Computation of Standardized Residuals for MARSS Models

E. E. Holmes*

December 9, 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 normal 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 way 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 data), which is needed for leave-one-out cross-validation tests and to compute the conditional variance of the missing data. 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. The main report is on the smoothation residuals (conditioned on all the data). The final section shows the innovations residuals (conditioned on the data up to $t - 1$).

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

---

*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:

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

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 equation. The unconditional[1] variance of the model residuals is

$$\text{var}_{XY}[\boldsymbol{Y}_t - (\mathbf{Z}_t\boldsymbol{X}_t + \mathbf{a}_t)] = \text{var}[\boldsymbol{V}_t] = \mathbf{R}_t \tag{2}$$

based on the definition of $\boldsymbol{V}_t$. var_{XY} indicates that the integration is over the joint-unconditional distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$ (over all time not just t).

Once we have data, $\mathbf{R}_t$ is not the variance 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 the 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. 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.

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]. In this report, the distribution over which the integration is done in an expectation $\text{E}[]$ or variance $\text{var}[]$ is given by the subscript, e.g. $\text{E}_A[]$ indicates an unconditional expectation over the distribution of $\mathbf{A}$ without conditioning on another random variable while $\text{E}_{A|b}[]$ would indicate an expectation over the distribution of $\mathbf{A}$ conditioned on $\mathbf{B} = \mathbf{b}$. A joint expectation of $f(\mathbf{A}, \mathbf{B})$ would be $\text{E}_{AB}[f(\mathbf{A}, \mathbf{B})]$, where 'AB' denotes the joint probability distribution. This can be written in terms of the conditional expectation using the law of total expectation $\text{E}_B[\text{E}_{A|b}[f(\mathbf{A}, \mathbf{B})|\mathbf{B} = \mathbf{b}]]$.

2.1 Model residuals conditioned on all the data

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

$$\hat{\boldsymbol{v}}_t = \boldsymbol{y}_t - \mathbf{Z}_t\widetilde{\boldsymbol{x}}_t - \mathbf{a}_t, \tag{3}$$

where $\widetilde{\boldsymbol{x}}_t$ is $\text{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ and 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 $\boldsymbol{Y}^{(2)}$

---

[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 $\boldsymbol{Y}(1) = \boldsymbol{y}(1)$. Expectations $\text{E}[]$ and variances $\text{var}[]$ 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

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 $\boldsymbol{y}^{(2)}$. $\hat{\boldsymbol{v}}_t$ is sample from the random variable $\hat{\boldsymbol{V}}_t$ since $\boldsymbol{Y}^{(1)}$ is a random variable and the data we have collected $\boldsymbol{y}^{(1)}$ is a sample from that. 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 $\boldsymbol{y}$, both $\boldsymbol{y}^{(1)}$ and $\boldsymbol{y}^{(2)}$, might take. The mean is 0 and we are concerned only with computing the variance:

$$\text{var}[\hat{\boldsymbol{V}}_t] = \text{var}_{Y^{(1)}}[\boldsymbol{Y}_t - \mathbf{Z}_t \, \mathrm{E}[\boldsymbol{X}_t | \boldsymbol{Y}^{(1)}) = \boldsymbol{y}^{(1)}] - \mathbf{a}_t] \tag{4}$$

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

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

$$\text{var}[A] = \text{var}_B[\mathrm{E}_{A|b}[A|B = b]] + \mathrm{E}_B[\text{var}_{A|b}[A|B = b]] \tag{5}$$

The subscripts on the inner expectations make it explicit that the expectations are being taken over the conditional distributions. However, going forward, I will write this more succinctly as

$$\text{var}[A] = \text{var}_B[\mathrm{E}[A|b]] + \mathrm{E}_B[\text{var}[A|b]] \tag{6}$$

It is understood that $\mathrm{E}[A|b]$ is the conditional expectation of A conditioned on $B = b$ and $\text{var}[A|b]$ is the conditional variance[4].

From the law of total variance, we can write

$$\text{var}[\hat{\boldsymbol{V}}_t] = \text{var}_{Y^{(1)}}[\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)})]] + \mathrm{E}_{Y^{(1)}}[\text{var}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)})]] \tag{7}$$

The var and E are expectations over all possible values of $\boldsymbol{Y}^{(1)}$.

2.1.1 First term in Equation 7

Notice that

$$\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}] = \mathrm{E}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}] - \mathbf{Z}_t \, \mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] - \mathbf{a}_t = (\boldsymbol{Y}_t + \mathbf{Z}_t\boldsymbol{X}_t + \mathbf{a}_t)|\boldsymbol{y}^{(1)} = \mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}] \tag{8}$$

So the first term is $\text{var}_{Y^{(1)}}[\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]]$.

