Computation of Standardized Residuals for MARSS Models

E. E. Holmes*

December 11, 2019

Abstract

This report shows how to compute the variance of the joint conditional model and state residuals for multivariate autoregressive Gaussian state-space (MARSS) models. The bulk of the report focuses on 'smoothations', which are the residuals conditioned on all the data $t = 1$ to T. The final part of the report covers 'innovations', which are residuals conditioned on the data $t = 1$ to $t - 1$.

The MARSS model can be written: $\boldsymbol{x}_t = \mathbf{B}\boldsymbol{x}_{t-1} + \mathbf{u} + \boldsymbol{w}_t$, $\boldsymbol{y}_t = \mathbf{Z}\boldsymbol{x}_t + \mathbf{z} + \boldsymbol{v}_t$, where $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$ are multivariate Gaussian error-terms with variance-covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ respectively. The joint conditional residuals are the $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$ conditioned on the observed data, which may be incomplete (missing values). Harvey, Koopman and Penzer (1998) show a recursive algorithm for the smoothation residuals (conditioned on all the data). I show an alternate algorithm to compute these residuals using the conditional variances of the states and the conditional covariance between unobserved data and states. This allows one to compute the variance of un-observed residuals (residuals associated with missing or left-out data), which is needed for leave-one-out cross-validation tests. I show how to modify the Harvey et al. algorithm in the case of missing values and how to modify it to return the non-normalized conditional residuals.

Keywords: Time-series analysis, Kalman filter, residuals, maximum-likelihood, vector autoregressive model, dynamic linear model, parameter estimation, state-space

citation: Holmes, E. E. 2014. Computation of standardized residuals for (MARSS) models. Technical Report. arXiv:1411.0045

*Northwest Fisheries Science Center, NOAA Fisheries, Seattle, WA 98112, eli.holmes@noaa.gov, http://faculty.washington.edu/eeholmes

1 Overview

This report discusses the computation of the variance of the conditional model and state residuals for MARSS models of the form:

$$\begin{aligned} \boldsymbol{x}_t &= \mathbf{B}_t\boldsymbol{x}_{t-1} + \mathbf{u}_t + \boldsymbol{w}_t, \text{ where } \boldsymbol{W}_t \sim \text{MVN}(0, \mathbf{Q}_t) \\ \boldsymbol{y}_t &= \mathbf{Z}_t\boldsymbol{x}_t + \mathbf{a}_t + \boldsymbol{v}_t, \text{ where } \boldsymbol{V}_t \sim \text{MVN}(0, \mathbf{R}_t) \\ \boldsymbol{X}_0 &\sim \text{MVN}(\boldsymbol{\xi}, \boldsymbol{\Lambda}) \end{aligned} \tag{1}$$

The state and model residuals are respectively

$$\begin{aligned} \boldsymbol{w}_t &= \boldsymbol{x}_t - \mathbf{B}_t\boldsymbol{x}_{t-1} - \mathbf{u}_t \\ \boldsymbol{v}_t &= \boldsymbol{y}_t - \mathbf{Z}_t\boldsymbol{x}_t - \mathbf{a}_t \end{aligned} \tag{2}$$

Given a set of observed data $\boldsymbol{y}_t$ and states $\boldsymbol{x}_t$, the model residuals are $\boldsymbol{y}_t - (\mathbf{Z}_t\boldsymbol{x}_t + \mathbf{a}_t) = \boldsymbol{v}_t$. The model residual is a random variable since $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ are drawn from the joint multivariate distribution of $\boldsymbol{Y}_t$ and $\boldsymbol{X}_t$ defined by the MARSS equations (Equation 1). The unconditional[1] variance of the model residuals is

$$\text{var}_V[\boldsymbol{V}_t] = \text{var}_{XY}[\boldsymbol{Y}_t - (\mathbf{Z}_t\boldsymbol{X}_t + \mathbf{a}_t)] = \mathbf{R}_t \tag{3}$$

based on the distribution of $\boldsymbol{V}_t$ in Equation 1. var_{XY} indicates that the integration is over the joint-unconditional distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Once we have data, $\mathbf{R}_t$ is not the variance-covariance matrix of our model residuals because our residuals are now conditioned[2] on a set of observed data. There are two types of conditional model residuals used in MARSS analyses: innovations and smoothations. Innovations are the model residuals at time t using the expected value of $\boldsymbol{X}_t$ conditioned on the data from 1 to $t-1$. Smoothations are the model residuals using the expected value of $\boldsymbol{X}_t$ conditioned on all the data, $t = 1$ to T. Smoothations are used in computing standardized residuals for outlier and structural break detection (Harvey et al., 1998; de Jong and Penzer, 1998; Commandeur and Koopman, 2007).

2 Distribution of MARSS smoothation residuals

This section discusses computation of the variance of the model and state residuals conditioned on all the data from $t = 1$ to T. These MARSS residuals are often used for outlier detection and shock detection, and in this case you only need the distribution of the model residuals for the observed values. However if you wanted to do a leave-one-out cross-validation, you would need to know the distribution of the residuals for data points you left out (treated as unobserved). The equations in this report give you the former and the latter, while the algorithm by Harvey et al. (1998) gives only the former.

2.1 Notation

Throughout, I follow the convention that capital letters are random variables and small letters are a realization from the random variable. This only applies to random variables; parameters are not random variables[3]. Parameters are shown in Roman font while while random variables are bold slanted font. Parameters written as capital letters are matrices, while parameters written in small letters are strictly column matrices.

In this report, the distribution over which the integration is done in an expectation or variance is given by the subscript, e.g. $\text{E}_A[f(A)]$ indicates an unconditional expectation over the distribution of A without conditioning on another random variable while $\text{E}_{A|b}[f(A)|b]$ would indicate an expectation over the distribution of A conditioned on $B = b$; presumably A and B are not independent otherwise $B = b$ would have no effect on A. $\text{E}_{A|b}[f(A)|b]$ is a fixed value, not random. It is the expected value when $B = b$. In contrast, $\text{E}_{A|B}[f(A)|B]$ denotes the random variable of all the possible $\text{E}_{A|b}[f(A)|b]$ given all the possible b values that B might take. The variance of $\text{E}_{A|B}[f(A)|B]$ is the variance of this random variable. The variance of

[1] meaning not conditioning on any particular set of observed data but rather taking the expectation across all possible values of $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$.

[2] 'conditioned' means that the probability distribution of the residual has changed. The distribution is now the distribution given that $\boldsymbol{Y} = \boldsymbol{y}$, say. Expectations and variances var[] are integrals over the value that a random variable might take multiplied by the probability of that value. When presenting an 'expectation', the probability distribution is normally implicit but for derivations involving conditional expectiations, it is important to be explicit about the distribution that is being integrated over.

[3] in a frequentist framework

$\mathrm{E}_{A|b}[f(A)|b]$ in contrast is 0 since it is a fixed value. We will often be working with the random variables, $\mathrm{E}_{A|B}[f(A)|B$ or $\mathrm{var}_{A|B}[f(A)|B]$, inside an expectation or variance: such as $\mathrm{var}_B[\,\mathrm{E}_{A|B}[f(A)|B]]$.

