncated to zero for $p^* > 1$, any matrix P_s, where $s \in \{1, \ldots, p^* - 1\}$, can be truncated to zero.

An alternative inferential approach would have been to adopt a spike-and-slab prior for each unconstrained A_s matrix. This is the basis of the Bayesian stochastic search procedure that is often used for individual coefficients in the analysis of autoregressions (George et al., 2008; So et al., 2006). An increasing shrinkage prior within this class is the cumulative shrinkage process (CUSP) (Legramanti et al., 2020). However, our experience working with this prior suggests that, though sensible posterior inferences can be obtained in the analysis of simulated data, inference on the model order is very sensitive to the choice of prior hyperparameters in analyses involving real data. This suggests a lack of robustness to the kind of model misspecification that is inevitable in analyses of real time-series. Further discussion of these approaches can be found in Chapter 5 of Binks (2024).

3.2   Joint prior

Denoting the collection of unknown hyperparameters in the multiplicative gamma process prior by $\boldsymbol{\vartheta}$, we adopt an overall prior specification of the form

$$\pi(\Sigma, A_1, \ldots, A_{p_{\max}}, \boldsymbol{\vartheta}) = \pi(\Sigma)\pi(\boldsymbol{\vartheta}) \prod_{s=1}^{p_{\max}} \pi(A_s | \boldsymbol{\vartheta}). \tag{2}$$

Various options are available for the error variance matrix Σ and distributions which offer the property of invariance with respect to the order of the variables in the observation vector are discussed in Heaps (2023). In the applications in this paper, we use one such distribution, taking Σ to be inverse Wishart, with a scale matrix that has a common element on the diagonal and a common element off the diagonal.

4   Posterior inference

4.1   Posterior distribution

For $i \leq j$, denote by $\boldsymbol{y}_{i:j}$ the time-series $\boldsymbol{y}_i, \ldots, \boldsymbol{y}_j$. The likelihood for a series of n observations, $\boldsymbol{y}_{1:n}$, from a zero-mean $\mathrm{VAR}_m(p_{\max})$ process can be expressed as

$$p(\boldsymbol{y}_{1:n} \mid \Sigma, \Phi) = p(\boldsymbol{y}_{1:p_{\max}} \mid \Sigma, \Phi) \prod_{t=p_{\max}+1}^{n} p(\boldsymbol{y}_t \mid \boldsymbol{y}_{(t-p_{\max}):(t-1)}, \Sigma, \Phi)$$

in which $\boldsymbol{Y}_t \mid \boldsymbol{y}_{(t-p_{\max}):(t-1)}, \Sigma, \Phi \sim \mathrm{N}_m \left(\sum_{i=1}^{p_{\max}} \phi_i \boldsymbol{y}_{t-i}, \Sigma \right)$ and the initial distribution is $(\boldsymbol{Y}_1^{\mathrm{T}}, \ldots, \boldsymbol{Y}_{p_{\max}}^{\mathrm{T}})^{\mathrm{T}} \mid \Sigma, \Phi \sim \mathrm{N}_{mp_{\max}}(\boldsymbol{0}, G)$. Here G is given by

$$G = \begin{pmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{p_{\max}-1} \\ \Gamma_1^{\mathrm{T}} & \Gamma_0 & \cdots & \Gamma_{p_{\max}-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{p_{\max}-1}^{\mathrm{T}} & \Gamma_{p_{\max}-2}^{\mathrm{T}} & \cdots & \Gamma_0 \end{pmatrix},$$

where the matrices $\Gamma_s = \mathrm{Cov}(\boldsymbol{y}_t, \boldsymbol{y}_{t+s})$, $s = 0, \ldots, p_{\max}-1$, are available as by-products of the recursive mapping between the partial autocorrelation matrices and the original model parameters.

Regarding the likelihood as a function of the new parameters and combining it with the prior (2) via Bayes theorem yields the posterior distribution as

$$\pi(\Sigma, A_1, \ldots, A_{p_{\max}}, \boldsymbol{\vartheta} \mid \boldsymbol{y}_{1:n}) \propto \pi(\Sigma)\pi(\boldsymbol{\vartheta}) \prod_{s=1}^{p_{\max}} \pi(A_s \mid \boldsymbol{\vartheta}) p(\boldsymbol{y}_{1:n} \mid \Sigma, A_1, \ldots, A_{p_{\max}}). \tag{3}$$

As explained in Heaps (2023), the posterior distribution is a complicated function of the A_s, making it ill-suited to computational inference based on Gibbs sampling. Rather than appealing to conditional independence structure in the posterior for one-at-a-time

parameter updates, Hamiltonian Monte Carlo (HMC) (Neal, 2011) uses information on the slope of the logarithm of the posterior density to generate global proposals that update all parameters simultaneously. We have found it efficient in sampling from the posterior (3) and use `cmdstanr` (Gabry and Cesnovar, 2021), a lightweight R interface to the Stan software (Carpenter et al., 2017), to implement the HMC algorithm. Stan requires users to write a program in the probabilistic Stan modelling language, the role of which is to provide instructions for computing the logarithm of the kernel of the posterior density function. The Stan software then automatically sets up a Markov chain simulation to sample from the resulting posterior. This includes calculation of the gradient of the logarithm of the posterior density, random initialization of the chains, and the tuning of the sampler.

4.2   Truncation criterion

Following Bhattacharya and Dunson (2011), we choose to truncate P_s to a zero matrix if the absolute values of all of its elements lie below some threshold, say ε. In classical time-series analysis, the partial autocorrelation plot, with its associated confidence intervals, plays an important role in the choice of order for a univariate autoregression (AR). Under the hypothesis that the process is $\mathrm{AR}(p)$, the estimators for the partial autocorrelations of order $p+1, p+2, \ldots$ based on a sample of size n are approximately independent with mean equal to zero and variance equal to $1/n$. As a guide, we therefore approximate the posterior for the m^2 components $p_{s,ij}$ of P_s under this hypothesis as independent $\mathrm{N}(0, 1/n)$ random variables and then compute the quantile $q_{m,n}(\beta)$ such that $\Pr\{\max_{i,j} |p_{s,ij}| < q_{m,n}(\beta)\} = \beta$ for some large value of β and set the threshold

$$\varepsilon = q_{m,n}(\beta) = \Phi^{-1}\{(\beta^{1/m^2} + 1)/2\}/\sqrt{n}. \tag{4}$$

By choosing the threshold in this way, we account for both the length and dimension of the data, in addition to operating on a parameter which is unit-free. We typically set $\beta = 0.99$, however, as discussed further in Section 5, the posterior for the effective order p^* is largely insensitive to sensible variation in the value of β, for example, from $\beta = 0.95$ to $\beta = 0.999$.