From the law of total variance, we can write

$$\text{var}[\boldsymbol{V}_t] = \text{var}_{Y^{(1)}}[\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]] + \mathrm{E}_{Y^{(1)}}[\text{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]] \tag{9}$$

From Equation 9, we can solve for $\text{var}_{Y^{(1)}}[\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]]$:

$$\text{var}_{Y^{(1)}}[\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]] = \text{var}[\boldsymbol{V}_t] - \mathrm{E}_{Y^{(1)}}[\text{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]] \tag{10}$$

From Equation 2, $\text{var}[\boldsymbol{V}_t] = \mathbf{R}_t$. The second term to the right of the equal sign, $\text{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]$, is the variance of $\boldsymbol{V}_t$ holding $\boldsymbol{y}^{(1)}$ fixed but allowing $\boldsymbol{X}_t$ (and the rest of the $\boldsymbol{X}$) to be random variables:

$$\text{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}] = \text{var}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t - \mathbf{a}_t|\boldsymbol{y}^{(1)}]. \tag{11}$$

$\mathbf{a}_t$ is a fixed value and can be dropped.

Equation 11 can then be written as[5]

$$\begin{aligned}
\text{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}] &= \text{var}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] \\
&= \text{var}[-\mathbf{Z}_t\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \text{var}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}] + \text{cov}[\boldsymbol{Y}_t, -\mathbf{Z}_t\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \text{cov}[-\mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{Y}_t|\boldsymbol{y}^{(1)}] \\
&= \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{U}}_t - \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top - \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top
\end{aligned} \tag{12}$$

---

[4]Normally the law of total variance is written $\text{var}[A] = \text{var}[\mathrm{E}[A|B]] + \mathrm{E}[\text{var}[A|B]]$

[5]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.

$\widetilde{\boldsymbol{V}}_t = \text{var}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ and is output by the Kalman smoother. $\widetilde{\mathbf{U}}_t = \text{var}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}]$ and $\widetilde{\mathbf{S}}_t = \text{cov}[\boldsymbol{Y}_t, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$. The equations for these are given in Holmes (2012) and are output by the `MARSShatyt` function in the MARSS R package[6].

$\widetilde{\boldsymbol{V}}_t$, $\widetilde{\mathbf{U}}_t$ and $\widetilde{\mathbf{S}}_t$ are conditional multivariate Normal variances and their values do not depend on the values of $\boldsymbol{y}^{(1)}$. They only depend on the parameters values, $\mathbf{Q}$, $\mathbf{B}$, $\mathbf{R}$, etc., in the MARSS equation. Thus $\text{E}_{Y^{(1)}}[\text{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]] = \text{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}]$. Putting this all together, we can write the variance of $\hat{\boldsymbol{V}}_t$ as

$$\text{var}_{Y^{(1)}}[\text{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}]] = \text{var}[\boldsymbol{V}_t] - \text{var}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}] = \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top - \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top \tag{13}$$

2.1.2 Second term in Equation 7

Consider the second term in Equation 7. This term is

$$\text{E}_{Y^{(1)}}[\text{var}[(\boldsymbol{Y}_t - \mathbf{Z}_t\text{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] - \mathbf{a}_t)|\boldsymbol{y}^{(1)}]] \tag{14}$$

$\text{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ is a fixed value; it is not $\boldsymbol{X}_t$ but its expected value. $\mathbf{a}_t$ is also fixed. Thus the second term reduces to $\text{E}_{Y^{(1)}}[\text{var}[\boldsymbol{Y}|\boldsymbol{y}^{(1)}]] = \text{E}_{Y^{(1)}}[\widetilde{\mathbf{U}}_t]$. $\widetilde{\mathbf{U}}_t$ is a variance and is not a function of $\boldsymbol{y}^{(1)}$. It is only a function of the MARSS parameters. Thus the second term in Equation 7 is simply $\widetilde{\mathbf{U}}_t$.

2.1.3 Putting together the two terms

$$\begin{aligned}
\text{var}[\hat{\boldsymbol{V}}_t] &= \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top - \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top + \widetilde{\mathbf{U}}_t \\
&= \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top
\end{aligned} \tag{15}$$

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).

