Supplemental Materials

1 Simulation Study

We present a simulation study illustrating the performance of the model for a 4-d GARCH-SSM. We simulate a GARCH-SSM data using a sample size of $T = 1000$ and we estimate the parameters using both the standard state space model (SSM) approach and our GARCH-SSM formulation. The observation matrix, $\boldsymbol{F}'$, and state evolution matrix, $\boldsymbol{G}$, are constant throughout time and fixed as $\boldsymbol{F}' = \boldsymbol{G} = \boldsymbol{I}_4$, where $\boldsymbol{I}_4$ is the 4×4 identity matrix. With the same parameter values we simulate 100 replicate data sets from a 4-dimensional GARCH(1,1)-SSM, thus allowing us to assess the frequentist behaviour in the parameter estimates. The model parameters are the unknown components of the variance, that is the state variance and the components of the GARCH(1,1) error. We seek to estimate these parameters, the latent states and an estimate of the observation variance. The structure of $\boldsymbol{F}'$ and $\boldsymbol{G}$ in this case represent a random-walk plus noise structure for the mean level.

1.1 Choice of priors

To estimate the model using the Bayesian framework we need to specify the priors. For the GARCH parameter $\alpha_0^{(i)}$ we choose a half-Cauchy prior; for $(\alpha_1^{(i)}, \beta_1^{(i)})$ we choose a joint half-Cauchy prior restricted to the region such that $\alpha_1^{(i)} + \beta_1^{(i)} < 1$; for the state covariance matrix, $\boldsymbol{W}$, we choose an Inverse-Wishart prior, $IW(10, 10)$, where, $i, j = 1, 2, 3, 4$. The choice of priors reflects the positivity restriction that we have made for the GARCH parameters and the diagonal components of the correlation components. The half-Cauchy prior has a thicker tail than other priors such as the half-normal and this allows for larger values for α_0 to occur more frequently. This has the effect of allowing the baseline variance for the GARCH process to be high. We have also made a stationarity restriction for the GARCH parameters as given by the support $\alpha_1^{(i)} + \beta_1^{(i)} < 1$ for $i = 1, 2, 3, 4$. For the correlation parameters in the matrix $\boldsymbol{U}$ described below, we choose priors for the unnormalized version of $\boldsymbol{U}$ and then normalize the proposed quantity so that $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{U}^\top$ has unit diagonal entries. For the diagonal entries of the unnormalized version of $\boldsymbol{U}$ we choose a half-Cauchy prior, for the remainder of the correlation parameters, u_{ij}, we choose a $N(0, 1)$ prior. Our choice of the Inverse-Wishart prior for the state variance $\boldsymbol{W}$, allows for a conjugate analysis. However, for all other parameters in the model we obtain samples from the posterior distributions using the Metropolis-Hastings algorithm.

Model specification and Bias Estimation:

The 4d-GARCH(1,1)-SSM is given by the following specification.

$$\begin{aligned} \text{Observation equation:} \quad & \boldsymbol{y}_t = \boldsymbol{F}'\boldsymbol{\theta}_t + \boldsymbol{z}_t \\ \text{State equation:} \quad & \boldsymbol{\theta}_t = \boldsymbol{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad w_t \sim N(\boldsymbol{0}, \boldsymbol{W}) \\ \text{GARCH error specification :} \quad & \begin{cases} \boldsymbol{z}_t = \boldsymbol{V}_t^{1/2}\boldsymbol{\epsilon_t}, & \boldsymbol{\epsilon_t} \sim N(\boldsymbol{0}, \boldsymbol{I}_n) \\ \boldsymbol{V}_t = \boldsymbol{D}_t\boldsymbol{R}\boldsymbol{D}_t \end{cases} \end{aligned} \tag{1}$$

with $\boldsymbol{D_t} = \text{diag}(\sigma_{1,t}, \ldots, \sigma_{n,t})$ and, for $i = 1, \ldots, n$,

$$\sigma_{i,t}^2 = \alpha_0^{(i)} + \alpha_1^{(i)} z_{i,t-1}^2 + \beta_1^{(i)} \sigma_{i,t-1}^2$$

with $\alpha_0^{(i)}, \alpha_1^{(i)}, \beta_1^{(i)} \in [0, 1]$, and

$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,n} \\ \rho_{2,1} & 1 & \cdots & \rho_{2,n} \\ \vdots & & \ddots & \vdots \\ \rho_{n,1} & \rho_{n,2} & \cdots & 1 \end{bmatrix} = \boldsymbol{U}\boldsymbol{U}^\top, \tag{2}$$

and the parametrization $\boldsymbol{U}$ represents the upper Cholesky decomposition of $\boldsymbol{R}$.

$$\boldsymbol{U} = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{nn} \end{pmatrix}$$

In the simulation we used $\boldsymbol{F}' = \boldsymbol{G} = \boldsymbol{I}_{n\times n}$ so that we have a random walk plus noise model. Below we provide the estimation results from the 100 simulations summarized in the form of bias. The bias is computed based on the posterior median estimates from each of the 100 simulations. First, reported are the bias of the GARCH parameters; next, we report the bias for the estimates of the unknown state Variance matrix, $\boldsymbol{W}$; finally, we report the boxplot based on the posterior median estimates for the parameters of $\boldsymbol{U}$.

