Online Appendix for: Semi-Complete Data Augmentation for Efficient State Space Model Fitting

Agnieszka Borowska
University of Glasgow, School of Mathematics and Statistics, UK
and
Ruth King
School of Mathematics, University of Edinburgh, UK

In this online appendix we provide additional examples and results, technical details and a description of the lapwings dataset. The structure of the online appendix is as follows. Appendix A discusses an additional conceptual example of the basic SV model and its extensions. Appendix B presents technical details for the two applications from the paper. Appendix C provides specification details of the HMM-based approximations for three models (the motivating example from Section 4.2, the SV model and the lapwings population model). Appendix D discusses computing the effective sample size. Appendix E provides details on the lapwings dataset. Appendix F contains additional results for both applications from the main paper.

A Additional conceptual example: stochastic volatility model

In this appendix we present a conceptual exercise for a simple vertical integration introduced in Section 3. We consider the stochastic volatility (SV) model, which is a popular tool for modelling time-varying volatility. Originally, it has been applied to financial time series, see e.g. Taylor (1994), Ghysels et al. (1996) or Shephard (1996), however, it has also proven useful in macroeconometrics, in particular as a building block of the unobserved

component stochastic volatility (UCSV) model (Stock and Watson, 2007) considered in Section 5.1.

We emphasise that the aim of the following exercise is to illustrate on a simple example how to derive SCDL (7) and approximate is using an HMM embedding (16) from Section 4. We believe this facilitates driving more general and complex schemes, such as the one for the UCSV model. At the same time, we acknowledge that for the specific purpose of the SV model several bespoke approaches have been proposed, including optimised algorithms available as "off-the-shelf" packages such as the `stochvol R` package (Kastner, 2016).

A.1 Basic SV model and its challenges

The basic form of the SV model is given by

$$y_t|h_t, \boldsymbol{\theta} \sim \mathcal{N}\left(0, \exp(h_t)\right), \qquad t = 1, \ldots, T, \tag{S.1}$$

$$h_t|h_{t-1}, \boldsymbol{\theta} \sim \mathcal{N}\left(\mu + \phi(h_{t-1} - \mu), \sigma^2\right), \tag{S.2}$$

where $h_0|\boldsymbol{\theta} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right)$, $\boldsymbol{\theta} = (\mu, \phi, \sigma^2)^T$. We adopt standard priors (Kim et al., 1998):

$$\mu \sim \mathcal{N}(0, \sigma^2_{\mu 0}), \quad \frac{\phi + 1}{2} \sim \mathcal{B}(\alpha_{\phi 0}, \beta_{\phi 0}), \quad \sigma^2 \sim \mathcal{IG}(\alpha_{\sigma^2 0}, \beta_{\sigma^2 0}),$$

with $\sigma^2_{\mu 0} = 10$, $\alpha_{\phi 0} = 20$, $\beta_{\phi 0} = 1.5$, $\alpha_{\sigma^2 0} = 5/2$, $\beta_{\sigma^2 0} = 0.05/2$. The estimation of the SV model has been considered a challenging problem due to the intractable likelihood

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \int p(\boldsymbol{y}, \boldsymbol{h}) d\boldsymbol{h} = \int p(h_0) \prod_{t=1}^{T} p(y_t|h_t) p(h_t|h_{t-1}) dh_0 dh_1 \ldots dh_T. \tag{S.3}$$

Some of the previous approaches to tackle this issue include standard DA approach, in which the latent volatilities are imputed in an MCMC scheme, see Kim et al. (1998). The associated complete data likelihood admits a closed form $p(\boldsymbol{y}, \boldsymbol{h}|\boldsymbol{\theta}) = p(h_0) \prod_{t=1}^{T} p(y_t|h_t) p(h_t|h_{t-1})$.

Since the seminal paper by Kim et al. (1998), numerous approaches have been proposed to increase the estimation efficiency for the SV model. One class of techniques includes reparameterisations for more efficient sampling, relating to centred and non-centred parameterizations described in Papaspiliopoulos et al. (2007), see e.g. Strickland et al. (2008). In the ancillarity-sufficiency interweaving strategy (ASIS), originally proposed by Yu and Meng (2011) and adopted by e.g. Kastner and Frühwirth-Schnatter (2014); Hosszejni and

Kastner (2019); Bitto and Frühwirth-Schnatter (2019) to the SV model case, improved mixing is achieved by combine two DA schemes based on different parametrisations. The ASIS strategy has been implemented in the popular `stochvol R` package (Kastner, 2016). A different approach has been proposed by Barra et al. (2017), who use an independent MH sampler with the proposal distributions being approximations to the joint posterior density of the parameters and states. The efficient approximation is obtained by first approximating the marginal parameter posterior distribution by a mixture of Student's t distribution using the algorithm of Hoogerheide et al. (2012); second, conditionally sampling the states from a Gaussian proposal density from the numerically accelerated importance sampling method of Koopman et al. (2015).

An alternative strand of literature is concerned with numerical integration of the latent states, see e.g. Fridman and Harris (1998); Langrock et al. (2012). In particular, Langrock et al. (2012) approximate (S.3) using an HMM by discretizing the state space of $\boldsymbol{h}$. They perform a numerical integration of the latent states based on a grid of B equally sized bins $B_i = [b_{i-1}, b_i)$, $i = 1, \ldots, B$, with the corresponding midpoints b_i^*. The range of the admissible values for the demeaned volatility, b_0 and b_B, is set to $\pm 5\sigma_h$, where σ_h is the stationary standard deviation of the logvolatility process. This leads to an approximation of (S.3) via $p(\boldsymbol{y}|\boldsymbol{\theta}) \approx \boldsymbol{u}_0 \prod_{t=1}^{T} \Gamma_t Q_t \mathbf{1}$, where $\Gamma_t = \left[\gamma_{ij,t}\right]_{i,j=1,\ldots,B}$, with

$$\gamma_{ij,t} = \mathbb{P}(h_t - \mu \in B_j | h_{t-1} - \mu = b_i^*) = \Phi\left(\frac{b_j - \phi b_i^*}{\sigma}\right) - \Phi\left(\frac{b_{j-1} - \phi b_i^*}{\sigma}\right),$$

$$Q_t = \operatorname{diag}\left(\varphi\left(\frac{y_t}{\exp((\mu + b_i^*)/2)}\right)\right)_{i=1,\ldots,B},$$

where Φ and φ denote the cdf and the pdf of the standard normal density, respectively. Notice that the transition probabilities are time-constant so that the underlying Markov chain is homogeneous. Sandmann and Koopman (1998) point out that for the SV model such a form of numerical integration might not be always suitable since a fixed grid cannot efficiently accommodate for periods of low and high volatility. We address this issue by suggesting a more efficient, adaptive HMM-based approximation as an alternative to the fixed bins used by Langrock et al. (2012).

Finally, we note that for μ and σ^2 Gibbs updates can be performed based on full conditional densities, see Kim et al. (1998). Furthermore, numerous bespoke enhancements

for sampling of the hidden states has been devised, see e.g. Kim et al. (1998). However, we provide a general framework requiring only "vanilla" type updates (based on a MH RW algorithm) and consider the full DA as a comparison benchmark.

A.2 Dependence structure and SCDL

The basic SV specification concerns a single one-dimensional state on the real line and the sampling inefficiency originates from a high persistence of the logvolatility process. In order to break this dependence, we propose to impute $\boldsymbol{h_{2T}}$ and integrate out $\boldsymbol{h_{2T+1}}$, the states in even and odd time periods, respectively. This corresponds to the *vertical* integration scheme with $\boldsymbol{x}_{int} = \boldsymbol{h_{2T+1}}$ and $\boldsymbol{x}_{aug} = \boldsymbol{h_{2T}}$. Without loss of generality we assume that T is odd so that h_T is integrated out and we denote $T^* = \frac{T-1}{2}$; if T is even then we add one extra integration based on uniformly distributed h_{T+1}. The exact SCDL is given by

$$p(y, \boldsymbol{h_{2T}}) = p(h_0) \int p(h_1|h_0)p(y_1|h_0) \left(\prod_{t=1}^{T^*} p(y_{2t+1}|h_{2t+1})p(h_{2t+1}|h_{2t})p(y_{2t}|h_{2t})p(h_{2t}|h_{2t-1}) \right) dh_1 \dots dh_T,$$

and can be split (by conditioning on the even states) into a product $T^* + 1$ of integrals

$$p(y, \boldsymbol{h_{2T}}) = \underbrace{p(h_0)}_{=:C_0} \underbrace{\int p(h_1|h_0)p(y_1|h_0)dh_1}_{=:D_0} \prod_{t=1}^{T^*} \underbrace{p(y_{2t}|h_{2t})}_{=:C_t} \underbrace{\int p(y_{2t+1}|h_{2t+1})p(h_{2t+1}|h_{2t})p(h_{2t}|h_{2t-1})dh_{2t+1}}_{=:D_t}.$$

Since the integrals above are conditionally independent, we have

$$p(y, \boldsymbol{h_{2T}}) = C_0 D_0 \prod_{t=1}^{T^*} C_t D_t = \prod_{t=0}^{T^*} C_t D_t.$$

A.3 Hidden Markov model approximation

The integrals D_t cannot be evaluated analytically and need to be numerically approximated. We consider two different approaches.

Case (i) Fixed bins This approach follows Langrock et al. (2012) and consists in relating $z_t = k$, the Markov chain being in state k, to the event $h_{2t+1} - \mu \in \mathcal{B}_k$, the demeaned volatility in an odd time period $2t+1$ falling into the kth bin B_k. We take equally spaced bins, each of length λ. Falling into bin B_k can be specified as e.g. lying in the interval

$[b_{k-1}, b_k)$ or being equal to this interval's midpoint $b_k^* = \frac{b_{k-1}+b_k}{2}$. In particular, we consider approximation of the following form

$$D_t \approx \sum_{k=1}^{B} p(y_{2t+1}|h_{2t+1} - \mu = b_k^*) p(h_{2t+2}|h_{2t+1} - \mu = b_k^*) p(h_{2t+1} - \mu \in B_i|h_{2t}). \tag{S.4}$$

The last term in (S.11) can be approximated as

$$p(h_{2t+1} - \mu \in B_k|h_{2t}) \approx \Phi\left(\frac{b_k - \phi(h_{2t} - \mu)}{\sigma}\right) - \Phi\left(\frac{b_{k-1} - \phi(h_{2t} - \mu)}{\sigma}\right),$$

which is adopted in Langrock et al. (2012), or using a simpler midpoint approximation

$$p(h_{2t+1} - \mu \in B_k|h_{2t}) \approx \lambda\varphi\left(\frac{b_k^* - \phi(h_{2t} - \mu)}{\sigma}\right),$$

which we adopt in our application due to computing time.

Case (ii) Adaptive bins Instead of specifying the grid points, we can fix the probabilities for each bin (which in Case (i) needed to be determined) and consider quantiles corresponding to intervals of equal probability. The vector of mid-quantiles $\boldsymbol{q}^*$ determines the midpoints at time $2t+1$, which are given as $\beta_{k,2t+1}^* = \phi(h_{2t} - \mu) + \sigma \cdot \Phi^{-1}(q_k^*)$, $k = 1, \ldots, B$, where h_{2t} is the imputed volatility for the previous time period. We approximate D_t as

$$D_t \approx \sum_{k=1}^{B} \varphi\left(\frac{y_{2t+1}}{\exp((\beta_{k,2t+1}^* + \mu)/2)}\right) \varphi\left(\frac{h_{2t+2} - \mu - \phi\beta_{k,2t+1}^*}{\sigma}\right) \cdot \frac{1}{B},$$

since $\gamma_{ij,t} = \frac{1}{B}$, and these constant transition probabilities from an imputed state cancel out in the acceptance ratios.

A.4 Extensions of the basic SV model

Below, we show how the SCDA scheme for the basic SV model easily extends to more complex models: the SV in the mean model of Koopman and Uspensky (2002) or SV with leverage of Omori et al. (2007).

A.4.1 SV in the mean

The proposed SCDA scheme easily extends to more complex models, e.g. the popular Stochastic Volatility in the Mean (SVM) model of Koopman and Uspensky (2002) (see

also Chan, 2017). Its basic specification is given by

$$y_t|h_t, \boldsymbol{\theta} \sim \mathcal{N}\left(\beta \exp(h_t), \exp(h_t)\right), \tag{S.5}$$

$$h_{t+1}|h_t \sim \mathcal{N}\left(\mu + \phi(h_t - \mu), \sigma^2\right), \tag{S.6}$$

$$h_0 \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right), \tag{S.7}$$

for $t = 1, \ldots, T$. Hence, the latent volatility process h_t influences both the conditional variance and the conditional mean of the observation series y_t, which is additionally controlled by a scaling parameter β. For the volatility parameters μ, ϕ and σ^2 we adopt the prior specification as for the standard SV model, while for the mean-scaling parameter we specify $\beta \sim \mathcal{N}(0, \sigma^2_{\beta_0})$, with $\sigma^2_{\beta_0} = 10$.

