Ancillarity-Sufficiency or not; Interweaving to Improve MCMC Estimation of the Local Level DLM

Matt Simpson

Department of Statistics, Iowa State University

October 2, 2013

Outline

1. The model
2. MCMC estimation and its problems
3. Solutions: reparameterization and interweaving
4. Simulation results applying the solutions to the model

The Dynamic Linear Model

For $t = 1, 2, ..., T$ let y_t denote the data and θ_t denote the latent state in period t:

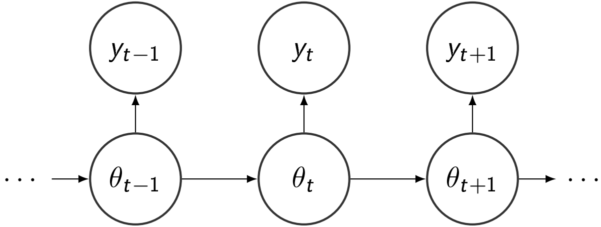
$$\begin{aligned} y_t =& F_t\theta_t + v_t && (1) \\ \theta_t =& G_t\theta_{t-1} + w_t && (2) \end{aligned}$$

with $v_t \overset{ind}{\sim} N(0, V_t)$, $w_t \overset{ind}{\sim} N(0, W_t)$ and $\theta_0 \sim N(m_0, C_0)$, all mutually independent.

- (1) is the observation equation; (2) is the system equation.
- v_t is the (observation) error; w_t is the (system) disturbance.
- F_t is the observation matrix and V_t is the observation covariance matrix. G_t is the system matrix and W_t is the system covariance matrix. All four possibly depend on unkown parameter ϕ.

The Dynamic Linear Model

The θ's form a markov chain. Conditional on the θ's, the y's are mutually independent.



The Univariate Local Level Model

For $t = 1, 2, ..., T$

$$\begin{aligned} y_t &= \theta_t + v_t \\ \theta_t &= \theta_{t-1} + w_t \end{aligned}$$

with $v_t \overset{iid}{\sim} N(0, V)$, $w_t \overset{iid}{\sim} N(0, W)$ and $\theta_0 \sim N(m_0, C_0)$, all mutually independent.

- $\theta_t = E[y_t|\theta_t]$; i.e. the “level” in period t, and evolves over time through a random walk.
- W is the system variance, a.k.a. the signal; V is the observational variance, a.k.a the noise.
- $\phi = (V, W)$ is the unknown parameter vector.

The Univariate Local Level Model

For $t = 1, 2, ..., T$

$$\begin{aligned} y_t =& \theta_t + v_t \\ \theta_t =& \theta_{t-1} + w_t \end{aligned}$$

with $v_t \overset{iid}{\sim} N(0, V)$, $w_t \overset{iid}{\sim} N(0, W)$ and $\theta_0 \sim N(m_0, C_0)$, all mutually independent.

- $\theta_t = E[y_t|\theta_t]$; i.e. the "level" in period t, and evolves over time through a random walk.
- W is the system variance, a.k.a. the signal; V is the observational variance, a.k.a the noise.
- $\phi = (V, W)$ is the unknown parameter vector.

The Univariate Local Level Model

For $t = 1, 2, ..., T$

$$\begin{aligned} y_t =& \theta_t + v_t \\ \theta_t =& \theta_{t-1} + w_t \end{aligned}$$

with $v_t \overset{iid}{\sim} N(0, V)$, $w_t \overset{iid}{\sim} N(0, W)$ and $\theta_0 \sim N(m_0, C_0)$, all mutually independent.

- $\theta_t = E[y_t|\theta_t]$; i.e. the "level" in period t, and evolves over time through a random walk.
- W is the system variance, a.k.a. the signal; V is the observational variance, a.k.a the noise.
- $\phi = (V, W)$ is the unknown parameter vector.

Estimating The Local Level Model: Priors

(V, W, θ_0) independent and:

$$\begin{aligned} \theta_0 &\sim N(m_0, C_0) \\ V &\sim IG(\alpha_v, \beta_v) \\ W &\sim IG(\alpha_w, \beta_w) \end{aligned}$$

where m_0, C_0, α_v, β_v, α_w, and β_w are known hyperparameters.

Uninformative priors: $m_0 = 0$, C_0 large, independent half-t priors for V and W.

Estimating The Local Level Model: Priors

(V, W, θ_0) independent and:

$$\begin{aligned}\theta_0 &\sim N(m_0, C_0)\\ V &\sim IG(\alpha_v, \beta_v)\\ W &\sim IG(\alpha_w, \beta_w)\end{aligned}$$

where m_0, C_0, α_v, β_v, α_w, and β_w are known hyperparameters.

Uninformative priors: $m_0 = 0$, C_0 large, independent half-t priors for V and W.