2.2 Model residuals conditioned on all the data

Define the smoothations $\hat{\boldsymbol{v}}_t$ as:

$$\hat{\boldsymbol{v}}_t = \boldsymbol{y}_t - \mathbf{Z}_t\widetilde{\boldsymbol{x}}_t - \mathbf{a}_t, \tag{4}$$

where $\widetilde{\boldsymbol{x}}_t$ is $\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$. The smoothation is different from $\boldsymbol{v}_t$ because it uses $\widetilde{\boldsymbol{x}}_t$ not $\boldsymbol{x}_t$; $\boldsymbol{x}_t$ is not known, $\widetilde{\boldsymbol{x}}_t$ is its estimate. $\widetilde{\boldsymbol{x}}_t$ is output by the Kalman smoother. $\boldsymbol{y}^{(1)}$ means all the observed data from $t = 1$ to T. $\boldsymbol{y}^{(1)}$ is a sample from the random variable $\boldsymbol{Y}^{(1)}$. The unobserved $\boldsymbol{y}$ will be termed $\boldsymbol{y}^{(2)}$ and is a sample from the random variable $\boldsymbol{Y}^{(2)}$. When $\boldsymbol{Y}$ appears without a superscript, it means both $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$ together. Similarly $\boldsymbol{y}$ means both $\boldsymbol{y}^{(1)}$ and $\boldsymbol{y}^{(2)}$ together—the observed data that we use to estimate $\widetilde{\boldsymbol{x}}_t$ and the unobserved data that we do not use and may or may not know. $\hat{\boldsymbol{v}}_t$ exists for both $\boldsymbol{y}^{(1)}$ and $\boldsymbol{y}^{(2)}$, though we might not know $\boldsymbol{y}^{(2)}$ and thus might not know its corresponding $\hat{\boldsymbol{v}}_t$. In some cases, however, we do know $\boldsymbol{y}^{(2)}$; they are data that we left out of our model fitting, in say a k-fold or leave-one-out cross-validation.

$\hat{\boldsymbol{v}}_t$ is sample from the random variable $\hat{\boldsymbol{V}}_t$ since $\boldsymbol{Y}$ is a random variable. We want to compute the mean and variance of this random variable $\hat{\boldsymbol{V}}_t$ over all possibles values that $\boldsymbol{X}$ and $\boldsymbol{Y}$ might take. The mean of $\hat{\boldsymbol{V}}_t$ is 0 and we are concerned only with computing the variance:

$$\mathrm{var}[\hat{\boldsymbol{V}}_t] = \mathrm{var}_{XY}[\boldsymbol{Y}_t - \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{Y}^{(1)}] - \mathbf{a}_t] \tag{5}$$

Notice we have an unconditional variance over XY on the outside (that subscript on var) and a conditional expectation over a specific value of $\boldsymbol{Y}^{(1)}$ on the inside (in the $\mathrm{E}[\,]$).

To compute this, I will use the "law of total variance":

$$\mathrm{var}_A[A] = \mathrm{var}_B[\,\mathrm{E}_{A|B}[A|B]] + \mathrm{E}_B[\,\mathrm{var}_{A|B}[A|B]] \tag{6}$$

The subscripts on the inner expectations make it explicit that the expectations are being taken over the conditional distributions. $\mathrm{var}_{A|B}[A|B]$ and $\mathrm{E}_{A|B}[A|B]$ are random variables because the B in the conditional is a random variable. We take the expectation or variance with it fixed at one value, b, but B can take other values of b also. Going forward, I will write this more succinctly as

$$\mathrm{var}[A] = \mathrm{var}_B[\,\mathrm{E}[A|B]] + \mathrm{E}_B[\,\mathrm{var}[A|B]] \tag{7}$$

It is important to note that $\mathrm{E}_{A|b}[A|b]$ and $\mathrm{var}_{A|b}[A|b]$ are fixed values while $\mathrm{E}_{A|B}[A|B]$ and $\mathrm{var}_{A|B}[A|B]$ are random variables. When computing $\mathrm{var}[\,\mathrm{E}_{A|B}[A|B]]$, we will compute $\mathrm{E}_{A|b}[A|b]$ and compute the variance of that over all possible values of b.

From the law of total variance, we can write the variance of the model residuals as

$$\mathrm{var}[\hat{\boldsymbol{V}}_t] = \mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{Y}^{(1)}]] + \mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\hat{\boldsymbol{V}}_t|\boldsymbol{Y}^{(1)}]] \tag{8}$$

2.2.1 First term in Equation 8

Notice that the random variable inside the $\mathrm{var}[\,]$ in the first term is

$$\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{Y}^{(1)}] == \mathrm{E}[(\boldsymbol{Y}_t + \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{Y}^{(1)}] + \mathbf{a}_t)|\boldsymbol{Y}^{(1)}] \tag{9}$$

Let's consider this for a specific value $\boldsymbol{Y}^{(1)} = \boldsymbol{y}^{(1)}$:

$$\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}] = \mathrm{E}[(\boldsymbol{Y}_t + \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathbf{a}_t)|\boldsymbol{y}^{(1)}] = \mathrm{E}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}] + \mathbf{Z}_t\,\mathrm{E}[\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]|\boldsymbol{y}^{(1)}] + \mathrm{E}[\mathbf{a}_t)|\boldsymbol{y}^{(1)}] \tag{10}$$

$\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ is a fixed value, and the expected value of a fixed value is itself. So $\mathrm{E}[\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]|\boldsymbol{y}^{(1)}] = \mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$. Thus,

$$\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}] = \mathrm{E}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}] + \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{E}[\mathbf{a}_t)|\boldsymbol{y}^{(1)}] \tag{11}$$

and this applies for all $\boldsymbol{y}^{(1)}$. So

$$\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{Y}^{(1)}] = \mathrm{E}[\boldsymbol{Y}_t|\boldsymbol{Y}^{(1)}] + \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{Y}^{(1)}] + \mathrm{E}[\mathbf{a}_t)|\boldsymbol{Y}^{(1)}]. \tag{12}$$

We can move the conditional out and write

$$\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{Y}^{(1)}] = \mathrm{E}[(\boldsymbol{Y}_t + \mathbf{Z}_t\boldsymbol{X}_t + \mathbf{a}_t)|\boldsymbol{Y}^{(1)}]. \tag{13}$$

The right side is $\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)}]$, no 'hat' on the $\boldsymbol{V}_t$. This means that the first term in Equation 8 can be written with no hat on $\boldsymbol{V}$:

$$\mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{Y}^{(1)}]] = \mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)}]] \tag{14}$$

Using the law of total variance, we can re-write $\mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)}]]$ as:

$$\mathrm{var}[\boldsymbol{V}_t] = \mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)}]] + \mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)}]] \tag{15}$$

From Equation 15, we can solve for $\mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)}]]$:

$$\mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)}]] = \mathrm{var}[\boldsymbol{V}_t] - \mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)}]] \tag{16}$$

From Equation 3, we know that $\mathrm{var}[\boldsymbol{V}_t] = \mathbf{R}_t$ (this is the unconditional variance).

The second term in Equation 16 to the right of the equal sign and inside the expectation, $\mathrm{var}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)}]$, is the variance of $\boldsymbol{V}_t$ with $\boldsymbol{Y}^{(1)}$ held at a specific fixed $\boldsymbol{y}^{(1)}$. The variability in $\mathrm{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]$ (notice $\boldsymbol{y}^{(1)}$ not $\boldsymbol{Y}^{(1)}$ now) comes from $\boldsymbol{X}_t$ (and the rest of the $\boldsymbol{X}$) and $\boldsymbol{Y}^{(2)}$ which are random variables. Let's compute this variance for a specific $\boldsymbol{y}^{(1)}$ value.

$$\mathrm{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}] = \mathrm{var}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t - \mathbf{a}_t|\boldsymbol{y}^{(1)}]. \tag{17}$$

Notice that there is no E (expectation) on the $\boldsymbol{X}_t$; this is $\boldsymbol{V}_t$ not $\hat{\boldsymbol{V}}_t$. $\mathbf{a}_t$ is a fixed value and can be dropped; the $\boldsymbol{Y}_t$ and $\boldsymbol{X}_t$ are random variables.