Series		$\boldsymbol{\alpha_0}$	$\boldsymbol{\alpha_1}$	$\boldsymbol{\beta_1}$
	True	**1.00**	**0.10**	**0.80**
$\boldsymbol{Y_1}$	Bias	0.18	0.03	-0.04
	Std. Error	(0.17)	(0.01)	(0.02)
	True	**1.00**	**0.30**	**0.60**
$\boldsymbol{Y_2}$	Bias	0.29	0.12	-0.15
	Std. Error	(0.17)	(0.03)	(0.03)
	True	**2.00**	**0.10**	**0.40**
$\boldsymbol{Y_3}$	Bias	0.51	0.04	-0.17
	Std. Error	(0.45)	(0.02)	(0.11)
	True	**2.00**	**0.20**	**0.70**
$\boldsymbol{Y_4}$	Bias	0.55	0.07	-0.09
	Std. Error	(0.28)	(0.02)	(0.03)

Table 1: GARCH Parameter Estimates and Biases

$$Bias(\boldsymbol{W}) = \begin{pmatrix} \mathbf{-0.028} & \mathbf{0.000} & \mathbf{0.007} & \mathbf{0.008} \\ (0.014) & (0.004) & (0.004) & (0.003) \\ & \mathbf{0.005} & \mathbf{0.000} & \mathbf{0.001} \\ - & (0.009) & (0.004) & (0.003) \\ - & - & \mathbf{0.014} & \mathbf{0.005} \\ - & - & (0.009) & (0.005) \\ - & - & - & \mathbf{0.020} \\ - & - & - & (0.009) \end{pmatrix} \tag{3}$$

The true value of $\boldsymbol{W}$ used to generate the model was $\boldsymbol{W} = diag(0.1, 0.1, 0.1, 0.1)$. From (3) we can see that with respect to estimating the unknown state variance $\boldsymbol{W}$ the model is doing well. Next, for the correlation coefficient matrix, $\boldsymbol{U}$, we see from the boxplots in Figure 1 that the model is also able to estimate this parameter well. Table 1 shows that the GARCH parameters are being estimated reasonably well. Overall owing to the latent structure and the multidimensionality of the model it becomes difficult to estimate the GARCH parameters. It must be noted again that the latent state also must be estimated in order to calculate the GARCH paramaters.

Finally, it still remains to be checked that for each of these simulated data, the GARCH-SSM is being selected as the most appropriate model based on our WAIC selection criteria. Figure 2 shows the boxplots of the WAIC estimated when each of the simulated data sets are estimated using a GARCH-SSM versus a standard-SSM. We see that the WAIC is correctly selecting the GARCH-SSM over the standard-SSM. This can be understood from the fact that the WAIC of the GARCH model is greater than that of the standard-SSM. The panel on the right in Figure 2 is a histogram of the difference in the GARCH and the standard-SSM based WAICs computed for each data set. This also reinforces the finding that the WAIC is correctly selecting the appropriate model, since for almost all of the simulated data sets we have that the difference in the WAICs is positive, thus showing that the GARCH based WAIC is greater than the standard SSM based WAIC.

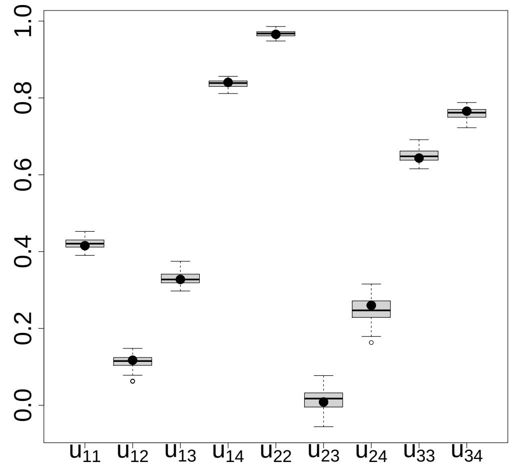


Figure 1: Boxplot: Distribution of the correlation components of $\boldsymbol{U}$ based on posterior median estimates of the 100 simulated data sets. The bold black solid circles represent the true values.

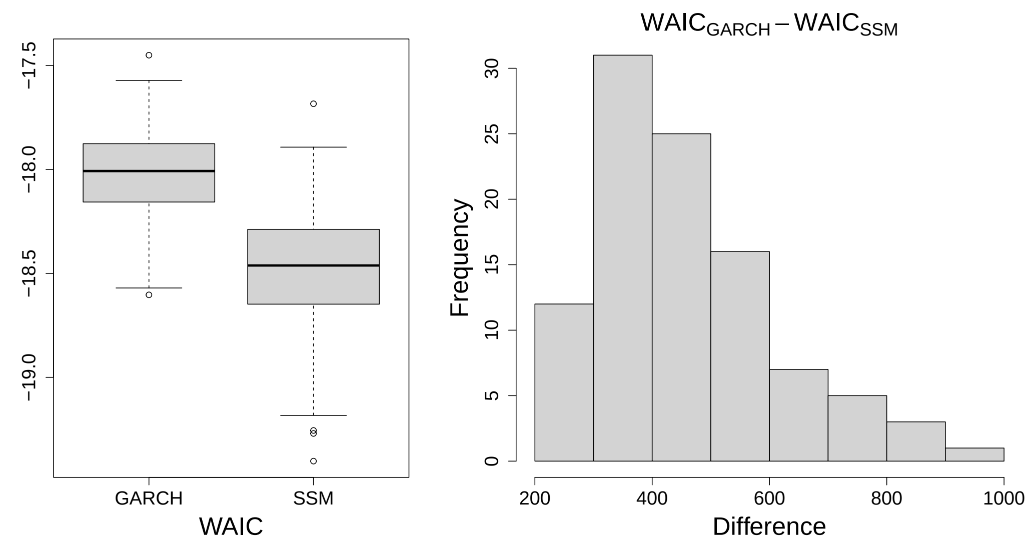


Figure 2: Left: The boxplot shows the comparison of the WAIC estimates of the 100 simulated data when each data is estimated using a GARCH-SSM versus a standard-SSM. Y-axis is scaled down by a factor of a 1000. Right: A histogram of the difference between the WAIC from the GARCH estimates and the standard SSM.