Estimating The Local Level Model: Data Augmentation

In general, let y denote the data vector and ϕ the parameter vector. Goal: obtain a markov chain with stationary distribution $p(\phi|y)$.

Data Augmentation (DA) algorithm: let θ denote an *augmented data vector* such that

$$\int_{\Theta} p(\theta, \phi|y) d\theta = p(\phi|y)$$

Given iteration $\phi^{(k)}$:

1. Draw $\theta^{(k+1)}$ from $p(\theta|\phi^{(k)}, y)$
2. Draw $\phi^{(k+1)}$ from $p(\phi|\theta^{(k+1)}, y)$

Side effect: we obtain joint draws from $p(\theta, \phi|y)$

In the local level model:

$$\theta = \theta_{0:T}, \; y = y_{1:T}, \text{ and } \phi = (V, W).$$

Estimating The Local Level Model: Data Augmentation

In general, let y denote the data vector and ϕ the parameter vector. Goal: obtain a markov chain with stationary distribution $p(\phi|y)$.

Data Augmentation (DA) algorithm: let θ denote an *augmented data vector* such that

$$\int_\Theta p(\theta,\phi|y)d\theta = p(\phi|y)$$

Given iteration $\phi^{(k)}$:

1. Draw $\theta^{(k+1)}$ from $p(\theta|\phi^{(k)},y)$
2. Draw $\phi^{(k+1)}$ from $p(\phi|\theta^{(k+1)},y)$

Side effect: we obtain joint draws from $p(\theta,\phi|y)$

In the local level model:

$\theta = \theta_{0:T}$, $y = y_{1:T}$, and $\phi = (V,W)$.

Estimating The Local Level Model: Data Augmentation

Step 1: Use Forward Filtering, Backward Sampling (FFBS) to draw $(\theta_{0:T}|V, W, y_{1:T})$:

1. Run the Kalman filter to obtain a draw from $p(\theta_T|V, W, y_{1:T})$.
2. Recursively sample θ_t from $p(\theta_t|\theta_{t+1:T}, V, W, y_{1:T})$.

Step 2: Draw $(V, W|\theta_{0:T}, y_{1:T})$:

1. Draw V from $p(V|\theta_{0:T}, y_{1:T}) = IG(a_V, b_V)$ where

$$\begin{aligned} a_v =& \alpha_v + T/2 \\ b_v =& \beta_v + \textstyle\sum_{t=1}^{T}(y_t - \theta_t)^2/2 \end{aligned}$$

2. Draw W from $p(W|\theta_{0:T}, V, y_{1:T}) = IG(a_W, b_W)$ where

$$\begin{aligned} a_w =& \alpha_w + T/2 \\ b_w =& \beta_w + \textstyle\sum_{t=1}^{T}(\theta_t - \theta_{t-1})^2/2 \end{aligned}$$

Estimating The Local Level Model: Data Augmentation

Step 1: Use Forward Filtering, Backward Sampling (FFBS) to draw $(\theta_{0:T}|V, W, y_{1:T})$:

1. Run the Kalman filter to obtain a draw from $p(\theta_T|V, W, y_{1:T})$.
2. Recursively sample θ_t from $p(\theta_t|\theta_{t+1:T}, V, W, y_{1:T})$.

Step 2: Draw $(V, W|\theta_{0:T}, y_{1:T})$:

1. Draw V from $p(V|\theta_{0:T}, y_{1:T}) = IG(a_V, b_V)$ where
$$\begin{aligned} a_v =& \alpha_v + T/2 \\ b_v =& \beta_v + \textstyle\sum_{t=1}^T (y_t - \theta_t)^2/2 \end{aligned}$$
2. Draw W from $p(W|\theta_{0:T}, V, y_{1:T}) = IG(a_W, b_W)$ where
$$\begin{aligned} a_w =& \alpha_w + T/2 \\ b_w =& \beta_w + \textstyle\sum_{t=1}^T (\theta_t - \theta_{t-1})^2/2 \end{aligned}$$

Estimating The Local Level Model: Problems & Solutions

This sampler is called the *state sampler*.

Problems:

1. Kalman filter effectively requires drawing $\theta_t|\theta_{0:t-1}, V, W, y_{1:T}$ for $t = 0, 1, ..., T$.
2. Mixing is often awful, requiring large posterior samples.
3. All problems seem to get worse for large T.

Solutions:

1. Alternate parameterizations / data augmentations.
2. (Ancillarity-Sufficiency) Interweaving

Estimating The Local Level Model: Problems & Solutions

This sampler is called the *state sampler*.

Problems:

1. Kalman filter effectively requires drawing $\theta_t|\theta_{0:t-1}, V, W, y_{1:T}$ for $t = 0, 1, ..., T$.
2. Mixing is often awful, requiring large posterior samples.
3. All problems seem to get worse for large T.