A.4.2 SV with leverage

The basic SV or SVM models can be extended to allow for *leverage* effects, i.e. a feedback from past logreturns to the current value of the volatility process. This effect is typically modeled as a negative correlation between the last period logreturns and the current value of volatility. The motivation behind the leverage effect is that the volatility in financial markets may adapt differently to positive and negative shocks/news (affecting logreturns), where large negative shocks are likely to increase the volatility. The SV model with leverage (SVL) has been frequently analyzed in the literature, see Jungbacker and Koopman (2007), Meyer and Yu (2000), Yu (2005), Durbin and Koopman (2012, Section 9.5.5.) or Zucchini et al. (2016, Section 20.2.3). For convenience, we rewrite the basic SV model (S.1)–(S.2) as

$$\begin{aligned} y_t &= \exp(h_t/2)\varepsilon_t, & \varepsilon_t &\sim \mathcal{N}\left(0, 1\right), \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, & \eta_t &\sim \mathcal{N}\left(0, \sigma^2\right), \\ h_1 &\sim \mathcal{N}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right), \end{aligned}$$

for $t = 1, \ldots, T$. The only difference between the SVL model and the basic specification of the SV model is that now the error terms ε_t and η_t are assumed to be correlated: $\text{corr}[\varepsilon_t, \eta_t] = \rho \neq 0$, with ρ typically estimated to be negative. This apparently slight modification has, however, substantial effect on the dependence structure in the model

(see Figure 1) and hence the conditional distribution of h_t. To derive the latter several reformulations of the model has been proposed (Jungbacker and Koopman, 2007 or Meyer and Yu, 2000), however we will use the treatment provided by Zucchini et al. (2016, Section 20.2.3). These authors use the basic regression lemma for normal variables to show that

$$h_t|h_{t-1}, y_{t-1}, \mu, \phi, \sigma^2, \rho \sim \mathcal{N}\left(\mu + \phi(h_{t-1} - \mu) + \frac{\rho\sigma y_{t-1}}{\exp(h_{t-1}/2)}, \sigma^2(1-\rho^2)\right) \tag{S.8}$$

(the details of the derivation are provided in the next subsection). Formulation (S.8) is particularly convenient for "reusing" the derived integration scheme for the basic SV model, as we only need to adjust the transition probabilities in the approximation to D_t.

A.4.3 Modifications to the HMM-based approximation

The proposed HMM-based approximation to SCDL can be easily adapted to allow for both extension by simply modifying the components of the matrices Γ_t, P_t and Q_t specified in (12)–(14). Notice that for the SVM model the dependence structure of the state is the same as for the basic SV model, hence the core of the integration/imputation scheme remains unchanged. What needs to be adjusted is the observation density, which can be done is a straightforward manner. The modification for the SVL model requires adjusting of the transition probabilities and the pdfs of the augmented states. Below we present the required modifications for the largest model, allowing for both SV in the mean and for the leverage effect (which we refer to as the SVML model).

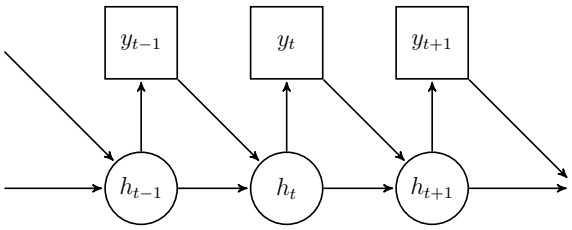


Figure 1: SV model with leverage: modified dependence structure due to feedback from the logreturns y_{t-1} to logvolatilities h_t.

Below, we skip $\boldsymbol{\theta}$ in the conditioning to simplify notation. Following Zucchini et al. (2016), we aim at deriving $p(h_{t+1}|h_t, y_t)$, the conditional distribution of h_{t+1} given h_t and y_t. Since $y_t = \exp(h_t/2)\varepsilon_t$ we can replace conditioning on y_t by conditioning on ε_t. Moreover,

conditioning on h_t is equivalent to conditioning on η_t and adding the mean $\mu + \phi(h_t - \mu)$. Hence, we are interested in the the distribution of η_t given ε_t.

The distribution of $\eta_t|\varepsilon_t$ can be obtained using the basic result from multivariate normal regression, which we recall below for convenience:

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \right) \Rightarrow x|y \sim \mathcal{N}\left(\mu_x + \frac{\sigma_{xy}}{\sigma_y^2}(y - \mu_y), \sigma_x^2 - \frac{\sigma_{xy}^2}{\sigma_y^2} \right).$$

Hence, we obtain

$$\eta_t|\varepsilon_t \sim \mathcal{N}\left(0 + \frac{\rho\sigma}{1}(y - 0), \sigma^2 - \frac{\rho^2\sigma^2}{1} \right) = \mathcal{N}(\rho\sigma\varepsilon_t, \sigma^2(1 - \rho^2))$$

so that

$$h_{t+1}|h_t, \varepsilon_t \sim \mathcal{N}\big(\mu + \rho(h_t - \mu) + \rho\sigma\varepsilon_t, \sigma^2(1 - \rho^2)\big).$$

Finally, we can express the latter in terms of the actual observation y_t rather than the unobserved disturbance ε_t. For the basic SV model this becomes

$$h_{t+1}|h_t, y_t \sim \mathcal{N}\left(\mu + \rho(h_t - \mu) + \rho\sigma\frac{y_t}{\exp(h_t/2)}, \sigma^2(1 - \rho^2) \right),$$

which is the result reported in Section A.4.2, while for the SVM we have

$$h_{t+1}|h_t, y_t \sim \mathcal{N}\left(\mu + \rho(h_t - \mu) + \rho\sigma\frac{y_t - \beta\exp(h_t)}{\exp(h_t/2)}, \sigma^2(1 - \rho^2) \right).$$

B Technical details for the two applications

B.1 UCSV model from Section 5.1

Figure 2 presents the dependence structure of the UCSV model. We observe that both log-volatilities, $\boldsymbol{h}$ and $\boldsymbol{g}$, are independent in the model. Moreover, g_t feeds into another latent state, i.e. the trend component τ_t, via which it affects y_t.

B.1.1 SCDL for two integration schemes

Without loss of generality we assume that T is odd so that g_T is integrated out and we denote $T^* = \frac{T-1}{2}$; if T is even then we add one extra integration based on uniformly distributed g_{T+1}.

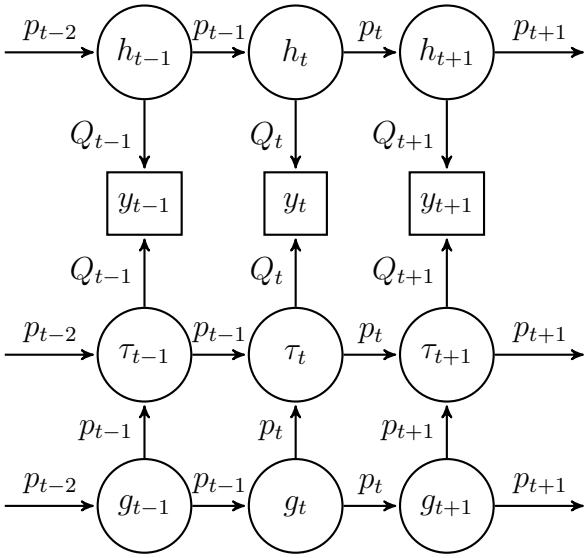


Figure 2: **UCSV model structure**. Squares represent the data, circles – the unknown variables.

SCDL for Scheme 1 We start with integrating out g_t at odd time periods. The corresponding exact SCDL is given by

$$p(\boldsymbol{y}, \boldsymbol{\tau}, \boldsymbol{h}, \boldsymbol{g}_{2T}) = \int p(\tau_1)p(h_1)p(g_1)\prod_{t=1}^{T^*} p(y_{2t+1}|\tau_{2t+1}, h_{2t+1})p(\tau_{2t+1}|\tau_{2t}, g_{2t+1})p(h_{2t+1}|h_{2t})$$
$$p(g_{2t+1}|g_{2t})p(y_{2t}|\tau_{2t}, h_{2t})p(\tau_{2t}|\tau_{2t-1}, g_{2t})p(h_{2t}|h_{2t-1})p(g_{2t}|g_{2t-1})dg_1 dg_3 \ldots dg_T$$

and can be split (by conditioning on even g_ts) into a product of $T^* + 1$ integrals

$$\underbrace{p(\tau_1)p(h_1)}_{=:C_0}\underbrace{\int p(g_1)dg_1}_{=:D_0}\prod_{t=1}^{T^*}\underbrace{p(y_{2t+1}|\tau_{2t+1}, h_{2t+1})p(h_{2t+1}|h_{2t})p(y_{2t}|\tau_{2t}, h_{2t})p(\tau_{2t}|\tau_{2t-1}, g_{2t})p(h_{2t}|h_{2t-1})}_{=:C_t}$$
$$\underbrace{\int p(\tau_{2t+1}|\tau_{2t}, g_{2t+1})p(g_{2t+1}|g_{2t})p(g_{2t+2}|g_{2t+1})dg_{2t+1}}_{=:D_t}.$$

Since the integrals above are conditionally independent, we have

$$p(\boldsymbol{y}, \boldsymbol{\tau}, \boldsymbol{h}, \boldsymbol{g}_{2T}) = C_0 D_0 \prod_{t=1}^{T^*} C_t D_t. \tag{S.9}$$

SCDL for Scheme 2 We next move to adding h_t at even time periods to the previous integration scheme. The corresponding exact SCDL is given by

$$p(\boldsymbol{y}, \boldsymbol{\tau}, \boldsymbol{h}_{2T+1}, \boldsymbol{g}_{2T}) = \int p(\tau_1)p(h_1)p(g_1)\prod_{t=1}^{T^*} p(y_{2t+1}|\tau_{2t+1}, h_{2t+1})p(\tau_{2t+1}|\tau_{2t}, g_{2t+1})p(h_{2t+1}|h_{2t})$$
$$p(g_{2t+1}|g_{2t})p(y_{2t}|\tau_{2t}, h_{2t})p(\tau_{2t}|\tau_{2t-1}, g_{2t})p(h_{2t}|h_{2t-1})p(g_{2t}|g_{2t-1})dg_1 \dots dg_T dh_2 \dots dh_{T-1},$$

which again can be split (by conditioning on even g_ts and odd h_ts) into a product of T^*+1 integrals of the same form as in (S.9)

$$\underbrace{p(\tau_1)p(h_1)}_{=:C_1}\underbrace{\int p(g_1)dg_1}_{=:D_1}\prod_{t=1}^{T^*}\underbrace{p(y_{2t+1}|\tau_{2t+1}, h_{2t+1})p(\tau_{2t}|\tau_{2t-1}, g_{2t})}_{=:C_t} \tag{S.10}$$
$$\underbrace{\int p(y_{2t}|\tau_{2t}, h_{2t})p(h_{2t}|h_{2t-1})p(h_{2t+1}|h_{2t})dh_{2t}\int p(\tau_{2t+1}|\tau_{2t}, g_{2t+1})p(g_{2t+1}|g_{2t})p(g_{2t+2}|g_{2t+1})dg_{2t+1}}_{=:D_t}.$$

This factorisation of integrals is possible due to the independence of the processes for g_t and h_t and allows us to update each of them, as well as the corresponding parameters (e.g. g_0 and ω_g^2 for g_t), using only the relevant part of SCDL.

Updating τ_t becomes interesting in Scheme 2 as it involves conditioning on both g_t and h_t: the standard update for τ_t requires computing $p(\tau_t|\tau_{t-1}, g_t)p(\tau_{t+1}|\tau_t, g_{t+1})p(y_t|\tau_t, h_t)$ (i.e. these are the only terms in the complete data likelihood which involve τ_t). For even t this means that $g(t+1)$ and $h(t)$ are integrated, while for odd t only $g(t)$ is integrated. Comparing to Scheme 1, the odd t update is unchanged, however the even t update requires computing two one-dimensional integrals.

t **odd** In $p(\tau_t|\tau_{t-1}, g_t)p(\tau_{t+1}|\tau_t, g_{t+1})p(y_t|\tau_t, h_t)$ only g_t is integrated (g_{t+1} and h_t are imputed), hence the part of SCDL relevant for the update becomes

$$\left(\int p(\tau_t|\tau_{t-1}, g_t)p(g_t|g_{t-1})p(g_{t+1}|g_t)dg_t\right)p(\tau_{t+1}|\tau_t, g_{t+1})p(y_t|\tau_t, h_t),$$

which is the same as in Scheme 1.

t **even** In $p(\tau_t|\tau_{t-1}, g_t)p(\tau_{t+1}|\tau_t, g_{t+1})p(y_t|\tau_t, h_t)$ now g_{t+1} and h_t are integrated (and g_t imputed). Thus, the update-relevant part of SCDL now involves two one-dimensional

integrals

$$p(\tau_t|\tau_{t-1}, g_t)\left(\int p(\tau_{t+1}|\tau_t, g_{t+1})p(g_{t+1}|g_t)p(g_{t+2}|g_{t+1})dg_{t+1}\right) \times \left(\int p(y_t|\tau_t, h_t)p(h_t|h_{t-1})p(h_{t+1}|h_t)dh_t\right).$$

B.1.2 Hidden Markov model approximations

The integrals D_t cannot be evaluated analytically and need to be numerically approximated. We consider two different approaches.

