Computation of Standardized Residuals for MARSS Models

E. E. Holmes[*]

November 1, 2018

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

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

Given a set of observed data $\boldsymbol{y}_t$ and states $\boldsymbol{x}_t$, the model residuals are $\boldsymbol{y}_t - (\mathbf{Z}_t\boldsymbol{x}_t + \mathbf{a}_t) = \mathbf{v}_t$. 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

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

based on the definition of $\mathbf{V}_t$.

Once we have data, $\mathbf{R}_t$ is not the variance of our model residuals because our residuals are now conditioned 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).

1 Distribution of the MARSS conditional residuals

This report 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 later, 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[2].

1.1 Model residuals conditioned on all the data

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

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

where $\widetilde{\mathbf{x}}_t$ 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; the unobserved $\boldsymbol{y}$ will be termed $\boldsymbol{y}^{(2)}$. $\hat{\mathbf{v}}_t$ is sample from the random variable $\hat{\mathbf{V}}_t$ since $\boldsymbol{Y}^{(1)}$ is a random

---

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

[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]in a frequentist framework

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 $\hat{\mathbf{V}}_t$; unconditional here means we take the expectations over all possibles values that $\boldsymbol{y}$, both $\boldsymbol{y}^{(1)}$ and $\boldsymbol{y}^{(1)}$, might take. The mean is 0 and we are concerned only with computing the variance:

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

Notice we have an unconditional variance over $\boldsymbol{Y}$ on the outside and a conditional variance over a specific value of $\boldsymbol{y}^{(1)}$ on the inside.

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

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

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

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

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

From the law of total variance , we can write

$$\text{var}[\hat{\mathbf{V}}_t] = \text{var}_Y[\,\text{E}[\hat{\mathbf{V}}_t|\boldsymbol{y}^{(1)}]\,] + \text{E}_Y[\,\text{var}[\hat{\mathbf{V}}_t|\boldsymbol{y}^{(1)}]\,] \tag{7}$$

var_Y and E_Y are expectations over both $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$, so all possible values of $\boldsymbol{Y}$.

1.1.1 First term in Equation 7

Notice that $\text{E}[\hat{\mathbf{V}}_t|\boldsymbol{y}^{(1)}] = \text{E}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}] - \mathbf{Z}_t\,\text{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] - \mathbf{a}_t = \text{E}[\mathbf{V}_t|\boldsymbol{y}^{(1)}]$. So the first term is $\text{var}_Y[\,\text{E}[\mathbf{V}_t|\boldsymbol{y}^{(1)}]\,]$. From the law of total variance, we can write

$$\text{var}[\mathbf{V}_t] = \text{var}_Y[\,\text{E}[\mathbf{V}_t|\boldsymbol{y}^{(1)}]\,] + \text{E}_Y[\,\text{var}[\mathbf{V}_t|\boldsymbol{y}^{(1)}]\,] \tag{8}$$

From Equation 8, we can solve for $\text{var}_Y[\,\text{E}[\mathbf{V}_t|\boldsymbol{y}^{(1)}]\,]$:

$$\text{var}_Y[\,\text{E}[\mathbf{V}_t|\boldsymbol{y}^{(1)}]\,] = \text{var}[\mathbf{V}_t] - \text{E}_Y[\,\text{var}[\mathbf{V}_t|\boldsymbol{y}^{(1)}]\,] \tag{9}$$

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

$$\text{var}[\mathbf{V}_t|\boldsymbol{y}^{(1)}] = \text{var}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t - \mathbf{a}_t|\boldsymbol{y}^{(1)}]. \tag{10}$$

where $\mathbf{a}_t$ is a fixed value and can be dropped. Equation 10 can then be written as

$$\begin{aligned}
\text{var}[\mathbf{V}_t|\boldsymbol{y}^{(1)}] &= \text{var}[\boldsymbol{Y}_t - \mathbf{Z}_t\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] \\
&= \text{var}[-\mathbf{Z}_t\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \text{var}[\boldsymbol{Y}_t|\boldsymbol{y}^{(1)}] + \text{cov}[\boldsymbol{Y}_t, -\mathbf{Z}_t\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \text{cov}[-\mathbf{Z}_t\boldsymbol{X}_t, \boldsymbol{Y}_t|\boldsymbol{y}^{(1)}] \\
&= \mathbf{Z}_t\widetilde{\mathbf{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{U}}_t - \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top - \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top
\end{aligned} \tag{11}$$

$\widetilde{\mathbf{V}}_t = \text{var}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ 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.