Solutions:

1. Alternate parameterizations / data augmentations.
2. (Ancillarity-Sufficiency) Interweaving

Alternative Data Augmentations

In general, let y denote the data, ϕ denote the model parameters, and θ denote a data augmentation. Then:

- θ is a *sufficient augmentation or SA* if $p(y|\theta,\phi) = p(y|\theta)$. A.K.A. the centered parameterization (CP).
- θ is an *ancillary augmentation or AA* if $p(\theta|\phi) = p(\theta)$. A.K.A. the non-centered parameterization (NCP).

Papaspiliopoulos et al. [3]: DA based on a SA and DA based on an AA will typically have poor mixing and convergence in opposite regions of the parameter space.

- Option 1: Figure out what region of the parameter space you're usually in, use the appropriate parameterization.
- Option 2: Parameter expanded data augmentation. See e.g. Van Dyk and Meng [4].
- Option 3: Ancillarity-Sufficiency Interweaving Strategies (ASIS) of Yu and Meng [5].

Alternative Data Augmentations

In general, let y denote the data, ϕ denote the model parameters, and θ denote a data augmentation. Then:

- θ is a *sufficient augmentation or SA* if $p(y|\theta,\phi) = p(y|\theta)$. A.K.A. the centered parameterization (CP).
- θ is an *ancillary augmentation or AA* if $p(\theta|\phi) = p(\theta)$. A.K.A. the non-centered parameterization (NCP).

Papaspiliopoulos et al. [3]: DA based on a SA and DA based on an AA will typically have poor mixing and convergence in opposite regions of the parameter space.

- Option 1: Figure out what region of the parameter space you're usually in, use the appropriate parameterization.
- Option 2: Parameter expanded data augmentation. See e.g. Van Dyk and Meng [4].
- Option 3: Ancillarity-Sufficiency Interweaving Strategies (ASIS) of Yu and Meng [5].

Alternative Data Augmentations

In general, let y denote the data, ϕ denote the model parameters, and θ denote a data augmentation. Then:

- θ is a *sufficient augmentation or SA* if $p(y|\theta, \phi) = p(y|\theta)$. A.K.A. the centered parameterization (CP).
- θ is an *ancillary augmentation or AA* if $p(\theta|\phi) = p(\theta)$. A.K.A. the non-centered parameterization (NCP).

Papaspiliopoulos et al. [3]: DA based on a SA and DA based on an AA will typically have poor mixing and convergence in opposite regions of the parameter space.

- Option 1: Figure out what region of the parameter space you're usually in, use the appropriate parameterization.
- Option 2: Parameter expanded data augmentation. See e.g. Van Dyk and Meng [4].
- Option 3: Ancillarity-Sufficiency Interweaving Strategies (ASIS) of Yu and Meng [5].

Weaving Together the Beauty and the Beast

Suppose we have data y, parameter ϕ, and two augmented data vectors θ and $\tilde{\theta}$ such that the joint distribution of $(\phi, \theta, \tilde{\theta})$ is well defined. Then given iteration $\phi^{(k)}$, a *global interweaving strategy (GIS)* obtains $\phi^{(k+1)}$ by:

1. Draw θ from $p(\theta|\phi^{(k)}, y)$
2. Draw $\tilde{\theta}$ from $p(\tilde{\theta}|\theta, y)$
3. Draw $\phi^{(k+1)}$ from $p(\phi|\tilde{\theta}, y)$

Often $\tilde{\theta} = M(\theta|\phi, y)$ where M is a one-to-one function of θ and step 2 is most easily accomplished with two steps:

1. Draw θ from $p(\theta|\phi^{(k)}, y)$
2. Draw ϕ from $p(\phi|\theta, y)$
3. Update $\tilde{\theta} = M(\theta|\phi, y)$
4. Draw $\phi^{(k+1)}$ from $p(\phi|\tilde{\theta}, y)$

An *alternating* algorithm would replace step 3 with a draw from $p(\tilde{\theta}|\phi, y)$. Is alternating or interweaving better?

Weaving Together the Beauty and the Beast

Suppose we have data y, parameter ϕ, and two augmented data vectors θ and $\tilde{\theta}$ such that the joint distribution of $(\phi, \theta, \tilde{\theta})$ is well defined. Then given iteration $\phi^{(k)}$, a *global interweaving strategy (GIS)* obtains $\phi^{(k+1)}$ by:

1. Draw θ from $p(\theta|\phi^{(k)}, y)$
2. Draw $\tilde{\theta}$ from $p(\tilde{\theta}|\theta, y)$
3. Draw $\phi^{(k+1)}$ from $p(\phi|\tilde{\theta}, y)$

Often $\tilde{\theta} = M(\theta|\phi, y)$ where M is a one-to-one function of θ and step 2 is most easily accomplished with two steps:

1. Draw θ from $p(\theta|\phi^{(k)}, y)$
2. Draw ϕ from $p(\phi|\theta, y)$
3. Update $\tilde{\theta} = M(\theta|\phi, y)$
4. Draw $\phi^{(k+1)}$ from $p(\phi|\tilde{\theta}, y)$

An *alternating* algorithm would replace step 3 with a draw from $p(\tilde{\theta}|\phi, y)$. Is alternating or interweaving better?

