Enforcing Stationarity through the Prior in Vector Autoregressions

Sarah Heaps

Durham University

Bayesian Inference in High Energy Physics, Durham University

May 25–27, 2022

Outline

1. Time series analysis
2. Stationarity
3. Vector autoregressions
4. Prior construction
5. Posterior computation
6. Application
7. Conclusions

Time series analysis

- A time series is a set of observations collected sequentially in time.
- A time series process is a collection of random variables y_t indexed in time.
- A process is a Gaussian process if (and only if) any finite subcollection $(y_{t_1}, \dots, y_{t_n})$ has a multivariate normal distribution.
- Of fundamental interest is the dependence between the sequence of random variables.
- If we can form a (reasonable) model for a time series, then we can learn about its properties which can be useful in a variety of settings.

Why model a time series?

- Description. Summarise what has occurred in the past in a simple way.
- Forecasting. Prediction of future values.
- Measure the effect of interventions.
- Control. Monitor a time series and take action to influence its future behaviour.

Stationary Gaussian processes

- Let $\{\boldsymbol{y}_t\}$ denote a Gaussian process whose components represent m univariate time series.
- The process is stationary if and only if (iff)
  1. The mean is constant over time, i.e. for all t
  $$\mathrm{E}(\boldsymbol{y}_t) = \boldsymbol{\mu}.$$
  2. The cross-covariance function depends only on the lag
  $$\boldsymbol{\Gamma}_i = \mathrm{Cov}(\boldsymbol{y}_t, \boldsymbol{y}_{t+i}) = \mathrm{E}\{(\boldsymbol{y}_t - \boldsymbol{\mu})(\boldsymbol{y}_{t+i} - \boldsymbol{\mu})^T\}$$
  for $i = 0, 1, 2, \dots$ with $\boldsymbol{\Gamma}_{-i} = \boldsymbol{\Gamma}_i^T$.

Why is stationarity important?

- Stationarity is a very common assumption in time-series analysis.
- Generally not plausible for the raw time series but often appropriate after differencing, “detrending” or as a model for particular components of a time-series.
- Stationarity prevents the predictive variance increasing without bound as the forecast horizon increases.
- This is often a desirable property, e.g. when goal is long-term forecasting or characterising long-run behaviour.

The $\text{VAR}_m(p)$ model

- Any stationary Gaussian process can be approximated by a finite-order, vector autoregressive moving average (VARMA) model.
- Our main focus is the subclass of vector autoregressive models. Consider a zero-mean process of order p ($\text{VAR}_m(p)$):

$$\boldsymbol{y}_t = \phi_1 \boldsymbol{y}_{t-1} + \ldots + \phi_p \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \overset{iid}{\sim} \mathrm{N}_m(\boldsymbol{0}_m, \boldsymbol{\Sigma}).$$

- The parameters comprise the autoregressive coefficient matrices

$$\phi_i \in M_{m\times m}(\mathbb{R}), \quad i = 1, \ldots, p$$

  and the error variance matrix

$$\boldsymbol{\Sigma} \in \mathcal{S}_m^+.$$

- We denote the collection $(\phi_1, \ldots, \phi_p)$ by $\boldsymbol{\Phi} \in M_{m\times m}(\mathbb{R})^p$.

The stationary region

- The characteristic polynomial of a $\text{VAR}_m(p)$ model is given by

$$\phi(u) = \boldsymbol{I}_m - \phi_1 u - \ldots - \phi_p u^p, \qquad u \in \mathbb{C}.$$

- The process is stationary iff all the roots of $\det\{\phi(u)\} = 0$ lie outside the unit circle.
- This subset of $M_{m\times m}(\mathbb{R})^p$ is the stationary region, denoted $\mathcal{C}_{p,m}$. It has a very complex geometry.

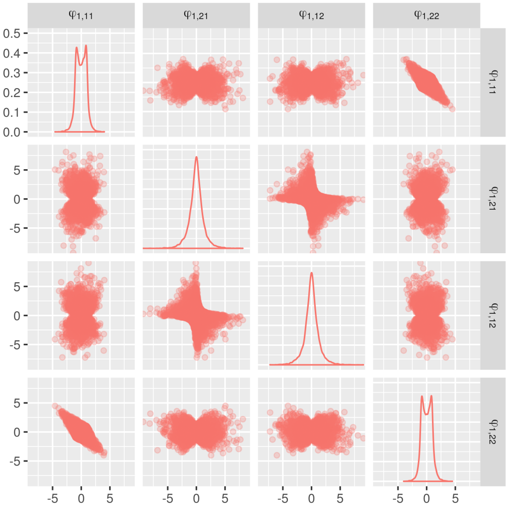
Stationary region of $\text{VAR}_2(1)$

- Consider the simplest case where $m > 1$, i.e.