Case (i) Fixed bins This approach is based on the method developed by Langrock et al. (2012) for the standard SV model and for Scheme 1 consists in relating $z_t = k$, the Markov chain being in state k, to the event $g_{2t+1} \in \mathcal{B}_k$, the permanent log-volatility in an odd time period $2t+1$ falling into the kth bin B_k. For Scheme 2, we consider a two dimensional Markov chain $\boldsymbol{z}_t$ and relate $\boldsymbol{z}_t = [k, l]^T$ to the events $g_{2t+1} \in \mathcal{B}_k^g$ and $h_{2t} \in \mathcal{B}_l^h$, where we allow for different bins for $g(t)$ and $h(t)$. For the sake of exposition, we focus below on Scheme 1, but generalising to more integrated states is straightforward.

We take equally spaced bins, each of length λ. Falling into bin B_k can be specified as e.g. lying in the interval $[b_{k-1}, b_k)$ or being equal to this interval's midpoint $b_k^* = \frac{b_{k-1}+b_k}{2}$. In particular, we consider approximation of the following form

$$D_t \approx \sum_{k=1}^{B} p(\tau_{2t+1}|\tau_{2t}, g_{2t+1} = b_k^*)p(g_{2t+2}|g_{2t+1} = b_k^*)p(g_{2t+1} \in B_i|g_{2t}). \tag{S.11}$$

The last term in (S.11) can be approximated as

$$p(g_{2t+1} \in B_k|g_{2t}) \approx \Phi\left(\frac{b_k - g_{2t}}{\sqrt{\omega_g^2}}\right) - \Phi\left(\frac{b_{k-1} - g_{2t}}{\sqrt{\omega_g^2}}\right),$$

which is adopted in Langrock et al. (2012), or using a simpler midpoint approximation

$$p(g_{2t+1} \in B_k|g_{2t}) \approx \lambda\varphi\left(\frac{b_k^* - g_{2t}}{\sqrt{\omega_g^2}}\right),$$

which we adopt in our application due to computing time.

Case (ii) Adaptive bins Instead of specifying the grid points, we can fix the probabilities for each bin (which in Case (i) needed to be determined) and consider quantiles

corresponding to intervals of equal probability. The vector of mid-quantiles $\boldsymbol{q}^*$ determines the midpoints at time $2t+1$, which are given as $\beta^*_{k,2t+1} = g_{2t} + \sqrt{\omega_g^2}\Phi^{-1}(q^*_k)$, $k = 1, \ldots, B$, where g_{2t} is the imputed permanent log-volatility for the previous time period. We approximate D_t as

$$D_t \approx \sum_{k=1}^{B} \varphi\left(\frac{\tau_{2t+1} - \tau_{2t}}{\exp(\beta^*_{k,2t+1}/2)}\right) \varphi\left(\frac{g_{2t+2} - \beta^*_{k,2t+1}}{\sqrt{\omega_g^2}}\right) \cdot \frac{1}{B},$$

since $\gamma_{ij,t} = \frac{1}{B}$, and these constant transition probabilities from an imputed state cancel out in the acceptance ratios.

B.2 Lapwings data model from Section 5.2

SCDL for the augmented data $(\boldsymbol{y}^T, \boldsymbol{N}_{\boldsymbol{a}}^T)^T$ is

$$p(\boldsymbol{y}, \boldsymbol{N_a}|\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{N_a}, \boldsymbol{\theta}) p(\boldsymbol{N_a}|\boldsymbol{\theta}), \tag{S.12}$$

which is still intractable, thus we approximate it using the HMM embedding in (16). Since $N_{1,t}$ follows a Poisson distribution, we only need to specify a truncation value N^* for the maximum population size for first-years for a fixed bin approach (i.e. we set $b_B = N^*$, with $b_0 = 0$). Since the observations $\boldsymbol{y}$ are conditionally independent from $\boldsymbol{N_1}$ given $\boldsymbol{N_a}$, summing over of $\boldsymbol{N_1}$ is necessary for the second term on the right hand side of (S.12), to obtain the marginal pmf for $\boldsymbol{N_a}$. The marginal pmf of $\boldsymbol{N_a}$ is

$$p(\boldsymbol{N_a}) = \sum_{\boldsymbol{N_1}} p(N_{a,0}) p(N_{1,0}) \prod_{t=1}^{T} p(N_{1,t}|N_{a,t-1})\, p(N_{a,t}|N_{a,t-1}, N_{1,t-1}). \tag{S.13}$$

We consider the "exact" approximation with the bin size equal to one, in which the approximation error is only due to truncating to N^*. A typical element in the product in (S.13) can be approximated as (for $t \geq 2$)

$$p(N_{a,t}|\boldsymbol{N}_{a,0:t-1}) = \sum_{k=0}^{N^*} \underbrace{\mathbb{P}(N_{1,t-1} = k|N_{a,t-2})}_{=:u_{k,t-1}} \underbrace{p(N_{a,t}|N_{a,t-1}, N_{1,t-1} = k)}_{=:p_{k,t}}, \tag{S.14}$$

where $p_{k,t}$ denotes the conditional pmf of $N_{a,t}$ given $N_{1,t-1} = k$ and $N_{a,t-1}$. Next, $u_{k,t}$ (collected in $\boldsymbol{u}_t = [u_{k,t}]_{k=1}^{N^*}$) denotes the "quasi-unconditional" probability of $N_{1,t} = k$ (in

the sense of the Markov structure but not in terms of $N_{a,t-1}$) and can be derived as

$$u_{k,t} = \mathbb{P}(N_{1,t} = k|N_{a,t-1}) = \sum_{l=0}^{N^*} \mathbb{P}(N_{1,t-1} = l|\boldsymbol{N}_{a,0:t-1})\mathbb{P}(N_{1,t} = k|N_{1,t-1} = l, \boldsymbol{N}_{a,0:t-1})$$
$$= \sum_{l=0}^{N^*} \underbrace{\mathbb{P}(N_{1,t-1} = l|N_{a,t-2})}_{=u_{l,t-1}} \underbrace{\mathbb{P}(N_{1,t} = k|N_{a,t-1})}_{=:\gamma_{lk,t}}.$$

In general, we have $\boldsymbol{u}_t = \boldsymbol{u}_{t-1}\Gamma_t$ with $\Gamma_t = [\gamma_{lk,t}]_{l,k=1}^{N^*}$ and $\gamma_{lk,t} = \mathbb{P}(N_{1,t} = k|N_{1,t-1} = l, \boldsymbol{N}_{a,0:t-1})$. However, since in our case $N_{1,t}$'s are mutually independent given $N_{a,t-1}$, the transition probabilities simplify to $\gamma_{lk,t} = \mathbb{P}(N_{1,t} = k|N_{a,t-1})$ for $k = 0, \ldots, N^* - 1$, while for $k = N^*$ we need $\gamma_{lk,t} = 1 - \sum_{j=0}^{N^*-1} \gamma_{lj,t}$ to ensure a valid probability distribution. Thus, the time varying state transition matrix Γ_t takes a simple form with rows equal to $(\boldsymbol{\gamma}_{11,t}, \ldots, \boldsymbol{\gamma}_{(N^*-1)(N^*-1),t}, \boldsymbol{\gamma}_{N^*N^*,t})$. Next, we express (S.14) in a convenient matrix notation

$$p(N_{a,t}|\boldsymbol{N}_{a,0:t-1}) = \underbrace{\begin{bmatrix} \gamma_{11,t-1} & \cdots & \gamma_{1N^*,t-1} \\ \vdots & \ddots & \vdots \\ \gamma_{11,t-1} & \cdots & \gamma_{1N^*,t-1} \end{bmatrix}}_{=\Gamma_{t-1}} \underbrace{\begin{bmatrix} p_{1,t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_{N^*,t} \end{bmatrix}}_{=:P_t} \underbrace{\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}}_{\mathbf{1}} = \Gamma_{t-1}P_t\mathbf{1}.$$

Combining (S.13) and (S.14) yields the HMM form for the joint pmf of the imputed states

$$p(\boldsymbol{N_a}) = \boldsymbol{u}_0 p(N_{a,0}) \left(\prod_{t=1}^{T} P_t\Gamma_t \right) \mathbf{1},$$

where $\boldsymbol{u}_0 = [p(N_{1,0} = 0), \ldots, p(N_{1,0} = N^*)]^T$ is the initial state distribution.

Since the real observations y_t, conditionally on $N_{a,t}$, are independent of $N_{1,t}$, the observation matrix becomes the identity matrix $\mathbb{I}$ scaled by $p(y_t|N_{a,t})$, i.e. $Q_t = p(y_t|N_{a,t})\mathbb{I}$.

Finally, the approximation to the SCDL (S.12) can be expressed as

$$p(\boldsymbol{y}, \boldsymbol{N_a}|\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{N_a})p(\boldsymbol{N_a}) = \boldsymbol{u}_0 p(N_{a,0}) \left(\prod_{t=1}^{T} P_t\Gamma_t Q_t \right) \mathbf{1}.$$

C Specification details of the HMM approximations

In this section we present how the general formulation of the HMM-based approximation to SCDL can be applied to the examples discussed in Sections 4 and 5.

C.1 Motivating example from Section 4.2

SSM from Section 4.2 is given by

$$\begin{aligned} y_t|x_{1,t}, x_{2,t} &\sim p(x_{1,t}, x_{2,t}), \\ x_{1,t+1}|x_{1,t}, x_{2,t} &\sim p(x_{1,t}, x_{2,t}), \\ x_{2,t+1}|x_{1,t}, x_{2,t} &\sim p(x_{1,t}, x_{2,t}), \\ x_{i,0} &\sim p(x_{i,0}), i = 1, 2 \end{aligned}$$

and we aim at imputing $x_{1,t}$ and integrating out $x_{2,t}$, which implies $T_{int} = T_{aug} = \{0, 1, \ldots, T\}$. Hence, the index functions $\tau(t)$, $a(t)$ and $o(t)$ are simply identities and we skip them below to simplify the exposition. Figure 3 illustrates the corresponding dependencies.

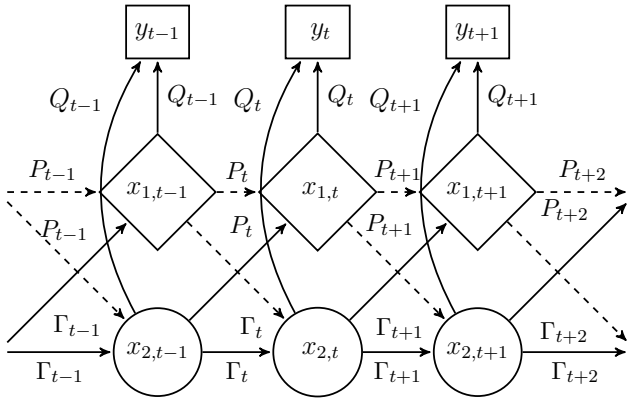


Figure 3: Illustration of combining DA and HMM structure. Conditionally independent observations added to the state specification from Figure 2a. Diamonds represent the imputed states, circles – the integrated states. Dashed lines used for the relations *from* the imputed (known) states.

The "quasi marginal" distribution[1] of the imputed state $x_{1,t}$ can be approximated as

$$p(x_{1,t}|\boldsymbol{x}_{1,0:t-1}) \approx \sum_{j=1}^{B} \underbrace{\mathbb{P}(x_{2,t-1} \in \mathcal{B}_j|\boldsymbol{x}_{1,0:t-2})}_{=:u_{j,t-1}} \underbrace{p(x_{1,t}|x_{1,t-1}, x_{2,t-1} \in \mathcal{B}_j)}_{=:p_{j,t}}, \tag{S.15}$$

where $p_{j,t}$ is the likelihood of the augmented state at t given the imputed state at $t-1$ was in the jth bin (and previous realizations of x_1, but these are treated as known) and $u_{j,t-1}$

[1]I.e. marginal in the sense of the Markov structure, not the augmented states which we treat as known.

is the unconditional probability of the hidden process x_2 falling into the jth bin at $t-1$. This "quasi-unconditional" probability can be expressed as

$$u_{k,t} = \mathbb{P}(x_{2,t} \in \mathcal{B}_k | \boldsymbol{x}_{1,0:t-1}) = \sum_{j=1}^{B} \underbrace{\mathbb{P}(x_{2,t-1} \in \mathcal{B}_j | \boldsymbol{x}_{1,0:t-2})}_{=u_{j,t-1}} \underbrace{\mathbb{P}(x_{2,t} \in \mathcal{B}_k | \boldsymbol{x}_{1,0:t-1}, x_{2,t-1} \in \mathcal{B}_j)}_{=:\gamma_{jk,t}}, \tag{S.16}$$

which corresponds to the standard result in HMM that the unconditional distributions in subsequent periods are related via the transition matrix $\Gamma_t = [\gamma_{jk,t}]_{k,j=1,\dots,B}$ as follows (see Zucchini et al., 2016, p.16, 32)

$$\boldsymbol{u}_t = \boldsymbol{u}_{t-1}\Gamma_t.$$

Next, the observations are conditionally independent, hence

$$p(y_t | \boldsymbol{x}_{1,0:t}) \approx \sum_{k=1}^{B} u_{k,t} \underbrace{p(y_t | x_{1,t}, x_{2,t} \in \mathcal{B}_k)}_{=:q_{k,t}}, \tag{S.17}$$

with $q_{k,t}$ denoting the likelihood of the observation at t given the hidden state in the same period t falling into bin k (and the imputed state at t).