For each draw from the posterior, we can apply this criterion to determine the effective order p^* of the process. This can be summarized to yield a numerical approximation of the posterior for p^* which provides a proxy for the posterior for p. We emphasize that this truncation criterion does not depend on the choice of prior and could be applied in conjunction with any prior for an overparameterized stationary vector autoregression in order to learn the order of the process.

5   Simulation experiments

Consider the idealized setting in which we know that the data were generated from a stationary vector autoregression of known order, p. In order to explore the behaviour of the posterior distribution for p^* in this context, we carried out simulation experiments

that considered data generated from processes whose orders took various values. Our choice of truncation criterion makes allowance for the dimension of the observation vector m and the length of the time-series n. We might therefore expect some degree of robustness in the more challenging inferential situations when n is small or, in particular, when m is large. This was investigated by considering simulations under a variety of values of m and n.

For each $m \in \{1, 3, 5, 7\}$ and $p \in \{1, 2, 3, 4\}$ we simulated ten sets of $m \times m$ matrices $A_1, \ldots, A_p$ with elements sampled independently from a standard normal distribution. Taking the error variance matrix to be $\Sigma = I_m$, these were used to simulate ten stationary $\text{VAR}_m(p)$ processes of length $n = 1000$. Conditional on each data set, we then generated samples from the posterior distribution using Stan, as described in Section 4, setting the maximum possible order as $p_{\max} = 8$. Values for the other hyperparameters in the prior are detailed in Section S3.1 of the Supplementary Materials (Binks et al., 2024). In all cases, we used four chains each with 1,000 iterations of warm-up followed by 4,000 sampling iterations. Using the truncation criterion with $\beta = 0.99$, we calculated the limits $q_{m,n}(\beta)$ as 0.081, 0.103, 0.112 and 0.117 for $m = 1, 3, 5$, and 7 respectively, and obtained a posterior mass function for the effective order p^* of each process. The posterior mass functions are summarized in Figure 1 across all simulation experiments. For a given (m, p), the posteriors for the ten data sets are presented as a collection of overlaid bar charts. In all but three (of 160) cases, the true order p of the process is the mode in the posterior for p^*, with considerable posterior support. The results are similar across different values of m and p, suggesting robustness to the dimension of the data through our choice of truncation criterion.

Fixing $m = 3$, considering $p \in \{1, 2, 3, 4\}$ and using the same ten sets of matrices $A_1, \ldots, A_p$ as in the previous experiment, we then simulated ten stationary $\text{VAR}_m(p)$ processes of length $n = 100$ and another ten of length $n = 500$, facilitating comparison across $n \in \{100, 500, 1000\}$. Retaining the same prior specification in the new experiments, we fit the model using HMC via Stan, as discussed above. Again, using the truncation criterion with $\beta = 0.99$ led to limits $q_{m,n}(\beta)$ equal to 0.326, 0.146 and 0.103 for $n = 100, 500$ and 1000, respectively. This yielded the posterior mass functions for p^* which are displayed in Figure 2. In all but one experiment for the different values of n, the posterior mode for the effective order p^* recovers the true order p of the process, again, with considerable support. The similarity in the behaviour of the posterior for all values of n suggests robustness through the choice of truncation criteria, even for short time-series.