$$\boldsymbol{y}_t = \phi_1 \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} \mathrm{N}_2(\boldsymbol{0}_2, \boldsymbol{\Sigma}),$$

  so that $\boldsymbol{\Phi} = \phi_1$.
- The constraint $\boldsymbol{\Phi} \in \mathcal{C}_{1,2}$ is equivalent to saying the spectral radius, $\rho(\phi_1)$, must be less than one.
- What does this look like?

Stationary region of VAR$_2$(1)



General approach to building a prior over $\mathcal{C}_{p,m}$

- The goal is to develop a prior that:
  - Encodes genuine initial beliefs, e.g. exchangeability with respect to the order of the elements in $\boldsymbol{y}_t$ (c.f. Ansley and Kohn (1986)).
  - Facilitates routine computational inference using probabilistic programming software.
- The solution is to specify a reparameterisation of $(\boldsymbol{\Sigma}, \boldsymbol{\Phi}) \in \mathcal{S}_m^+ \times \mathcal{C}_{p,m}$ in which the new parameters are:
  - Less constrained;
  - Interpretable;
  - Amenable to Monte Carlo sampling.

  } c.f. Roy et al. (2019).
- A prior for $\boldsymbol{\Phi}$ over $\mathcal{C}_{p,m}$ is induced through specification of a prior for the new parameters.

Reparameterisation 1: partial autocorrelation matrices

Ansley and Kohn (1986) extend univariate results, establishing a bijection between

$$\{\boldsymbol{\Sigma}, (\phi_1, \dots, \phi_p)\} \in \mathcal{S}_m^+ \times \mathcal{C}_{p,m} \quad \text{and} \quad \{\boldsymbol{\Sigma}, (\boldsymbol{P}_1, \dots, \boldsymbol{P}_p)\} \in \mathcal{S}_m^+ \times \mathcal{V}_m^p.$$

- $\boldsymbol{P}_{s+1}$ is the $(s+1)$-th partial autocorrelation matrix – "a" conditional cross-correlation matrix between $\boldsymbol{y}_{t+1}$ and $\boldsymbol{y}_{t-s}$ given $\boldsymbol{y}_t, \dots, \boldsymbol{y}_{t-s+1}$ (written $\boldsymbol{y}_{t:t-s+1}$):

$$\boldsymbol{P}_{s+1} = \boldsymbol{S}_s^{-1} \mathrm{Cov}(\boldsymbol{y}_{t+1}, \boldsymbol{y}_{t-s} | \boldsymbol{y}_{t:t-s+1}) (\boldsymbol{S}_s^{*\,-1})^T, \quad s = 0, \dots, p-1,$$

in which

$$\boldsymbol{\Sigma}_s = \boldsymbol{S}_s \boldsymbol{S}_s^T = \mathrm{Var}(\boldsymbol{y}_{t+1} | \boldsymbol{y}_{t:t-s+1}), \quad \boldsymbol{\Sigma}_s^* = \boldsymbol{S}_s^* \boldsymbol{S}_s^{*\,T} = \mathrm{Var}(\boldsymbol{y}_{t-s} | \boldsymbol{y}_{(t-s+1):t}).$$

- We take the symmetric matrix-square roots: $\boldsymbol{S}_s = \boldsymbol{\Sigma}_s^{1/2}$, $\boldsymbol{S}_s^* = \boldsymbol{\Sigma}_s^{*\,1/2}$.
- $\mathcal{V}_m$ denotes the subset of matrices in $M_{m \times m}(\mathbb{R})$ whose singular values are all less than one.
- The mapping and its inverse proceeds by recursion (Heaps, in press).