2 Model Estimation and Residual Analysis

2.1 Model Estimation

For illustration, the following results are for a single 4-dimensional series of length $T = 1000$. Figure 3 shows the posterior estimates of the one-step-ahead forecast, $\boldsymbol{f}_t$. We can see that this basically traces out the mean level of the observational series for each series. Figure 4 shows the posterior estimates of the latent state for each of the series in thick black and the true state vector in the broken lines. We can see that we are able to estimate and recover the unobserved states accurately and the true value of the unobserved state almost always lies in the 95% credible interval. In the data analysis of real data, when we apply this model to the HR and BP data, the posterior estimates of the latent state will allow us to study the variability of the HR and BP over the time period of the data.

The panels in Figure 5 shows the posterior estimates of the standard deviation in black and in gray we have the true standard deviation. This plot shows the flexibility our proposed model provides over the standard Gaussian state-space model. We know that for the data the variance is not constant and being able to estimate it properly allows us to draw better inference on the process at a particular time, t. We can see that our estimate for the dynamic standard deviation is quite close to the actual standard deviation of the observed series and the true values are recovered.

The 95% credible intervals of both the estimated state vector and the variances contains the true value of the state vector and the dynamic variances. This shows that our model is able to accurately capture both latent state and the unknown dynamic variance.

2.2 Residual Analysis

Next, we carry out a residual analysis. Recall that by assumption from (1) that the GARCH errors when normalized with respect to the dynamic variance should result in a standard normal error distribution. Replicating this result empirically would be evidence that our model is able to capture the correct heteroskedastic structure of the errors. From Figure (5) we can see that for the second series, there is quite a bit of heteroskedasticity. Thus one would suspect that compared to the quantiles of the standard normal distribution, this series would show large deviations. And this is exactly what can be seen in Figure 6 (top right panel). Consequently, normalizing all the series with their true heterskedasticity should result in normalized errors that adhere strongly to the quantiles of the standard normal distribution. This second result can be seen in Figure 7.

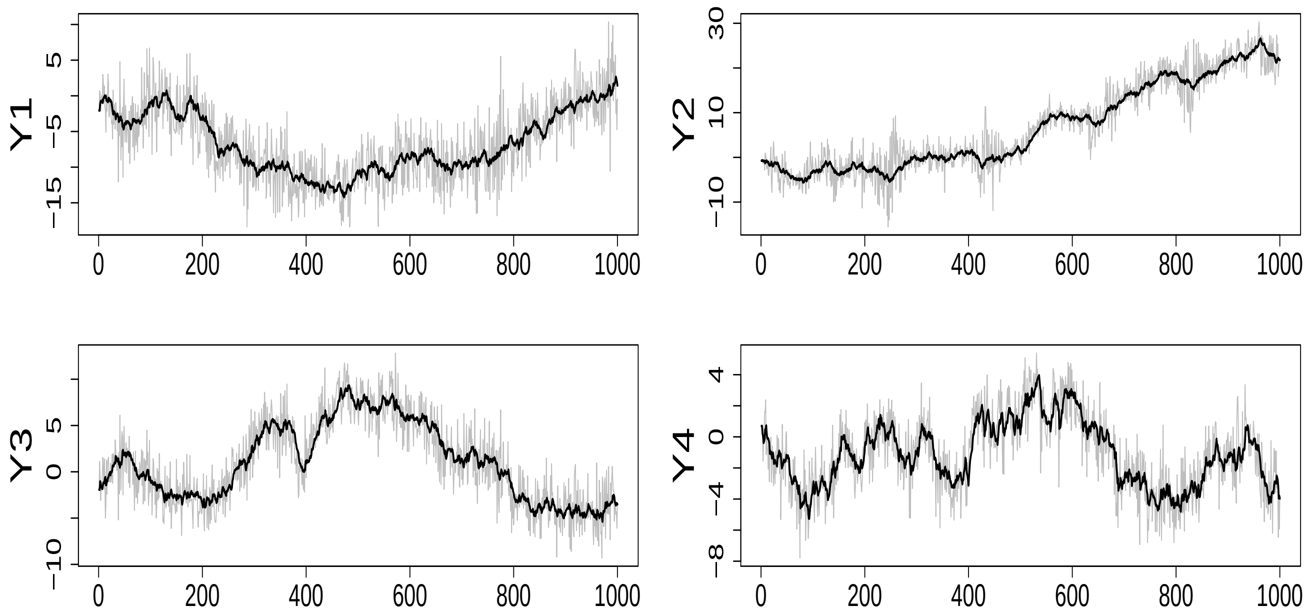


Figure 3: Simulated data: the posterior estimates of the one step ahead forecasts of each series. In gray are the observed data.