In order to assess sensitivity to our choice of truncation criterion we considered a selection of other choices for the probability β in (4). The results, presented in full in Section S3.2 of the Supplementary Materials (Binks et al., 2024), revealed that the posterior for the effective order p^* of the process was largely insensitive to variation in the value of β over a sensible range, with the posterior mode remaining equal to the known order of the process and an increasing degree of concentration about the mode as β increased. Indeed, posterior mass only started to concentrate above the known order when the value of β was purposely chosen to be unreasonably small, leading to overestimation of p. These observations held across all values of m, n and p which
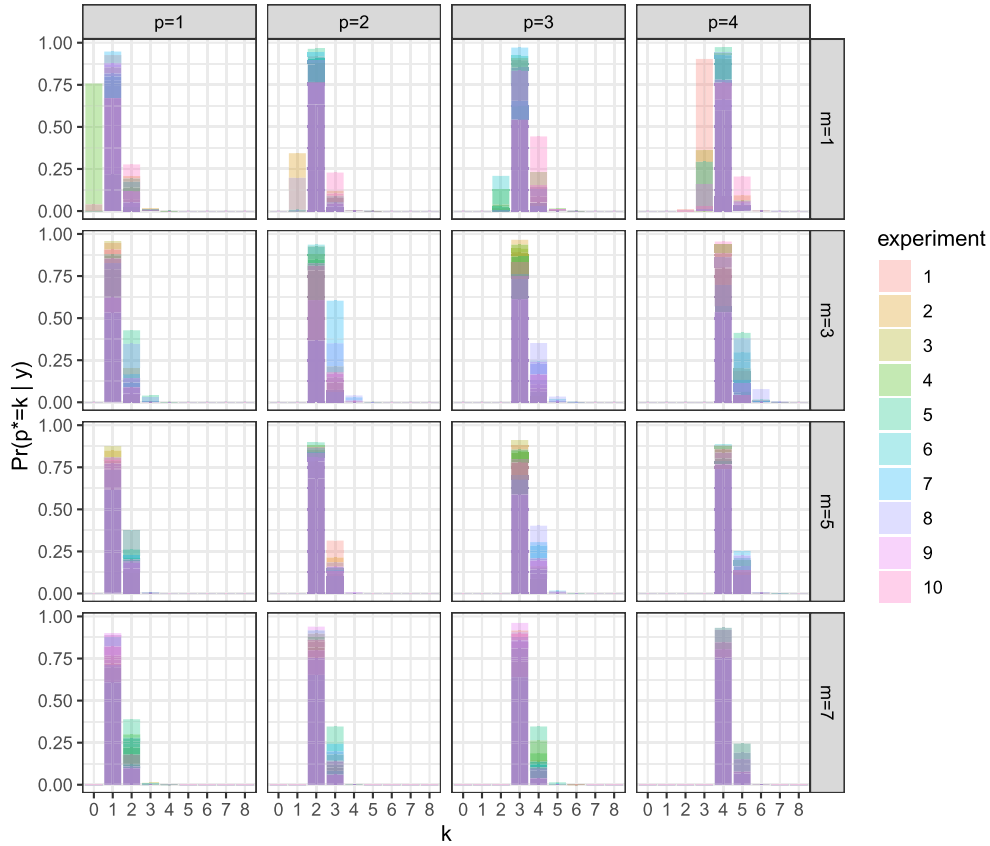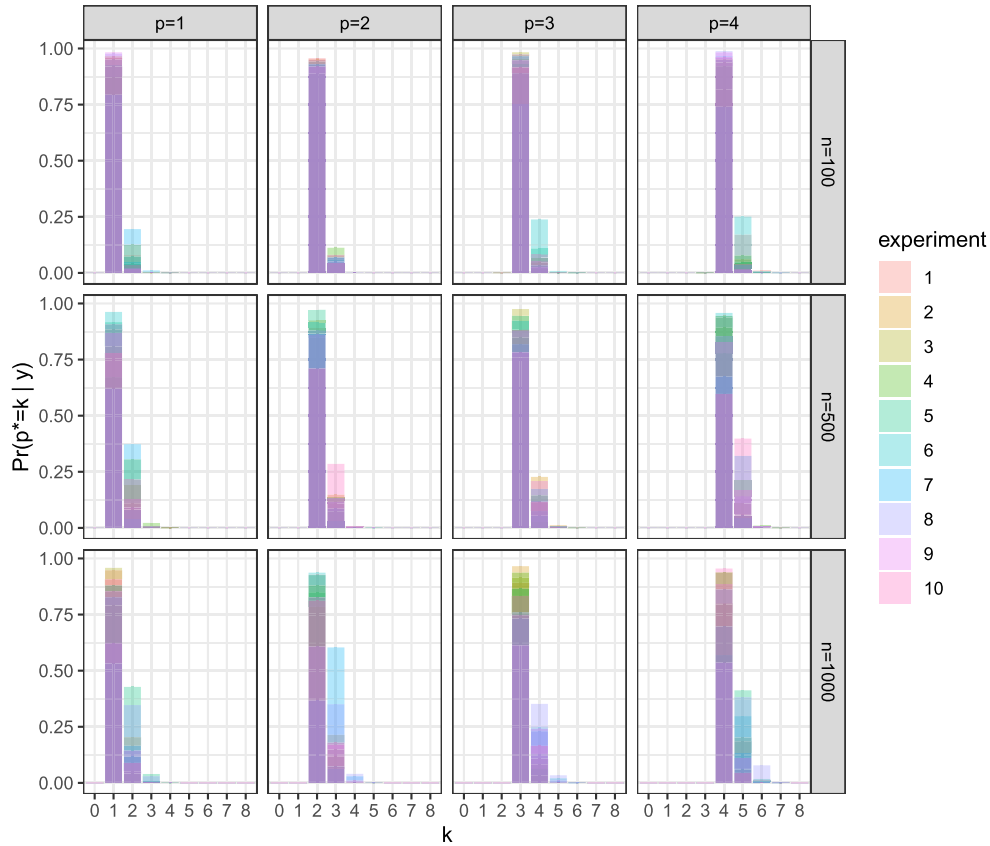
Figure 1: Overlaid posterior mass functions for the effective order p^* from 10 experiments for each combination of $m \in \{1, 3, 5, 7\}$ and $p \in \{1, 2, 3, 4\}$, with $n = 1000$.

provides further empirical evidence of the utility of our principled approach for selecting the truncation criterion.

6   Application: understanding brain rhythms

6.1   Background

As an example application, we will apply our model and inferential procedures to a dataset of long-term intracranial electroencephalography (EEG) recordings to understand biological rhythms in the brain. Biological rhythms on ultradian, circadian, and longer timescales have been demonstrated in human physiology but particularly the ultradian rhythms remain elusive in mechanism and function in the brain (Goh et al., 2019; Lloyd and Stupfel, 1991). Multiple lines of evidence suggest that some prominent ultradian rhythms exist in brain activity as measured by EEG (Hayashi et al.,

Figure 2: Overlaid posterior mass functions for the effective order p^* from 10 experiments for each combination of $n \in \{100, 500, 1000\}$ and $p \in \{1, 2, 3, 4\}$, with $m = 3$.

1994; Panagiotopoulou et al., 2022), and may be related to rest-activity cycles, or even modulate disease symptoms.

In this exploratory application we investigate the properties that such ultradian biological rhythms may display in human brain activity. We use band power in two common frequency bands (delta and beta) as our features of interest.

6.2   Data preprocessing

Intracranial EEG (iEEG) recordings are considered from four subjects with refractory focal epilepsy from the University College London Hospital (UCLH). We give the individuals the anonymous identities of A, B, C and D. The nature of the recording was chosen for its high signal-to-noise ratio without the need for extensive artefact detection and removal.

Firstly, we divided each subject's iEEG data into non-overlapping, consecutive segments of length 30 seconds. All channels within each segment were re-referenced to a common average reference. In the common average calculation, channels with extreme amplitude values were excluded. A notch filter was then applied at 50 Hz for each 30 second time window to remove line noise, after which the time windows were band-pass filtered from 0.5–80 Hz using a fourth order zero-phase Butterworth filter (second order forward and backward filter applied) and downsampled to 200 Hz.

Next, the iEEG data were decomposed into commonly studied frequency bands, including delta and beta (Taylor et al., 2022). We calculated the iEEG band power for each 30 second segment for all channels in two frequency bands (δ: 1–4 Hz, β: 13–30 Hz) using Welch's method with three-second non-overlapping windows. After taking logarithms to base 10 of the band power recordings in each channel, the channels were averaged into the brain regions from which they were recorded based on the Desikan-Killiany atlas; see Wang et al. (2023) for further details. The number of brain regions varied between individuals, with $m = 8, 8, 9$ and 13 for individuals A, B, C and D respectively. Finally, the data were mean-centered prior to analysis.