2.2 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 $\text{E}[\boldsymbol{X}_t - (\mathbf{B}_t\boldsymbol{X}_{t-1} + \mathbf{u}_t)] = \text{E}[\boldsymbol{W}_t] = 0$ and the unconditional variance of the state residuals is

$$\text{var}[\boldsymbol{X}_t - (\mathbf{B}_t\boldsymbol{X}_{t-1} + \mathbf{u}_t)] = \text{var}[\boldsymbol{W}_t] = \mathbf{Q}_t \tag{16}$$

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

$$\hat{\boldsymbol{w}}_t = \widetilde{\boldsymbol{x}}_t - \mathbf{B}_t\widetilde{\boldsymbol{x}}_{t-1} - \mathbf{u}_t. \tag{17}$$

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 $\text{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t]$ from the law of total variance.

$$\text{var}[\boldsymbol{W}_t] = \text{var}_{Y^{(1)}}[\text{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] + \text{E}_{Y^{(1)}}[\text{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{18}$$

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

$$\text{var}[\boldsymbol{W}_t] = \text{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t] + \text{E}_{Y^{(1)}}[\text{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{19}$$

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

$$\text{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t] = \text{var}[\boldsymbol{W}_t] - \text{E}_{Y^{(1)}}[\text{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{20}$$

The variance in the expectation on the far right is

$$\text{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}] = \text{var}[(\boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1} - \mathbf{u}_t)|\boldsymbol{y}^{(1)}] \tag{21}$$

---

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

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

$$
\begin{aligned}
\mathrm{var}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}] &= \mathrm{var}[(\boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1})|\boldsymbol{y}^{(1)}] \\
&= \mathrm{var}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{var}[\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[\boldsymbol{X}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[-\mathbf{B}_t\boldsymbol{X}_{t-1}, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}] \quad (22)\\
&= \widetilde{\boldsymbol{V}}_t + \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1}\mathbf{B}_t^\top - \widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1,t}
\end{aligned}
$$

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 20, the variance of the conditional state residuals is

$$
\mathrm{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t] = \mathbf{Q}_t - \widetilde{\boldsymbol{V}}_t - \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1}\mathbf{B}_t^\top + \widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1,t} \quad (23)
$$

2.3 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

$$
\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = \mathrm{cov}_{Y^{(1)}}[\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}], \mathrm{E}[\hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}]] + \mathrm{E}_{Y^{(1)}}[\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}]] \quad (24)
$$

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

$$
\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}] = \mathrm{E}_{Y^{(1)}}[\mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] - \mathbf{a}_t, \mathrm{E}[\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] - \mathbf{B}_t\,\mathrm{E}[\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] - \mathbf{u}_t|\boldsymbol{y}^{(1)}]] \quad (25)
$$

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 24 reduces to

$$
\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = \mathrm{cov}_{Y^{(1)}}[\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}^{(1)}], \mathrm{E}[\hat{\boldsymbol{W}}_t|\boldsymbol{y}^{(1)}]] + 0 = \mathrm{cov}_{Y^{(1)}}[\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}], \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \quad (26)
$$

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

$$
\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t] = \mathrm{E}_{Y^{(1)}}[\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] + \mathrm{cov}_{Y^{(1)}}[\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}], \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \quad (27)
$$

The unconditional covariance of $\boldsymbol{V}_t$ and $\boldsymbol{W}_t$ is 0. Thus the right side of Equation 27 is 0 and combining Equation 26 and 27, we get

$$
\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = -\mathrm{E}_{Y^{(1)}}[\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \quad (28)
$$

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

$$
\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] = \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t - \mathbf{a}_t, \boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1} - \mathbf{u}_t|\boldsymbol{y}^{(1)}] \quad (29)
$$

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

$$
\begin{aligned}
\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] &= \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] \\
&= \mathrm{cov}[\boldsymbol{Y}_t, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{cov}[\boldsymbol{Y}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[-\mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{cov}[-\mathbf{Z}_t\boldsymbol{X}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] \quad (30)\\
&= \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top
\end{aligned}
$$

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 $\boldsymbol{y}^{(1)}$. Thus

$$
\mathrm{E}_{Y^{(1)}}[\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] = \mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] = \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \quad (31)
$$

$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t]$ is the negative of this (Equation 28), thus

$$
\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = -\widetilde{\mathbf{S}}_t + \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \quad (32)
$$

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

$$\text{cov}_{Y^{(1)}}[\,\text{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}],\,\text{E}[\boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]\,] = -\,\text{E}_{Y^{(1)}}[\,\text{cov}[\boldsymbol{V}_t,\boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]\,] \tag{33}$$

and

$$
\begin{aligned}
\text{cov}[\boldsymbol{V}_t,\boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}] &= \text{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t - \mathbf{a}_t, \boldsymbol{X}_{t+1} - \mathbf{B}_{t+1}\boldsymbol{X}_t - \mathbf{u}_{t+1}|\boldsymbol{y}^{(1)}] \\
&= \text{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_{t+1} - \mathbf{B}_{t+1}\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] \\
&= \widetilde{\mathbf{S}}_{t,t+1} - \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top
\end{aligned}
\tag{34}
$$