Equation 17 can then be written as[4]:

$$\begin{aligned}
\mathrm{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}] &= \mathrm{var}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] \\
&= \mathrm{var}[-\mathbf{Z}_t\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{var}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}] + \mathrm{cov}[\boldsymbol{Y}_t, -\mathbf{Z}_t\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{cov}[-\mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{Y}_t|\boldsymbol{y}^{(1)}] \\
&= \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{U}}_t - \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top - \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top
\end{aligned} \tag{18}$$

If there were no missing data, i.e. if $\boldsymbol{y}^{(1)} = \boldsymbol{y}$, then $\widetilde{\mathbf{U}}_t$ and $\widetilde{\mathbf{S}}_t$ would be zero and this would reduce to $\mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top$. But we are concerned with the case where there are missing values. Those missing values need not be for all t. That is, there may be some observed y at time t and some missing y. $\boldsymbol{y}_t$ is multivariate.

$\widetilde{\boldsymbol{V}}_t = \mathrm{var}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ and is output by the Kalman smoother. $\widetilde{\mathbf{U}}_t = \mathrm{var}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}]$ and $\widetilde{\mathbf{S}}_t = \mathrm{cov}[\boldsymbol{Y}_t, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$. The equations for these are given in Holmes (2012) and are output by the `MARSShatyt` function in the MARSS R package[5].

So we now know $\mathrm{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]$ for a specific $\boldsymbol{y}^{(1)}$. We want $\mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)}]]$ which is its expected value of all possible values of $\boldsymbol{y}^{(1)}$. $\widetilde{\boldsymbol{V}}_t$, $\widetilde{\mathbf{U}}_t$ and $\widetilde{\mathbf{S}}_t$ are multivariate normal random variables. The conditional variance does not depend on the value that you are conditioning on[6]

Because $\widetilde{\boldsymbol{V}}_t$, $\widetilde{\mathbf{U}}_t$ and $\widetilde{\mathbf{S}}_t$ only depend on the MARSS parameters values, $\mathbf{Q}$, $\mathbf{B}$, $\mathbf{R}$, etc., the second term in Equation 16, $\mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)}]]$, is equal to $\mathrm{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]$. Putting this into Equation 16, we have

$$\mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)})]] = \mathrm{var}[\boldsymbol{V}_t] - \mathrm{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}] = \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top - \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top \tag{19}$$

Since $\mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{Y}^{(1)})]] = \mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{Y}^{(1)})]]$ (Equation 14), this means that the first term in Equation 8 is

$$\mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{Y}^{(1)})]] = \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top - \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top \tag{20}$$

[4] $var(A+B) = var(A) + var(B) + cov(A,B) + cov(B,A)$

[5] $\widetilde{\mathbf{U}}_t$ is `ytT` and $\widetilde{\mathbf{S}}_t$ is `OtT - tcrossprod(ytT)` in the `MARSShatyt()` output.

[6] Let the $\mathbf{A}$ be a N-dimensional multivariate normal random variable partitioned into $\mathbf{A}_1$ and $\mathbf{A}_2$ with variance-covariance matrix $\Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_2 1 & \Sigma_2 \end{bmatrix}$. The variance-covariance matrix of $\mathbf{A}$ conditioned on $\mathbf{A}_1 = \mathbf{a}$ is $\Sigma = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_2 - \Sigma_{12}\Sigma_1\Sigma_{21} \end{bmatrix}$. Notice that $\mathbf{a}$ does not appear in the conditional variance matrix.

2.2.2 Second term in Equation 8

Consider the second term in Equation 8. This term is

$$\mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[(\boldsymbol{Y}_t - \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{Y}^{(1)}] - \mathbf{a}_t)|\boldsymbol{Y}^{(1)}]] \tag{21}$$

The middle term is a conditional expectation inside a conditional variance inside an unconditional expectation:

$$\mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{Y}^{(1)}]|\boldsymbol{Y}^{(1)}]] \tag{22}$$

Let's solve the inner part for a specific $\boldsymbol{Y}^{(1)} = \boldsymbol{y}^{(1)}$. $\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ is a fixed value. Thus $\mathrm{var}[\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]|\boldsymbol{y}^{(1)}] = 0$. The variance of a fixed value is 0. This is true for all $yy^{(1)}$ so the middle term reduces to 0. $\mathbf{a}_t$ is also fixed and its variance is also 0. Thus the second term reduces to $\mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\boldsymbol{Y}|\boldsymbol{Y}^{(1)}]]$.

$\boldsymbol{Y}$ is a shortcut for $(\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)})$, two (multivariate) random variables. The unconditional distribution of $\boldsymbol{Y}$ can be written

$$\begin{bmatrix} \Sigma_2 & \Sigma_{12} \\ \Sigma_2 1 & \Sigma_1 \end{bmatrix} \tag{23}$$

The conditional variance of $\boldsymbol{Y}$ conditioned on $\boldsymbol{Y}^{(1)} = \boldsymbol{y}^{(1)}$, is

$$\begin{bmatrix} \Sigma_2 - \Sigma_{12}\Sigma_1\Sigma_2 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{24}$$

Notice that $\boldsymbol{y}^{(1)}$ does not appear in the conditional variance. It is the same value regardless of the value of $\boldsymbol{y}^{(1)}$. Thus

$$\mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\boldsymbol{Y}|\boldsymbol{Y}^{(1)}]] = \mathrm{var}[\boldsymbol{Y}|\boldsymbol{y}^{(1)}]] = \widetilde{\mathbf{U}}_t$$

(25)

$\widetilde{\mathbf{U}}_t$ is only a function of the MARSS parameters.

Thus the second term in Equation 8 is simply $\widetilde{\mathbf{U}}_t$.

2.2.3 Putting together the first and second terms

We can now put the first and second terms in Equation 8 together and write out the equation for the variance of the model residuals:

$$\begin{aligned} \mathrm{var}[\hat{\boldsymbol{V}}_t] &= \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top - \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top + \widetilde{\mathbf{U}}_t \\ &= \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top \end{aligned} \tag{26}$$

This will reduce to $\mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top$ if $\boldsymbol{y}_t$ has no missing values and to $\mathbf{R}_t + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top$ if $\boldsymbol{y}_t$ is all missing values. The behavior if $\boldsymbol{y}_t$ has some missing and some not missing values depends on whether $\mathbf{R}_t$ is a diagonal matrix or not (i.e. if the $\boldsymbol{y}_t^{(1)}$ and $\boldsymbol{y}_t^{(2)}$ are correlated).