For each individual we analysed the longest possible contiguous time period of their band power time-series for which graphical interrogation of the data suggested stationarity was a plausible assumption; therefore the length of the recording chosen for further analysis varied across subjects. The number of observations in the recordings used were $n = 685, 622, 651$ and 231 for individuals A, B, C and D respectively, equivalent to 5.7, 5.175, 5.417 and 1.917 hours. These recordings were obtained during day-time hours. We apply our model and inferential procedures to the time-series of both the beta and delta band power values for each individual, with a maximum order of $p_{\max} = 8$. The choices of hyperparameters in the prior are provided in Section S4.1 of the Supplementary Materials (Binks et al., 2024).

6.3 MCMC implementation

For each individual and frequency band, we used the `cmdstanr` interface to the Stan software (version 2.29.2) to run 4 chains, initialized at different starting points. The chains were run for 17,000 iterations, discarding the first 1,000 as burn-in, and thinning to retain every fourth draw in order to reduce computational overheads. The usual graphical and numerical diagnostics gave no evidence of any lack of convergence and, after pooling the chains, the effective sample size was at least 2,023 for every model parameter. Using a CentOS Linux 7 (Core) 64bit operating system with an Intel Xeon E5-2699 v4 processor (2.2 GHz), the average run time across chains for the beta frequency band in individual A was 31.3 hours, with similar run times obtained for the other data sets.

6.4 Order determination

For each of the individuals, the posterior distributions for p^* for both the beta and delta series were calculated using the truncation criteria described in Section 4.2 with $\beta = 0.99$. For example, in Figure 3, the posterior mass functions for the data pertaining
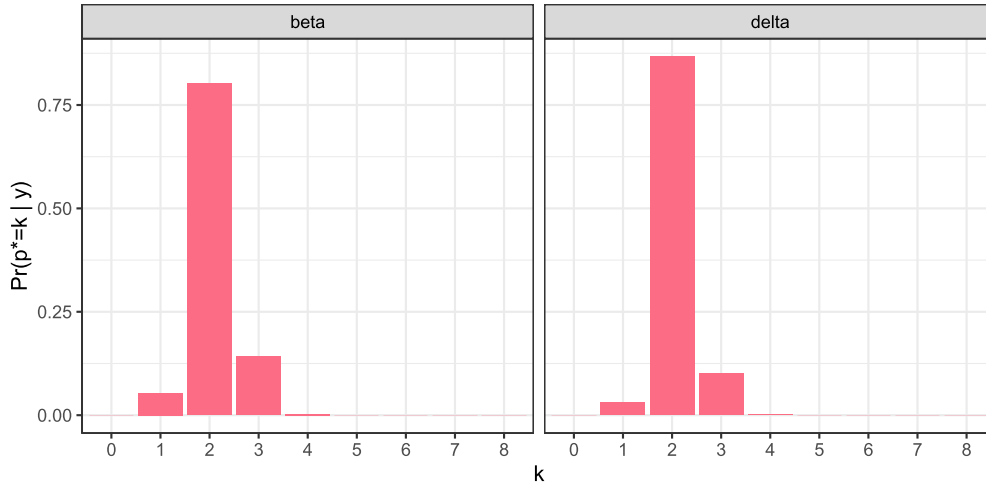
Figure 3: Posterior mass function for the order of process for the beta (left) and delta (right) series for individual A.

to individual A are shown. For both the delta and beta series, the posterior mode is 2, with posterior support exceeding 2/3. Similar results were obtained using the commonly adopted frequentist strategy of selecting model order based on the Akaike's Information Criterion (AIC) or Schwartz Criterion (SC) (Lütkepohl, 2006, Chapter 4), with the AIC selecting an order of 2 for the data in the beta and delta bands and the SC selecting 1 in each case. These results are quantitatively similar across all individuals, with figures provided in Section S4.2 of the Supplementary Materials (Binks et al., 2024), possibly indicating similar generative processes for their ultradian rhythms.

6.5 Granger causality

Conditioning on the modal order of the process for both series in each subject, we obtain samples from the posterior distributions of the autoregressive coefficient matrices. The (i, j)-th element in the autoregressive matrix at lag-s, $\phi_{s,ij}$, governs the effect of the j-th variable at time $t - s$ on the i-th variable at time t. If $\phi_{s,ij}$ is non-zero we say that variable j Granger-causes variable i at lag s; this Granger-causal connection can be represented in a directed network, called a Granger causality plot, through an edge from vertex j to vertex i. Conditional on the posterior modal order, $p^* = 2$, Figures 4 and 5 show the Granger causality plots at lags 1 and 2 for individual A in the beta and delta bands, respectively. In these plots, an autoregressive coefficient is visualized as non-zero whenever zero lies outside the 50% equi-tailed Bayesian credible interval; the thickness of the edges representing non-zero coefficients are representations of the absolute value of the posterior mean. The coordinates of the vertices, representing the different brain regions, correspond to the x and y coordinates of the centre of the region using the Desikan-Killiany atlas. A noticeable feature of these Granger causality plots
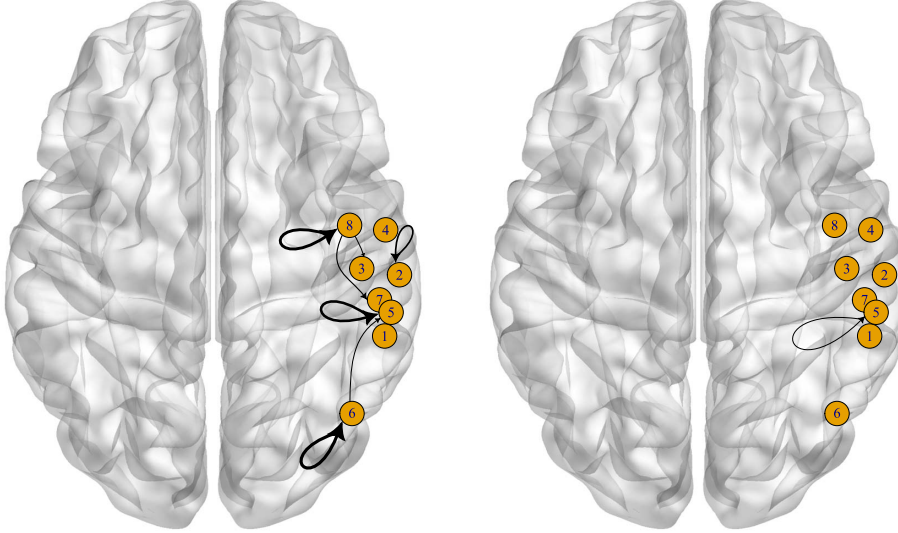
Figure 4: Granger causality plots of the posterior mean of the autoregressive coefficient matrices for the VAR process of individual A in the beta band at lag 1 (left) and lag 2 (right), overlaid on glass brains showing the locations of the regions. Region names: 1 – r.bankssts, 2 – r.middletemporal, 3 – r.postcentral, 4 – r.superiortemporal, 5 – r.supramarginal, 6 – r.inferiorparietal, 7 – r.inferiortemporal, 8 – r.precentral.