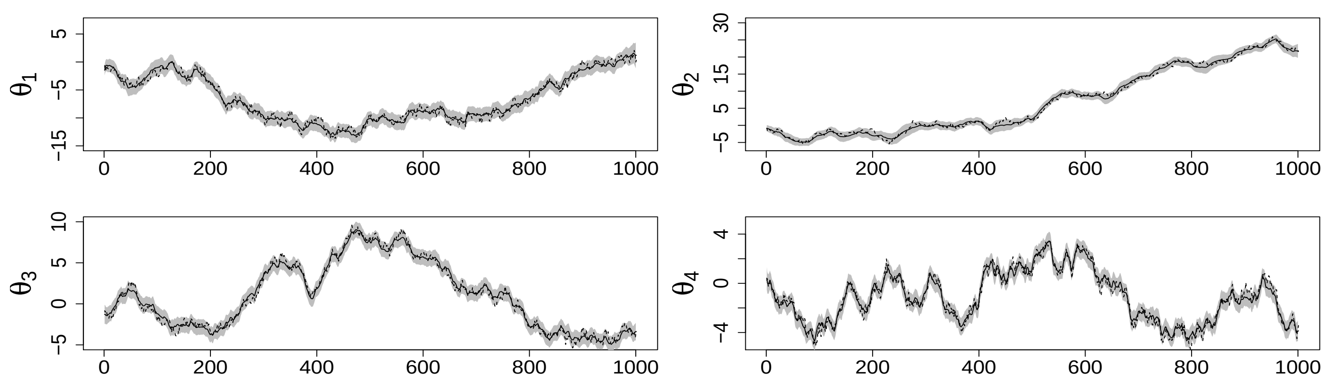
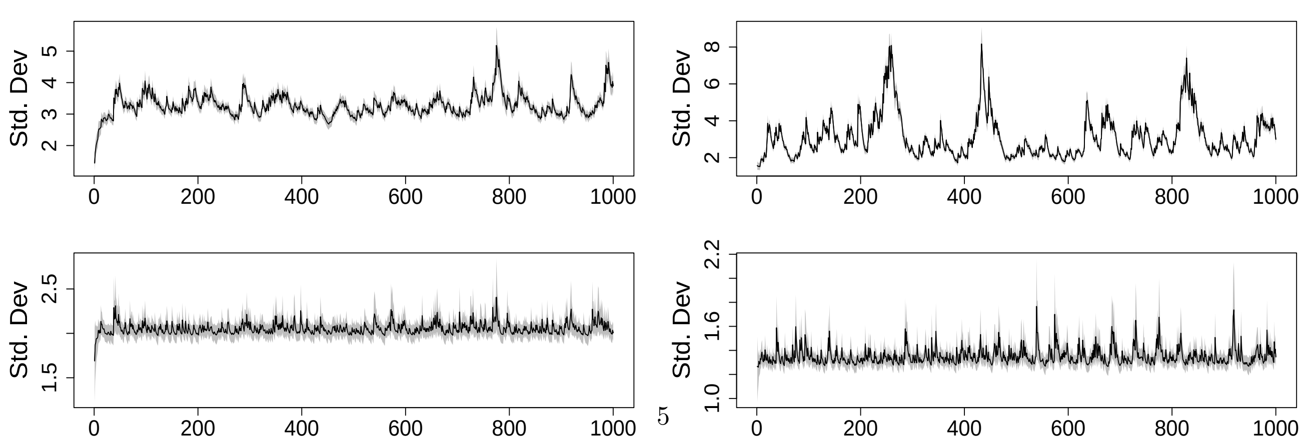


Figure 4: Simulated data: The thick dark line are the posterior median estimates of the states. The dashed lines are the true value of the states. The gray band is the 95% posterior credible interval.



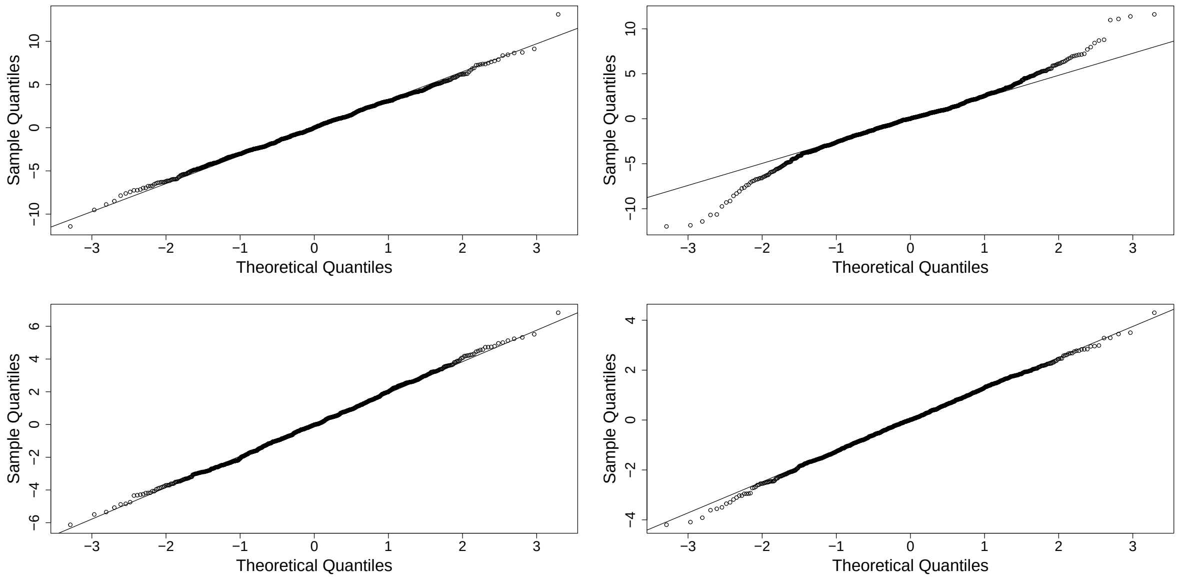


Figure 6: Normal QQ-plots for the un-normalized estimated error terms. Top (left to right): Y1 and Y2. Bottom row (left to right): Y3 and Y4.