Thus,

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

2.4   Joint distribution of the conditional residuals

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

$$\hat{\varepsilon}_t = \begin{bmatrix} \hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{w}}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_t - \mathbf{Z}_t\widetilde{\boldsymbol{x}}_t - \mathbf{a}_t \\ \widetilde{\boldsymbol{x}}_t - \mathbf{B}_t\widetilde{\boldsymbol{x}}_{t-1} - \mathbf{u}_t \end{bmatrix}. \tag{35}$$

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

$$\left[\begin{array}{c|c} \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top & \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \\ \hline (\widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t)^\top & \mathbf{Q}_t - \widetilde{\boldsymbol{V}}_t - \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1}\mathbf{B}_t^\top + \widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1,t} \end{array} \right] \tag{36}$$

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

$$\hat{\varepsilon}_t = \begin{bmatrix} \hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{w}}_{t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_t - \mathbf{Z}_t\widetilde{\boldsymbol{x}}_t - \mathbf{a}_t \\ \widetilde{\boldsymbol{x}}_{t+1} - \mathbf{B}_{t+1}\widetilde{\boldsymbol{x}}_t - \mathbf{u}_{t+1} \end{bmatrix} \tag{37}$$

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

$$\left[\begin{array}{c|c} \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top & -\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top \\ \hline (-\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top)^\top & \mathbf{Q}_{t+1} - \widetilde{\boldsymbol{V}}_{t+1} - \mathbf{B}_{t+1}\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top + \widetilde{\boldsymbol{V}}_{t+1,t}\mathbf{B}_{t+1}^\top + \mathbf{B}_{t+1}\widetilde{\boldsymbol{V}}_{t,t+1} \end{array} \right] \tag{38}$$

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:

$$
\begin{aligned}
\boldsymbol{x}_{t+1} &= \mathbf{B}_{t+1}\boldsymbol{x}_t + \mathbf{u}_{t+1} + \mathbf{G}_{t+1}\epsilon_t, \\
\boldsymbol{y}_t &= \mathbf{Z}_t\boldsymbol{x}_t + \mathbf{a}_t + \mathbf{H}_t\epsilon_t,
\end{aligned}
\tag{39}
$$
$$\text{where } \epsilon_t \sim \text{MVN}(0, \mathbf{I}_{m+n \times m+n}), \mathbf{G}_t\mathbf{G}_t^\top = \mathbf{Q}_t \text{ and } \mathbf{H}_t\mathbf{H}_t^\top = \mathbf{R}_t$$

$\mathbf{G}_t$ 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 output from the Kalman filter: the one-step ahead prediction covariance $\mathbf{F}_t$ and the Kalman gain $\mathbf{K}_t$. Starting from $t = T$ and working backwards to $t = 1$ and using $r_T = 0$ and $N_T = 0$, the algorithm is

$$\mathbf{Q}_{t+1}^* = \begin{bmatrix} \mathbf{Q}_{t+1} & 0_{m \times n} \end{bmatrix}, \ \mathbf{R}_t^* = \begin{bmatrix} 0_{n \times m} & \mathbf{R}_t^* \end{bmatrix}$$

$$\mathbf{F}_t = \mathbf{Z}_t^* \widetilde{V}_t \mathbf{Z}_t^{*\top} + \mathbf{R}_t^*, \ K_t = \mathbf{B}_{t+1} \mathbf{K}_t$$

$$L_t = \mathbf{B}_{t+1} - K_t \mathbf{Z}_t^*, \ J_t = \mathbf{Q}_{t+1}^* - K_t \mathbf{R}_t^*, \ u_t = \mathbf{F}_t^{-1} - K_t^\top r_t \tag{40}$$

$$r_{t-1} = \mathbf{Z}_t^{*\top} u_t + \mathbf{B}_{t+1}^\top r_t, \ N_{t-1} = K_t^\top N_t K_t + L_t^\top N_t L_t$$

Bolded terms are the same as in Equation 39. 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): 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; for example 0 on diagonal replaced with 1, inverse taken, and 1 on diagonal replaced back with 0.