is the higher number of connections in the delta band compared to the beta band. This was common across all individuals (see Section S4.3 of the Supplementary Materials Binks et al., 2024), and may indicate more localized processes underpinning the delta rhythms that interact with each other, whereas the beta rhythms in each region may be more driven by common processes. However future work has to confirm if this feature is a result of the epilepsy, or medication.

6.6 Decomposition into latent series

Using classic theory of time-series decompositions (Prado, 1998), a $\text{VAR}_m(p)$ process can be decomposed into pm latent series. These series correspond to the pm distinct eigenvalues of the companion matrix which arises from the representation of the model as a $\text{VAR}_{mp}(1)$ process. Suppose there are c complex conjugate pairs of eigenvalues denoted $r_j e^{\pm i\omega_j}$, $j = 1, \ldots, c$, and $pm - 2c$ real eigenvalues denoted r_j, $j = 2c+1, \ldots, pm$ where $r_j > 0$ and $\omega_j \in [0, \pi)$. The latent decomposition of $\boldsymbol{y}_t = (y_{t1}, \ldots, y_{tm})^{\text{T}}$ then takes the form

$$y_{ti} = \sum_{j=1}^{c} z_{tij} + \sum_{j=2c+1}^{pm} x_{tij},$$

where z_{tij} and x_{tij} are real-valued processes corresponding to the jth pair of complex eigenvalues and the jth real eigenvalue, respectively. The process z_{tij} has similar
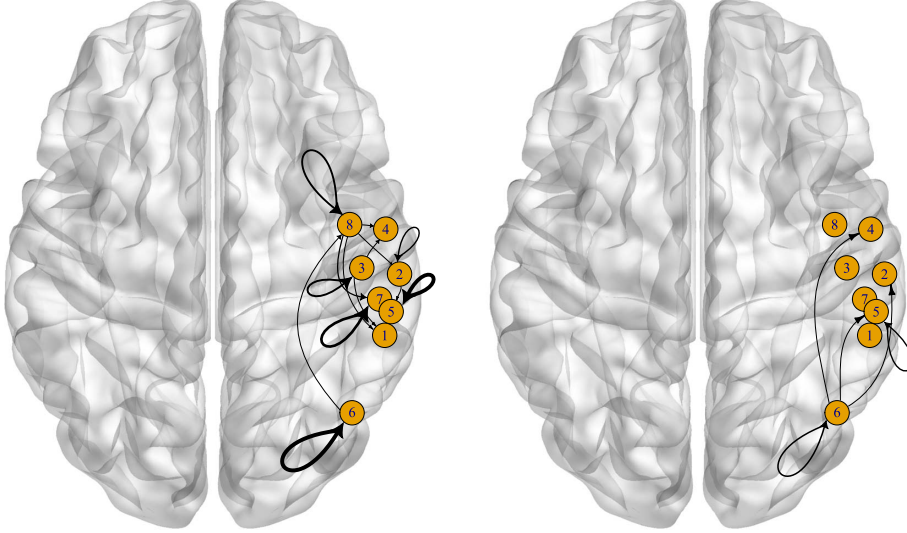
Figure 5: Granger causality plots of the posterior mean of the autoregressive coefficient matrices for the VAR process of individual A in the delta band at lag 1 (left) and lag 2 (right), overlaid on glass brains showing the locations of the regions. Region names: 1 – r.bankssts, 2 – r.middletemporal, 3 – r.postcentral, 4 – r.superiortemporal, 5 – r.supramarginal, 6 – r.inferiorparietal, 7 – r.inferiortemporal, 8 – r.precentral.

behaviour to an AR(2) process with coefficients $2r_j \cos \omega_j$ and $-r_j^2$ and is therefore quasi-periodic with characteristic frequency ω_j and modulus r_j. This holds for all dimensions $i = 1, \ldots, m$, though the time-varying amplitude and phase are different for each i. Similarly, the process x_{tij} follows an AR(1) structure with coefficient r_j for all $i = 1, \ldots, m$. The innovations that drive the z_{tij} and x_{tij} processes are correlated and arise from the error terms ϵ_t in the original model.

The quasi-periodic series arising from the complex conjugate pairs of eigenvalues are of particular interest as they can capture the cyclical patterns that are key to understanding variation in brain activity. The pairs of complex eigenvalues, $r_j e^{\pm i \omega_j}$, $j = 1, \ldots, c$, are not identifiable as the model remains unchanged under any permutation of their labelling. However, identification can be achieved by applying an ordering constraint, for example, based on the modulus or the argument. Imposing the constraint $\omega_1 < \omega_2 < \cdots < \omega_c$, the quasi-periodic series z_{tij} are ordered by decreasing period $2\pi/\omega_j$.