Weaving Together the Beauty and the Beast

Yu and Meng [5]:

1. GIS has a geometric rate of convergence no worse than the worst of the two underlying DA algorithms, and the bound gets better the less θ and $\tilde{\theta}$ are correlated in the posterior.

2. Often, but not always, GIS has better convergence than the associated alternating algorithm.

3. If θ and $\tilde{\theta}$ are independent in the posterior, then GIS results in ***iid*** draws from the posterior.

4. If θ is a SA, $\tilde{\theta}$ is an AA, $\tilde{\theta}$ (so the algorithm is ASIS) is a one-to-one transformation of θ, and the priors are "nice," then the GIS algorithm is the same as the optimal PX-DA algorithm of Liu and Wu [2].

Weaving Together the Beauty and the Beast

General intuition behind interweaving:

- Fundamental problem is that the chain $\{\phi^{(k)}\}$ is highly autocorrelated.
- If θ and ϕ are highly correlated in the posterior, drawing $p(\theta|\phi, y)$ then $p(\phi|\theta, y)$ won't move the chain much.
- Interweaving draws from $p(\theta|\phi, y)$, then $p(\tilde{\theta}|\theta, y)$, then $p(\phi|\tilde{\theta}, y)$. The less correlated θ and $\tilde{\theta}$ are in the posterior, the less correlated the chain is.
- If the DA algorithms based on θ and $\tilde{\theta}$ yield chains with high autocorrelations in opposite regions of the parameter space, interweaving algorithm ensures that at least one step moves the chain significantly.

In this way, interweaving takes advantage of the fact that one chain is a "beauty" and the other is a "beast."

Weaving Together the Beauty and the Beast

General intuition behind interweaving:

- Fundamental problem is that the chain $\{\phi^{(k)}\}$ is highly autocorrelated.
- If θ and ϕ are highly correlated in the posterior, drawing $p(\theta|\phi, y)$ then $p(\phi|\theta, y)$ won't move the chain much.
- Interweaving draws from $p(\theta|\phi, y)$, then $p(\tilde{\theta}|\theta, y)$, then $p(\phi|\tilde{\theta}, y)$. The less correlated θ and $\tilde{\theta}$ are in the posterior, the less correlated the chain is.
- If the DA algorithms based on θ and $\tilde{\theta}$ yield chains with high autocorrelations in opposite regions of the parameter space, interweaving algorithm ensures that at least one step moves the chain significantly.

In this way, interweaving takes advantage of the fact that one chain is a "beauty" and the other is a "beast."

Weaving Together the Beauty and the Beast

General intuition behind interweaving:

- Fundamental problem is that the chain $\{\phi^{(k)}\}$ is highly autocorrelated.
- If θ and ϕ are highly correlated in the posterior, drawing $p(\theta|\phi, y)$ then $p(\phi|\theta, y)$ won't move the chain much.
- Interweaving draws from $p(\theta|\phi, y)$, then $p(\tilde{\theta}|\theta, y)$, then $p(\phi|\tilde{\theta}, y)$. The less correlated θ and $\tilde{\theta}$ are in the posterior, the less correlated the chain is.
- If the DA algorithms based on θ and $\tilde{\theta}$ yield chains with high autocorrelations in opposite regions of the parameter space, interweaving algorithm ensures that at least one step moves the chain significantly.

In this way, interweaving takes advantage of the fact that one chain is a "beauty" and the other is a "beast."

Componentwise Interweaving Strategy (CIS)

Sometimes it's not possible to find an SA-AA pair of augmentations for ϕ, but if $\phi = (\phi_1, \phi_2)$, it's possible to do componentwise interweaving while still taking advantage of SA-AA pairs.

Suppose there are four DAs θ, $\tilde{\theta}$, γ, and $\tilde{\gamma}$. Then the CIS algorithm is:

1. Draw θ from $p(\theta|\phi_1^{(k)}, \phi_2^{(k)}, y)$.
2. Draw $\tilde{\theta}$ from $p(\tilde{\theta}|\theta, \phi_1^{(k)}, \phi_2^{(k)}, y)$
3. Draw $\phi_1^{(k+1)}$ from $p(\phi_1|\tilde{\theta}, \phi_2^{(k)}, y)$
4. Draw γ from $p(\gamma|\phi_1^{(k+1)}, \phi_2^{(k)}, y)$.
5. Draw $\tilde{\gamma}$ from $p(\tilde{\gamma}|\gamma, \phi_1^{(k+1)}, \phi_2^{(k)}, y)$
6. Draw $\phi_2^{(k+1)}$ from $p(\phi_2|\tilde{\gamma}, \phi_1^{(k+1)}, y)$

Componentwise Interweaving Strategy (CIS)

Sometimes it's not possible to find an SA-AA pair of augmentations for ϕ, but if $\phi = (\phi_1, \phi_2)$, it's possible to do componentwise interweaving while still taking advantage of SA-AA pairs.