The residuals are

$$\hat{\varepsilon}_t^* = \begin{bmatrix} \hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{w}}_{t+1} \end{bmatrix} = (\mathbf{R}_t^*)^\top u_t + (\mathbf{Q}_{t+1}^*)^\top r_t \tag{41}$$

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

$$\Sigma_t^* = \mathrm{var}_{Y^{(1)}}(\hat{\varepsilon}_t) = \mathbf{R}_t^{*\top} \mathbf{F}_t^{-1} \mathbf{R}_t^* + J_t^\top N_t J_t \tag{42}$$

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

If you compare their state equation (their equation 20) with my state equation, you will notice that my time indexing on $\mathbf{B}$ matches the left $\boldsymbol{x}$ while in theirs, it matches the right $\boldsymbol{x}$. Thus $\mathbf{B}_{t+1}$ (and $\mathbf{Q}_{t+1}$) appears in my implementation of their algorithm instead of $\mathbf{B}_t$. Harvey et al. (1998, eqns. 19, 20) use G_t to refer to the $\mathrm{chol}(\mathbf{R}_t)^\top$ (essentially) and H_t to refer to $\mathrm{chol}(\mathbf{Q}_t)^\top$. I've replaced these with $\mathbf{R}_t^*$ and $\mathbf{Q}_t^*$, respectively, which causes my variant of their algorithm to give the 'non-normalized' variance of the residuals. Their T_t is my $\mathbf{B}_{t+1}$. $\mathbf{K}_t$ is the Kalman gain output by the MARSS package. The Kalman gain as used in the Harvey et al. (1998) algorithm is $K_t = \mathbf{B}_{t+1} \mathbf{K}_t$.

3.1 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":

$$(\Sigma_t^*)^{-1/2} \hat{\varepsilon}_t^* \tag{43}$$

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.

4 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'.

4.0.1 Variance of the one-step-ahead model residuals

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

$$\hat{\boldsymbol{v}}_t = \boldsymbol{y}_t - \mathbf{Z}_t \widetilde{\boldsymbol{x}}_t - \mathbf{a}_t, \tag{44}$$

where $\widetilde{x}_t$ is now $\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}]$ and is output by the Kalman filter. The difference between the smoothation (Equation 3) 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:

$$\mathrm{var}[\hat{\boldsymbol{V}}_t] = \mathrm{var}_{Y_{1:t-1}}[\boldsymbol{Y}_t - \boldsymbol{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}] - \mathbf{a}_t] \tag{45}$$

While this looks similar to the smoothation case (Equation 4), it is different in that $\boldsymbol{y}_t$ does not appear in $\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}]$; the conditioning is only on data up to $t - 1$. Looking at Equation 45, you might guess that $\mathrm{var}[\hat{\boldsymbol{V}}_t] = \mathbf{R}_t - \boldsymbol{Z}_t\,\mathrm{var}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}]]\boldsymbol{Z}_t^\top$ 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":

$$\mathrm{var}_A[A] = \mathrm{var}_B[\,\mathrm{E}_{A|b}[A|b]] + \mathrm{E}_B[\,\mathrm{var}_{A|b}[A|b]] \tag{46}$$

which again I write more succinctly as

$$\mathrm{var}_A[A] = \mathrm{var}_B[\,\mathrm{E}[A|b]] + \mathrm{E}_B[\,\mathrm{var}[A|b]] \tag{47}$$

From the law of total variance , we can write

$$\mathrm{var}_{V_t}[\hat{\boldsymbol{V}}_t] = \mathrm{var}_{Y_{1:t-1}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}_{1:t-1}^{(1)}]] + \mathrm{E}_{Y_{1:t-1}}[\,\mathrm{var}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}_{1:t-1}^{(1)}]] \tag{48}$$

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

4.0.2   First term in Equation 48

Because $\boldsymbol{y}_t$ does not appear in the conditional of $\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}_{1:t-1}^{(1)}]$, 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 44), we can write

$$\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}_{1:t-1}^{(1)}] = \mathrm{E}[\boldsymbol{Y}_t|\boldsymbol{y}_{1:t-1}^{(1)}] - \boldsymbol{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}] - \mathbf{a}_t \tag{49}$$

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

$$\mathrm{E}[\boldsymbol{Y}_t|\boldsymbol{y}_{1:t-1}^{(1)}] = \boldsymbol{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}] + \mathbf{a}_t \tag{50}$$

So the right side of Equation 49 is 0 and the first term in Equation 48 is 0:

$$\mathrm{var}_{Y_{1:t-1}}[\,\mathrm{E}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}_{1:t-1}^{(1)}]] = 0 \tag{51}$$