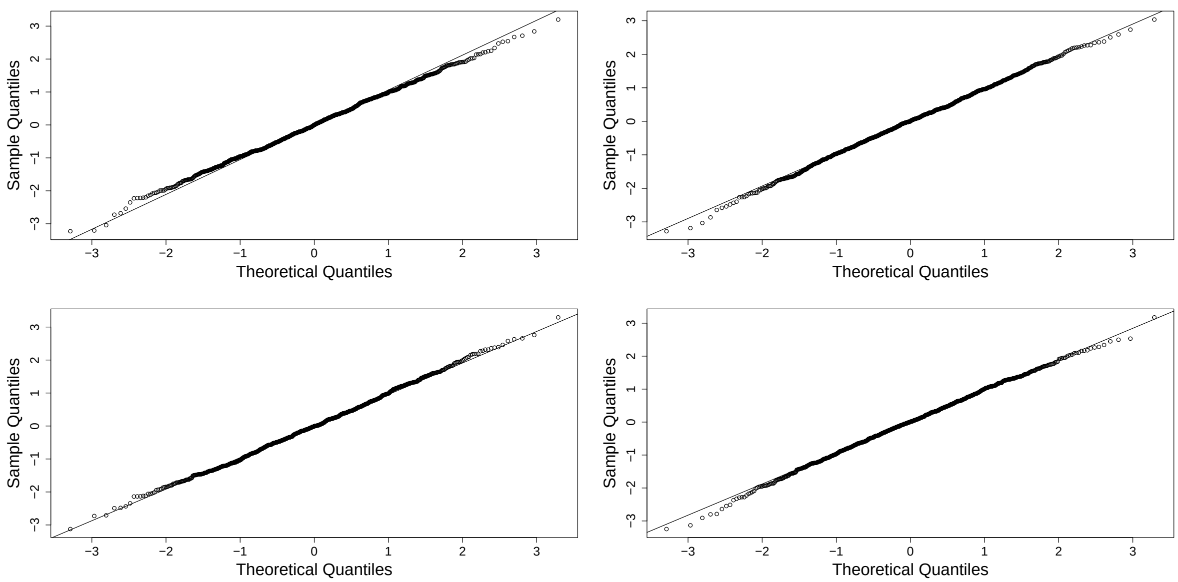


Figure 7: Heteroskedasdicity adjusted normal QQ-plots for the estimated error terms. Top (left to right): Y1 and Y2. Bottom (left to right): Y3 and Y4.

3 Secondary Analysis

In this section we analyze data from some more patient data from the MUHC data set. Below we have the estimates for a second patient using the GARCH-model and the following that the estimates using a standard SSM. We see that in this case, the standard SSM is more appropriate than the the GARCH-model when using the WAIC.

	Estimates for HR series			
	$\boldsymbol{\alpha}_0^{HR}$	$\boldsymbol{\alpha}_1^{HR}$	$\boldsymbol{\beta}_1^{HR}$	$\boldsymbol{W}^{HR}$
Post. est.	6.321	0.376	0.429	4.150
s.d.	(2.278)	(0.093)	(0.122)	(0.754)
CI. 95%	(2.95, 11.71)	(0.23, 0.58)	(0.18, 0.65)	(2.97,5.88)
	Estimates for BP series			
	$\boldsymbol{\alpha}_0^{BP}$	$\boldsymbol{\alpha}_1^{BP}$	$\boldsymbol{\beta}_1^{BP}$	$\boldsymbol{W}^{BP}$
Post. est.	9.574	0.278	0.163	0.923
s.d.	(2.162)	(0.088)	(0.136)	(0.227)
CI. 95%	(4.92, 13.31)	(0.14, 0.48)	(0.01, 0.52)	(0.58, 1.47)
	Estimates of Correlation Parameters			
	$\boldsymbol{\rho}$	$\boldsymbol{\rho}_s$		
Post. est.	0.0197	0.251		
s.d.	(0.017)	(0.923)		
CI 95%	$(0.0008, 0.065)$	$(-0.026, 0.59)$		
WAIC	-5508			

Table 2: Montreal ICU data: Bivariate GARCH(1,1)-SSM parameter estimates. ρ is the estimate of the correlation between the observation errors of the HRT series and the BP series and ρ_s is the estimate of the correlation between the state errors of both series. The estimates are the posterior mean values and in the parenthes is below is their standard error and below that the 95% credible interval.

	Observation Level Parameter Estimates		
	$\boldsymbol{V}_{HR}$	$\boldsymbol{V}_{BP}$	$\boldsymbol{\rho}_{obs}$
Post. est.	13.578	12.684	0.074
s.d.	(3.38)	(1.381)	(1.092)
	State Level Parameter Estimates		
	$\boldsymbol{W}_{HR}$	$\boldsymbol{W}_{BP}$	$\boldsymbol{\rho}_s$
Post. est.	9.927	2.177	0.225
s.d.	(4.879)	(0.288)	(0.333)
WAIC	-5483		

Table 3: Montreal ICU data: Parameter estimates of the bivariate Gaussian-SSM. Here ρ_{obs} is the estimate of the correlation between the observation errors of the HRT series and the BP series and ρ_s is the estimate of the correlation between the state errors of both series. The values displayed are the posterior means of the parameters and the values in the parenthesis are the standard errors.