Comparing (S.16) and (S.17) shows that the distributions of the same period t augmented states $x_{1,t}$ and "real" observations y_t are conditioned on the latent states from different periods, i.e. $t-1$ and t, respectively. This is a consequence of the general dependence structure in SSMs. The transition matrix at t captures this change of the underlying state so that combining of all there parts (S.15), (S.16) and (S.17) results in

$$p(y_t, x_{1,t} | \boldsymbol{x}_{1,0:t-1}) \approx \sum_{j=1}^{B} \sum_{k=1}^{B} u_{j,t-1} p_{j,t} \gamma_{jk,t} q_{k,t}.$$

To compute the HMM-based approximation to the SCDL we consider *forward probabilities* $\alpha_{\mathbf{t}}$ of the imputed states $x_{1,t}$ and observations y_t (Zucchini et al., 2016, Sec. 2.3.2) defined as

$$\begin{aligned} \boldsymbol{\alpha}_t &= p(x_{1,0}) \boldsymbol{u}_0 \prod_{s=1}^{t} P_s \Gamma_s Q_s, \qquad t = 1, 2, \dots T^*, \\ \boldsymbol{\alpha}_0 &= p(x_{1,0}) \boldsymbol{u}_0 Q_0, \end{aligned}$$

with $\boldsymbol{u}_0 = (\mathbb{P}(x_{2,0} \in \mathcal{B}_1), \dots, \mathbb{P}(x_{2,0} \in \mathcal{B}_B))$ being the initial distribution of the latent state and $Q_0 = \mathbb{I}$. It follows from this definition that the forward probabilities can be expressed

recursively as

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} P_t \Gamma_t Q_t,$$

so that the required approximation to the SCDL is given by

$$\hat{p}_B(\boldsymbol{y}, \boldsymbol{x_1}) = p(x_{1,0}) \boldsymbol{u}_0 \boldsymbol{\alpha}_{T^*} \mathbb{I}.$$

Notice that the transition matrix Γ_t is a full matrix, however in some cases, e.g. the lapwing population model, the transition matrix can take a simpler form e.g. it is "column-wise constant": $\gamma_{lk,t} = \mathbb{P}(x_{2,t} = k | \boldsymbol{x}_{1,0:t-1}),\ \forall l$ (each row is the same). On the other hand, the augmented observation matrix and the real observation matrix have diagonal forms $P_t = \text{diag}\ (p_{j,t})_{j=1,\dots,B}$ and $Q_t = \text{diag}\ (q_{k,t})_{j=k,\dots,B}$, respectively. Using the notation introduced in Section 4.2 we can write

$$\hat{p}_B(\boldsymbol{y}, \boldsymbol{x_1}) = p(x_{1,0}) \boldsymbol{u}_0 Q_0 \prod_{t=1}^{T^*} (P_t \Gamma_t Q_t)\, \mathbf{1}.$$

We can verify the above results be explicitly calculating

$$\begin{aligned}
\boldsymbol{\alpha}_{t-1} P_t \Gamma_t Q_t \mathbf{1} &= \begin{bmatrix} \alpha_{1,t-1} & \dots & \alpha_{B,t-1} \end{bmatrix} \begin{bmatrix} p_{1,t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & p_{B,t} \end{bmatrix} \begin{bmatrix} \gamma_{11,t} & \dots & \gamma_{1B,t} \\ \vdots & \ddots & \vdots \\ \gamma_{B1,t} & \dots & \gamma_{BB,t} \end{bmatrix} \begin{bmatrix} q_{1,t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & q_{B,t} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \alpha_{1,t-1} & \dots & \alpha_{B,t-1} \end{bmatrix} \begin{bmatrix} p_{1,t}\gamma_{11,t} & \dots & p_{1,t}\gamma_{1B,t} \\ \vdots & \ddots & \vdots \\ p_{B,t}\gamma_{B1,t} & \dots & p_{B,t}\gamma_{BB,t} \end{bmatrix} \begin{bmatrix} q_{1,t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & q_{B,t} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \alpha_{1,t-1} & \dots & \alpha_{B,t-1} \end{bmatrix} \begin{bmatrix} p_{1,t}\gamma_{11,t}q_{1,t} & \dots & p_{1,t}\gamma_{1B,t}q_{B,t} \\ \vdots & \ddots & \vdots \\ p_{B,t}\gamma_{B1,t}q_{1,t} & \dots & p_{B,t}\gamma_{BB,t}q_{B,t} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \underbrace{\sum_{j=1}^{B} \alpha_{j,t-1} p_{j,t} \gamma_{j1,t} q_{1,t}}_{=\alpha_{1,t}} & \dots & \underbrace{\sum_{j=1}^{B} \alpha_{j,t-1} p_{j,t} \gamma_{jB,t} q_{B,t}}_{=\alpha_{B,t}} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \boldsymbol{\alpha}_t \mathbf{1}
\end{aligned}$$

and expressing

$$\hat{p}_B(\boldsymbol{y}, \boldsymbol{x_1}) = \sum_{k_0=1}^{B} \sum_{k_1=1}^{B} \cdots \sum_{k_{T^*}=1}^{B} p(x_{1,0}) u_{k_0,0} \prod_{t=1}^{T^*} p_{k_{t-1},t} \gamma_{k_{t-1}k_t,t} q_{k_t,t}.$$

C.2 SV model

Basic SV model The SCDL for the basic SV model can be expressed as

$$p(y, \boldsymbol{h_{2T}}|\boldsymbol{\theta}) = p(h_0) \int p(h_1|h_0)p(y_1|h_0) \Bigg(\prod_{t=1}^{T^*} p(h_{2t+1}|h_{2t})p(y_{2t+1}|h_{2t+1}) \times p(h_{2t}|h_{2t-1})p(y_{2t}|h_{2t}) \Bigg) dh_1 \dots dh_{T^*}, \quad \text{(S.18)}$$

where $T^* = \frac{T-1}{2}$ (we assume T odd). Since we impute volatilities at even time periods the Markov chain is given by $\{z_t\} = \{h_{2t+1}\}$ for $t = 1, \dots, T^*$ and its transition matrix has the form

$$\begin{aligned} \Gamma_t &= \begin{bmatrix} \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t-1} \in \mathcal{B}_1, h_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t-1} \in \mathcal{B}_1, h_{2t}) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t-1} \in \mathcal{B}_B, h_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t-1} \in \mathcal{B}_B, h_{2t}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t}) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t}) \end{bmatrix}. \end{aligned}$$

We can see that the rows of Γ_t are the same, which means that the hidden states are conditionally independent given the imputed states. For the augmented observation matrix P_t we have

$$P_t = \text{diag } \left(p(h_{2t}|h_{2t-1} \in \mathcal{B}_j) \right)_{j=1,\dots,B}.$$

The observation matrix has the form

$$Q_t = \text{diag } \left(p(y_{2t}, y_{2t+1}|h_{2t+1} \in \mathcal{B}_j, h_{2t}) \right)_{j=1,\dots,B}.$$

Inserting Γ_t, P_t and Q_t in (16) with $\tau(t) = 2t+1$, $a(t) = 2t$ and $o(t) = \{2t, 2t+1\}$ leads to

$$\hat{p}_B(y, \boldsymbol{h_{2T}}) = p(h_0)\mathbf{u}_0 Q_0 \prod_{t=1}^{T^*} P_t \Gamma_t Q_t \mathbf{1}, \quad \text{(S.19)}$$

where $\mathbf{u}_0 = \begin{bmatrix} \mathbb{P}(h_1 \in \mathcal{B}_k|h_0) & \dots & \mathbb{P}(h_1 \in \mathcal{B}_B|h_0) \end{bmatrix}$ and $Q_0 = \text{diag } (y_1|h_1 \in \mathcal{B}_k)_{k=1,\dots,B}$. Then (S.19) is an HMM-based approximation to (S.18) converging to its true value in $B \to \infty$ and $b_0 \to -\infty$, $b_B \to \infty$.

Figure 4 illustrates the HMM approximation and shows dependencies relevant for a single imputation problem.

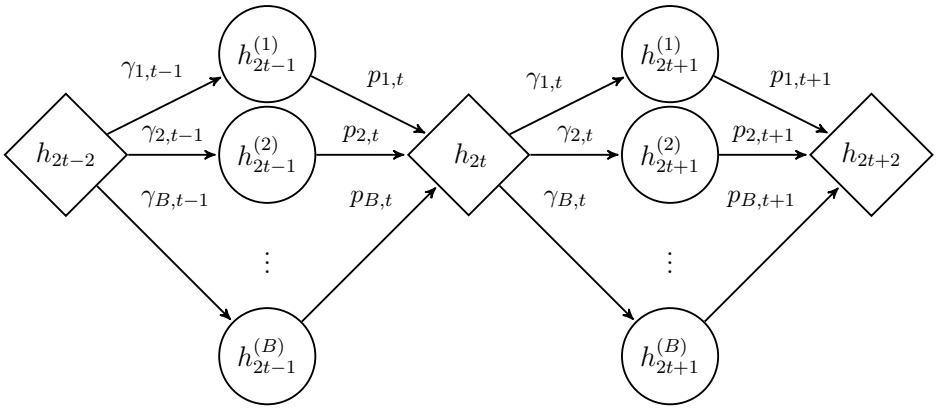


Figure 4: **SV model**: combining DA and the HMM-based integration. Diamonds represent the imputed states, circles – the states being integrated out. $h_t^{(k)}$ denotes $h_t \in \mathcal{B}_k$. The graph presents a single imputation problem of h_{2t} with the associated integrations.

SVML model For the SVML model we only need to adjust the matrices P_t and Q_t as the dependence structure of the observations remains unchanged

$$\Gamma_t = \begin{bmatrix} \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t}, y_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t}, y_{2t}) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(h_{2t+1} \in \mathcal{B}_1 | h_{2t}, y_{2t}) & \dots & \mathbb{P}(h_{2t+1} \in \mathcal{B}_B | h_{2t}, y_{2t}) \end{bmatrix},$$

$$P_t = \text{diag}\left(p(h_{2t} | h_{2t-1} \in \mathcal{B}_j, y_{2t-1})\right)_{j=1,\dots,B}.$$

C.3 Lapwing population model

The approximation for the lapwings model is a special case of scheme used for the general model discussed in the Section C.1, with the transition matrix Γ_t having equal rows. The HMM is here given as $\{z_t\} = \{N_{1,t}\}$ for $t = 0, \dots, T$ and we again set $T_{int} = T_{aug} = \{0, 1, \dots, T\}$, so that the index functions $\tau(t)$, $a(t)$ and $o(t)$ are simply identities. The

transition matrix has the form

$$\Gamma_t = \begin{bmatrix} \mathbb{P}(N_{1,t} = b_1^* | N_{1,t-1} = b_1^*, \boldsymbol{N}_{a,0:t-1}) & \dots & \mathbb{P}(N_{1,t} = b_B^* | N_{1,t-1} = b_1^*, \boldsymbol{N}_{a,0:t-1}) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(N_{1,t} = b_1^* | N_{1,t-1} = b_B^*, \boldsymbol{N}_{a,0:t-1}) & \dots & \mathbb{P}(N_{1,t} = b_B^* | N_{1,t-1} = b_B^*, \boldsymbol{N}_{a,0:t-1}) \end{bmatrix}$$
$$= \begin{bmatrix} \mathbb{P}(N_{1,t} = b_1^* | N_{a,t-1}) & \dots & \mathbb{P}(N_{1,t} = b_B^* | N_{a,t-1}) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(N_{1,t} = b_1^* | N_{a,t-1}) & \dots & \mathbb{P}(N_{1,t} = b_B^* | N_{a,t-1}) \end{bmatrix},$$

with $b_k^* = k$, for $k = 0, \dots, N^*$. We can see that for each column of Γ_t its elements are the same. For the augmented observation matrix P_t we have

$$P_t = \text{diag}\left(p(N_{a,t} | N_{a,t-1}, N_{1,t-1} = b_j^*)\right)_{j=1,\dots,B},$$

so P_t and Γ_t condition on the same hidden state. The observation matrix has a simple form

$$Q_t = p(y_t | N_{a,t}) \mathbb{I}.$$

Inserting Q_t, P_t and Q_t in (16) leads to

$$\hat{p}_B(y, \boldsymbol{N_a}) = p(N_{a,0}) \mathbf{u}_0 Q_0 \prod_{t=1}^{T^*} P_t \Gamma_t Q_t \mathbf{1}, \tag{S.20}$$

where $\mathbf{u}_0 = \begin{bmatrix} \mathbb{P}(N_{1,0} \in \mathcal{B}_k) & \dots & \mathbb{P}(N_{1,0} \in \mathcal{B}_B) \end{bmatrix}$ and $Q_0 = \mathbb{I}$. Then (S.20) is an HMM-based approximation to (S.12) converging to its true value in $B \to \infty$ and $b_B \to \infty$.