2.3 State residuals conditioned on the data

The state residuals are $\boldsymbol{x}_t - (\mathbf{B}_t\boldsymbol{x}_{t-1} + \mathbf{u}_t) = \boldsymbol{w}_t$. The unconditional expected value of the state residuals is $\mathrm{E}[\boldsymbol{X}_t - (\mathbf{B}_t\boldsymbol{X}_{t-1} + \mathbf{u}_t)] = \mathrm{E}[\boldsymbol{W}_t] = 0$ and the unconditional variance of the state residuals is

$$\mathrm{var}[\boldsymbol{X}_t - (\mathbf{B}_t\boldsymbol{X}_{t-1} + \mathbf{u}_t)] = \mathrm{var}[\boldsymbol{W}_t] = \mathbf{Q}_t \tag{27}$$

based on the definition of $\boldsymbol{W}_t$. The conditional state residuals (conditioned on the data from $t = 1$ to $t = T$) are defined as

$$\hat{\boldsymbol{w}}_t = \widetilde{\boldsymbol{x}}_t - \mathbf{B}_t\widetilde{\boldsymbol{x}}_{t-1} - \mathbf{u}_t. \tag{28}$$

It is a sample from the random variable $\hat{\boldsymbol{W}}_t$; random over different possible data sets. The expected value of $\hat{\boldsymbol{W}}_t$ is 0, and we can compute $\mathrm{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t]$ from the law of total variance.

$$\mathrm{var}[\boldsymbol{W}_t] = \mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] + \mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{29}$$

Using the observation that $\hat{\boldsymbol{W}}_t = \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{Y}^{(1)} = \boldsymbol{y}^{(1)}]$, we can write:

$$\mathrm{var}[\boldsymbol{W}_t] = \mathrm{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t] + \mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{30}$$

Solve for $\mathrm{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t]$ to get:

$$\mathrm{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t] = \mathrm{var}[\boldsymbol{W}_t] - \mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{31}$$

The variance in the expectation on the far right is

$$\mathrm{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}] = \mathrm{var}[(\boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1} - \mathbf{u}_t)|\boldsymbol{y}^{(1)}] \tag{32}$$

$\mathbf{u}_t$ is not a random variable and can be dropped. Thus,

$$\begin{aligned}\mathrm{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}] &= \mathrm{var}[(\boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1})|\boldsymbol{y}^{(1)}] \\ &= \mathrm{var}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{var}[\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[\boldsymbol{X}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[-\mathbf{B}_t\boldsymbol{X}_{t-1}, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}] \\ &= \widetilde{\boldsymbol{V}}_t + \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1}\mathbf{B}_t^\top - \widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1,t}\end{aligned} \tag{33}$$

This conditional variance does not depend on the actual values of $\boldsymbol{y}$. It depends only on the parameters values, $\mathbf{Q}$, $\mathbf{B}$, $\mathbf{R}$, etc. Putting together the above and $\mathrm{var}[\boldsymbol{W}_t] = \mathbf{Q}_t$ into Equation 31, the variance of the conditional state residuals is

$$\mathrm{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t] = \mathbf{Q}_t - \widetilde{\boldsymbol{V}}_t - \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1}\mathbf{B}_t^\top + \widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1,t} \tag{34}$$

2.4 Covariance of the conditional model and state residuals

The unconditional model and state residuals, $\boldsymbol{V}_t$ and $\boldsymbol{W}_t$, are independent (by definition), i.e. $\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t] = 0$. However the conditional model and state residuals, $\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t]$, are not independent since both depend on $\boldsymbol{y}^{(1)}$. Using the law of total covariance, we can write

$$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = \mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}],\, \mathrm{E}[\hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}]] + \mathrm{E}_{Y^{(1)}}[\,\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}]] \tag{35}$$

The covariance in the second term on the right can be written out as

$$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}] = \mathrm{E}_{Y^{(1)}}[\,\mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] - \mathbf{a}_t,\, \mathrm{E}[\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] - \mathbf{B}_t\,\mathrm{E}[\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] - \mathbf{u}_t|\boldsymbol{y}^{(1)}]] \tag{36}$$

The $\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ are fixed values for a given set of data. The covariance of a random variable with a fixed value is 0, thus $\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}]$ is 0. Thus Equation 35 reduces to

$$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = \mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}],\, \mathrm{E}[\hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}]] + 0 = \mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}],\, \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{37}$$

since $\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}] = \mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]$ and $\mathrm{E}[\hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}] = \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]$.

In the same way we used the law of total variance, we can use the law of total covariance to obtain $\mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}],\, \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]]$:

$$\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t] = \mathrm{E}_{Y^{(1)}}[\,\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] + \mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}],\, \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{38}$$

The unconditional covariance of $\boldsymbol{V}_t$ and $\boldsymbol{W}_t$ is 0. Thus the right side of Equation 38 is 0 and combining Equation 37 and 38, we get

$$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = -\,\mathrm{E}_{Y^{(1)}}[\,\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{39}$$

and our problem reduces to solving for the conditional covariance of the model and state residuals.

The conditional covariance $\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]$ can be written out as

$$\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] = \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t - \mathbf{a}_t, \boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1} - \mathbf{u}_t|\boldsymbol{y}^{(1)}] \tag{40}$$

$\mathbf{a}_t$ and $\mathbf{u}_t$ are fixed values and can be dropped. Thus

$$\begin{aligned}\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] &= \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] \\ &= \mathrm{cov}[\boldsymbol{Y}_t, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{cov}[\boldsymbol{Y}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[-\mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{cov}[-\mathbf{Z}_t\boldsymbol{X}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] \\ &= \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top\end{aligned} \tag{41}$$

where $\widetilde{\mathbf{S}}_t = \text{cov}[\boldsymbol{Y}_t, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ and $\widetilde{\mathbf{S}}_{t,t-1} = \text{cov}[\boldsymbol{Y}_t, \boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}]$; the equations for $\widetilde{\mathbf{S}}_t$ and $\widetilde{\mathbf{S}}_{t,t-1}$ are given in Holmes (2012) and are output by the `MARSShatyt` function in the MARSS R package.

$\widetilde{\boldsymbol{V}}_t$, $\widetilde{\boldsymbol{V}}_{t,t-1}$, $\widetilde{\mathbf{S}}_t$ and $\widetilde{\mathbf{S}}_{t,t-1}$ are only functions of the MARSS parameters not of $\boldsymbol{y}^{(1)}$. Thus

$$\text{E}_{Y^{(1)}}[\,\text{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] = \text{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] = \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \tag{42}$$

$\text{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t]$ is the negative of this (Equation 39), thus

$$\text{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = -\widetilde{\mathbf{S}}_t + \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \tag{43}$$

The Harvey et al. algorithm shown below gives the joint distribution of the model residuals at time t and state residuals at time $t+1$. Using the law of total covariance as above, the covariance in this case is

$$\text{cov}_{Y^{(1)}}[\,\text{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}],\, \text{E}[\boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]] = -\,\text{E}_{Y^{(1)}}[\,\text{cov}[\boldsymbol{V}_t, \boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]] \tag{44}$$

and

$$\begin{aligned}\text{cov}[\boldsymbol{V}_t, \boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}] &= \text{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t - \mathbf{a}_t, \boldsymbol{X}_{t+1} - \mathbf{B}_{t+1}\boldsymbol{X}_t - \mathbf{u}_{t+1}|\boldsymbol{y}^{(1)}] \\ &= \text{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_{t+1} - \mathbf{B}_{t+1}\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] \\ &= \widetilde{\mathbf{S}}_{t,t+1} - \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top\end{aligned} \tag{45}$$

Thus,

$$\text{cov}_{Y^{(1)}}[\,\text{E}[\boldsymbol{V}_t|\boldsymbol{y}],\, \text{E}[\boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]] = -\,\text{E}_{Y^{(1)}}[\,\text{cov}[\boldsymbol{V}_t, \boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]] = -\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top.$$

2.5 Joint distribution of the conditional residuals

We now the write the variance of the joint distribution of the conditional residuals. Define

$$\hat{\varepsilon}_t = \begin{bmatrix}\hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{w}}_t\end{bmatrix} = \begin{bmatrix}\boldsymbol{y}_t - \mathbf{Z}_t\widetilde{\boldsymbol{x}}_t - \mathbf{a}_t \\ \widetilde{\boldsymbol{x}}_t - \mathbf{B}_t\widetilde{\boldsymbol{x}}_{t-1} - \mathbf{u}_t\end{bmatrix}. \tag{46}$$

where $\widetilde{\boldsymbol{x}}_t$ and $\widetilde{\boldsymbol{x}}_{t-1}$ are conditioned on $\boldsymbol{y}(1)$, the observed $\boldsymbol{y}$. $\hat{\varepsilon}_t$ is a sample drawn from the distribution of $\hat{\mathcal{E}}_t$ conditioned on observations at the (1) locations in $\boldsymbol{Y}$. The expected value of $\hat{\mathcal{E}}_t$ over all possible $\boldsymbol{y}$ is 0 and the variance of $\hat{\mathcal{E}}_t$ is

$$\left[\begin{array}{c|c}\mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top & \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \\ \hline (\widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t)^\top & \mathbf{Q}_t - \widetilde{\boldsymbol{V}}_t - \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1}\mathbf{B}_t^\top + \widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1,t}\end{array}\right] \tag{47}$$