Suppose there are four DAs θ, $\tilde{\theta}$, γ, and $\tilde{\gamma}$. Then the CIS algorithm is:

1. Draw θ from $p(\theta|\phi_1^{(k)}, \phi_2^{(k)}, y)$.
2. Draw $\tilde{\theta}$ from $p(\tilde{\theta}|\theta, \phi_1^{(k)}, \phi_2^{(k)}, y)$
3. Draw $\phi_1^{(k+1)}$ from $p(\phi_1|\tilde{\theta}, \phi_2^{(k)}, y)$
4. Draw γ from $p(\gamma|\phi_1^{(k+1)}, \phi_2^{(k)}, y)$.
5. Draw $\tilde{\gamma}$ from $p(\tilde{\gamma}|\gamma, \phi_1^{(k+1)}, \phi_2^{(k)}, y)$
6. Draw $\phi_2^{(k+1)}$ from $p(\phi_2|\tilde{\gamma}, \phi_1^{(k+1)}, y)$

Componentwise Interweaving Strategy (CIS): Notes

- Often $\theta = \gamma$ or $\theta = \tilde{\gamma}$, which simplifies the algorithm.
- We want θ & $\tilde{\theta}$ to be an SA-AA pair for ϕ_1 given ϕ_2 and γ & $\tilde{\gamma}$ to be an SA-AA pair for ϕ_2 given ϕ_1.
- The order we draw the DA vectors matters here (as well as in GIS) and so does the order we draw the parameter subvectors, but not much.
- Steps 2 and/or 4 might be update steps if, e.g., θ is a one-to-one function of $\tilde{\theta}$ (conditional on ϕ and y).
- Yu and Meng [5]: The CIS sampler does at least as well (in some sense) as the Gibbs sampler that integrates out the augmented data vectors.

Parameterizations in the Local Level Model

Recall the model: for $t = 1, 2, ..., T$:

$$y_t|\theta_{0:T} \overset{iid}{\sim} N(\theta_t, V)$$
$$\theta_t|\theta_{0:t-1} \sim N(\theta_{t-1}, W)$$

$\theta_{0:T}$ is neither SA nor AA for (V, W), but it is SA for $W|V$ and AA for $V|W$.

Recall the sampler based on $\theta_{0:T}$ is called the *state sampler*.

Other obvious DAs:

1. The disturbances $(w_{0:T})$ where $w_0 \equiv \theta_0$. Results in the state sampler because $p(V, W|\theta_{0:T}, y_{1:T}) = p(V, W|w_{0:T}, y_{1:T})$.
2. The errors $(v_{0:T})$ where $v_0 \equiv \theta_0$. Also results in the state sampler.

Parameterizations in the Local Level Model

Recall the model: for $t = 1, 2, ..., T$:

$$y_t|\theta_{0:T} \overset{iid}{\sim} N(\theta_t, V)$$
$$\theta_t|\theta_{0:t-1} \sim N(\theta_{t-1}, W)$$

$\theta_{0:T}$ is neither SA nor AA for (V, W), but it is SA for $W|V$ and AA for $V|W$.

Recall the sampler based on $\theta_{0:T}$ is called the *state sampler*.

Other obvious DAs:

1. The disturbances $(w_{0:T})$ where $w_0 \equiv \theta_0$. Results in the state sampler because $p(V, W|\theta_{0:T}, y_{1:T}) = p(V, W|w_{0:T}, y_{1:T})$.
2. The errors $(v_{0:T})$ where $v_0 \equiv \theta_0$. Also results in the state sampler.

Parameterizations in the Local Level Model

Less obvious: try scaling the errors/disturbances by their standard deviation. Define the scaled disturbances: $\gamma_0 = \theta_0$ and for $t = 1, 2, ..., T$

$$\gamma_t = \frac{\theta_t - \theta_{t-1}}{\sqrt{W}} = \frac{w_t}{\sqrt{W}}$$

Under the scaled disturbance parameterization, we can write the model as

$$y_t|\gamma_{0:T}, V, W \overset{ind}{\sim} N(\gamma_0 + \sqrt{W}\textstyle\sum_{s=1}^{t}\gamma_s, V)$$
$$\gamma_t \overset{iid}{\sim} N(0, 1)$$

We immediately see that $\gamma_{0:T}$ is an AA for (V, W) but not an SA for either V or W.