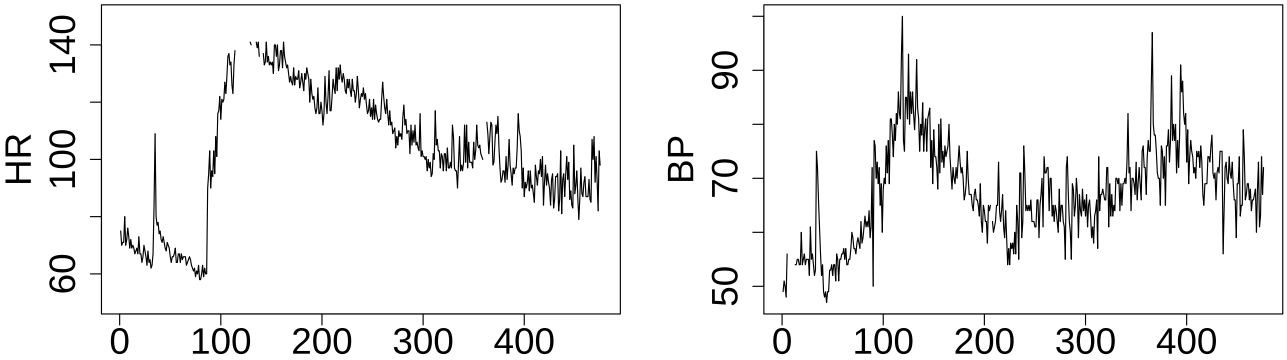


Figure 8: Left: Heart rate series of a patient; Right: Blood pressure series of same patient. Time axis is number of seconds since start of monitoring

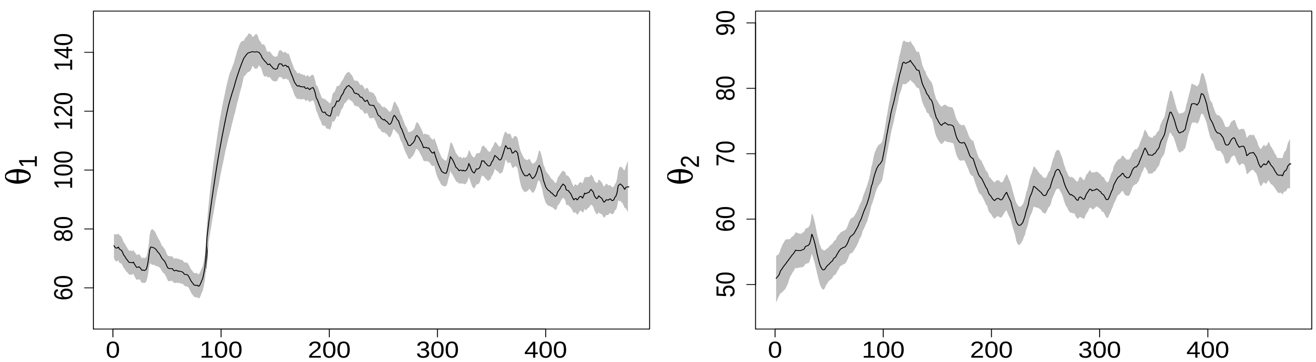


Figure 9: Montreal ICU data. Top row: posterior estimates of the mean level of each series with the credible intervals given by the dashed lines and the observed data in gray (HR left, BP right).

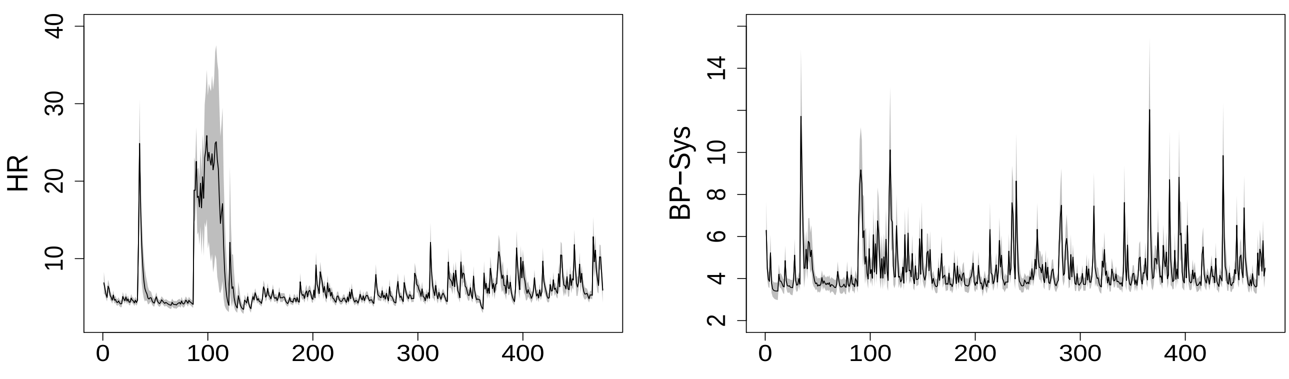


Figure 10: Posterior mean estimates of the dynamic standard deviation (HR left, BP right).

The following is an analysis from a third patient. The estimates and plots are provided below. In this case, based on the WAIC we see that the GARCH model is the most appropriate.

	Estimates for HR series			
	$\boldsymbol{\alpha}_0^{HR}$	$\boldsymbol{\alpha}_1^{HR}$	$\boldsymbol{\beta}_1^{HR}$	$\boldsymbol{W}^{HR}$
Post. est.	0.181	0.192	0.789	0.327
s.d.	(0.044)	(0.031)	(0.031)	(0.048)
CI. 95%	(0.11, 0.28)	(0.14, 0.26)	(0.72, 0.84)	(0.246,0.428)
	Estimates for BP series			
	$\boldsymbol{\alpha}_0^{BP}$	$\boldsymbol{\alpha}_1^{BP}$	$\boldsymbol{\beta}_1^{BP}$	$\boldsymbol{W}^{BP}$
Post. est.	0.700	0.456	0.510	0.388
s.d.	(0.142)	(0.055)	(0.053)	(0.051)
CI. 95%	(0.47, 1.03)	(0.35, 0.56)	(0.41, 0.62)	(0.30,0.50)
	Estimates of Correlation Parameters			
	$\boldsymbol{\rho}$	$\boldsymbol{\rho}_s$		
Post. est.	0.0512	0.031		
s.d.	(0.012)	(0.046)		
CI 95%	$(0.0295, 0.078)$	$(-0.081, 0.10)$		
WAIC	-20434			

Table 4: Montreal ICU data: Bivariate GARCH(1,1)-SSM parameter estimates. ρ is the estimate of the correlation between the observation errors of the HRT series and the BP series and ρ_s is the estimate of the correlation between the state errors of both series. The estimates are the posterior mean values and in the parenthes is below is their standard error and below that the 95% credible interval.