If the residuals are defined as in Harvey et al. (1998),

$$\hat{\varepsilon}_t = \begin{bmatrix}\hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{w}}_{t+1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{y}_t - \mathbf{Z}_t\widetilde{\boldsymbol{x}}_t - \mathbf{a}_t \\ \tilde{\boldsymbol{x}}_{t+1} - \mathbf{B}_{t+1}\widetilde{\boldsymbol{x}}_t - \mathbf{u}_{t+1}\end{bmatrix} \tag{48}$$

and the variance of $\hat{\mathcal{E}}_t$ is

$$\left[\begin{array}{c|c}\mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top & -\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top \\ \hline (-\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top)^\top & \mathbf{Q}_{t+1} - \widetilde{\boldsymbol{V}}_{t+1} - \mathbf{B}_{t+1}\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top + \widetilde{\boldsymbol{V}}_{t+1,t}\mathbf{B}_{t+1}^\top + \mathbf{B}_{t+1}\widetilde{\boldsymbol{V}}_{t,t+1}\end{array}\right] \tag{49}$$

The above gives the variance of both 'observed' model residuals (the ones associated with $\boldsymbol{y}^{(1)}$) and the unobserved model residuals (the ones associated with $\boldsymbol{y}^{(2)}$). When there are no missing values in $\boldsymbol{y}_t$, the $\widetilde{\mathbf{S}}_t$ and $\widetilde{\mathbf{S}}_{t,t-1}$ terms equal 0 and drop out.

3 Harvey et al. 1998 algorithm for the conditional residuals

Harvey et al. (1998, pgs 112-113) give a recursive algorithm for computing the variance of the conditional residuals when the time-varying MARSS equation is written as:

$$\begin{aligned} \boldsymbol{x}_{t+1} &= \mathbf{B}_{t+1}\boldsymbol{x}_t + \mathbf{u}_{t+1} + \mathbf{G}_{t+1}\epsilon_t, \\ \boldsymbol{y}_t &= \mathbf{Z}_t\boldsymbol{x}_t + \mathbf{a}_t + \mathbf{H}_t\epsilon_t, \\ \text{where } \epsilon_t &\sim \text{MVN}(0, \mathbf{I}_{m+n\times m+n}), \mathbf{G}_t\mathbf{G}_t^\top = \mathbf{Q}_t \text{ and } \mathbf{H}_t\mathbf{H}_t^\top = \mathbf{R}_t \end{aligned} \tag{50}$$

$\mathbf{G}_t$ has m rows and $m+n$ columns with the last n columns all 0; $\mathbf{H}_t$ has n rows and $m+n$ columns with the last m columns all zero. The algorithm in Harvey et al. (1998) gives the variance of the 'normalized' residuals, the ϵ_t. I have modified their algorithm so it returns the 'non-normalized' residuals:

$$\varepsilon_t = \begin{bmatrix} \mathbf{H}_t\epsilon_t \\ \mathbf{G}_{t+1}\epsilon_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_t \\ \boldsymbol{w}_{t+1} \end{bmatrix}.$$

The Harvey et al. algorithm is a backwards recursion using the following output from the Kalman filter: the one-step ahead prediction covariance $\mathbf{F}_t$, the Kalman gain $\mathbf{K}_t$, and $\widetilde{\boldsymbol{V}}_t^{t-1} = \text{var}[\boldsymbol{X}_t|\boldsymbol{y}^{(1,1:t-1}]$. These are output from `MARSSkfss()` in `Sigma`, `Kt`, and `Vtt1`.

4 Algorithm

Start from $t = T$ and work backwards to $t = 1$. At time T, $r_T = 0_{1\times m}$ and $N_T = 0_{m\times m}$. $\mathbf{B}_{t+1}$ and $\mathbf{Q}_{t+1}$ can be set to NA or 0. They will not appear in the algorithm at time T since $r_T = 0$ and $N_T = 0$. Note that the $\boldsymbol{w}$ residual and its associated variance and covariance with $\boldsymbol{v}$ at time T is NA since this residual would be for $\boldsymbol{x}_T$ to $\boldsymbol{x}_{T+1}$.

$$\begin{gathered} \mathbf{Q}^*_{t+1} = \begin{bmatrix}\mathbf{Q}_{t+1} & 0_{m\times n}\end{bmatrix}, \ \mathbf{R}^*_t = \begin{bmatrix}0_{n\times m} & \mathbf{R}^*_t\end{bmatrix} \\ \mathbf{F}_t = \mathbf{Z}^*_t\widetilde{\boldsymbol{V}}_t^{t-1}\mathbf{Z}^{*\top}_t + \mathbf{R}^*_t, \ K_t = \mathbf{B}_{t+1}\mathbf{K}_t = \mathbf{B}_{t+1}\widetilde{\boldsymbol{V}}_t^{t-1}\mathbf{Z}^{*\top}_t\mathbf{F}_t^{-1} \\ L_t = \mathbf{B}_{t+1} - K_t\mathbf{Z}^*_t, \ J_t = \mathbf{Q}^*_{t+1} - K_t\mathbf{R}^*_t, \ u_t = \mathbf{F}_t^{-1}\hat{\boldsymbol{v}}_t - K_t^\top r_t \\ r_{t-1} = \mathbf{Z}^{*\top}_t u_t + \mathbf{B}^\top_{t+1}r_t, \ N_{t-1} = \mathbf{Z}^{*\top}_t\mathbf{F}_t^{-1}\mathbf{Z}^*_t + L_t^\top N_t L_t \\ repeat for \text{t-1} \end{gathered} \tag{51}$$

Bolded terms are the same as in Equation 50 or output by `MARSSkfss()`. Unbolded terms are terms used in Harvey et al. (1998). The * on $\mathbf{Z}_t$ and $\mathbf{R}_t$, indicates that they are the missing value modified versions discussed in Shumway and Stoffer (2006, section 6.4) and **?**: the rows of $\mathbf{Z}_t$ corresponding to missing rows of $\boldsymbol{y}_t$ are set to zero and the (i,j) and (j,i) terms of $\mathbf{R}_t$ corresponding the missing rows of $\boldsymbol{y}_t$ are set to zero. For the latter, this means if the i-th row of $\boldsymbol{y}_t$ is missing, then then all the (i,j) and (j,i) terms, including (i,i) are set to 0. It is assumed that a missing values modified inverse of $\mathbf{F}_t$ is used: any 0's on the diagonal of $\mathbf{F}_t$ are replaced with 1, the inverse is taken, and 1s on diagonals is replaced back with 0s.

The residuals (Harvey et al., 1998, eqn 24) are

$$\hat{\varepsilon}^*_t = \begin{bmatrix} \hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{w}}_{t+1} \end{bmatrix} = (\mathbf{R}^*_t)^\top u_t + (\mathbf{Q}^*_{t+1})^\top r_t \tag{52}$$

with mean of 0 ($\text{E}_{Y^{(1)}}(\hat{\varepsilon}_t) = 0$) and variance

$$\Sigma^*_t = \text{var}_{Y^{(1)}}(\hat{\varepsilon}_t) = \mathbf{R}^{*\top}_t\mathbf{F}_t^{-1}\mathbf{R}^*_t + J_t^\top N_t J_t \tag{53}$$

The * signifies that these are the missing values modified $\hat{\varepsilon}_t$ and Σ_t; see comments above.