For individual A the posteriors for the moduli and periods of the first four quasi-periodic series are presented in Figures 6 and 7. We note that the z_{tij} with highest period also have highest modulus and might therefore be regarded as the dominating latent series. Corresponding figures for the other individuals are provided in Section S4.4 of the Supplementary Materials (Binks et al., 2024). Across individuals, a common feature is that the posterior for the period of the dominating latent series in each
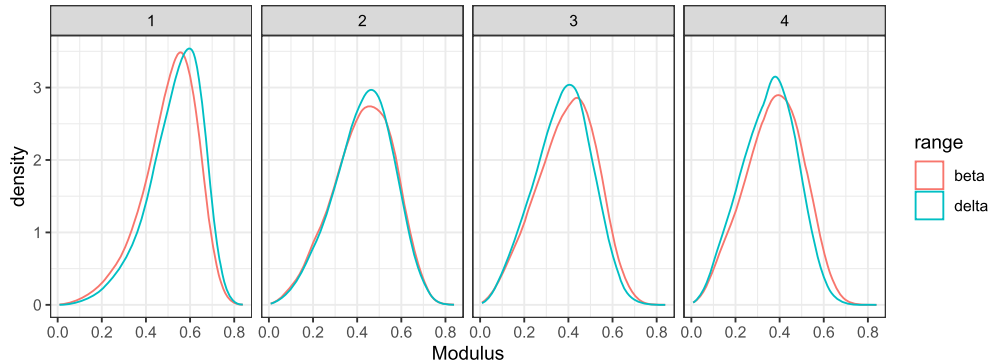
Figure 6: Posterior densities for the moduli of the first four quasi-periodic series for individual A.
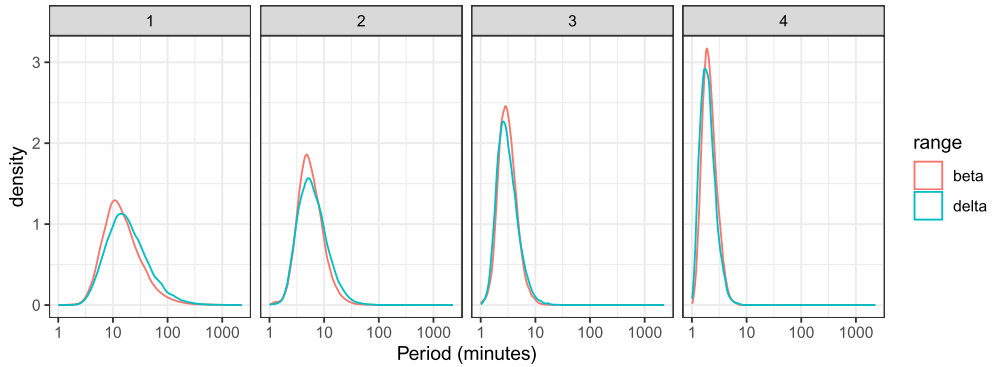

Figure 7: Posterior densities for the periods of the first four quasi-periodic series for individual A.

band has its mean at around 20 minutes; for example, for individual A, the posterior means in the beta and delta bands are 22.83 and 29.51 minutes, with 95% equi-tailed Bayesian credible intervals of $(4.04, 96.24)$ and $(4.32, 131.05)$ minutes, respectively. It is also noticeable that though there are some differences between the moduli of the series in the delta band compared to the beta band, there is very little difference between the corresponding periods. Again, this feature is replicated across all individuals. We elaborate further on this observation in Section 7.

7   Discussion

We have proposed a hierarchical Bayesian model, with accompanying model-fitting methodology, which allows inference on the order of a stationary vector autoregression. This is based on an unconstrained reparameterization of the stationary model

in terms of a set of transformed partial autocorrelation matrices (Heaps, 2023) whose properties can be exploited in the design of the prior. In particular, we capitalize on the nested structure of the new parameterization by constructing an overparameterized hierarchical model which shrinks unnecessary, high-order terms to zero; by identifying the lag beyond which the partial autocorrelation parameters become effectively equal to zero, we can then learn the order of the process. Further, using the relationship between the spectral norm of a partial autocorrelation matrix and its unconstrained counterpart, the prior is chosen to increasingly shrink the partial autocorrelation matrices at higher lags towards zero through a multiplicative gamma process for the unconstrained matrices.

An efficient Hamiltonian Monte Carlo sampler for computational inference was proposed and implemented through Stan, with accompanying code to allow easy dissemination into other fields. The interpretability of the reparameterization allowed use of classical theory on the distribution of the estimators of the partial autocorrelation function to make a judgement about which sampled partial autocorrelation matrices are approximately equal to zero. An associated truncation criterion determines the number of non-zero partial autocorrelation matrices, allowing posterior inference on the order of the process in a manner which accounts for the scale, dimension and length of the time-series. This truncation criterion is independent of the choice of prior and could be applied in conjunction with any prior for an overparameterized stationary vector autoregression.

We applied our methodology to a series of simulation experiments in which data sets of various lengths n were sampled from various stationary $\text{VAR}_m(p)$ models. For all values of m, p and n considered, the posterior for the effective order of the process was highly concentrated around the known model order. We then applied our methodology to iEEG data from recordings at various locations in the brain. Conditioning on the posterior modal order of these processes allowed physiological insight in a number of directions. By constructing Granger causality plots, we were able to highlight relationships between activity in different regions of the brain. Similarly, by constructing the latent decomposition of the series, we were able to identify underlying quasi-periodic structure. In particular, we found that the dominant latent component had a period that was around 20 minutes across all individuals in both the beta and the delta bands. This is consistent with ultradian rhythms of around 20 minutes which have previously been observed (Panagiotopoulou et al., 2022). The similarity in the periods across the beta and delta bands indicate that there is a global cycle in the band power pattern, rather than a local cycle within a specific band. The similarities between subjects are striking, particularly the period of 20 minutes, and warrant future investigations into the possible biological mechanisms and potentially endogenous drivers (Goh et al., 2019). However, as we only considered four subjects in this work, a larger study would be needed to confirm any biological interpretations, with a larger number of patients, longer recordings and accounting for the potential pathology present in these subjects.