Reparameterisation 2: unconstrained square matrices

- The space $\mathcal{V}_m^p$ is still fairly constrained and there are no standard distributions on $\mathcal{V}_m$.
- Ansley and Kohn (1986) establish a bijection between $\boldsymbol{P} \in \mathcal{V}_m$ and $\boldsymbol{A} \in M_{m\times m}(\mathbb{R})$.
- Forwards: let $\boldsymbol{B}^{-1}\boldsymbol{B}^{-1\,T} = \boldsymbol{I} - \boldsymbol{P}\boldsymbol{P}^T$ then write $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{P}$.
- Inverse: let $\boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{I} + \boldsymbol{A}\boldsymbol{A}^T$ then write $\boldsymbol{P} = \boldsymbol{B}^{-1}\boldsymbol{A}$.
- We take the symmetric matrix-square root factorisation.
- Intuition: mapping from $\boldsymbol{P}$ to $\boldsymbol{A}$ simply transforms the singular values of $\boldsymbol{P}$ from $r_i \in [0,1)$ to $\tilde{r}_i \in \mathbb{R}^+$:

$$\tilde{r}_i = r_i/\sqrt{1 - r_i^2} \quad \Longleftrightarrow \quad r_i = \tilde{r}_i/\sqrt{1 + \tilde{r}_i^2} \quad i = 1, \ldots, m$$

while left and right singular vectors are preserved.

Special structures

- The partial autocorrelation matrices $\boldsymbol{P}_s$ are interpretable.
- The mapping from $\boldsymbol{P}_s$ to $\boldsymbol{A}_s$ preserves various structured forms:
  1. Diagonal;
  2. Two-parameter exchangeable matrix, i.e. matrix of the form

$$\begin{pmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \cdots & a \end{pmatrix}$$

  3. Special cases of (2): scaled all-ones matrix, scaled identity matrix, the zero matrix.
- Zero matrix result is significant – the order of the autoregression is $k < p$ iff $\boldsymbol{A}_k \neq \boldsymbol{0}$ but $\boldsymbol{A}_{k+i} = \boldsymbol{0}$ for $i = 1, \ldots, p - k$.

Prior distribution

- Conditional on a set of unknown hyperparameters, we construct a prior of the form

$$\pi(\boldsymbol{\Sigma}, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_p) = \pi(\boldsymbol{\Sigma}) \prod_{s=1}^{p} \pi\{\mathrm{vec}(\boldsymbol{A}_s^T)\}.$$

- Then
  - $\boldsymbol{\Sigma}$ can be assigned an inverse Wishart distribution;
  - $\mathrm{vec}(\boldsymbol{A}_s^T)$, $s = 1, \ldots, p$, can be assigned a multivariate normal distribution.
- This prior has some nice properties.

Exchangeable prior

- Certain choices of the hyperparameters yield a prior which is invariant under permutation of the m elements in the observation vectors, e.g.

$$
\begin{aligned}
\boldsymbol{\Sigma} &\sim \mathrm{IW}(v\boldsymbol{W}), && \boldsymbol{W} \text{ is two-parameter exchangeable,} \\
a_{s,ii}|\mu_{s1},\omega_{s1} &\overset{iid}{\sim} \mathrm{N}(\mu_{s1},\omega_{s1}^{-1}), && i = 1,\ldots,m, \\
a_{s,ij}|\mu_{s2},\omega_{s2} &\overset{iid}{\sim} \mathrm{N}(\mu_{s2},\omega_{s2}^{-1}), && i \neq j = 1,\ldots,m, \\
\mu_{s1} &\sim \mathrm{N}(e_{s1}, f_{s1}^2), && \omega_{s1} \sim \gamma(g_{s1}, h_{s1}), \\
\mu_{s2} &\sim \mathrm{N}(e_{s2}, f_{s2}^2), && \omega_{s2} \sim \gamma(g_{s2}, h_{s2}).
\end{aligned}
$$

- This is useful because we often do not have prior information to distinguish between the m components of $\boldsymbol{y}_t$.

Exchangeable prior

- Certain choices of the hyperparameters yield a prior which is invariant under permutation of the m elements in the observation vectors, e.g.

$$\begin{aligned}
\boldsymbol{\Sigma} &\sim \mathrm{IW}(v\boldsymbol{W}), && \boldsymbol{W} \text{ is two-parameter exchangeable,} \\
a_{s,ii}|\mu_{s1}, \omega_{s1} &\overset{iid}{\sim} \mathrm{N}(\mu_{s1}, \omega_{s1}^{-1}), && i = 1, \dots, m, \\
a_{s,ij}|\mu_{s2}, \omega_{s2} &\overset{iid}{\sim} \mathrm{N}(\mu_{s2}, \omega_{s2}^{-1}), && i \neq j = 1, \dots, m, \\
\mu_{s1} &\sim \mathrm{N}(e_{s1}, f_{s1}^2), && \omega_{s1} \sim \gamma(g_{s1}, h_{s1}), \\
\mu_{s2} &\sim \mathrm{N}(e_{s2}, f_{s2}^2), && \omega_{s2} \sim \gamma(g_{s2}, h_{s2}).
\end{aligned}$$

- This is useful because we often do not have prior information to distinguish between the m components of $\boldsymbol{y}_t$.