4.1 Difference in notation

In Equation 20 in Harvey et al. (1998), their T_t is my $\mathbf{B}_{t+1}$ and their $H_tH_t^\top$ is my $\mathbf{Q}_{t+1}$. Notice the difference in the time indexing. My time indexing on $\mathbf{B}$ and $\mathbf{Q}$ matches the left $\boldsymbol{x}$ while in theirs, T and H indexing matches the right $\boldsymbol{x}$. Thus in my implementation of their algorithm (Harvey et al., 1998, eqns. 21-24), $\mathbf{B}_{t+1}$

appears in place of T_t and $\mathbf{Q}_{t+1}$ appears in place of H_t. See comments below on normalization and the difference between $\mathbf{Q}$ and H.

Harvey et al. (1998, eqns. 19, 20) use G_t to refer to the $\text{chol}(\mathbf{R}_t)^\top$ (essentially) and H_t to refer to $\text{chol}(\mathbf{Q}_t)^\top$. I have replaced these with $\mathbf{R}_t^*$ and $\mathbf{Q}_t^*$, respectively, which causes my variant of their algorithm (Equation 51) to give the 'non-normalized' variance of the residuals. The residuals function in the MARSS package has an option to give either normalized or non-normalized residuals.

$\mathbf{K}_t$ is the Kalman gain output by the MARSS package. The Kalman gain as used in the Harvey et al. (1998) algorithm is $K_t = \mathbf{B}_{t+1}\mathbf{K}_t$. Notice that Equation 21 in Harvey et al. (1998) has $H_t G_t^\top$ in the equation for K_t (Kalman gain). This is the covariance of the state and observation errors, which is allowed given the way they write the errors in their Equations 19 and 20. The way the MARSS package model is written, the state and observation errors are independent of each other. Thus $H_t G_t^\top = 0$ and this term drops out of the $\mathbf{K}_t$ equation.

4.2 Computing the standardized residuals

The standardized residuals are computed by multiplying $\hat{\varepsilon}_t$ by the inverse of the square root of the variance-covariance matrix from which $\hat{\varepsilon}_t$ is "drawn":

$$(\Sigma_t^*)^{-1/2}\hat{\varepsilon}_t^* \tag{54}$$

Notice that the missing values modified $\hat{\varepsilon}_t^*$ and Σ_t^* are used. if the i-th row of $\boldsymbol{y}_t$ is missing, the i-th row of $\hat{\varepsilon}_t$ is set to 0 and the i-th row and column of Σ_t is set to all 0. There will be 0s on the diagonal of Σ_t^* so your code will need to deal with these.

5 Distribution of the MARSS innovation residuals

One-step-ahead predictions (innovations) are often shown for MARSS models and these are used for likelihood calculations. Innovations are the prediction of $\mathbf{y}$ given data up to $t-1$. Although not termed 'innovations', we can also compute the predictions of $\mathbf{x}$ conditioned on the data up the $t-1$. This section gives the residual variance for these one-step-ahead predictions. I walk through the derivation of the variance matrix very pendantically following the procedures in Section 2. In texts on state-space models, this derivation would be written simply as 'it follows obviously that' because in this case the $\boldsymbol{y}_t$ is independent of the model preditions (which use only data up to $t-1$) and the variance-covariance matrix is indeed 'obvious'.

5.0.1 Variance of the one-step-ahead model residuals

Define the innovations $\hat{\boldsymbol{v}}_t$ as:

$$\hat{\boldsymbol{v}}_t = \boldsymbol{y}_t - \mathbf{Z}_t\widetilde{\boldsymbol{x}}_t - \mathbf{a}_t, \tag{55}$$

where $\widetilde{\boldsymbol{x}}_t$ is now $\text{E}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}]$ and is output by the Kalman filter. The difference between the smoothation (Equation 4) and innovation is that the former is conditioned on all the data $t=1$ to T and the latter is conditioned on $t=1$ to $t-1$ ($\boldsymbol{y}_{1:t-1}^{(1)}$). As before (1) denotes the observed data (non-missing) and (2) denotes the missing data. $\hat{\boldsymbol{v}}_t$ is sample from the random variable $\hat{\boldsymbol{V}}_t$ since $\boldsymbol{Y}_{1:t-1}^{(1)}$ is a random variable and the data we have collected $\boldsymbol{y}_{1:t-1}^{(1)}$ is a sample from that.

As before, we want to compute the unconditional mean and variance of this random variable $\hat{\boldsymbol{V}}_t$; unconditional here means we take the expectations over all possibles values that $\boldsymbol{y}_{1:t-1}$, both $\boldsymbol{y}_{1:t-1}^{(1)}$ and $\boldsymbol{y}_{1:t-1}^{(2)}$, might take:

$$\text{var}[\hat{\boldsymbol{V}}_t] = \text{var}_{Y_{1:t-1}}[\boldsymbol{Y}_t - \mathbf{Z}_t\,\text{E}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}] - \mathbf{a}_t] \tag{56}$$

While this looks similar to the smoothation case (Equation 5), it is different in that $\boldsymbol{y}_t$ does not appear in $\text{E}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}]$; the conditioning is only on data up to $t-1$. Looking at Equation 56, you might guess that $\text{var}[\hat{\boldsymbol{V}}_t] = \mathbf{R}_t - \mathbf{Z}_t\,\text{var}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}]]\mathbf{Z}_t^\top$ but that is not the case. The "-" should be a "+". Let's walk throught the derivation to see why that is.

As before, I will use the "law of total variance":

$$\text{var}_A[A] = \text{var}_B[\,\text{E}_{A|b}[A|b]] + \text{E}_B[\,\text{var}_{A|b}[A|b]] \tag{57}$$

which again I write more succinctly as

$$\mathrm{var}_A[A] = \mathrm{var}_B[\,\mathrm{E}[A|b]] + \mathrm{E}_B[\,\mathrm{var}[A|b]] \tag{58}$$

From the law of total variance , we can write

$$\mathrm{var}_{V_t}[\hat{\boldsymbol{V}}_t] = \mathrm{var}_{Y_{1:t-1}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}_{1:t-1}]] + \mathrm{E}_{Y_{1:t-1}}[\,\mathrm{var}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}_{1:t-1}]] \tag{59}$$

$\mathrm{var}_{Y_{1:t-1}}$ and $\mathrm{E}_{Y_{1:t-1}}$ are expectations over both $\boldsymbol{Y}^{(1)}_{1:t-1}$ and $\boldsymbol{Y}^{(2)}_{1:t-1}$, so all possible values of $\boldsymbol{Y}_{1:t-1}$. Notice that we are not conditioning on any data after $t-1$.

5.0.2 First term in Equation 59

Because $\boldsymbol{y}_t$ does not appear in the conditional of $\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}_{1:t-1}]$, it may be obvious that this is equal to $\mathrm{E}[\hat{\boldsymbol{V}}_t]$ and thus 0. However, let's walk through this since it is due to the structure of the model not a general property.