Although our focus has been learning the order of stationary vector autoregressions with independent, Gaussian errors, these assumptions about the error process could be relaxed in a number of ways. First, we could readily allow higher probabilities of extreme values by assuming the errors arose from a heavy tailed distribution

with finite first and second moments, such as the multivariate-t distribution with degree of freedom parameter greater than two. Second, it may also be possible to allow for heteroskedasticity by modelling the errors according to a multivariate generalized autoregressive conditional heteroskedasticity (GARCH) process, such as that described by Ling and McAleer (2003), in which additional constraints on the parameters of the error process guarantee second order stationarity. In this instance, however, because the mapping from the original parameter space (Σ, Φ) to $\{\Sigma, (P_1, \ldots, P_p)\}$ depends on the conditional variance Σ and this conditional variance would change at every time point, further work would be needed to investigate an alternative choice of positive definite matrix to replace $\boldsymbol{\Sigma}$, and to establish the interpretation of the resulting matrices $P_1, \ldots, P_p$. Finally, we could allow the errors to be autocorrelated by considering the more general class of stable and invertible vector autoregressive moving average (VARMA) models. A vector autoregressive moving average model of order (p, q) has the form

$$\boldsymbol{y}_t = \phi_1 \boldsymbol{y}_{t-1} + \ldots + \phi_p \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t + \psi_1 \boldsymbol{\epsilon}_{t-1} + \ldots + \psi_q \boldsymbol{\epsilon}_{t-q},$$

where $\boldsymbol{\epsilon}_t \sim \mathrm{N}(\boldsymbol{0}, \Sigma)$ for $t = q + 1, \ldots, n$, the parameters $\psi_i \in M_{m \times m}(\mathbb{R})$, $i = 1, \ldots, q$, are $m \times m$ moving average coefficient matrices and the parameters $\phi_1, \ldots, \phi_p$ are the autoregressive coefficient matrices which lie in $\mathcal{C}_{p,m}$ for a stable process. In backshift notation, this can be written as $\phi(B)\boldsymbol{y}_t = \psi(B)\boldsymbol{\epsilon}_t$ where $\psi(u) = (I_m + \psi_1 u + \ldots + \psi_q u^q)$, $u \in \mathbb{C}$, is the characteristic moving average polynomial. The VARMA process is invertible if and only if all roots of the equation $\det\{\psi(u)\} = 0$ lie outside the unit circle. The subset of $M_{m \times m}(\mathbb{R})$ in which the matrices $-\psi_1, \ldots, -\psi_q$ satisfy this condition is simply $\mathcal{C}_{q,m}$. To constrain the parameters to this region, we can therefore use the reparameterizations discussed in Section 2.2 to first map from the moving average coefficient matrices and variance matrix to a set of matrices $Q_1, \ldots, Q_q$ with singular values less than one, and then to map to a set of unconstrained square matrices; (see Heaps, 2023, Supplementary Materials). To allow inference on q, as well as p, a second multiplicative gamma process prior could then be used to increasingly shrink the unconstrained matrices at higher lags towards zero, again using a truncation criterion to determine which of the matrices are effectively zero. However, since the matrices $Q_1, \ldots, Q_q$ do not have a clear interpretation, making a principled choice of truncation criterion for the reparameterized moving average coefficients would not be straightforward. Moreover, since VARMA models can be approximated by higher order vector autoregressions, it is likely that posteriors would be multimodal, demanding development of a Markov chain Monte Carlo sampler that can move efficiently between modes. We defer further consideration of this challenging topic to future work.

Our ideas could also be extended to learn the order of processes that are only piecewise stationary. Indeed, an obvious limitation in the application to iEEG data was the necessity to pick out contiguous segments of data where stationarity was a plausible assumption. However, as remarked in Section 1, stationary autoregressions often serve as building blocks in the construction of more complex models. Motivated by applications involving iEEG data where subjects transition between states of wakefulness and sleep, or states of normal brain activity and seizure, we are currently exploring a hidden Markov model in which a (locally) stationary vector autoregression describes the within-state dynamics. Such a model would be ideally suited to a wide variety of time-series

where there are occasional step-changes in a process which otherwise appears to be mean reverting.

Supplementary Material

Supplement to "Bayesian inference on the order of stationary vector autoregressions" (DOI: 10.1214/24-BA1499SUPP; .pdf). Further discussion of the stationary region, a numerical example of the reparameterization over the stationary region, additional details on the prior specification for the simulated and EEG applications, more extensive discussion of posterior sensitivity to the choice of truncation criterion, and additional plots from the EEG application.

The simulated data from Section 5, R scripts for simulating data, running Stan and post-processing the Stan output, as well as the Stan program for fitting the model can be found at the GitHub repository `https://github.com/rachelbinks/Bayesian-VAR-order-determination`.

References

Barnett, G., Kohn, R., and Sheather, S. (1996). "Bayesian estimation of an autoregressive model using Markov chain Monte Carlo." *Journal of Econometrics*, 74: 237–254. MR1411599. doi: `https://doi.org/10.1016/0304-4076(95)01744-5`.   2

Bhattacharya, A. and Dunson, D. B. (2011). "Sparse Bayesian infinite factor models." *Biometrika*, 98(2): 291–306. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3419391/`. MR2806429. doi: `https://doi.org/10.1093/biomet/asr013`.   5, 6, 8

Binks, R. L. (2024). "Bayesian inference on the order of stationary vector autoregressions with application to multivariate modelling of electroencephalography data." Ph.D. thesis, Newcastle University.   6

Binks, R. L., Heaps, S. E., Panagiotopoulou, M., Wang, Y., and Wilkinson, D. J. (2024). "Supplement to "Bayesian inference on the order of stationary vector autoregressions"." doi: `https://doi.org/10.1214/24-BA1499SUPP`.   4, 5, 9, 12, 13, 14, 15

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). "Stan: A Probabilistic Programming Language." *Journal of Statistical Software*, 76.   8

Chiang, S., Guindani, M., Yeh, H. J., Haneef, Z., Stern, J. M., and Vannucci, M. (2016). "Bayesian vector autoregressive model for multi-subject effective connectivity inference using multi-modal neuroimaging data." *Human Brain Mapping*, 38(3): 1311–1332.   1

Durante, D. (2017). "A note on the multiplicative gamma process." *Statistics & Probability Letters*, 122: 198–204. MR3584158. doi: https://doi.org/10.1016/j.spl.2016.11.014.   6

Fan, J., Sitek, K., Chandrasekaran, B., and Sarkar, A. (2022). "Bayesian tensor factorized vector autoregressive models for inferring Granger causality patterns from high-dimensional multi-subject panel neuroimaging data." arXiv:2206.10757.   3