D Effective sample size

Since the samples generated by MCMC algorithms are not independent, the standard variance estimator cannot be used to measure the variance of the empirical average delivered by an MCMC algorithm. The asymptotic variance σ^2_{MCMC} of the (stationary) Markov chain $X_1, X_2, \dots$ is given by

$$\sigma^2_{\text{MCMC}} = \text{Var}[X_i] \underbrace{\left(1 + 2 \sum_{k=1}^{\infty} \rho(k)\right)}_{\text{IF}}, \tag{S.21}$$

where $\rho(k)$ is the kth order serial correlation (Geyer, 2011). The term in the parentheses in (S.21) is referred to as the *inefficiency factor* (IF, Pitt et al., 2012). High values of autocorrelation, typically reported for MCMC sampling, lead to the standard variance estimator underestimating the true variance σ^2_{MCMC}. A common measure to assess the deterioration in the sampling efficiency due to the draws autocorrelation is the *effective sample size* (ESS) defined as

$$\text{ESS} = \frac{M}{IF},$$

where M is the sample size (Robert and Casella, 2004, Ch. 12.3.5). It indicates what the size of an independent sample would be, had it the same variance as the MCMC sample.

In practice, one typically cannot compute the IF directly and needs to estimate it instead. As noted by Robert and Casella (2004, Ch. 12.3.5) estimation of IF is a "delicate issue", as it contains an infinite sum. A possible solution to this problem is set a cut-off value K for the autocorrelation terms being summed up: $\widehat{IF} = 1 + 2\sum_{k=1}^{K} \hat{\rho}(k)$. The choice of K poses the risk of subjectiveness; setting K to the lowest lag at which $\hat{\rho}(k)$ become insignificant seems to be a reasonable solution suggested by e.g. Kass et al. (1998) or Pitt et al. (2012) and this is the approach we take here.

E Lapwings dataset

The lapwings dataset plays an important role in statistical ecology and has served as an illustration in several handbooks (see King, 2011; King et al., 2010) and papers (e.g. Besbeas et al., 2002) in this field. It was also used as an example of a complex statistical model by e.g. Goudie et al. (2018). One of the main reasons for such a particular interest in this species is a sharp decline in its population in recent years: its European population is considered as *near threatened* by International Union for Conservation of Nature (2018), while in Britain in particular is has been moved to the *red list* of species of conservation concern, see The Royal Society for the Protection of Birds (2018) (i.e. of the highest conservation priority, with species needing urgent action) from the *amber list* (mentioned by previous literature, see Besbeas et al., 2002; Brooks et al., 2004). The population serves

as an indicator species for other farmland birds, giving us an insight into the dynamics of similar bird species.

We follow the approach of Besbeas et al. (2002) and use three datasets for the lapwings application: the count census data for the population index, the weather data on the number of frost days, and the ring-recovery data. Combining of independent sources of data underlies the integrated population modeling (IPM) framework and allows for a more precise parameter estimation. This is due to the survival parameters $\alpha_i, \beta_i, i \in 1, a$, being common to the state space model for the census data and to the ring-recovery model

Census data The census data are derived from the Common Birds Census (CBC) of the British Trust for Ornithology, which recently has been replaced by the Breeding Bird Survey. The dataset is constructed as annual estimates of the number of breeding female lapwings based on annual counts made at a number of sites around the UK. Since only a small fraction of sites are surveyed each year, the index can be seen as a proxy for the total population size. For comparability, we use the same time span as Brooks et al. (2004) and King (2011), i.e. from 1965 to 1998. The choice of the starting year is there motivated by the fact that in earlier years the index protocol was being standardized. Finally we note that year 1965 is associated with time index $t = 3$, for consistency with the ring-recovery data (to be discussed below) which start in 1963. The left panel in Figure 5 presents the census data.

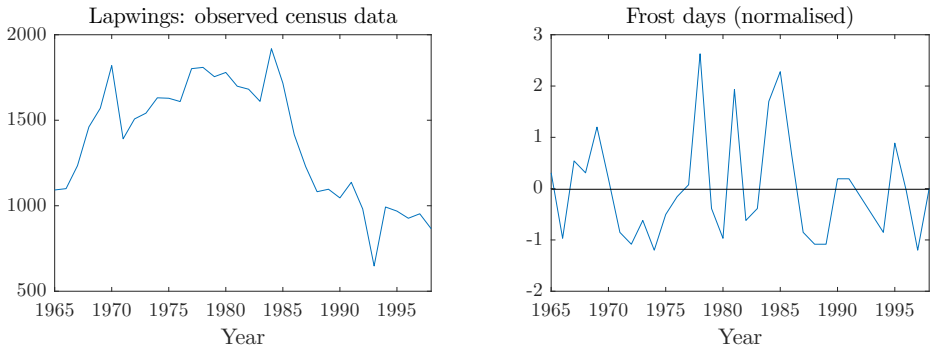


Figure 5: Lapwings census data and normalised frost days.

Weather data For bird species there is a natural relationship between the survival probabilities and the weather conditions, most importantly winter severity. Following Besbeas et al. (2002) we measure this factor for year t by the number of days between April of year t and March of year $(t+1)$ inclusive in which the temperature in Central England fell below freezing and denote it by $fdays_t$. We further normalize $fdays_t$ to obtain f_t which we use as a regressor in the logistic regression for the survival probabilities (see the right panel in Figure 5). As noted by King (2011), normalization of covariates is done to improve the mixing of the sampling scheme and to facilitate the interpretation of the parameters of the logistic regression (intercept and slope).

Ring–recovery data Ring–recovery studies aim at estimating demographic parameters of the population under consideration including first-year survival probabilities and adult survival probabilities. These studies consist in marking individuals (e.g. with a ring or a tag) at the beginning of period t and then releasing them. In subsequent periods $t+1, t+2, \ldots$ the number of dead animals is recorded, where it is assumed that any recovery of a dead animal is immediate. For lapwings, the ringed birds are chicks ("fist-years") and a "period" corresponds to a "bird year" i.e. 12 months from April to March. We analyze the ring-recovery data for the releases from 1963 to 1997, with the recoveries up to 1998.

Ring-recovery data are stored in an array, an example of which is provided in Table 1. The first column corresponds to the number of ringed animals in a given year R_t, $t = 1, \ldots, T-1$, and the subsequent columns report the number of rings $m_{t,s}$ recovered in the interval $(s, s+1]$, $s = 1, \ldots, T-1$, from animals released in year t. Obviously, $m_{t,s} = 0$ for $t > s$. Finally, we denote by $m_{t,T}$ the number of individuals ringed in year t but never seen again (their rings are not recovered), where $m_{t,T} = R_t - \sum_{s=1}^{T-1} m_{t,s}$.

The parameters of interest are $\phi_{1,t}$, $\phi_{a,t}$ and λ_t. The former two are the conditional probabilities of survival until year $t+1$ of a first-year and an adult, respectively, given such an individual is alive in year t. The latter one is the conditional probability of recovering a ring at time t, given the individual dies in the interval $(t-1, t]$. Each row $\boldsymbol{m}_t = \{m_{t,s}\}_{s=1}^{T}$, $t = 1, \ldots, T-1$, of the m-array is multinomially distributed: $\boldsymbol{m}_t \sim \mathcal{MN}(R_t, \boldsymbol{q}_t)$ ($\mathcal{MN}$ denotes the multinomial distribution), where $\boldsymbol{q}_t = \{q_{t,s}\}_{s=1}^{T}$ are the multinomial cell

Year of	Number	Year of Recovery s										
Ringing t	Ringed	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974
1963	1147	14	4	1	2	1	0	1	1	0	0	0
1964	1285		20	3	4	0	1	1	0	0	0	0
1965	1106			10	1	2	2	0	2	2	1	1
1966	1615				9	7	4	2	1	1	0	0
1967	1618					12	1	6	2	0	0	1
1968	2120						9	6	4	0	2	2
1969	2003							10	8	5	3	1
1970	1963								8	3	2	0
1971	2463									4	1	1
1972	3092										7	2
1973	3442											15

Table 1: A fragment of Ring-Recovery Data for lapwings for the years 1963-1973, table from King (2011).

probabilities specified for $t = 1, \ldots, T-1$ as[2]

$$q_{t,s} = \begin{cases} 0, & s = 1, \ldots, t-1, \\ (1-\phi_{1,t})\lambda_t & s = t, \\ \phi_{1,t}\left(\prod_{k=t+1}^{s-1}\phi_{a,k}\right)(1-\phi_{a,s})\lambda_s, & s = t+1, \ldots, T-1, \\ 1-\sum_{s=1}^{T+1} q_{t,s}, & s = T. \end{cases}$$

The likelihood of the m-array is then given by

$$p(\mathbf{m}|\boldsymbol{\phi_1}, \boldsymbol{\phi_a}, \boldsymbol{\lambda}) \propto \prod_{\mathbf{t=1}}^{\mathbf{T-1}} \prod_{\mathbf{s=t}}^{\mathbf{T}} \mathbf{q}_{\mathbf{t,s}}^{\mathbf{m_{t,s}}}.$$

The array $\boldsymbol{m} = [m_{t,s}]_{t=1,\ldots,T-1}^{s=1,\ldots,T}$ is a sufficient statistic for ring-recovery data.

Following Besbeas et al. (2002) we assume that the time-varying recovery rate λ_t follows a logistic regression given by

$$\operatorname{logit}\lambda_t = \log\left(\frac{\lambda_t}{1-\lambda_t}\right) = \alpha_\lambda + \beta_\lambda \tilde{t}_t,$$

where $\tilde{t}$ is the normalized time index.

[2]For $s = t+1$ we put $\prod_{k=t+1}^{s-1}$:=1.

F Additional results

F.1 Lapwings data

Figure 6 illustrates the posterior means and 95% credible intervals (CI) for the adult population comparing the accuracy of the full DA with that of the SCDA methods (separately for the adaptive intervals and fixed bins). We can see that all the methods deliver virtually the same posterior means and comparable 95% symmetric CI, with only the fixed bin case with 10 bins deviating slightly from all other methods. Interestingly, 10 adaptive bins give very comparable estimates to the other approaches in this case, indicating an increased accuracy of the adaptive approach.

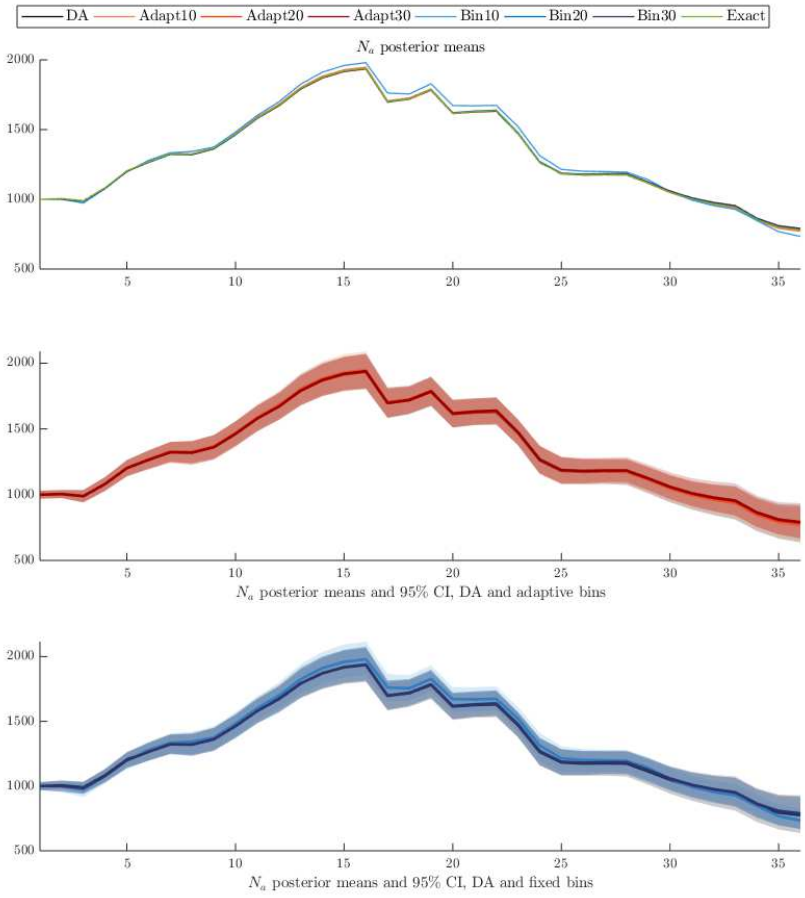


Figure 6: Lapwings data: the posterior means and 95% CI for the adult population.