Exchangeable prior

- Certain choices of the hyperparameters yield a prior which is invariant under permutation of the m elements in the observation vectors, e.g.

$$
\begin{aligned}
\boldsymbol{\Sigma} &\sim \mathrm{IW}(v\boldsymbol{W}), && \boldsymbol{W} \text{ is two-parameter exchangeable,} \\
a_{s,ii}|\mu_{s1},\omega_{s1} &\overset{iid}{\sim} \mathrm{N}(\mu_{s1},\omega_{s1}^{-1}), && i = 1,\dots,m, \\
a_{s,ij}|\mu_{s2},\omega_{s2} &\overset{iid}{\sim} \mathrm{N}(\mu_{s2},\omega_{s2}^{-1}), && i \neq j = 1,\dots,m, \\
\mu_{s1} &\sim \mathrm{N}(e_{s1}, f_{s1}^2), && \omega_{s1} \sim \gamma(g_{s1}, h_{s1}), \\
\mu_{s2} &\sim \mathrm{N}(e_{s2}, f_{s2}^2), && \omega_{s2} \sim \gamma(g_{s2}, h_{s2}).
\end{aligned}
$$

- This is useful because we often do not have prior information to distinguish between the m components of $\boldsymbol{y}_t$.

Posterior computation

- Given observations, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, the likelihood is a complicated function of $\{\boldsymbol{\Sigma}, (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)\}$.
- The posterior has no standard form and admits no simple factorisation; it is ill-suited to MCMC methods that are based on Gibbs sampling.
- We use Hamiltonian Monte Carlo (HMC) which generates global proposals that update all parameters simultaneously.
- `rstan` is used to implement the HMC algorithm.

Application

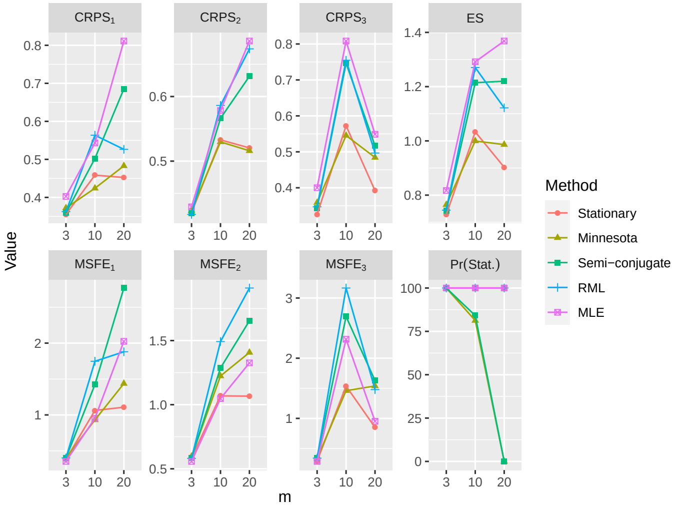
- Complete data are a quarterly time series of 168 US macroeconomic variables from 1959 to 2007, transformed to stationarity (Koop, 2013).
- Following earlier analyses:
  - Interest lies in forecasting three of the variables: real GDP, the consumer price index and an interest rate (Federal funds);
  - Consider three models: $\text{VAR}_3(4)$, $\text{VAR}_{10}(4)$ and $\text{VAR}_{20}(4)$.
- The last 40 observations are held back in model-fitting and used to assess forecast performance.

Comparison

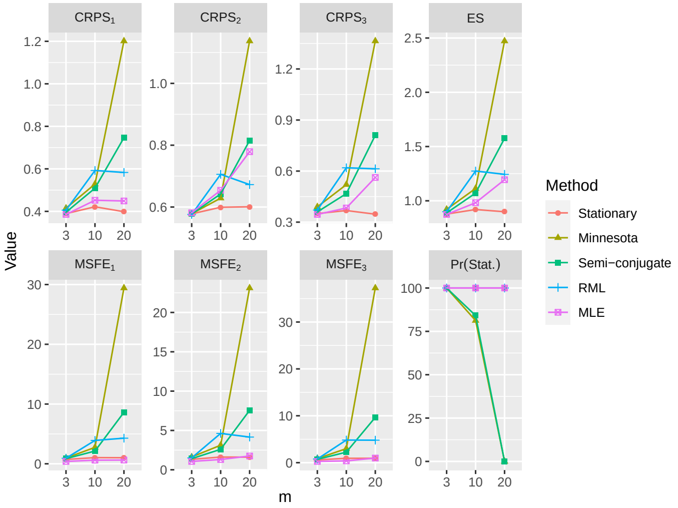
- We compare four priors:
  1. A stationary, exchangeable prior;
  2. A Minnesota prior; } stationarity not imposed
  3. A semi-conjugate prior; } stationarity not imposed
  4. A stationary, diffuse prior based on Roy et al. (2019).