Gabry, J. and Cesnovar, R. (2021). *cmdstanr: R Interface to 'CmdStan'*. URL https://mc-stan.org/cmdstanr, https://discourse.mc-stan.org.   8

George, E. I., Sun, D., and Ni, S. (2008). "Bayesian stochastic search for VAR model restrictions." *Journal of Econometrics*, 142: 553–580. MR2408749. doi: https://doi.org/10.1016/j.jeconom.2007.08.017.   6

Goh, G. H., Maloney, S. K., Mark, P. J., and Blache, D. (2019). "Episodic ultradian events – ultradian rhythms." *Biology*, 8(1): 15.   10, 17

Hannaford, N. E., Heaps, S. E., Nye, T. M., Curtis, T. P., Allen, B., Golightly, A., and Wilkinson, D. J. (2023). "A sparse Bayesian hierarchical vector autoregressive model for microbial dynamics in a wastewater treatment plant." *Computational Statistics & Data Analysis*, 179: 107659. MR4513686. doi: https://doi.org/10.1016/j.csda.2022.107659.   1

Hayashi, M., Sato, K., and Hori, T. (1994). "Ultradian rhythms in task performance, self-evaluation, and EEG activity." *Perceptual and Motor Skills*, 79(2): 791–800.   10

Heaps, S. E. (2023). "Enforcing stationarity through the prior in vector autoregressions." *Journal of Computational and Graphical Statistics*, 32(1): 74–83. MR4552938. doi: https://doi.org/10.1080/10618600.2022.2079648.   2, 3, 4, 5, 7, 17, 18

Heaps, S. E., Farrow, M., and Wilson, K. J. (2020). "Identifying the effect of public holidays on daily demand for gas." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(2): 471–492. MR4052786. doi: https://doi.org/10.1111/rssa.12504.   1

Huerta, G. and Prado, R. (2006). "Structured priors for multivariate time series." *Journal of Statistical Planning and Inference*, 136(11): 3802–3821. MR2299166. doi: https://doi.org/10.1016/j.jspi.2005.02.025.   3

Huerta, G. and West, M. (1999). "Priors and component structures in autoregressive time series models." *Journal of the Royal Statistical Society. Series B (Statistical*

Methodology), 61(4): 881–899. MR1722245. doi: https://doi.org/10.1111/1467-9868.00208.   3

Jiang, X., Hu, X., Xu, W., Li, G., and Wang, Y. (2013). "Inference of microbial interactions from time series data using vector autoregression model." In *2013 IEEE International Conference on Bioinformatics and Biomedicine*, 82–85.   1

Koop, G. and Korobilis, D. (2010). "Bayesian multivariate time series methods for empirical macroeconomics." *Foundations and Trends in Econometrics*, 3(4): 267–358. doi: https://doi.org/10.1561/0800000013.   1

Legramanti, S., Durante, D., and Dunson, D. B. (2020). "Bayesian cumulative shrinkage for infinite factorizations." *Biometrika*, 107(3): 745–752. MR4138988. doi: https://doi.org/10.1093/biomet/asaa008.   6

Ling, S. and McAleer, M. (2003). "Asymptotic theory for a vector ARMA-GARCH model." *Econometric Theory*, 19(2): 280–310. MR1966031. doi: https://doi.org/10.1017/S0266466603192092.   18

Lloyd, D. and Stupfel, M. (1991). "The occurrence and functions of ultradian rhythms." *Biological Reviews of the Cambridge Philosophical Society*, 66(3): 275–299.   10

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer. MR2172368. doi: https://doi.org/10.1007/978-3-540-27752-1.   13

Neal, R. M. (2011). "MCMC using Hamiltonian dynamics." In Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (eds.), *Handbook of Markov Chain Monte Carlo*, Handbooks of modern statistical methods, 113–162. Chapman & Hall/CRC. MR2858447.   8

Panagiotopoulou, M., Papasavvas, C. A., Schroeder, G. M., Thomas, R. H., Taylor, P. N., and Wang, Y. (2022). "Fluctuations in EEG band power at subject-specific timescales over minutes to days explain changes in seizure evolutions." *Human Brain Mapping*, 43(8): 2460–2477.   10, 17

Prado, R. (1998). "Latent structure in non-stationary time series." Ph.D. thesis, Duke University. MR2697749.   14

Rousseau, J. and Mengersen, K. (2011). "Asymptotic behaviour of the posterior distribution in overfitted mixture models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5): 689–710. MR2867454. doi: https://doi.org/10.1111/j.1467-9868.2011.00781.x.   5

So, M. K. P., Chen, C. W. S., and Liu, F.-C. (2006). "Best subset selection of autoregressive models with exogeneous variables and generalized autoregressive conditional heteroscedasticity errors." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(2): 201–224. MR2226545. doi: https://doi.org/10.1111/j.1467-9876.2006.00535.x.   6

Taylor, P. N., Papasavvas, C. A., Owen, T. W., Schroeder, G. M., Hutchings, F. E., Chowdhury, F. A., Diehl, B., Duncan, J. S., McEvoy, A. W., Miserocchi, A., de Tisi, J., Vos, S. B., Walker, M. C., and Wang, Y. (2022). "Normative brain mapping of

interictal intracranial EEG to localize epileptogenic tissue." *Brain*, 145(3): 939–949. 12

Vermaak, J., Andrieu, C., Doucet, A., and Godsill, S. J. (2004). "Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes." *Journal of Time Series Analysis*, 25(6): 785–809. MR2099640. doi: https://doi.org/10.1111/j.1467-9892.2004.00380.x.   3

Wang, Y., Schroeder, G. M., Horsley, J. J., Panagiotopoulou, M., Chowdhury, F. A., Diehl, B., Duncan, J. S., McEvoy, A. W., Miserocchi, A., de Tisi, J., and Taylor, P. N. (2023). "Temporal stability of intracranial electroencephalographic abnormality maps for localizing epileptogenic tissue." *Epilepsia*, 64(8): 2070–2080.   12

Zhang, W., Cribben, I., Petrone, S., and Guindani, M. (2021). "Bayesian time-varying tensor vector autoregressive models for dynamic effective connectivity." arXiv:2106. 14083.   3