Tables 2 and 3 report the estimation results for all the methods considered for the parameters and selected imputed states, respectively. Figure 7 shows the ACF plots for selected imputed states.

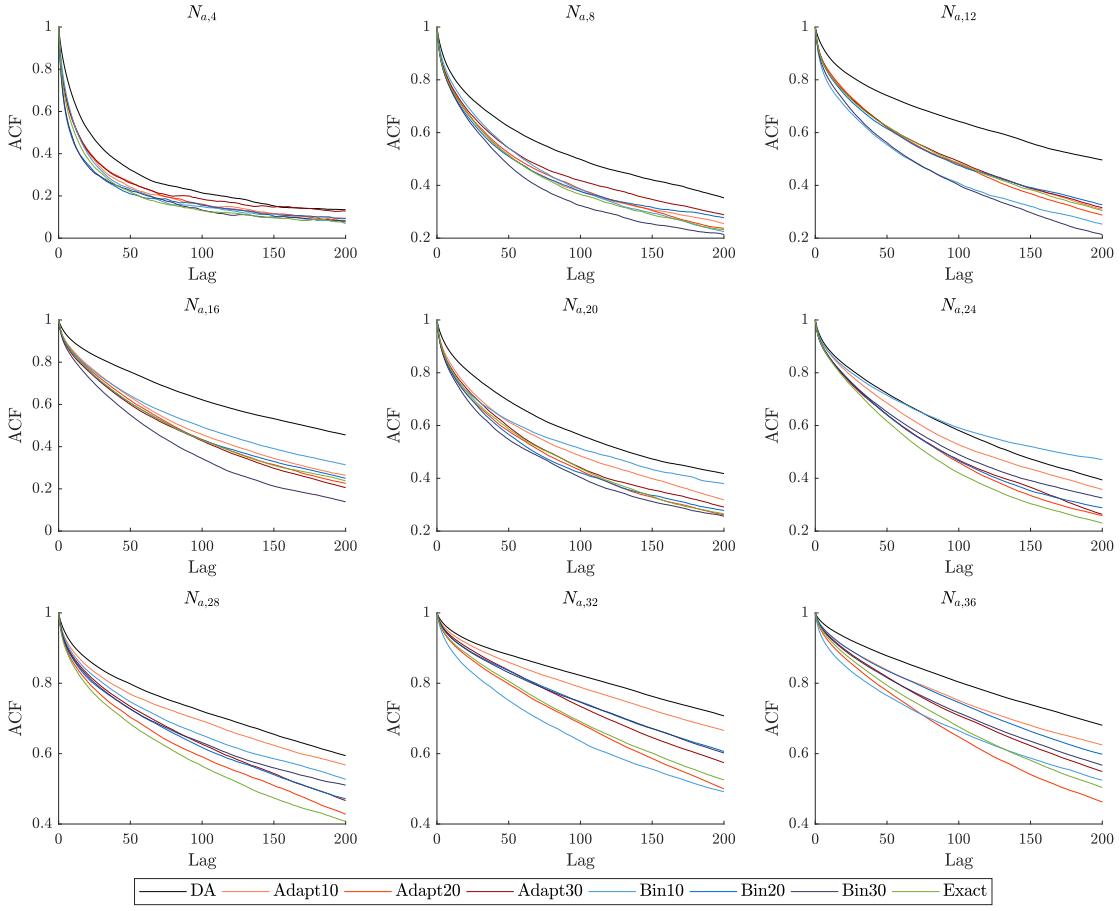


Figure 7: Lapwings data: ACF plots for the adult population.

Table 2: Lapwings data: parameter posterior means, standard deviations and ESSs. The highest ESS and ESS/s for each parameter in bold. Computing times (in s) in square brackets.

Method		α_1	α_a	α_ρ	α_λ	β_1	β_a	β_ρ	β_λ	σ_y^2
DA	Mean	0.547	1.574	-1.189	-4.578	-0.164	-0.240	-0.348	-0.364	30180
	(Std)	(0.068)	(0.071)	(0.091)	(0.035)	(0.062)	(0.039)	(0.043)	(0.04)	(8890)
	ESS	685	124	112	1089	1050	389	106	8206	1245
[1204 s]	ESS/sec.	0.57	0.10	0.09	0.90	0.87	0.32	0.09	6.82	1.03
Adapt10	Mean	0.547	1.564	-1.180	-4.580	-0.163	-0.239	-0.350	-0.364	30355
	(Std)	(0.068)	(0.070)	(0.092)	(0.035)	(0.061)	(0.040)	(0.040)	(0.040)	(8928)
	ESS	1490	390	316	**3035**	2777	527	126	7492	1852
[978 s]	ESS/sec.	1.52	**0.40**	**0.32**	**3.10**	**2.84**	0.54	0.13	7.66	1.89
Adapt20	Mean	0.544	1.564	-1.173	-4.581	-0.162	-0.238	-0.342	-0.363	30002
	(Std)	(0.069)	(0.072)	(0.094)	(0.035)	(0.060)	(0.039)	(0.039)	(0.040)	(8759)
	ESS	1359	395.695	324.918	2720.786	2685.188	586.964	243.425	8212	2075
[1068 s]	ESS/sec.	1.27	0.37	0.30	2.55	2.515	**0.55**	0.23	7.69	1.94
Adapt30	Mean	0.542	1.561	-1.166	-4.581	-0.162	-0.241	-0.339	-0.363	30311
	(Std)	(0.069)	(0.071)	(0.092)	(0.036)	(0.061)	(0.039)	(0.040)	(0.040)	(8888)
	ESS	1438	322	243	2736	2471	564	196	7146	2129
[1025 s]	ESS/sec.	1.40	0.31	0.24	2.67	2.41	0.55	0.19	6.97	2.08
Fixed10	Mean	0.512	1.441	-1.044	-4.599	-0.207	-0.205	-0.348	-0.353	29992
	(Std)	(0.070)	(0.055)	(0.063)	(0.034)	(0.050)	(0.039)	(0.022)	(0.040)	(8837)
	ESS	942	34	37	562	181	105	**282**	8771	1627
[1022 s]	ESS/sec.	0.92	0.03	0.04	0.55	0.18	0.10	**0.28**	**8.58**	1.59
Fixed20	Mean	0.546	1.570	-1.179	-4.579	-0.170	-0.240	-0.343	-0.364	30156
	(Std)	(0.069)	(0.069)	(0.090)	(0.035)	(0.061)	(0.039)	(0.040)	(0.040)	(8802)
	ESS	1250	270	210	2328	2566	525	139	8582	2270
[1060 s]	ESS/sec.	1.18	0.25	0.20	2.19	2.42	0.49	0.13	8.09	2.14
Fixed30	Mean	0.545	1.562	-1.170	-4.580	-0.162	-0.240	-0.342	-0.363	30012
	(Std)	(0.069)	(0.073)	(0.095)	(0.035)	(0.061)	(0.039)	(0.040)	(0.040)	(8698)
	ESS	**1758**	**438**	**329**	2902	**2873**	502	208	7613	2706
[1136 s]	ESS/sec.	**1.55**	0.39	0.29	2.55	2.52	0.44	0.18	6.70	**2.38**
Exact	Mean	0.545	1.564	-1.175	-4.580	-0.162	-0.240	-0.345	-0.363	30063
	(Std)	(0.068)	(0.069)	(0.090)	(0.035)	(0.060)	(0.039)	(0.042)	(0.040)	(8771)
	ESS	1632	433	362	3059	2659	**720**	191	**8860**	**2734**
[2855 s]	ESS/sec.	0.57	0.15	0.13	1.07	0.93	0.25	0.07	3.10	0.96

Table 3: Lapwings data: posterior means, standard deviations and ESSs for selected imputed states. The highest ESS and ESS/s for each state in bold. Computing times (in s) in square brackets.

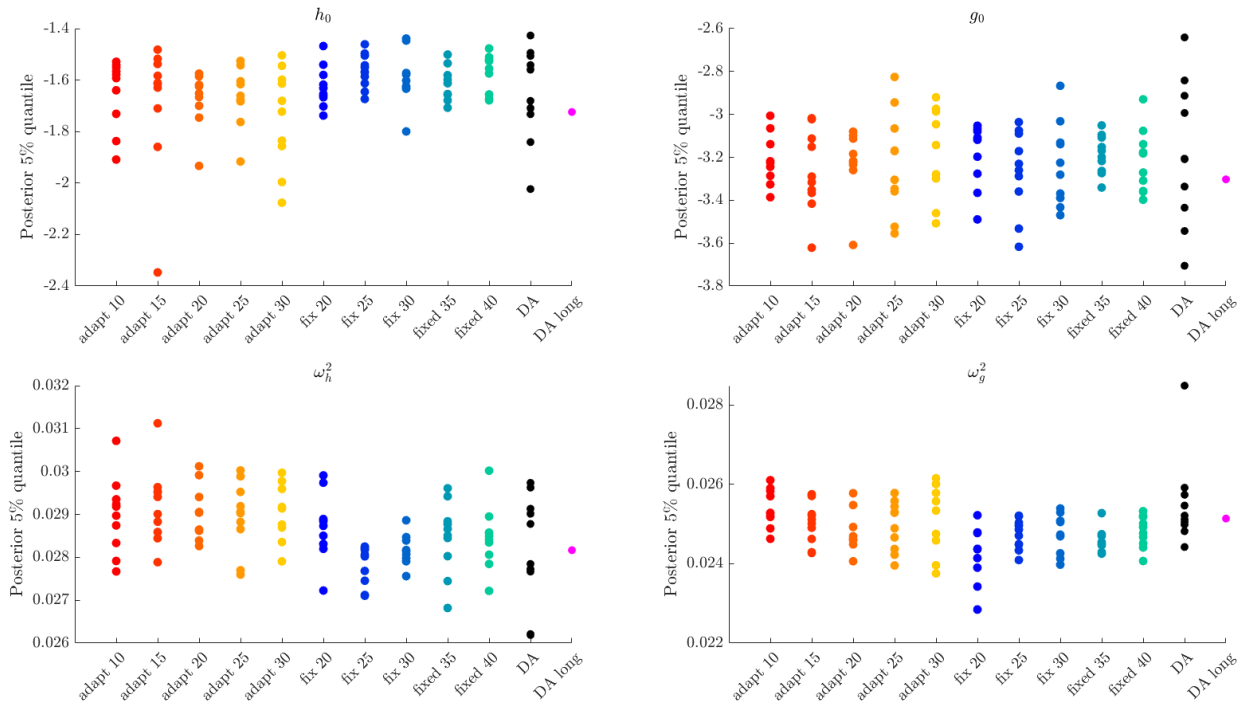
Method		Na_4	Na_8	Na_{12}	Na_{16}	Na_{20}	Na_{24}	Na_{28}	Na_{32}	Na_{36}	Na_{min}	Na_{max}
DA	Mean	1083.51	1325.45	1674.38	1935.84	1614.73	1264.61	1174.20	964.85	776.15	1113.76	1083.51
	(Std)	(26.31)	(43.08)	(52.06)	(67.99)	(53.97)	(53.68)	(49.46)	(59.57)	(72.07)	(50.97)	(26.31)
	ESS	460	179	120	134	147	154	59	62	68	55	460
[1204 s]	ESS/sec.	0.38	0.15	0.10	0.11	0.12	0.13	0.05	0.05	0.06	0.05	0.38
Adapt10	Mean	1083.46	1326.45	1681.57	1947.34	1621.63	1268.14	1174.99	962.14	770.63	1113.57	1083.46
	(Std)	(26.25)	(43.50)	(51.51)	(68.96)	(53.63)	(56.49)	(49.99)	(56.98)	(66.91)	(51.06)	(26.25)
	ESS	880	**393**	318	335	295	234	43	46	51	42	879
[978 s]	ESS/sec.	**0.90**	**0.40**	0.32	0.34	0.30	0.24	0.04	0.05	0.05	0.043	**0.90**
Adapt20	Mean	1081.74	1320.73	1670.04	1934.08	1615.29	1268.92	1181.09	973.61	785.42	1121.17	1081.74
	(Std)	(27.45)	(44.99)	(52.41)	(67.96)	(51.42)	(53.69)	(46.37)	(51.32)	(61.67)	(46.41)	(27.45)
	ESS	792	301	263	413	299	311	174	150	**161**	167	792
[1068 s]	ESS/sec.	0.74	0.28	0.25	0.39	0.28	0.29	0.16	0.14	0.15	0.16	0.74
Adapt30	Mean	1081.74	1319.29	1670.94	1938.90	1617.39	1268.43	1184.04	978.16	791.52	1124.32	1081.74
	(Std)	(27.37)	(45.49)	(51.54)	(65.25)	(53.62)	(54.57)	(48.40)	(56.59)	(68.24)	(49.64)	(27.37)
	ESS	597	278	247	434	326	306	181	155	164	**171**	597
[1025 s]	ESS/sec.	0.58	0.27	0.24	0.42	0.32	0.30	**0.18**	0.15	**0.16**	**0.17**	0.58
Fixed10	Mean	1075.91	1343.03	1699.24	1979.99	1671.88	1313.00	1194.68	954.96	733.69	1140.26	1075.91
	(Std)	(26.81)	(43.44)	(50.66)	(62.41)	(48.44)	(54.65)	(42.54)	(39.70)	(41.44)	(43.18)	(26.81)
	ESS	868	310	327	243	173	108	97	**163**	92	88	868
[1022 s]	ESS/sec.	0.85	0.30	0.32	0.24	0.17	0.10	0.10	**0.16**	0.09	0.09	0.85
Fixed20	Mean	1079.80	1319.96	1671.12	1939.17	1619.88	1270.62	1183.22	976.20	788.92	1123.72	1079.80
	(Std)	(25.97)	(44.14)	(51.98)	(67.01)	(52.05)	(54.00)	(46.78)	(53.32)	(64.39)	(47.5)	(25.97)
	ESS	785	279	225	331	344	328	74	58	64	67	785
[1060 s]	ESS/sec.	0.74	0.26	0.21	0.31	0.32	**0.31**	0.07	0.05	0.06	0.06	0.74
Fixed30	Mean	1079.48	1320.18	1671.22	1936.26	1615.844	1268.040	1181.895	975.230	787.55	1121.97	1079.48
	(Std)	(25.72)	(43.24)	(48.00)	(62.93)	(50.327)	(53.921)	(46.464)	(53.46)	(65.12)	(47.10)	(25.72)
	ESS	**911**	370	**374**	**505**	347	247	111	90	98	102	**911**
[1136 s]	ESS/sec.	0.80	0.33	**0.33**	**0.44**	**0.30**	0.22	0.10	0.08	0.09	0.09	0.80
Exact	Mean	1083.13	1324.13	1675.32	1939.63	1615.88	1265.87	1176.69	968.36	780.24	1116.33	1083.13
	(Std)	(27.19)	(44.55)	(52.06)	(67.38)	(53.86)	(54.55)	(46.09)	(54.11)	(66.99)	(47.25)	(27.19)
	ESS	902	349	293	366	**402**	**418**	**197**	122	117	168	902
[2855 s]	ESS/sec.	0.32	0.12	0.10	0.13	0.14	0.15	0.07	0.04	0.04	0.06	0.32