  and the MLE constrained to the stationary region (Ansley and Kohn, 1986).
- Out-of-sample forecasting performance compared at various horizons using
  - Continuous rank probability score for variable $j = 1, 2, 3$ (CRPS_j);
  - Energy score for variables 1–3 (ES_3);
  - Posterior for the mean-square-forecast-error for variable j (MSFE_j).
- Small values indicate better forecasts.
- Also computed: $\Pr(\mathrm{Stat.})$, which is $\Pr(\boldsymbol{\Phi} \in \mathcal{C}_{4,m} | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$.

One-step ahead scores for model-prior combinations



Eight-step ahead scores for model-prior combinations



Conclusions

- Prior (and hence posterior) inference for the parameters of a $\text{VAR}_m(p)$ process is constrained to the stationary region.
- The new parameters represent orientation-preserving transformations of partial autocorrelation matrices that retain the structure of numerous meaningful parametric forms.
- They are interpretable, unconstrained and facilitate specification of an exchangeable prior. Moreover, MCMC is routine.
- Current and future extensions:
  - Determination of model order using a cumulative shrinkage process for an overfitted model (Legramanti et al., 2020);
  - Computational inference under a uniform prior for the $\boldsymbol{P}_s$ using spherical augmentation (Lan and Shahbaba, 2016) and Lagrangian Monte Carlo (Lan et al., 2015).
  - Application to determine change points in multichannel electroencephalographic (EEG) data for epilsepsy patients.

References I

Ansley, C. F. and R. Kohn (1986). A note on reparameterizing a vector autoregressive moving average model to enforce stationarity. *Journal of Statistical Computation and Simulation 24*, 99–106.

Ansley, C. F. and P. Newbold (1979). Multivariate partial autocorrelations. In *Proceedings of the Business and Economics Section*, pp. 349–353. American Statistical Association.

Heaps, S. E. (in press). Enforcing stationarity through the prior in vector autoregressions. *Journal of Computational and Graphical Statistics*.

Koop, G. M. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics 28*(2), 177–203.

Lan, S. and B. Shahbaba (2016). Sampling constrained probability distributions using spherical augmentation. In H. Q. Minh and V. Murino (Eds.), *Algorithmic Advances in Riemannian Geometry and Applications*, Advances in Computer Vision and Pattern Recognition, pp. 25–71. Springer International Publishing.

References II

Lan, S., V. Stathopoulos, B. Shahbaba, and M. Girolami (2015). Markov chain Monte Carlo from Lagrangian dynamics. *Journal of Computational and Graphical Statistics 24*(2), 357–378.

Legramanti, S., D. Durante, and D. B. Dunson (2020). Bayesian cumulative shrinkage for infinite factorizations. *Biometrika 107*(3), 745–752.

Roy, A., T. S. McElroy, and P. Linton (2019). Constrained estimation of causal, invertible VARMA. *Statistica Sinica 29*, 455–478.

Skeleton Stan code

```
functions {
  /* Function to compute the matrix square root */
  matrix sqrtm(matrix A) {
    int m = rows(A);
    vector[m] root_root_evals = sqrt(sqrt(eigenvalues_sym(A)));
    matrix[m, m] evecs = eigenvectors_sym(A);
    matrix[m, m] eprod = diag_post_multiply(evecs, root_root_evals);
    return tcrossprod(eprod);
  }
  /* Function to transform A to P (inverse of part 2 of reparameterisation) */
  matrix AtoP(matrix A) {
    int m = rows(A);
    matrix[m, m] B = tcrossprod(A);
    for(i in 1:m) B[i, i] += 1.0;
    return mdivide_left_spd(sqrtm(B), A);
  }
```