Frühwirth-Schnatter [1] uses the analogue of $\gamma_{0:T}$ in a dynamic regression model.

Parameterizations in the Local Level Model

Less obvious: try scaling the errors/disturbances by their standard deviation. Define the scaled disturbances: $\gamma_0 = \theta_0$ and for $t = 1, 2, ..., T$

$$\gamma_t = \frac{\theta_t - \theta_{t-1}}{\sqrt{W}} = \frac{w_t}{\sqrt{W}}$$

Under the scaled disturbance parameterization, we can write the model as

$$y_t|\gamma_{0:T}, V, W \overset{ind}{\sim} N(\gamma_0 + \sqrt{W}\textstyle\sum_{s=1}^t \gamma_s, V)$$
$$\gamma_t \overset{iid}{\sim} N(0, 1)$$

We immediately see that $\gamma_{0:T}$ is an AA for (V, W) but not an SA for either V or W.

Frühwirth-Schnatter [1] uses the analogue of $\gamma_{0:T}$ in a dynamic regression model.

Parameterizations in the Local Level Model

Define the scaled errors:: $\psi_0 = \theta_0$ and for $t = 1, 2, ..., T$

$$\psi_t = \frac{y_t - \theta_t}{\sqrt{V}} = \frac{v_t}{\sqrt{V}}$$

Under the scaled error parameterization, we can write the model as

$$y_t | y_{1:t-1}, \psi_{0:T}, V, W \overset{ind}{\sim} N(y_{t-1} + \sqrt{V}(\psi_t - \psi_{t-1}), W)$$
$$\psi_t \overset{iid}{\sim} N(0, 1)$$

for $t = 2, 3, ..., T$, and for $t = 1$ the observation equation has the mean $\sqrt{V}\psi_1 - \psi_0$, but the system equation is unchanged.

$\psi_{0:T}$ is also an AA for (V, W) and not a SA for V nor W

Parameterizations in the Local Level Model

Define the scaled errors:: $\psi_0 = \theta_0$ and for $t = 1, 2, ..., T$

$$\psi_t = \frac{y_t - \theta_t}{\sqrt{V}} = \frac{v_t}{\sqrt{V}}$$

Under the scaled error parameterization, we can write the model as

$$\begin{aligned} y_t|y_{1:t-1}, \psi_{0:T}, V, W &\overset{ind}{\sim} N(y_{t-1} + \sqrt{V}(\psi_t - \psi_{t-1}), W) \\ \psi_t &\overset{iid}{\sim} N(0, 1) \end{aligned}$$

for $t = 2, 3, ..., T$, and for $t = 1$ the observation equation has the mean $\sqrt{V}\psi_1 - \psi_0$, but the system equation is unchanged.

$\psi_{0:T}$ is also an AA for (V, W) and not a SA for V nor W

$\theta_{0:T}$ as a sort of SA for (V, W)

Consider the augmented data vector $(\theta_0, v_{1:T}, w_{1:T})$. Then

$$y_t = \theta_0 + \sum_{s=1}^{T} w_s + v_t$$

so that $y_{1:T}|\theta_0, v_{1:T}, w_{1:T}$ has a singular distribution. However, this distribution is free of (V, W) — i.e. $(\theta_0, v_{1:T}, w_{1:T})$ is a SA for (V, W).

It turns out that $p(V, W|\theta_{0:T}, y_{1:T}) = p(V, W|\theta_0, v_{1:T}, w_{1:T}, y_{1:T})$ so conditioning on $\theta_{0:T}$ is the same as conditioning on $(\theta_0, v_{1:T}, w_{1:T})$.

In this way, we can view $\theta_{0:T}$ as a SA for (V, W) even though, strictly speaking, this isn't true.

$\theta_{0:T}$ as a sort of SA for (V, W)

Consider the augmented data vector $(\theta_0, v_{1:T}, w_{1:T})$. Then

$$y_t = \theta_0 + \sum_{s=1}^{T} w_s + v_t$$

so that $y_{1:T}|\theta_0, v_{1:T}, w_{1:T}$ has a singular distribution. However, this distribution is free of (V, W) — i.e. $(\theta_0, v_{1:T}, w_{1:T})$ is a SA for (V, W).

It turns out that $p(V, W|\theta_{0:T}, y_{1:T}) = p(V, W|\theta_0, v_{1:T}, w_{1:T}, y_{1:T})$ so conditioning on $\theta_{0:T}$ is the same as conditioning on $(\theta_0, v_{1:T}, w_{1:T})$.

In this way, we can view $\theta_{0:T}$ as a SA for (V, W) even though, strictly speaking, this isn't true.