$\widetilde{\mathbf{V}}_t$, $\widetilde{\mathbf{U}}_t$ and $\widetilde{\mathbf{S}}_t$ do not depend on the actual values of $\boldsymbol{y}$; the conditional is merely indicating that we are concerned with the variances of the conditional values. They depend instead on the parameters values, $\mathbf{Q}$, $\mathbf{B}$, $\mathbf{R}$, etc., in the MARSS equation. Thus $\text{E}_Y[\,\text{var}[\mathbf{V}_t|\boldsymbol{y}^{(1)}]\,] = \text{var}[\mathbf{V}_t|\boldsymbol{y}^{(1)}]$ and

$$\text{var}[\hat{\mathbf{V}}_t] = \text{var}[\mathbf{V}_t] - \text{var}[\mathbf{V}_t|\boldsymbol{y}^{(1)}] = \mathbf{R}_t - \mathbf{Z}_t\widetilde{\mathbf{V}}_t\mathbf{Z}_t^\top - \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top \tag{12}$$

1.1.2 Second term in Equation 7

Consider the second term in Equation 7. This term is

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

$\mathrm{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ is a fixed value; it is not $\boldsymbol{X}_t$ but its expected value. Thus the second term reduces to $\mathrm{E}_Y[\, \mathrm{var}[\boldsymbol{Y}|\boldsymbol{y}^{(1)}]\,] = \mathrm{E}_Y[\widetilde{\mathbf{U}}_t]$. $= \widetilde{\mathbf{U}}_t$ 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 $\widetilde{\mathbf{U}}_t$.

1.1.3 Putting together the two terms

$$
\begin{aligned}
\mathrm{var}[\hat{\mathbf{V}}_t] &= \mathbf{R}_t - \mathbf{Z}_t\widetilde{\mathbf{V}}_t\mathbf{Z}_t^\top - \widetilde{\mathbf{U}}_t + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top + \widetilde{\mathbf{U}}_t \\
&= \mathbf{R}_t - \mathbf{Z}_t\widetilde{\mathbf{V}}_t\mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t\mathbf{Z}_t^\top + \mathbf{Z}_t\widetilde{\mathbf{S}}_t^\top
\end{aligned}
\tag{13}
$$

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

1.2 State residuals conditioned on the data

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

$$\mathrm{var}[\boldsymbol{X}_t - (\mathbf{B}_t\boldsymbol{X}_{t-1} + \mathbf{u}_t)] = \mathrm{var}[\mathbf{W}_t] = \mathbf{Q}_t \tag{14}$$

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

$$\hat{\mathbf{w}}_t = \widetilde{\mathbf{x}}_t - \mathbf{B}_t\widetilde{\mathbf{x}}_{t-1} - \mathbf{u}_t. \tag{15}$$

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

$$\mathrm{var}[\mathbf{W}_t] = \mathrm{var}_Y[\, \mathrm{E}[\mathbf{W}_t|\boldsymbol{y}^{(1)}]\,] + \mathrm{E}_Y[\, \mathrm{var}[\mathbf{W}_t|\boldsymbol{y}^{(1)}]\,] \tag{16}$$

Thus,

$$\mathrm{var}_Y[\hat{\mathbf{w}}_t] = \mathrm{var}_Y[\, \mathrm{E}[\mathbf{W}_t|\boldsymbol{y}^{(1)}]\,] = \mathrm{var}(\mathbf{W}_t) - \mathrm{E}_Y[\, \mathrm{var}[\mathbf{W}_t|\boldsymbol{y}^{(1)}]\,] \tag{17}$$

The variance in the expectation on the far right is

$$
\begin{aligned}
\mathrm{var}[\mathbf{W}_t|\boldsymbol{y}^{(1)}] &= \mathrm{var}[\boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1} - \mathbf{u}_t|\boldsymbol{y}^{(1)}] \\
&\quad \mathbf{u} \text{ is not a random variable and can be dropped} \\
&= \mathrm{var}[\boldsymbol{X}_t - \mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] \\
&= \mathrm{var}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \mathrm{var}[\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[\boldsymbol{X}_t, -\mathbf{B}_t\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \mathrm{cov}[-\mathbf{B}_t\boldsymbol{X}_{t-1}, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}] \\
&= \widetilde{\mathbf{V}}_t + \mathbf{B}_t\widetilde{\mathbf{V}}_{t-1}\mathbf{B}_t^\top - \widetilde{\mathbf{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{B}_t\widetilde{\mathbf{V}}_{t-1,t}
\end{aligned}
\tag{18}
$$