Skeleton Stan code cont'd

```
functions {
  /* Function to perform the reverse mapping from Appendix A.2.
     Returned: a (2 x p) array of (m x m) matrices; the (1, s)-th component
               of the array is phi_s and the (2, s)-th component of the array
               is Gamma_{s-1}*/
  matrix[,] rev_mapping(matrix[] P, matrix Sigma) {
    // ... details ...
  }
}
data {
  // ... as you would expect ...
}
parameters {
  matrix[m, m] A[p];
  cov_matrix[m] Sigma;
  vector[p] Amu[2];
  vector<lower=0>[p] Aomega[2];
}
```

Skeleton Stan code cont'd

```
transformed parameters {
  matrix[m, m] phi[p];
  matrix[p*m, p*m] Gamma; // (Stationary) variance of (y_1, ..., y_p)
  {
    /* ... construct phi and Gamma from the A_s and Sigma using
           the AtoP and rev_mapping functions ... */
  }
}
model {
  // ... likelihood in terms of phi_s, Sigma and Gamma ...
  // ... prior for A_s, Sigma, Amu, Aomega ...
}
```

Defining the partial autocorrelation matrices

- For each $s = 1, \dots, p$ define forward and reverse sub-processes:

$$\boldsymbol{y}_{t+1} = \sum_{i=1}^{s} \phi_{si} \boldsymbol{y}_{t-i+1} + \boldsymbol{\epsilon}_{s,t+1}, \quad \boldsymbol{\epsilon}_{s,t+1} \sim \mathrm{N}_m(\boldsymbol{0}, \boldsymbol{\Sigma}_s)$$

and

$$\boldsymbol{y}_{t-s} = \sum_{i=1}^{s} \phi_{si}^* \boldsymbol{y}_{t-s+i} + \boldsymbol{\epsilon}_{s,t-s}^*, \quad \boldsymbol{\epsilon}_{s,t-s}^* \sim \mathrm{N}_m(\boldsymbol{0}, \boldsymbol{\Sigma}_s^*).$$

- The ϕ_{si} (ϕ_{si}^*) are coefficients in the conditional expectations of $\boldsymbol{y}_t$ given its s predecessors (successors).
- $\boldsymbol{\Sigma}_s = \mathrm{Var}(\boldsymbol{y}_{t+1} | \boldsymbol{y}_{t:t-s+1})$ and $\boldsymbol{\Sigma}_s^* = \mathrm{Var}(\boldsymbol{y}_{t-s} | \boldsymbol{y}_{(t-s+1):t})$ are the corresponding conditional variances.
- Let $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0^* = \boldsymbol{\Gamma}_0$ where $\boldsymbol{\Gamma}_i = \mathrm{Cov}(\boldsymbol{y}_t, \boldsymbol{y}_{t+i}) = \mathrm{E}(\boldsymbol{y}_t \boldsymbol{y}_{t+i}^T)$.

Defining the partial autocorrelation matrices cont'd

- Let $\boldsymbol{\Sigma}_s = \boldsymbol{S}_s \boldsymbol{S}_s^T$ and $\boldsymbol{\Sigma}_s^* = \boldsymbol{S}_s^* \boldsymbol{S}_s^{*\,T}$ for $s = 0, \dots, p$.
- We take the symmetric matrix-square-root factorisation so $\boldsymbol{S}_s = \boldsymbol{S}_s^T = \boldsymbol{\Sigma}_s^{1/2}$ and $\boldsymbol{S}_s^* = \boldsymbol{S}_s^{*\,T} = \boldsymbol{\Sigma}_s^{*\,1/2}$.
- Let $\boldsymbol{z}_{s,t+1} = \boldsymbol{S}_s^{-1} \boldsymbol{\epsilon}_{s,t+1}$ and $\boldsymbol{z}_{s,t-s}^* = \boldsymbol{S}_s^{*\,-1} \boldsymbol{\epsilon}_{s,t-s}^*$ be standardised versions of the forward and reverse error series, then

$$
\begin{aligned}
\boldsymbol{P}_{s+1} &= \mathrm{Cov}(\boldsymbol{z}_{s,t+1}, \boldsymbol{z}_{s,t-s}^*) \\
&= \boldsymbol{S}_s^{-1} \mathrm{Cov}(\boldsymbol{y}_{t+1}, \boldsymbol{y}_{t-s} | \boldsymbol{y}_t, \dots, \boldsymbol{y}_{t-s+1}) (\boldsymbol{S}_s^{*\,-1})^T \\
&= \boldsymbol{S}_s^{-1} \phi_{s+1,s+1} \boldsymbol{S}_s^*
\end{aligned}
$$

for $s = 0, \dots, p-1$.