Two New DA Algorithms or the Local Level Model

The *scaled-disturbance sampler*:

1. Draw $\gamma_{0:T}$ from $p(\gamma_{0:T}|V^{(k)}, W^{(k)}, y_{1:T})$, e.g. by using FFBS to draw $\theta_{0:T}$ then transforming.
2. Draw $V^{(k+1)}$ from $p(V|\gamma_{0:T}, W^{(k)}, y_{1:T})$ (same inverse gamma draw as for V in step 2 of the state sampler)
3. Draw $W^{(k+1)}$ from $p(W|\gamma_{0:T}, V^{(k+1)}, y_{1:T})$ (complicated density, can use adaptive rejection sampling sometimes)

The *scaled-error sampler*:

1. Draw $\psi_{0:T}$ from $p(\psi_{0:T}|V^{(k)}, W^{(k)}, y_{1:T})$, e.g. by using FFBS to draw $\theta_{0:T}$ then transforming.
2. Draw $V^{(k+1)}$ from $p(V|\psi_{0:T}, W^{(k)}, y_{1:T})$ (same type of complicated density as in step 3 of the scaled-disturbance sampler)
3. Draw $W^{(k+1)}$ from $p(W|\psi_{0:T}, V^{(k+1)}, y_{1:T})$ (same inverse gamma draw as for W in step 2 of the state sampler)

Two New DA Algorithms or the Local Level Model

The *scaled-disturbance sampler*:

1. Draw $\gamma_{0:T}$ from $p(\gamma_{0:T}|V^{(k)}, W^{(k)}, y_{1:T})$, e.g. by using FFBS to draw $\theta_{0:T}$ then transforming.
2. Draw $V^{(k+1)}$ from $p(V|\gamma_{0:T}, W^{(k)}, y_{1:T})$ (same inverse gamma draw as for V in step 2 of the state sampler)
3. Draw $W^{(k+1)}$ from $p(W|\gamma_{0:T}, V^{(k+1)}, y_{1:T})$ (complicated density, can use adaptive rejection sampling sometimes)

The *scaled-error sampler*:

1. Draw $\psi_{0:T}$ from $p(\psi_{0:T}|V^{(k)}, W^{(k)}, y_{1:T})$, e.g. by using FFBS to draw $\theta_{0:T}$ then transforming.
2. Draw $V^{(k+1)}$ from $p(V|\psi_{0:T}, W^{(k)}, y_{1:T})$ (same type of complicated density as in step 3 of the scaled-disturbance sampler)
3. Draw $W^{(k+1)}$ from $p(W|\psi_{0:T}, V^{(k+1)}, y_{1:T})$ (same inverse gamma draw as for W in step 2 of the state sampler)

Simulation Setup

To test these parameterizations, I simulated a fake data set from the local level model for each (V, W) pair over a grid and for each of $T = 10$, $T = 100$, $T = 1000$. Then the model was fit using all three samplers with starting values $(\tilde{V}, \tilde{W})$ where $\tilde{V}$ and $\tilde{W}$ are the true values used to simulate the fake dataset.

Independent Priors:

- $\theta_0 \sim N(0, 10^7)$
- $V \sim IG(\alpha_V, \beta_V)$ with $\alpha_V = 5$ and $\beta_V = (\alpha_V - 1)\tilde{V}$.
- $W \sim IG(\alpha_W, \beta_W)$ with $\alpha_W = 5$ and $\beta_W = (\alpha_W - 1)\tilde{W}$.

Priors for V and W are centered on the true value and not too flat to avoid well known issues with the $IG(\epsilon, \epsilon)$ prior where $\epsilon \to 0$.

Effective Sample Size and Effective Sample Proportion

Suppose we want to estimate $\mathrm{E}[V|y_{1:T}]$ with

$$\bar{V} \equiv \frac{\sum_{k=1}^{K} V^{(k)}}{K}$$

If we had K iid draws from the posterior, the CLT would give

$$\mathrm{Var}(\bar{V}) = \frac{\sigma_V^2}{K}$$

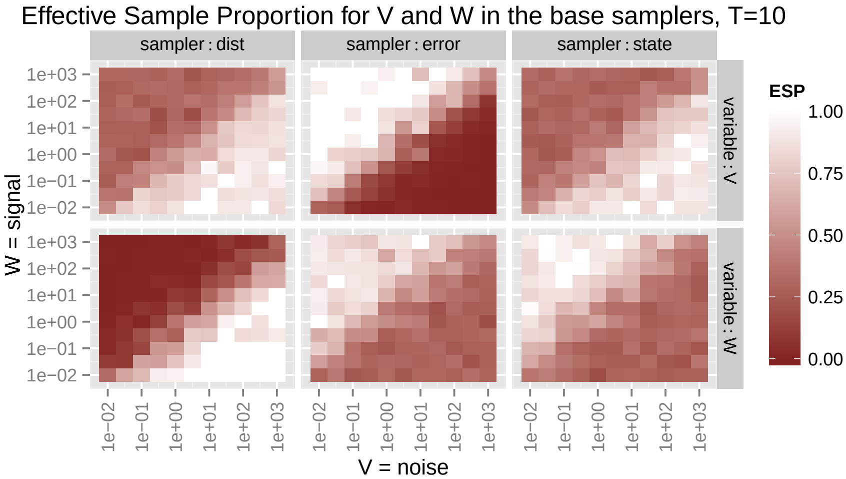
where $\sigma_V^2 = \mathrm{Var}(V|y_{1:T})$. For our markov chain