4.0.3   Second term in Equation 48

This term is

$$\mathrm{E}_{Y_{1:t-1}}[\,\mathrm{var}[\hat{\boldsymbol{V}}_t|\boldsymbol{y}_{1:t-1}^{(1)}] = \mathrm{E}_{Y_{1:t-1}}[\,\mathrm{var}[\boldsymbol{Y}_t - \boldsymbol{Z}_t\,\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}] - \mathbf{a}_t|\boldsymbol{y}_{1:t-1}^{(1)}]] \tag{52}$$

$\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}]$ 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 $\mathrm{E}_{Y_{1:t-1}}[\,\mathrm{var}[\boldsymbol{Y}_t|\boldsymbol{y}_{1:t-1}^{(1)}]]$. The conditional variance of $\boldsymbol{Y}_t$ is from Equation 1

$$\mathrm{var}[\boldsymbol{Y}_t|\boldsymbol{y}_{1:t-1}^{(1)}] = \mathrm{var}[\boldsymbol{Z}_t\boldsymbol{X}_t + \mathbf{a}_t + \boldsymbol{V}_t|\boldsymbol{y}_{1:t-1}^{(1)}] = \boldsymbol{Z}_t\,\mathrm{var}[\boldsymbol{X}_t|\boldsymbol{y}_{1:t-1}^{(1)}]\boldsymbol{Z}_t^\top + \mathrm{var}[\boldsymbol{V}_t|\boldsymbol{y}_{1:t-1}^{(1)}]. \tag{53}$$

$\widetilde{\mathbf{U}}_t$ is not a function of $\boldsymbol{y}$ is is only a function of the MARSS parameters. Thus the second term in Equation 7 is simply $\widetilde{\mathbf{U}}_t$.

4.0.4 Putting together the two terms

$$\text{var}[\hat{\boldsymbol{V}}_t] = \mathbf{R}_t - \mathbf{Z}_t \widetilde{\boldsymbol{V}}_t \mathbf{Z}_t^\top - \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{S}}_t \mathbf{Z}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{S}}_t^\top + \widetilde{\mathbf{U}}_t$$
$$= \mathbf{R}_t - \mathbf{Z}_t \widetilde{\boldsymbol{V}}_t \mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t \mathbf{Z}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{S}}_t^\top \tag{54}$$

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.

4.1 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 $\text{E}(\boldsymbol{X}_t - (\mathbf{B}_t \boldsymbol{X}_{t-1} + \mathbf{u}_t)) = \text{E}(\boldsymbol{W}_t) = 0$ and the unconditional variance of the state residuals is

$$\text{var}[\boldsymbol{X}_t - (\mathbf{B}_t \boldsymbol{X}_{t-1} + \mathbf{u}_t)] = \text{var}[\boldsymbol{W}_t] = \mathbf{Q}_t \tag{55}$$

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

$$\hat{\boldsymbol{w}}_t = \widetilde{\boldsymbol{x}}_t - \mathbf{B}_t \widetilde{\boldsymbol{x}}_{t-1} - \mathbf{u}_t. \tag{56}$$

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 $\text{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t]$ from the law of total variance using the observation that $\hat{\boldsymbol{w}}_t = \text{E}[\boldsymbol{W}_t | \boldsymbol{y}^{(1)}]$.

$$\text{var}[\boldsymbol{W}_t] = \text{var}_{Y^{(1)}}[\text{E}[\boldsymbol{W}_t | \boldsymbol{y}^{(1)}]] + \text{E}_{Y^{(1)}}[\text{var}[\boldsymbol{W}_t | \boldsymbol{y}^{(1)}]] \tag{57}$$

Thus,

$$\text{var}_{Y^{(1)}}[\hat{\boldsymbol{w}}_t] = \text{var}_{Y^{(1)}}[\text{E}[\boldsymbol{W}_t | \boldsymbol{y}^{(1)}]] = \text{var}[\boldsymbol{W}_t] - \text{E}_{Y^{(1)}}[\text{var}[\boldsymbol{W}_t | \boldsymbol{y}^{(1)}]] \tag{58}$$