Forward mapping

The mapping from $\{\boldsymbol{\Sigma}, (\phi_1, \dots, \phi_p)\} \in \mathcal{S}_m^+ \times \mathcal{C}_{p,m}$ to $\{\boldsymbol{\Sigma}, (\boldsymbol{P}_1, \dots, \boldsymbol{P}_p)\} \in \mathcal{S}_m^+ \times \mathcal{V}_m^p$, described in Ansley and Newbold (1979), proceeds in two main stages.

1. From $\{\boldsymbol{\Sigma}, (\phi_1, \dots, \phi_p)\}$, compute the autocovariances $\boldsymbol{\Gamma}_i = \mathrm{Cov}(\boldsymbol{y}_t, \boldsymbol{y}_{t+i})$ for $i = 0, \dots, p$.
2. From $(\phi_1, \dots, \phi_p)$ and $(\boldsymbol{\Gamma}_0, \dots, \boldsymbol{\Gamma}_p)$ compute the partial autocorrelation matrices $(\boldsymbol{P}_1, \dots, \boldsymbol{P}_p)$ as follows.
   a. Initialise: construct $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0^* = \boldsymbol{\Gamma}_0$ and then calculate their matrix-square-root factorisations, $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0^* = \boldsymbol{S}_0 \boldsymbol{S}_0^T = \boldsymbol{S}_0^* \boldsymbol{S}_0^{*\,T}$.
   b. Recursion: for each $s = 0, \dots, p-1$
      i. Compute $\phi_{s+1,s+1}$ and $\phi_{s+1,s+1}^*$ using

$$\phi_{s+1,s+1} = \left(\boldsymbol{\Gamma}_{s+1}^T - \phi_{s1}\boldsymbol{\Gamma}_s^T - \dots - \phi_{ss}\boldsymbol{\Gamma}_1^T\right) \boldsymbol{\Sigma}_s^{*\,-1}$$

$$\phi_{s+1,s+1}^* = \left(\boldsymbol{\Gamma}_{s+1} - \phi_{s1}^*\boldsymbol{\Gamma}_s - \dots - \phi_{ss}^*\boldsymbol{\Gamma}_1\right) \boldsymbol{\Sigma}_s^{-1}.$$

Forward mapping cont'd

ii. If $s > 0$, for $i = 1, \ldots, s$, compute $\phi_{s+1,i}$ and $\phi^*_{s+1,i}$ using

$$\phi_{s+1,i} = \phi_{si} - \phi_{s+1,s+1}\phi^*_{s,s-i+1}$$
$$\phi^*_{s+1,i} = \phi^*_{si} - \phi^*_{s+1,s+1}\phi_{s,s-i+1}$$

iii. Compute the $(s+1)$-th partial autocorrelation $\boldsymbol{P}_{s+1}$ using

$$\boldsymbol{P}_{s+1} = \boldsymbol{S}_s^{-1}\phi_{s+1,s+1}\boldsymbol{S}^*_s,$$
$$\text{or} \quad \boldsymbol{P}_{s+1} = \left(\boldsymbol{S}^{*\,-1}_s\phi^*_{s+1,s+1}\boldsymbol{S}_s\right)^T.$$

iv. If $s < p-1$, compute $\boldsymbol{\Sigma}_{s+1}$ and $\boldsymbol{\Sigma}^*_{s+1}$ using

$$\boldsymbol{\Sigma}_{s+1} = \boldsymbol{\Gamma}_0 - \phi_{s+1,1}\boldsymbol{\Gamma}_1 - \ldots - \phi_{s+1,s+1}\boldsymbol{\Gamma}_{s+1}$$
$$\boldsymbol{\Sigma}^*_{s+1} = \boldsymbol{\Gamma}_0 - \phi^*_{s+1,1}\boldsymbol{\Gamma}_1^T - \ldots - \phi^*_{s+1,s+1}\boldsymbol{\Gamma}_{s+1}^T$$

and then calculate their matrix-square-root factorisations, $\boldsymbol{\Sigma}_{s+1} = \boldsymbol{S}_{s+1}\boldsymbol{S}_{s+1}^T$ and $\boldsymbol{\Sigma}^*_{s+1} = \boldsymbol{S}^*_{s+1}\boldsymbol{S}^{*\,T}_{s+1}$.