F.2 UCSV model

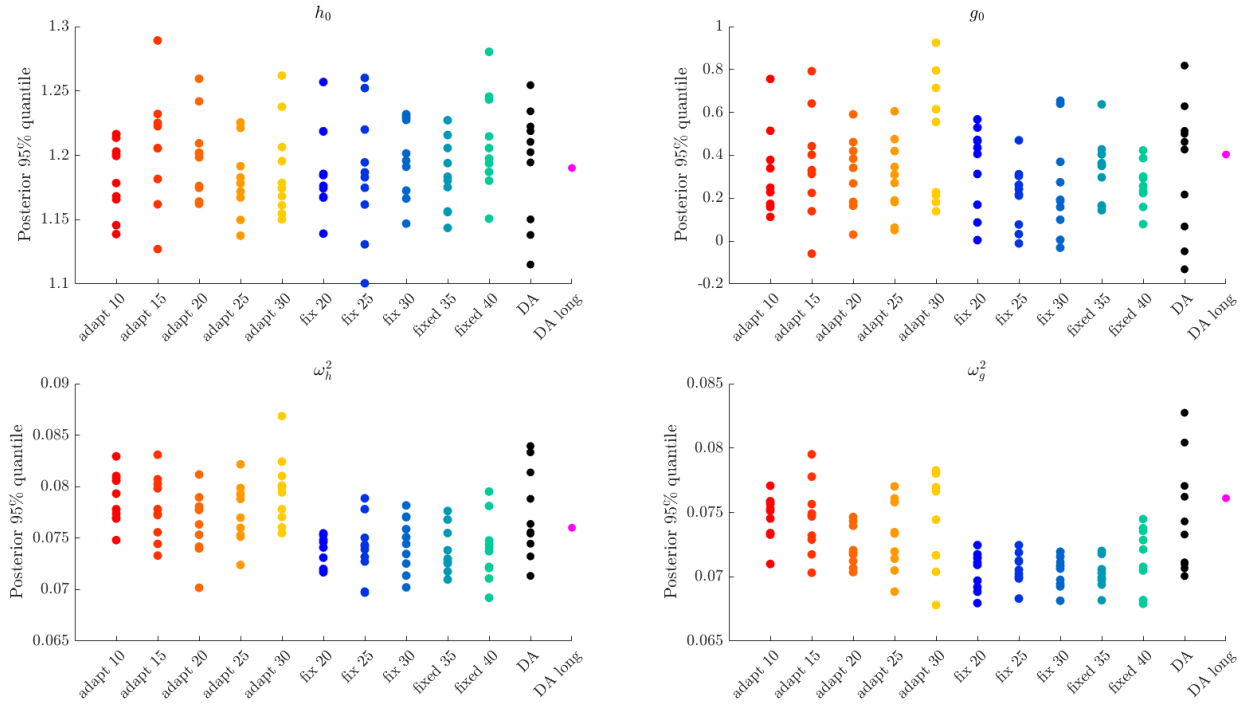
Table 4 quantitatively summarises Figure 4. We observe that the SCDA approaches typically deliver much more precise posterior mean estimates. Figures 8 and 9 present "quantile counterparts" of Figure 4, i.e. they compare SCDA and DA in terms of the corresponding quantile estimates (for 10 replications of each algorithm). In Figure 8 we consider deep quantiles of 5% and 95%, while in Figure 9 we present 25% and 75% quantiles. Again, the precision of the SCDA approach is typically noticeably higher than that of the full DA algorithm.

Table 4: UCSV model: parameter posterior means (and their standard deviations) averaged over 10 replications for SCDA Scheme 2 and DA together with the benchmark from DA long.

Param.		Adapt10	Adapt15	Adapt20	Adapt25	Adapt30	Fix20	Fix25	Fix30	Fix35	Fix40	DA	DA long
h_0	Mean	-0.183	-0.167	-0.186	-0.193	-0.209	-0.183	-0.158	-0.149	-0.173	-0.135	-0.171	-0.200
	(Std)	(0.062)	(0.104)	(0.060)	(0.069)	(0.082)	(0.053)	(0.080)	(0.071)	(0.046)	(0.062)	(0.112)	(–)
g_0	Mean	-1.443	-1.449	-1.430	-1.466	-1.377	-1.392	-1.521	-1.487	-1.426	-1.478	-1.408	-1.461
	(Std)	(0.101)	(0.195)	(0.125)	(0.189)	(0.228)	(0.183)	(0.146)	(0.203)	(0.117)	(0.113)	(0.261)	(–)
ω_h^2	Mean	0.049	0.049	0.048	0.049	0.049	0.047	0.047	0.047	0.047	0.047	0.048	0.048
	(Std)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(–)
ω_g^2	Mean	0.049	0.048	0.046	0.047	0.048	0.044	0.045	0.045	0.044	0.046	0.049	0.049
	(Std)	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.005)	(–)

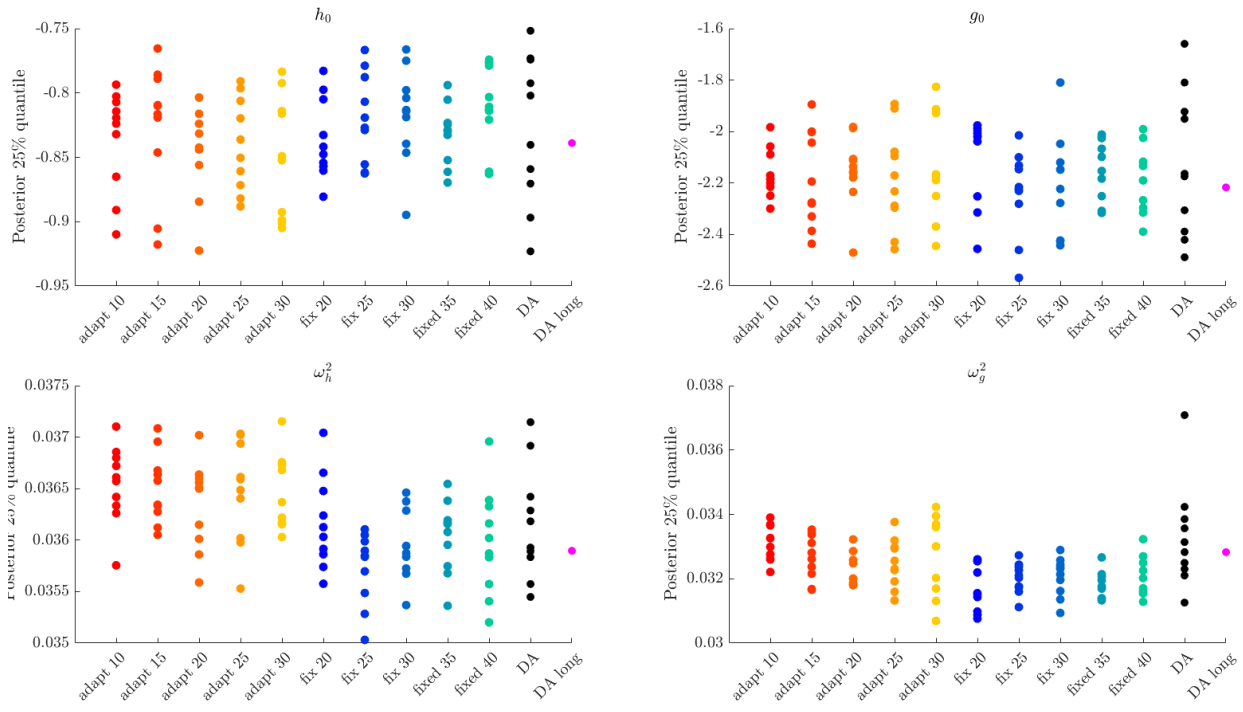


5% quantiles

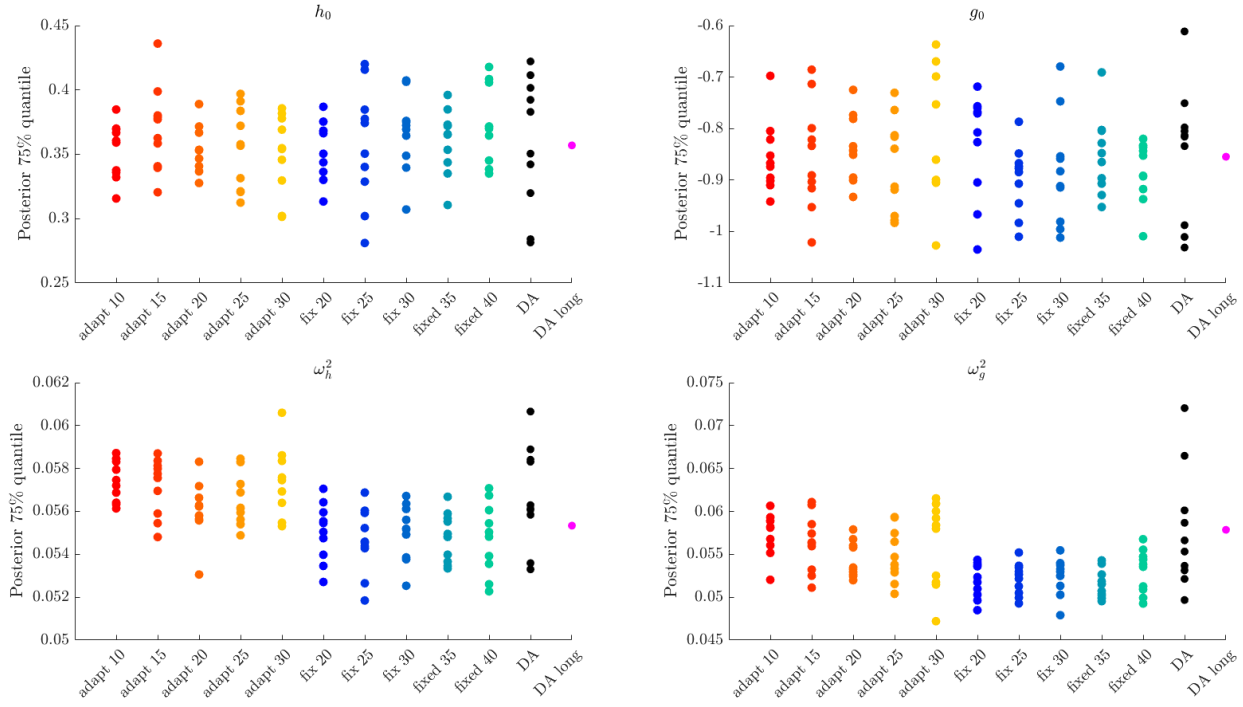


95% quantiles

Figure 8: UCSV model: parameter posterior 5% and 95% quantiles from 10 replications for SCDA Scheme 2 and DA together with the benchmark from DA long.



25% quantiles



75% quantiles

Figure 9: UCSV model: parameter posterior 25% and 75% quantiles from 10 replications for SCDA Scheme 2 and DA together with the benchmark from DA long.