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 $\mathrm{var}[\mathbf{W}_t] = \mathbf{Q}_t$ in Equation 17, the variance of the conditional state residuals is

$$\mathrm{var}_Y[\hat{\mathbf{W}}_t] = \mathbf{Q}_t - \widetilde{\mathbf{V}}_t - \mathbf{B}_t\widetilde{\mathbf{V}}_{t-1}\mathbf{B}_t^\top + \widetilde{\mathbf{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{B}_t\widetilde{\mathbf{V}}_{t-1,t} \tag{19}$$

1.3 Covariance of the conditional model and state residuals

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

$$\mathrm{cov}[\hat{\mathbf{V}}_t, \hat{\mathbf{W}}_t] = \mathrm{cov}_Y[\, \mathrm{E}[\hat{\mathbf{V}}_t|\boldsymbol{y}^{(1)}], \mathrm{E}[\hat{\mathbf{W}}_t|\boldsymbol{y}^{(1)}]\,] + \mathrm{E}_Y[\, \mathrm{cov}[\hat{\mathbf{V}}_t, \hat{\mathbf{W}}_t|\boldsymbol{y}^{(1)}]\,] \tag{20}$$

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

$$\text{cov}[\hat{\mathbf{V}}_t, \hat{\mathbf{W}}_t | \boldsymbol{y}^{(1)}] = \text{E}_Y[\text{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t \text{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}] - \mathbf{a}_t, \text{E}[\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] - \mathbf{B}_t \text{E}[\boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] - \mathbf{u}_t|\boldsymbol{y}^{(1)}]] \tag{21}$$

The $\text{E}[\boldsymbol{X}_t|\boldsymbol{y}^{(1)}]$ are fixed values for a given set of data. The covariance of a random variable with a fixed value is 0, thus $\text{cov}[\hat{\mathbf{V}}_t, \hat{\mathbf{W}}_t | \boldsymbol{y}^{(1)}]$ is 0. Thus Equation 20 reduces to

$$\text{cov}[\hat{\mathbf{V}}_t, \hat{\mathbf{W}}_t] = \text{cov}_Y[\text{E}[\hat{\mathbf{V}}_t|\boldsymbol{y}^{(1)}], \text{E}[\hat{\mathbf{W}}_t|\boldsymbol{y}^{(1)}]] + 0 = \text{cov}_Y[\text{E}[\mathbf{V}_t|\boldsymbol{y}^{(1)}], \text{E}[\mathbf{W}_t|\boldsymbol{y}^{(1)}]] \tag{22}$$

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

$$\text{cov}[\mathbf{V}_t, \mathbf{W}_t] = \text{E}_Y[\text{cov}[\mathbf{V}_t, \mathbf{W}_t|\boldsymbol{y}^{(1)}]] + \text{cov}_Y[\text{E}[\mathbf{V}_t|\boldsymbol{y}^{(1)}], \text{E}[\mathbf{W}_t|\boldsymbol{y}^{(1)}]] \tag{23}$$

The unconditional covariance of $\mathbf{V}_t$ and $\mathbf{W}_t$ is 0. Thus the right side of Equation 23 is 0 and combining Equation 22 and 23,

$$\text{cov}[\hat{\mathbf{V}}_t, \hat{\mathbf{W}}_t] = -\text{E}_Y[\text{cov}[\mathbf{V}_t, \mathbf{W}_t|\boldsymbol{y}^{(1)}]] \tag{24}$$