Reverse mapping

The inverse mapping from $\{\boldsymbol{\Sigma}, (\boldsymbol{P}_1, \dots, \boldsymbol{P}_p)\} \in \mathcal{S}_m^+ \times \mathcal{V}_m^p$ to $\{\boldsymbol{\Sigma}, (\phi_1, \dots, \phi_p)\} \in \mathcal{S}_m^+ \times \mathcal{C}_{p,m}$, proceeds in two main stages, the second of which is based on Lemma 2.1 of Ansley and Kohn (1986).

1. From $\{\boldsymbol{\Sigma}, (\boldsymbol{P}_1, \dots, \boldsymbol{P}_p)\}$ compute the stationary variance matrix $\boldsymbol{\Gamma}_0$.
   a. Initialise: let $\boldsymbol{\Sigma}_p = \boldsymbol{\Sigma}$ with corresponding matrix-square-root factorisation, $\boldsymbol{\Sigma}_p = \boldsymbol{S}_p \boldsymbol{S}_p^T$.
   b. Recursion: for each $s = p-1, \dots, 0$ construct the symmetric (or lower triangular) matrix $\boldsymbol{S}_s$ such that

$$\boldsymbol{\Sigma}_{s+1} = \boldsymbol{S}_s(\boldsymbol{I}_m - \boldsymbol{P}_{s+1}\boldsymbol{P}_{s+1}^T)\boldsymbol{S}_s^T$$

   then compute $\boldsymbol{\Sigma}_s = \boldsymbol{S}_s \boldsymbol{S}_s^T$.
   c. Output: take $\boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma}_0$.

Reverse mapping cont'd

2. From $(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_p)$ and $\boldsymbol{\Gamma}_0$ compute the matrices of autoregressive coefficients $(\phi_1, \ldots, \phi_p)$ as follows.
   a. Initialise: let $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0^* = \boldsymbol{\Gamma}_0$ with corresponding matrix-square-root factorisation, $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0^* = \boldsymbol{S}_0 \boldsymbol{S}_0^T = \boldsymbol{S}_0^* \boldsymbol{S}_0^{*\,T}$.
   b. Recursion: for each $s = 0, \ldots, p-1$
      i. Compute $\phi_{s+1,s+1}$ and $\phi_{s+1,s+1}^*$ using

$$\phi_{s+1,s+1} = \boldsymbol{S}_s \boldsymbol{P}_{s+1} \boldsymbol{S}_s^{*\,-1}$$
$$\phi_{s+1,s+1}^* = \boldsymbol{S}_s^* \boldsymbol{P}_{s+1}^T \boldsymbol{S}_s^{-1}$$

      ii. If $s > 0$, for $i = 1, \ldots, s$, compute $\phi_{s+1,i}$ and $\phi_{s+1,i}^*$ using

$$\phi_{s+1,i} = \phi_{si} - \phi_{s+1,s+1} \phi_{s,s-i+1}^*,$$
$$\phi_{s+1,i}^* = \phi_{si}^* - \phi_{s+1,s+1}^* \phi_{s,s-i+1}.$$

Reverse mapping cont'd

iii. Compute $\boldsymbol{\Sigma}_{s+1}$ and $\boldsymbol{\Sigma}^*_{s+1}$ using

$$\boldsymbol{\Sigma}_{s+1} = \boldsymbol{\Sigma}_s - \phi_{s+1,s+1} \boldsymbol{\Sigma}^*_s \phi^T_{s+1,s+1}$$
$$\boldsymbol{\Sigma}^*_{s+1} = \boldsymbol{\Sigma}^*_s - \phi^*_{s+1,s+1} \boldsymbol{\Sigma}_s \phi^{*\,T}_{s+1,s+1}$$

and then calculate their matrix-square-root factorisations, $\boldsymbol{\Sigma}_{s+1} = \boldsymbol{S}_{s+1}\boldsymbol{S}^T_{s+1}$ and $\boldsymbol{\Sigma}^*_{s+1} = \boldsymbol{S}^*_{s+1}\boldsymbol{S}^{*\,T}_{s+1}$.

iv. Compute $\boldsymbol{\Gamma}_{s+1}$ using

$$\boldsymbol{\Gamma}^T_{s+1} = \phi_{s+1,s+1} \boldsymbol{\Sigma}^*_s + \phi_{s1} \boldsymbol{\Gamma}^T_s + \ldots + \phi_{ss} \boldsymbol{\Gamma}^T_1$$

c. Output: take $\phi_i = \phi_{pi}$ for $i = 1, \ldots, p$. By construction, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_p$.