The variance in the expectation on the far right is

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

$\mathbf{u}$ is not a random variable and can be dropped

$$= \text{var}[\boldsymbol{X}_t - \mathbf{B}_t \boldsymbol{X}_{t-1} | \boldsymbol{y}^{(1)}]$$
$$= \text{var}[\boldsymbol{X}_t | \boldsymbol{y}^{(1)}] + \text{var}[\mathbf{B}_t \boldsymbol{X}_{t-1} | \boldsymbol{y}^{(1)}] + \text{cov}[\boldsymbol{X}_t, -\mathbf{B}_t \boldsymbol{X}_{t-1} | \boldsymbol{y}^{(1)}] + \text{cov}[-\mathbf{B}_t \boldsymbol{X}_{t-1}, \boldsymbol{X}_t | \boldsymbol{y}^{(1)}] \tag{59}$$
$$= \widetilde{\boldsymbol{V}}_t + \mathbf{B}_t \widetilde{\boldsymbol{V}}_{t-1} \mathbf{B}_t^\top - \widetilde{\boldsymbol{V}}_{t,t-1} \mathbf{B}_t^\top - \mathbf{B}_t \widetilde{\boldsymbol{V}}_{t-1,t}$$

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 $\text{var}[\boldsymbol{W}_t] = \mathbf{Q}_t$ in Equation 20, the variance of the conditional state residuals is

$$\text{var}_{Y^{(1)}}[\hat{\boldsymbol{W}}_t] = \mathbf{Q}_t - \widetilde{\boldsymbol{V}}_t - \mathbf{B}_t \widetilde{\boldsymbol{V}}_{t-1} \mathbf{B}_t^\top + \widetilde{\boldsymbol{V}}_{t,t-1} \mathbf{B}_t^\top + \mathbf{B}_t \widetilde{\boldsymbol{V}}_{t-1,t} \tag{60}$$

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

$$\text{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = \text{cov}_{Y^{(1)}}[\text{E}[\hat{\boldsymbol{V}}_t | \boldsymbol{y}^{(1)}], \text{E}[\hat{\boldsymbol{W}}_t | \boldsymbol{y}^{(1)}]] + \text{E}_{Y^{(1)}}[\text{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t | \boldsymbol{y}^{(1)}]] \tag{61}$$

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

$$\text{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t | \boldsymbol{y}^{(1)}] = \text{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t \text{E}[\boldsymbol{X}_t | \boldsymbol{y}^{(1)}] - \mathbf{a}_t, \text{E}[\boldsymbol{X}_t | \boldsymbol{y}^{(1)}] - \mathbf{B}_t \text{E}[\boldsymbol{X}_{t-1} | \boldsymbol{y}^{(1)}] - \mathbf{u}_t | \boldsymbol{y}^{(1)}] \tag{62}$$

The $\text{E}[\boldsymbol{X}_t | \boldsymbol{y}^{(1)}]$, $\text{E}[\boldsymbol{X}_{t-1} | \boldsymbol{y}^{(1)}]$ 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 $\text{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t | \boldsymbol{y}^{(1)}]$ is 0. Thus Equation 61 reduces to

$$\text{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = \text{cov}_{Y^{(1)}}[\text{E}[\hat{\boldsymbol{V}}_t | \boldsymbol{y}^{(1)}], \text{E}[\hat{\boldsymbol{W}}_t | \boldsymbol{y}^{(1)}]] + 0 = \text{cov}_{Y^{(1)}}[\text{E}[\boldsymbol{V}_t | \boldsymbol{y}^{(1)}], \text{E}[\boldsymbol{W}_t | \boldsymbol{y}^{(1)}]] \tag{63}$$

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

$$\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t] = \mathrm{E}_{Y^{(1)}}[\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] + \mathrm{cov}_{Y^{(1)}}[\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}], \mathrm{E}[\boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{64}$$

The unconditional covariance of $\boldsymbol{V}_t$ and $\boldsymbol{W}_t$ is 0. Thus the right side of Equation 27 is 0 and combining Equation 26 and 27,

$$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = -\mathrm{E}_{Y^{(1)}}[\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] \tag{65}$$

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

$$\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] = \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t - \mathbf{a}_t, \boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1} - \mathbf{u}_t|\boldsymbol{y}^{(1)}] \tag{66}$$

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

$$\begin{aligned}
\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] &= \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}y] \\
&= \mathrm{cov}[\boldsymbol{Y}_t, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{cov}[\boldsymbol{Y}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[-\mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_t] + \mathrm{cov}[-\mathbf{Z}_t\boldsymbol{X}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}] \\
&= \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top
\end{aligned} \tag{67}$$

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

$$\mathrm{E}_{Y^{(1)}}[\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}]] = \mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_t|\boldsymbol{y}^{(1)}] = \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \tag{68}$$

$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t]$ is the negative of this (Equation 28), thus

$$\mathrm{cov}[\hat{\boldsymbol{V}}_t, \hat{\boldsymbol{W}}_t] = -\widetilde{\mathbf{S}}_t + \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \tag{69}$$