	Observation Level Parameter Estimates		
	$\boldsymbol{V}_{HR}$	$\boldsymbol{V}_{BP}$	$\boldsymbol{\rho}_{obs}$
Post. est.	8.94	3.611	0.050
s.d.	(0.367)	(0.249)	(0.342)
	State Level Parameter Estimates		
	$\boldsymbol{W}_{HR}$	$\boldsymbol{W}_{BP}$	$\boldsymbol{\rho}_s$
Post. est.	1.323	3.456	0.062
s.d.	(0.196)	(0.314)	(0.532)
WAIC	-22792		

Table 5: Montreal ICU data: Parameter estimates of the bivariate Gaussian-SSM. Here ρ_{obs} is the estimate of the correlation between the observation errors of the HRT series and the BP series and ρ_s is the estimate of the correlation between the state errors of both series. The values displayed are the posterior means of the parameters and the values in the parenthesis are the standard errors.

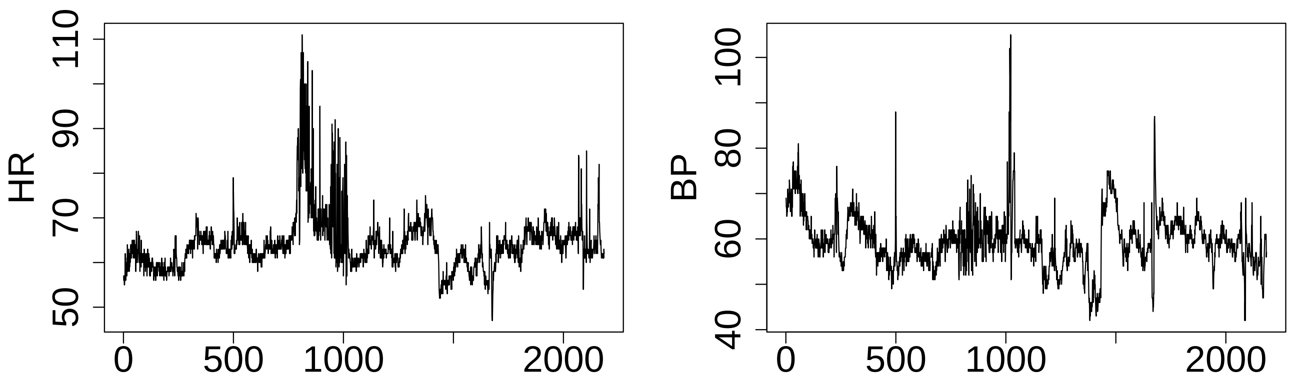


Figure 11: Left: Heart rate series of a patient; Right: Blood pressure series of same patient. Time axis is number of seconds since start of monitoring

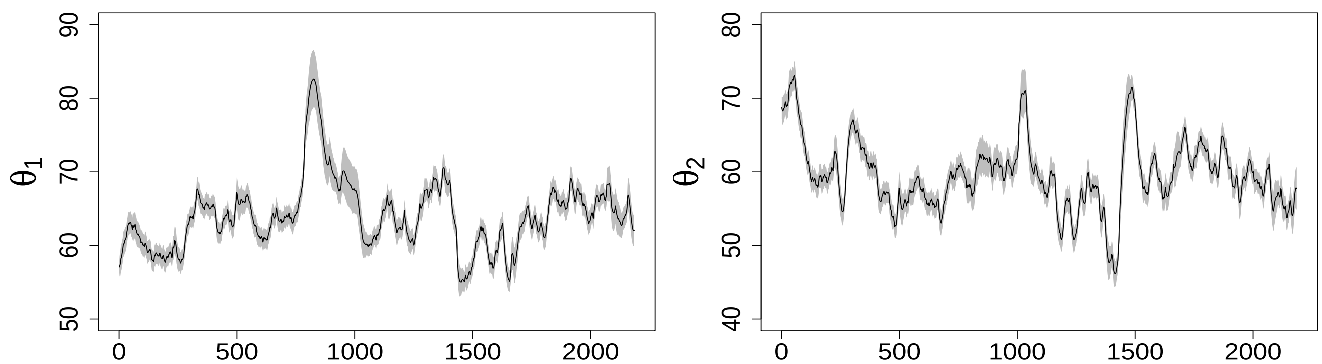


Figure 12: Montreal ICU data. Top row: posterior estimates of the mean level of each series with the credible intervals given by the dashed lines and the observed data in gray (HR left, BP right).

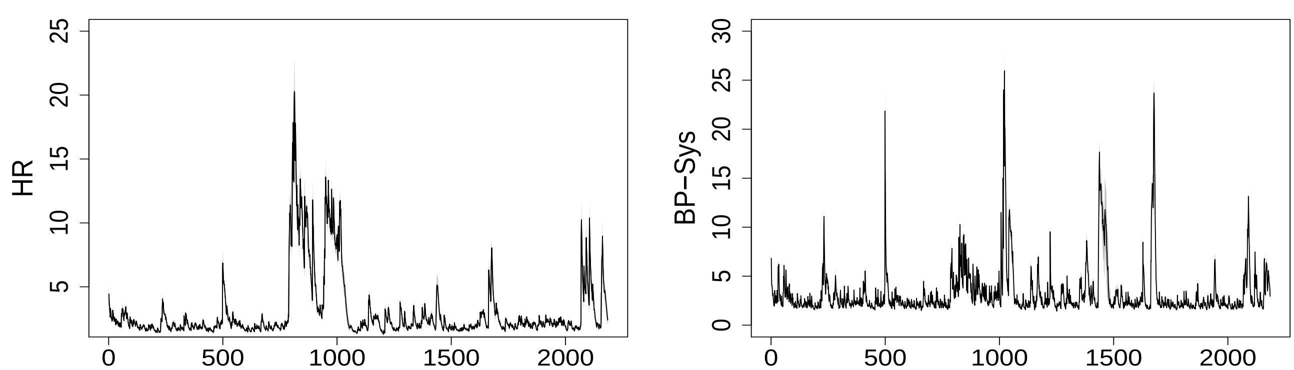


Figure 13: Posterior mean estimates of the dynamic standard deviation (HR left, BP right).