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

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

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

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

$$\begin{aligned}
\text{cov}[\mathbf{V}_t, \mathbf{W}_t|\boldsymbol{y}^{(1)}] &= \text{cov}[\boldsymbol{Y}_t - \mathbf{Z}_t \boldsymbol{X}_t, \boldsymbol{X}_t - \mathbf{B}_t \boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}y] \\
&= \text{cov}[\boldsymbol{Y}_t, \boldsymbol{X}_t|\boldsymbol{y}^{(1)}] + \text{cov}[\boldsymbol{Y}_t, -\mathbf{B}_t \boldsymbol{X}_{t-1}|\boldsymbol{y}^{(1)}] + \text{cov}[-\mathbf{Z}_t \boldsymbol{X}_t, \boldsymbol{X}_t] + \text{cov}[-\mathbf{Z}_t \boldsymbol{X}_t, -\mathbf{B}_t \boldsymbol{X}_{t-1}] \\
&= \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t \widetilde{\mathbf{V}}_t + \mathbf{Z}_t \widetilde{\mathbf{V}}_{t,t-1}\mathbf{B}_t^\top
\end{aligned} \tag{26}$$

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

$$\text{E}_Y[\text{cov}[\mathbf{V}_t, \mathbf{W}_t|\boldsymbol{y}^{(1)}]] = \text{cov}[\mathbf{V}_t, \mathbf{W}_t|\boldsymbol{y}^{(1)}] = \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{V}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t \widetilde{\mathbf{V}}_t \tag{27}$$

$\text{cov}[\hat{\mathbf{V}}_t, \hat{\mathbf{W}}_t]$ is the negative of this (Equation 24), thus

$$\text{cov}[\hat{\mathbf{V}}_t, \hat{\mathbf{W}}_t] = -\widetilde{\mathbf{S}}_t + \widetilde{\mathbf{S}}_{t,t-1}\mathbf{B}_t^\top - \mathbf{Z}_t \widetilde{\mathbf{V}}_{t,t-1}\mathbf{B}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{V}}_t \tag{28}$$

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

$$\text{cov}_Y[\text{E}[\mathbf{V}_t|\boldsymbol{y}^{(1)}], \text{E}[\mathbf{W}_{t+1}|\boldsymbol{y}^{(1)}]] = -\text{E}_Y[\text{cov}[\mathbf{V}_t, \mathbf{W}_{t+1}|\boldsymbol{y}^{(1)}]] \tag{29}$$

and

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

Thus,

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

1.4 Joint distribution of the conditional residuals

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

$$\hat{\varepsilon}_t = \begin{bmatrix} \hat{\mathbf{v}}_t \\ \hat{\mathbf{w}}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_t - \mathbf{Z}_t \widetilde{\mathbf{x}}_t - \mathbf{a}_t \\ \widetilde{\mathbf{x}}_t - \mathbf{B}_t \widetilde{\mathbf{x}}_{t-1} - \mathbf{u}_t \end{bmatrix}. \tag{31}$$

where $\widetilde{\mathbf{x}}_t$ and $\widetilde{\mathbf{x}}_{t-1}$ 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

$$\left[\begin{array}{c|c} \mathbf{R}_t - \mathbf{Z}_t \widetilde{\mathbf{V}}_t \mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t \mathbf{Z}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{S}}_t^\top & \widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1} \mathbf{B}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{V}}_{t,t-1} \mathbf{B}_t^\top - \mathbf{Z}_t \widetilde{\mathbf{V}}_t \\ \hline (\widetilde{\mathbf{S}}_t - \widetilde{\mathbf{S}}_{t,t-1} \mathbf{B}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{V}}_{t,t-1} \mathbf{B}_t^\top - \mathbf{Z}_t \widetilde{\mathbf{V}}_t)^\top & \mathbf{Q}_t - \widetilde{\mathbf{V}}_t - \mathbf{B}_t \widetilde{\mathbf{V}}_{t-1} \mathbf{B}_t^\top + \widetilde{\mathbf{V}}_{t,t-1} \mathbf{B}_t^\top + \mathbf{B}_t \widetilde{\mathbf{V}}_{t-1,t} \end{array} \right] \tag{32}$$

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

$$\hat{\varepsilon}_t = \begin{bmatrix} \hat{\mathbf{v}}_t \\ \hat{\mathbf{w}}_{t+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_t - \mathbf{Z}_t \widetilde{\mathbf{x}}_t - \mathbf{a}_t \\ \widetilde{\boldsymbol{x}}_{t+1} - \mathbf{B}_{t+1} \widetilde{\mathbf{x}}_t - \mathbf{u}_{t+1} \end{bmatrix} \tag{33}$$