Using the equation for $\hat{\boldsymbol{v}}_t$ (Equation 55), we can write

$$\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}_{1:t-1}] = \mathrm{E}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}_{1:t-1}] - \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}_{1:t-1}] - \mathbf{a}_t \tag{60}$$

Because $\boldsymbol{y}_t$ does not appear in $\mathrm{E}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}_{1:t-1}]$,

$$\mathrm{E}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}_{1:t-1}] = \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}_{1:t-1}] + \mathbf{a}_t \tag{61}$$

So the right side of Equation 60 is 0 and the first term in Equation 59 is 0:

$$\mathrm{var}_{Y_{1:t-1}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}_{1:t-1}]] = 0 \tag{62}$$

5.0.3 Second term in Equation 59

This term is

$$\mathrm{E}_{Y_{1:t-1}}[\,\mathrm{var}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}_{1:t-1}] = \mathrm{E}_{Y_{1:t-1}}[\,\mathrm{var}[\boldsymbol{Y}_t - \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}_{1:t-1}] - \mathbf{a}_t|\boldsymbol{y}^{(1)}_{1:t-1}]] \tag{63}$$

$\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}_{1:t-1}]$ is a fixed value, it is not $\boldsymbol{X}_t$ but its expected value, and it appears inside a conditional variance not the unconditional variance. If the variance was over $\boldsymbol{Y}_{1:t-1}$, the expectation would be a random variable but the variance is conditioned on $\boldsymbol{Y}_{1:t-1} = \boldsymbol{y}_{1:t-1}$ so the expectation is a fixed value. Thus the second term reduces to $\mathrm{E}_{Y_{1:t-1}}[\,\mathrm{var}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}_{1:t-1}]]$. The conditional variance of $\boldsymbol{Y}_t$ is from Equation 1

$$\mathrm{var}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}_{1:t-1}] = \mathrm{var}[\mathbf{Z}_t\boldsymbol{X}_t + \mathbf{a}_t + \boldsymbol{V}_t|\boldsymbol{y}^{(1)}_{1:t-1}] = \mathbf{Z}_t\,\mathrm{var}[\boldsymbol{X}_t||\boldsymbol{y}^{(1)}_{1:t-1}]\mathbf{Z}_t^\top + \mathrm{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}_{1:t-1}]. \tag{64}$$

$\widetilde{\mathbf{U}}_t$ is not a function of $\boldsymbol{y}$ is is only a function of the MARSS parameters. Thus the second term in Equation **??** is simply $\widetilde{\mathbf{U}}_t$.

5.0.4 Putting together the two terms

$$\begin{aligned}\mathrm{var}[\hat{\boldsymbol{V}}_t] &= \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top - \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top + \widetilde{\mathbf{U}}_t \\ &= \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top\end{aligned} \tag{65}$$

This will reduce to $\mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top$ if $\boldsymbol{y}_t$ has no missing values and to $\mathbf{R}_t + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top$ if $\boldsymbol{y}_t$ is all missing values.

5.1 State residuals conditioned on the data

The state residuals are $\boldsymbol{x}_t - (\mathbf{B}_t\boldsymbol{x}_{t-1} + \mathbf{u}_t) = \boldsymbol{w}_t$. The unconditional expected value of the state residuals is $\mathrm{E}(\boldsymbol{X}_t - (\mathbf{B}_t\boldsymbol{X}_{t-1} + \mathbf{u}_t)) = \mathrm{E}(\boldsymbol{W}_t) = 0$ and the unconditional variance of the state residuals is

$$\mathrm{var}[\boldsymbol{X}_t - (\mathbf{B}_t\boldsymbol{X}_{t-1} + \mathbf{u}_t)] = \mathrm{var}[\boldsymbol{W}_t] = \mathbf{Q}_t \tag{66}$$

based on the definition of $\boldsymbol{W}_t$. The conditional state residuals (conditioned on the full data) are defined as

$$\hat{\boldsymbol{w}}_t = \widetilde{\boldsymbol{x}}_t - \mathbf{B}_t\widetilde{\boldsymbol{x}}_{t-1} - \mathbf{u}_t. \tag{67}$$

It is a sample from the random variable $\hat{\boldsymbol{W}}_t$; random over different possible data sets. The expected value of $\hat{\boldsymbol{W}}_t$ is 0, and we can compute $\mathrm{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t]$ from the law of total variance using the observation that $\hat{\boldsymbol{w}}_t = \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]$.

$$\mathrm{var}[\boldsymbol{W}_t] = \mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] + \mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{68}$$

Thus,

$$\mathrm{var}_{Y^{(1)}}[\hat{\boldsymbol{w}}_t] = \mathrm{var}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] = \mathrm{var}[\boldsymbol{W}_t] - \mathrm{E}_{Y^{(1)}}[\,\mathrm{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{69}$$

The variance in the expectation on the far right is

$$\begin{aligned}
\mathrm{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}] &= \mathrm{var}[\boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1} - \mathbf{u}_t|\boldsymbol{y}^{(1)}] \\
&\quad \mathbf{u} \text{ is not a random variable and can be dropped} \\
&= \mathrm{var}[\boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] \\
&= \mathrm{var}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{var}[\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[\boldsymbol{X}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[-\mathbf{B}_t\boldsymbol{X}_{t-1}, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}] \\
&= \widetilde{\boldsymbol{V}}_t + \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1}\mathbf{B}_t^\top - \widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1,t}
\end{aligned} \tag{70}$$

This conditional variance does not depend on the actual values of $\boldsymbol{y}$. It depends only on the parameters values, $\mathbf{Q}$, $\mathbf{B}$, $\mathbf{R}$, etc. Using the above and $\mathrm{var}[\boldsymbol{W}_t] = \mathbf{Q}_t$ in Equation 31, the variance of the conditional state residuals is

$$\mathrm{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t] = \mathbf{Q}_t - \widetilde{\boldsymbol{V}}_t - \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1}\mathbf{B}_t^\top + \widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1,t} \tag{71}$$

5.2 Covariance of the conditional model and state residuals

The unconditional model and state residuals, $\boldsymbol{V}_t$ and $\boldsymbol{W}_t$, are independent (by definition), i.e. $\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t] = 0$. However the conditional model and state residuals, $\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t]$, are not independent since both depend on $\boldsymbol{y}^{(1)}$. Using the law of total covariance, we can write

$$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = \mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}],\, \mathrm{E}[\hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}]] + \mathrm{E}_{Y^{(1)}}[\,\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}]] \tag{72}$$

The covariance in the second term on the right can be written out as

$$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}] = \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] - \mathbf{a}_t,\, \mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] - \mathbf{B}_t\,\mathrm{E}[\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] - \mathbf{u}_t|\boldsymbol{y}^{(1)}] \tag{73}$$

The $\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$, $\mathrm{E}[\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}]$ and $\mathbf{u}_t$ in the second term are fixed values for a given set of data, thus the second term in the covariance is a fixed value. The covariance of a random variable with a fixed value is 0, thus $\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}]$ is 0. Thus Equation 72 reduces to

$$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = \mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}],\, \mathrm{E}[\hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}]] + 0 = \mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}],\, \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{74}$$

Since $\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}] = \mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]$ and $\mathrm{E}[\hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}] = \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]$. In the same way we used the law of total variance, we can use the law of total covariance to obtain $\mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}],\, \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]]$:

$$\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t] = \mathrm{E}_{Y^{(1)}}[\,\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] + \mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}],\, \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{75}$$

The unconditional covariance of $\boldsymbol{V}_t$ and $\boldsymbol{W}_t$ is 0. Thus the right side of Equation 38 is 0 and combining Equation 37 and 38,

$$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = -\,\mathrm{E}_{Y^{(1)}}[\,\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{76}$$

and our problem reduces to solving for the conditional covariance of the model and state residuals.