4 Calculation of the WAIC

The *lppd* is given by,

$$
\begin{aligned}
lppd &= \sum_{i=1}^{n} \log \int p(y_i|\boldsymbol{\psi}) dp_{post}(\boldsymbol{\psi}) = \sum_{i=1}^{n} \Big(\log[E_{post} p(y_i|\boldsymbol{\psi})] \Big) \\
p_{waic} &= 2 \sum_{i=1}^{n} \Bigg\{ \Big(\log[E_{post} p(y_i|\boldsymbol{\psi})] \Big) - \Big(E_{post} \log[p(y_i|\boldsymbol{\psi})] \Big) \Bigg\} \\
WAIC &= 2 \sum_{i=1}^{n} \Big(E_{post} \log[p(y_i|\boldsymbol{\psi})] \Big) - \sum_{i=1}^{n} \Big(\log[E_{post} p(y_i|\boldsymbol{\psi})] \Big)
\end{aligned}
\tag{4}
$$

We have that -2 times the $WAIC$ in (4) brings it on the deviance scale. This allows for a more straight forward comparison with the AIC and DIC. We do not however do this in our case. For us, a larger WAIC value implies a better model fit.

Here $\boldsymbol{\psi}$ is the unknown set of parameters. In our case we have that the unknown parameters is the set observation variances and the state variance. The integrations in (4) may be intractable since we have to integrate over the posterior distribution. However, the consistency of the sample means and the continuity of the *log*-function gives us that,

$$
\begin{aligned}
\log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i|\boldsymbol{\theta}_s)\right) &\xrightarrow{p} \log\Big(E_{post} p(y_i|\boldsymbol{\theta})\Big) \\
\left(\frac{1}{S}\sum_{s=1}^{S} \log(p\big(y_i|\boldsymbol{\theta}_s)\big)\right) &\xrightarrow{p} E_{post} \log\Big(p(y_i|\boldsymbol{\theta})\Big)
\end{aligned}
$$

This gives the computed $WAIC$ as,

$$
\begin{aligned}
cWAIC = 2\sum_{i=1}^{n}\left(\frac{1}{S}\sum_{s=1}^{S} \log(p\big(y_i|\boldsymbol{\theta}_s)\big)\right) - \\
\sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i|\boldsymbol{\theta}_s)\right)
\end{aligned}
$$

and $cWAIC$ converges in probability to $WAIC$ as S goes to infinity., where S is the size of the posterior sample.

Algorithms for Sampling

Algorithm 1: FFBS

```
Given $\boldsymbol{m}_0, \boldsymbol{C}_0$, run the Kalman Filter.
Draw $\boldsymbol{\theta}_n \sim N(\boldsymbol{m}_n, \boldsymbol{C}_n)$, where $\boldsymbol{m}_n, \boldsymbol{C}_n$.
for $t$ in $t = n-1, ..., 0$ do
    Draw $\boldsymbol{\theta}_t \sim N(\boldsymbol{h}_t, \boldsymbol{H}_t)$, ($\boldsymbol{h}_t, \boldsymbol{H}_t$ are defined in the main paper)
end for.
```

Algorithm 2: Posterior Sampling with FFBS step

```
Initialize the unknown parameters, $\psi_1^{(0)}, \psi_2^{(0)}, ..., \psi_p^{(0)}$
Initialize state parameters using FFBS using, $\boldsymbol{\theta}_{0:n}^{(0)} \sim p(\boldsymbol{\theta}_{0:n}^{(0)} | \boldsymbol{y}_{1:n}, \psi_1^{(0)}, ..., \psi_p^{(0)})$.
for i in $1 : nsims$ do
    Draw $\psi_1^{(i)} \sim p(\psi_1^{(i)} | \boldsymbol{y}_{1:n}, \boldsymbol{\theta}_{0:n}^{(i-1)}, \psi_2^{(i-1)}, ..., \psi_p^{(i-1)})$
    Draw $\psi_2^{(i)} \sim p(\psi_2^{(i)} | \boldsymbol{y}_{1:n}, \boldsymbol{\theta}_{0:n}^{(i-1)}, \psi_1^{(i)}, \psi_3^{(i-1)}, ..., \psi_p^{(i-1)})$
    ⋮        ⋮        ⋮
    Draw $\psi_j^{(i)} \sim p(\psi_j^{(i)} | \boldsymbol{y}_{1:n}, \boldsymbol{\theta}_{0:n}^{(i-1)}, \psi_1^{(i)}, ..., \psi_{j-1}^{(i)}, \psi_{j+1}^{(i-1)}, ..., \psi_p^{(i-1)})$
    Draw the state parameters using FFBS using, $\boldsymbol{\theta}_{0:n}^{(i)} \sim p(\boldsymbol{\theta}_{0:n}^{(i)} | \boldsymbol{y}_{1:n}, \psi_1^{(i)}, ..., \psi_p^{(i)})$.
end for.
```