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

$$\left[\begin{array}{c|c} \mathbf{R}_t - \mathbf{Z}_t \widetilde{\mathbf{V}}_t \mathbf{Z}_t^\top + \widetilde{\mathbf{S}}_t \mathbf{Z}_t^\top + \mathbf{Z}_t \widetilde{\mathbf{S}}_t^\top & -\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t \mathbf{B}_{t+1}^\top + \mathbf{Z}_t \widetilde{\mathbf{V}}_{t,t+1} - \mathbf{Z}_t \widetilde{\mathbf{V}}_t \mathbf{B}_{t+1}^\top \\ \hline (-\widetilde{\mathbf{S}}_{t,t+1} + \widetilde{\mathbf{S}}_t \mathbf{B}_{t+1}^\top + \mathbf{Z}_t \widetilde{\mathbf{V}}_{t,t+1} - \mathbf{Z}_t \widetilde{\mathbf{V}}_t \mathbf{B}_{t+1}^\top)^\top & \mathbf{Q}_{t+1} - \widetilde{\mathbf{V}}_{t+1} - \mathbf{B}_{t+1} \widetilde{\mathbf{V}}_t \mathbf{B}_{t+1}^\top + \widetilde{\mathbf{V}}_{t+1,t} \mathbf{B}_{t+1}^\top + \mathbf{B}_{t+1} \widetilde{\mathbf{V}}_{t,t+1} \end{array} \right] \tag{34}$$

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.

2 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:

$$\boldsymbol{x}_{t+1} = \mathbf{B}_{t+1} \boldsymbol{x}_t + \mathbf{u}_{t+1} + \mathbf{G}_{t+1} \epsilon_t,$$
$$\boldsymbol{y}_t = \mathbf{Z}_t \boldsymbol{x}_t + \mathbf{a}_t + \mathbf{H}_t \epsilon_t, \tag{35}$$
$$\text{where } \epsilon_t \sim \text{MVN}(0, \mathbf{I}_{m+n \times m+n}), \mathbf{G}_t \mathbf{G}_t^\top = \mathbf{Q}_t \text{ and } \mathbf{H}_t \mathbf{H}_t^\top = \mathbf{R}_t$$

$\mathbf{G}_t$ 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:

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

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

$$\mathbf{Q}_{t+1}^* = \begin{bmatrix} \mathbf{Q}_{t+1} & 0_{m \times n} \end{bmatrix}, \ \mathbf{R}_t^* = \begin{bmatrix} 0_{n \times m} & \mathbf{R}_t^* \end{bmatrix}$$
$$\mathbf{F}_t = \mathbf{Z}_t^* \widetilde{\mathbf{V}}_t \mathbf{Z}_t^{*\top} + \mathbf{R}_t^*, \ K_t = \mathbf{B}_{t+1} \mathbf{K}_t$$
$$L_t = \mathbf{B}_{t+1} - K_t \mathbf{Z}_t^*, \ J_t = \mathbf{Q}_{t+1}^* - K_t \mathbf{R}_t^*, \ u_t = \mathbf{F}_t^{-1} - K_t^\top r_t \tag{36}$$
$$r_{t-1} = \mathbf{Z}_t^{*\top} u_t + \mathbf{B}_{t+1}^\top r_t, \ N_{t-1} = K_t^\top N_t K_t + L_t^\top N_t L_t$$

Bolded terms are the same as in Equation 35. 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

$$\hat{\varepsilon}_t^* = \begin{bmatrix} \hat{\mathbf{v}}_t \\ \hat{\mathbf{w}}_{t+1} \end{bmatrix} = (\mathbf{R}_t^*)^\top u_t + (\mathbf{Q}_{t+1}^*)^\top r_t \tag{37}$$

with mean of 0 ($\mathrm{E}_Y(\hat{\varepsilon}_t) = 0$) and variance

$$\Sigma_t^* = \mathrm{var}_Y(\hat{\varepsilon}_t) = {\mathbf{R}_t^*}^\top \mathbf{F}_t^{-1} \mathbf{R}_t^* + J_t^\top N_t J_t \tag{38}$$

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 $K_t = \mathbf{B}_{t+1}\mathbf{K}_t$.

2.1 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":

$$(\Sigma_t^*)^{-1/2}\hat{\varepsilon}_t^* \tag{39}$$

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.

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.