The conditional covariance $\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]$ can be written out as

$$\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] = \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t - \mathbf{a}_t, \boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1} - \mathbf{u}_t|\boldsymbol{y}^{(1)}] \tag{77}$$

$\mathbf{a}_t$ and $\mathbf{u}_t$ are fixed values and can be dropped. Thus

$$\begin{aligned}\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] &= \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}y] \\ &= \mathrm{cov}[\boldsymbol{Y}_t, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{cov}[\boldsymbol{Y}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[-\mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_t] + \mathrm{cov}[-\mathbf{Z}_t\boldsymbol{X}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}] \\ &= \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top\end{aligned} \tag{78}$$

where $\widetilde{\mathbf{S}}_t = \mathrm{cov}[\boldsymbol{Y}_t, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ and $\widetilde{\mathbf{S}}_{t,t-1} = \mathrm{cov}[\boldsymbol{Y}_t, \boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}]$; the equations for $\widetilde{\mathbf{S}}_t$ and $\widetilde{\mathbf{S}}_{t,t-1}$ are given in Holmes (2012) and are output by the `MARSShatyt` function in the MARSS R package. $\widetilde{\boldsymbol{V}}_t$, $\widetilde{\boldsymbol{V}}_{t,t-1}$, $\widetilde{\mathbf{S}}_t$ and $\widetilde{\mathbf{S}}_{t,t-1}$ are only functions of the MARSS parameters not of $\boldsymbol{y}$. Thus

$$\mathrm{E}_{Y^{(1)}}[\,\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] = \mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] = \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \tag{79}$$

$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t]$ is the negative of this (Equation 39), thus

$$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = -\widetilde{\mathbf{S}}_t + \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \tag{80}$$

The Harvey et al. algorithm shown below gives the joint distribution of the model residuals at time t and state residuals at time $t+1$. Using the law of total covariance as above The covariance in this case is

$$\mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}],\, \mathrm{E}[\boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]] = -\,\mathrm{E}_{Y^{(1)}}[\,\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]] \tag{81}$$

and

$$\begin{aligned}\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}] &= \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t - \mathbf{a}_t, \boldsymbol{X}_{t+1} - \mathbf{B}_{t+1}\boldsymbol{X}_t - \mathbf{u}_{t+1}|\boldsymbol{y}^{(1)}] \\ &= \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_{t+1} - \mathbf{B}_{t+1}\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] \\ &= \widetilde{\mathbf{S}}_{t,t+1} - \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top\end{aligned} \tag{82}$$

Thus,

$$\mathrm{cov}_{Y^{(1)}}[\,\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}],\, \mathrm{E}[\boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]] = -\,\mathrm{E}_{Y^{(1)}}[\,\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]] = -\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top.$$

5.3 Joint distribution of the conditional residuals

We now the write the variance of the joint distribution of the conditional residuals. Define

$$\hat{\varepsilon}_t = \begin{bmatrix}\hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{w}}_t\end{bmatrix} = \begin{bmatrix}\boldsymbol{y}_t - \mathbf{Z}_t\widetilde{\boldsymbol{x}}_t - \mathbf{a}_t \\ \widetilde{\boldsymbol{x}}_t - \mathbf{B}_t\widetilde{\boldsymbol{x}}_{t-1} - \mathbf{u}_t\end{bmatrix}. \tag{83}$$

where $\widetilde{\boldsymbol{x}}_t$ and $\widetilde{\boldsymbol{x}}_{t-1}$ are conditioned on $\boldsymbol{y}(1)$, the observed $\boldsymbol{y}$. $\hat{\varepsilon}_t$ is a sample drawn from the distribution of $\hat{\mathcal{E}}_t$ conditioned on observations at the (1) locations in $\boldsymbol{Y}$. The expected value of $\hat{\mathcal{E}}_t$ over all possible $\boldsymbol{y}$ is 0 and the variance of $\hat{\mathcal{E}}_t$ is

$$\left[\begin{array}{c|c}\mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top & \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \\ \hline (\widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t)^\top & \mathbf{Q}_t - \widetilde{\boldsymbol{V}}_t - \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1}\mathbf{B}_t^\top + \widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1,t}\end{array}\right] \tag{84}$$

If the residuals are defined as in Harvey et al. (1998),

$$\hat{\varepsilon}_t = \begin{bmatrix}\hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{w}}_{t+1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{y}_t - \mathbf{Z}_t\widetilde{\boldsymbol{x}}_t - \mathbf{a}_t \\ \tilde{\boldsymbol{x}}_{t+1} - \mathbf{B}_{t+1}\widetilde{\boldsymbol{x}}_t - \mathbf{u}_{t+1}\end{bmatrix} \tag{85}$$

and the variance of $\hat{\mathcal{E}}_t$ is

$$\left[\begin{array}{c|c}\mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top & -\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top \\ \hline (-\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top)^\top & \mathbf{Q}_{t+1} - \widetilde{\boldsymbol{V}}_{t+1} - \mathbf{B}_{t+1}\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top + \widetilde{\boldsymbol{V}}_{t+1,t}\mathbf{B}_{t+1}^\top + \mathbf{B}_{t+1}\widetilde{\boldsymbol{V}}_{t,t+1}\end{array}\right] \tag{86}$$

The above gives the variance of both 'observed' model residuals (the ones associated with $\boldsymbol{y}^{(1)}$) and the unobserved model residuals (the ones associated with $\boldsymbol{y}^{(2)}$). When there are no missing values in $\boldsymbol{y}_t$, the $\widetilde{\mathbf{S}}_t$ and $\widetilde{\mathbf{S}}_{t,t-1}$ terms equal 0 and drop out.

Thus we write Equation 55 directly as:

$$\mathrm{var}[\hat{\boldsymbol{V}}_t] = \mathrm{var}_{Y^{(1)}}[\boldsymbol{Y}_t] - \mathbf{Z}_t \, \mathrm{E}[\boldsymbol{X}_t | \boldsymbol{y}^{(1,1:t-1)}] \tag{87}$$

$\mathbf{a}_t$ drops out because it is not a random variable.

$$\begin{aligned} \mathrm{var}[\hat{\boldsymbol{V}}_t] &= \mathbf{R}_t - \mathbf{Z}_t \widetilde{\boldsymbol{V}}_t \mathbf{Z}_t^\top - \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{S}}_t \mathbf{Z}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{S}}_t^\top + \widetilde{\mathbf{U}}_t \\ &= \mathbf{R}_t - \mathbf{Z}_t \widetilde{\boldsymbol{V}}_t \mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t \mathbf{Z}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{S}}_t^\top \end{aligned} \tag{88}$$

This will reduce to $\mathbf{R}_t - \mathbf{Z}_t \widetilde{\boldsymbol{V}}_t \mathbf{Z}_t^\top$ if $\boldsymbol{y}_t$ has no missing values and to $\mathbf{R}_t + \mathbf{Z}_t \widetilde{\boldsymbol{V}}_t \mathbf{Z}_t^\top$ if $\boldsymbol{y}_t$ is all missing values.

References

Commandeur, J. J. and Koopman, S. J. (2007). *An introduction to state space time series analysis*. Practical Econometrics. Oxford University Press, Oxford.

de Jong, P. and Penzer, J. (1998). Diagnosing shocks in time series. *Journal of the American Statistical Association*, 93(442):796–806.

Harvey, A., Koopman, S. J., and Penzer, J. (1998). Messy time series: a unified approach. *Advances in Econometrics*, 13:103–143.

Holmes, E. E. (2012). Derivation of the EM algorithm for constrained and unconstrained MARSS models. Technical report, arXiv:1302.3919 [stat.ME].

Shumway, R. and Stoffer, D. (2006). *Time series analysis and its applications*. Springer-Science+Business Media, LLC, New York, New York, 2nd edition.