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

$$\mathrm{cov}_{Y^{(1)}}[\mathrm{E}[\boldsymbol{V}_t|\boldsymbol{y}^{(1)}], \mathrm{E}[\boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]] = -\mathrm{E}_{Y^{(1)}}[\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}]] \tag{70}$$

and

$$\begin{aligned}
\mathrm{cov}[\boldsymbol{V}_t, \boldsymbol{W}_{t+1}|\boldsymbol{y}^{(1)}] &= \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t - \mathbf{a}_t, \boldsymbol{X}_{t+1} - \mathbf{B}_{t+1}\boldsymbol{X}_t - \mathbf{u}_{t+1}|\boldsymbol{y}^{(1)}] \\
&= \mathrm{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{X}_{t+1} - \mathbf{B}_{t+1}\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] \\
&= \widetilde{\mathbf{S}}_{t,t+1} - \widetilde{\mathbf{S}}_t\mathbf{B}_{t+1}^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t+1} + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{B}_{t+1}^\top
\end{aligned} \tag{71}$$

Thus,

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

4.3 Joint distribution of the conditional residuals

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

$$\hat{\varepsilon}_t = \begin{bmatrix} \hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{w}}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_t - \mathbf{Z}_t\widetilde{\boldsymbol{x}}_t - \mathbf{a}_t \\ \widetilde{\boldsymbol{x}}_t - \mathbf{B}_t\widetilde{\boldsymbol{x}}_{t-1} - \mathbf{u}_t \end{bmatrix}. \tag{72}$$

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

$$\begin{bmatrix} \mathbf{R}_t - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top & \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t \\ (\widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t\widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t\widetilde{\boldsymbol{V}}_t)^\top & \mathbf{Q}_t - \widetilde{\boldsymbol{V}}_t - \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1}\mathbf{B}_t^\top + \widetilde{\boldsymbol{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{B}_t\widetilde{\boldsymbol{V}}_{t-1,t} \end{bmatrix} \tag{73}$$

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

$$\hat{\varepsilon}_t = \begin{bmatrix} \hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{w}}_{t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_t - \mathbf{Z}_t \widetilde{\boldsymbol{x}}_t - \mathbf{a}_t \\ \widetilde{\boldsymbol{x}}_{t+1} - \mathbf{B}_{t+1} \widetilde{\boldsymbol{x}}_t - \mathbf{u}_{t+1} \end{bmatrix} \tag{74}$$

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

$$\left[\begin{array}{c|c} \mathbf{R}_t - \mathbf{Z}_t \widetilde{\boldsymbol{V}}_t \mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t \mathbf{Z}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{S}}_t^\top & -\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t \mathbf{B}_{t+1}^\top + \mathbf{Z}_t \widetilde{\boldsymbol{V}}_{t,t+1} - \mathbf{Z}_t \widetilde{\boldsymbol{V}}_t \mathbf{B}_{t+1}^\top \\ \hline (-\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t \mathbf{B}_{t+1}^\top + \mathbf{Z}_t \widetilde{\boldsymbol{V}}_{t,t+1} - \mathbf{Z}_t \widetilde{\boldsymbol{V}}_t \mathbf{B}_{t+1}^\top)^\top & \mathbf{Q}_{t+1} - \widetilde{\boldsymbol{V}}_{t+1} - \mathbf{B}_{t+1} \widetilde{\boldsymbol{V}}_t \mathbf{B}_{t+1}^\top + \widetilde{\boldsymbol{V}}_{t+1,t} \mathbf{B}_{t+1}^\top + \mathbf{B}_{t+1} \widetilde{\boldsymbol{V}}_{t,t+1} \end{array} \right] \tag{75}$$

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 44 directly as:

$$\text{var}[\hat{\boldsymbol{V}}_t] = \text{var}_{Y^{(1)}}[\boldsymbol{Y}_t] - \mathbf{Z}_t \, \text{E}[\boldsymbol{X}_t | \boldsymbol{y}^{(1,1:t-1)}] \tag{76}$$

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

$$\begin{aligned} \text{var}[\hat{\boldsymbol{V}}_t] &= \mathbf{R}_t - \mathbf{Z}_t \widetilde{\boldsymbol{V}}_t \mathbf{Z}_t^\top - \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{S}}_t \mathbf{Z}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{S}}_t^\top + \widetilde{\mathbf{U}}_t \\ &= \mathbf{R}_t - \mathbf{Z}_t \widetilde{\boldsymbol{V}}_t \mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t \mathbf{Z}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{S}}_t^\top \end{aligned} \tag{77}$$

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.