F.3 TVP model

Figures 10 and 11 present ESSs for a selection of h_t and τ_t at 10 time point that were selected to roughly uniformly cover the time domain $t = 1, \ldots, T = 215$. Time points for τ_t and h_t are different, as the former states are imputed at even time points, while the latter are imputed at odd time points.

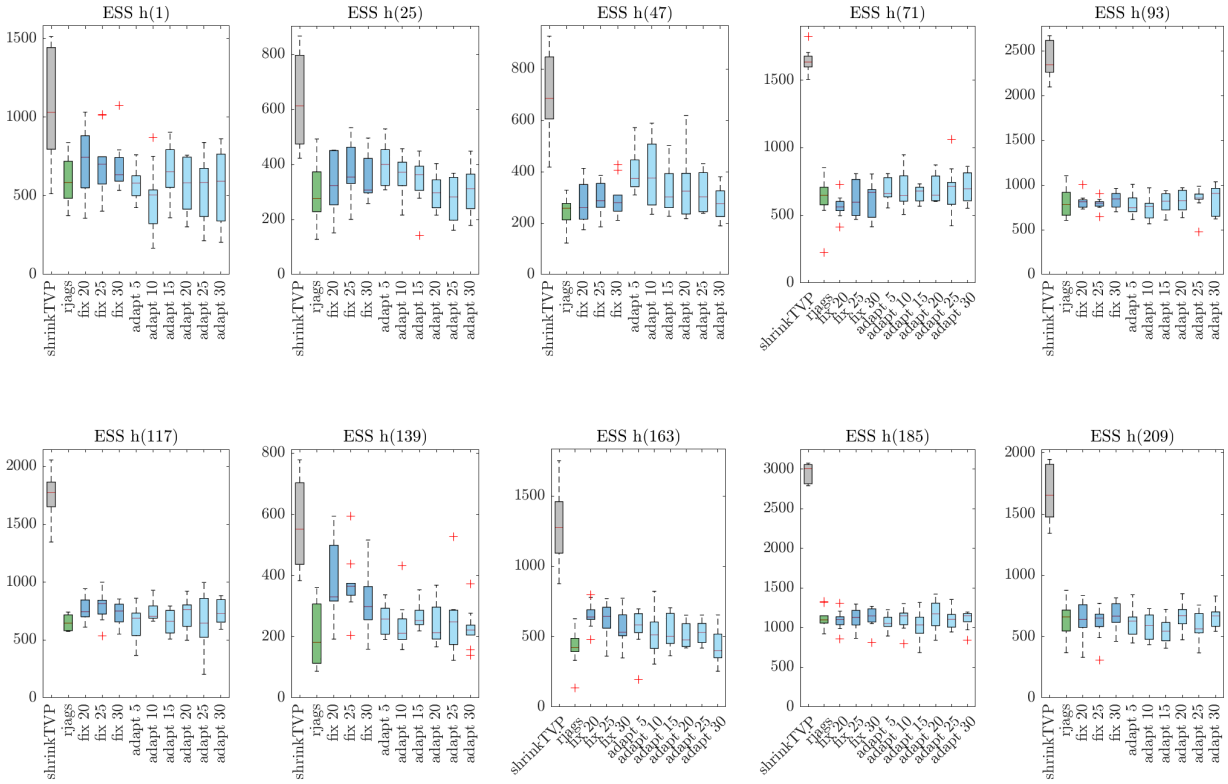


Figure 10: TVP model: ESSs for a selection of h_t (roughly every 23rd t – roughly, as h_t are imputed only for odd t).

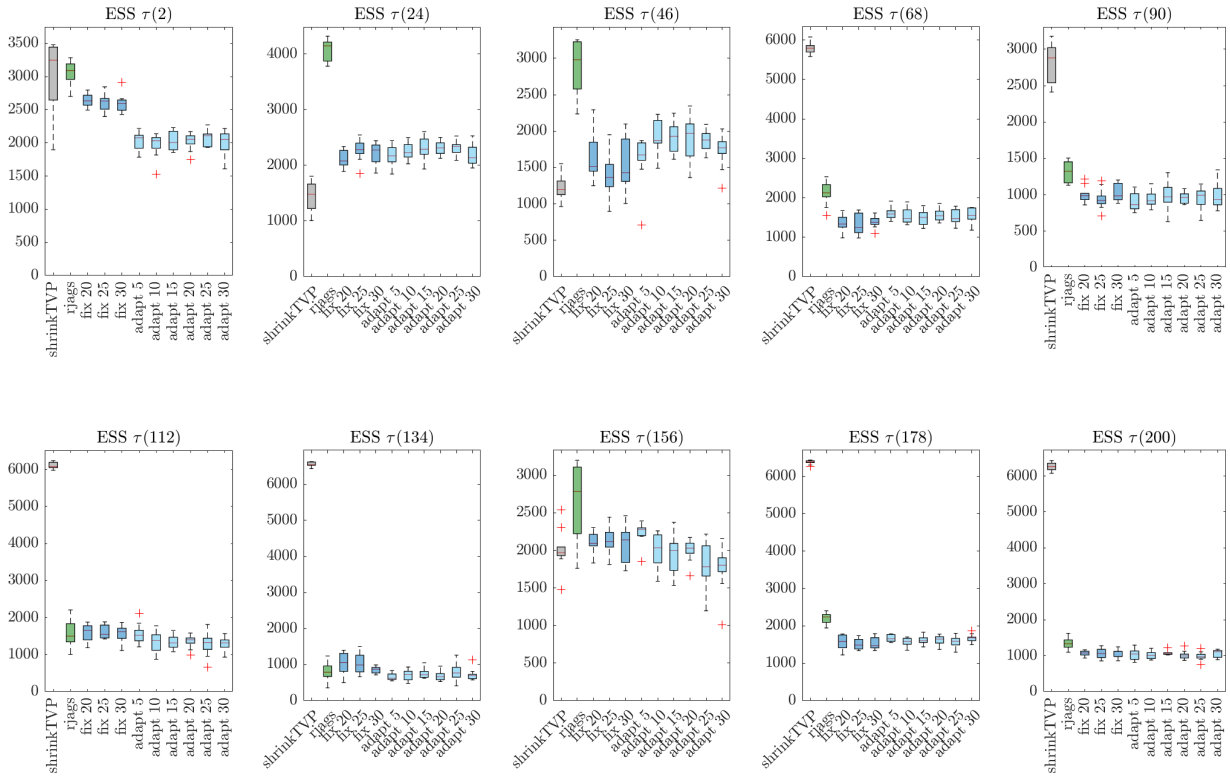


Figure 11: TVP model: ESSs for a selection of τ_t (every 22nd t).

Bibliography

Barra, I., Hoogerheide, L., Koopman, S. J., and Lucas, A. (2017), "Joint Bayesian analysis of parameters and states in nonlinear non-Gaussian state space models," *Journal of Applied Econometrics*, 32, 1003–1026.

Besbeas, P., Freeman, S. N., Morgan, B. J. T., and Catchpole, E. A. (2002), "Integrating Mark–Recapture–Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters," *Biometrics*, 58, 540–547.

Bitto, A. and Frühwirth-Schnatter, S. (2019), "Achieving shrinkage in a time-varying parameter model framework," *Journal of Econometrics*, 210, 75–97.

Brooks, S. P., King, R., and Morgan, B. J. T. (2004), "A Bayesian Approach to Combining Animal Abundance and Demographic Data," *Animal Biodiversity and Conservation*, 27, 515–529.

Chan, J. C. C. (2017), "The Stochastic Volatility in Mean Model with Time-Varying Parameters: An Application to Inflation Modeling," *Journal of Business & Economic Statistics*, 35, 17–28.

Durbin, J. and Koopman, S. J. (2012), *Time Series Analysis by State Space Methods: Second Edition*, Oxford Statistical Science Series, OUP Oxford.

Fridman, M. and Harris, L. (1998), "A Maximum Likelihood Approach for non-Gaussian Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 87, 284—-291.

Geyer, C. J. (2011), "Introduction to Markov Chain Monte Carlo," in *Handbook of Markov Chain Monte Carlo*, eds. S. Brooks, J. G. A. Gelman, and X. L. Meng, chap. 4, CRC Press, pp. 3–48.

Ghysels, E., Harvey, A. C., and Renault, E. (1996), "Stochastic volatility," in *Handbook of Statistics*, eds. G. S. Maddala and C. R. Rao, vol. 14, Elsevier, pp. 119–191.

Goudie, R. J. B., Presanis, A. M., Lunn, D., Angelis, D. D., and Wernisch, L. (2018), "Joining and Splitting Models with Markov Melding," *Bayesian Analysis*.

Hoogerheide, L., Opschoor, A., and Van Dijk, H. K. (2012), "A class of adaptive importance sampling weighted EM algorithms for efficient and robust posterior and predictive simulation," *Journal of Econometrics*, 171, 101–120.

Hosszejni, D. and Kastner, G. (2019), "Approaches toward the bayesian estimation of the stochastic volatility model with leverage," in *Bayesian Statistics and New Generations*, eds. R. Argiento, D. Durante, and S. Wade, Springer International Publishing, pp. 75–83.

International Union for Conservation of Nature (2018), "The IUCN Red List of Threatened Species: lapwing," `http://www.iucnredlist.org/details/22693949/0`. Accessed: 2021-01-30.

Jungbacker, B. and Koopman, S. J. (2007), "Monte Carlo Estimation for Nonlinear Non-Gaussian State Space Models," *Biometrika*, 94, 827–839.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998), "Markov Chain Monte Carlo in Practice: A Roundtable Discussion," *The American Statistician*, 52, 93–100.

Kastner, G. (2016), "Dealing with Stochastic Volatility in Time Series Using the R Package stochvol," *Journal of Statistical Software, Articles*, 69, 1–30.

Kastner, G. and Frühwirth-Schnatter, S. (2014), "Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models," *Computational Statistics & Data Analysis*, 76, 408–423.

Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65, 361–393.

King, R. (2011), "Statistical Ecology," in *Handbook of Markov Chain Monte Carlo*, eds. S. Brooks, J. G. A. Gelman, and X. L. Meng, chap. 17, CRC Press, pp. 419–447.

King, R., Morgan, B., Gimenez, O., and Brooks, S. (2010), *Bayesian Analysis for Population Ecology*, CRC Press.

Koopman, S. J., Lucas, A., and Scharth, M. (2015), "Numerically accelerated importance sampling for nonlinear non-Gaussian state-space models," *Journal of Business & Economic Statistics*, 33, 114–127.

Koopman, S. J. and Uspensky, E. H. (2002), "The Stochastic Volatility in Mean Model: Empirical Evidence from International Stock Markets," *Journal of Applied Econometrics*, 17, 667–689.

Langrock, R., MacDonald, I. L., and Zucchini, W. (2012), "Some Nonstandard Stochastic Volatility Models and their Estimation using Structured Hidden Markov Models," *Journal of Empirical Finance*, 147–161.

Meyer, R. and Yu, J. (2000), "BUGS for a Bayesian Analysis of Stochastic Volatility Models," *Econometrics Journal*, 3, 198–215.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007), "Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference," *Journal of Econometrics*, 140, 425–449.

Papaspiliopoulos, O., Roberts, G. O., and Sköld, M. (2007), "A general framework for the parametrization of hierarchical models," *Statistical Science*, 59–73.

Pitt, M. K., Silva, R. S., Giordani, P., and Kohn, R. (2012), "On Some Properties of Markov Chain Monte Carlo Simulation Methods Based on the Particle Filter," *Journal of Econometrics*, 171, 134–151.

Robert, C. P. and Casella, G. (2004), *Monte Carlo Statistical Methods: Second Edition*, Springer Texts in Statistics.

Sandmann, G. and Koopman, S. J. (1998), "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood," *Journal of Econometrics*, 87, 271–301.

Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, eds. D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Neilsen, Chapman & Hall, pp. 1–67.

Stock, J. H. and Watson, M. W. (2007), "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit and Banking*, 39, 3–33.

Strickland, C. M., Martin, G. M., and Forbes, C. S. (2008), "Parameterisation and efficient MCMC estimation of non-Gaussian state space models," *Computational Statistics & Data Analysis*, 52, 2911–2930.

Taylor, S. J. (1994), "Modeling Stochastic Volatility: A Review and Comparative Study," *Mathematical Finance*, 4, 183–204.

The Royal Society for the Protection of Birds (2018), "The Red List of Conservation Concern: lapwing," `https://www.rspb.org.uk/birds-and-wildlife/wildlife-guides/bird-a-z/lapwing/`. Accessed: 2021-01-30.

Yu, J. (2005), "On Leverage in a Stochastic Volatility Models," *Journal of Econometrics*, 127, 165–178.

Yu, Y. and Meng, X.-L. (2011), "To Center or Not to Center: That Is Not the Question—An Ancillarity–Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency," *Journal of Computational and Graphical Statistics*, 20, 531–570.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016), *Hidden Markov Models for Time Series: An Introduction Using R, Second Edition*, Monographs on Statistics and Applied Probability 150, CRC Press.