$$\mathrm{Var}(\bar{V}) = \frac{\sigma_V^2}{ESS}$$

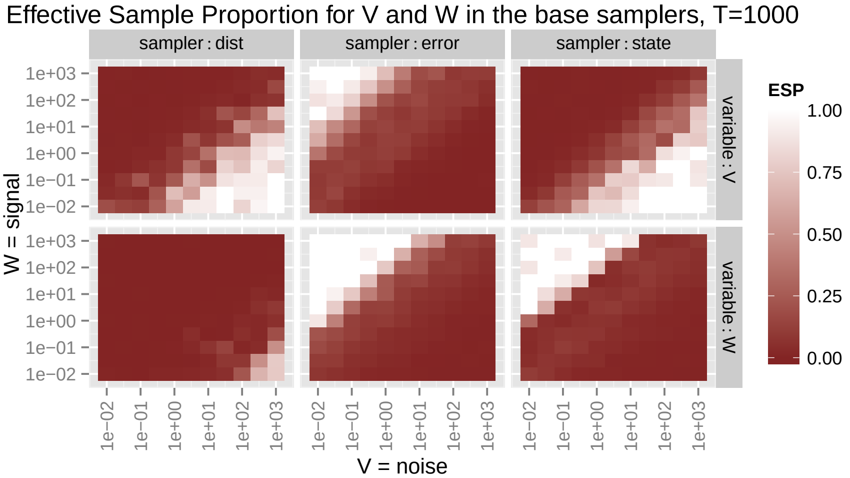
where ESS is the "effective sample size." ESS is estimated by fitting an $AR(p)$ model to the chain (to estimate the spectral density at 0).

The effective sample proportion is the effective sample size as a proportion of the actual sample size, i.e. $ESP = ESS/K$.

Simulation Results for the Base Algorithms, $T = 10$



Simulation Results for the Base Algorithms, $T = 1000$



Takeaways from Base Algorithm Simulations

When the signal-to-noise ratio (W/V) is low (< 1), the scaled disturbance sampler has high *ESP* for both V and W; when it's high (> 1), the scaled disturbance sampler has low *ESP* for both V and W.

When the signal-to-noise ratio (W/V) is low (< 1), the scaled error sampler has high *ESP* for both V and W; when it's high (> 1), the scaled error sampler has low *ESP* for both V and W.

The state sampler agrees with the scaled disturbance sampler about V and with the scaled error sampler about W. It's at it's best for (V, W) when the signal-to-noise ratio is near 1.

Despite not being an SA-AA pair, $\gamma_{0:T}$ and $\psi_{0:T}$ make a nice beauty and the beast pair.

GIS and Alternating Algorithms for the Local Level Model

There are four possible GIS algorithms and four corresponding alternating algorithms:

Alternating:	State + Dist	State + Error	Dist + Error	Triple
GIS:	State + Dist	State + Error	Dist + Error	Triple

The State-Dist interweaving sampler, for example:

1. Draw $\theta_{0:T}$ from $p(\theta_{0:T}|V^{(k)}, W^{(k)}, y_{1:T})$ using FFBS.
2. Draw (V, W) from $p(V, W|\theta_{0:T}, y_{1:T})$.
3. Update $\gamma_{0:T}$ using $\theta_{0:T}$ and (V, W).
4. Draw $V^{(k+1)}$ from $p(V|\gamma_{0:T}, W, y_{1:T})$.
5. Draw $W^{(k+1)}$ from $p(W|\gamma_{0:T}, V^{(k+1)}, y_{1:T})$, i.e. the same difficult density from before.

The State-Dist alternating sampler would replace step 3 with a draw from $p(\gamma_{0:T}|V, W, y_{1:T})$.

GIS and Alternating Algorithms for the Local Level Model

There are four possible GIS algorithms and four corresponding alternating algorithms:

Alternating:	State + Dist	State + Error	Dist + Error	Triple
GIS:	State + Dist	State + Error	Dist + Error	Triple

The State-Dist interweaving sampler, for example:

1. Draw $\theta_{0:T}$ from $p(\theta_{0:T}|V^{(k)}, W^{(k)}, y_{1:T})$ using FFBS.
2. Draw (V, W) from $p(V, W|\theta_{0:T}, y_{1:T})$.
3. Update $\gamma_{0:T}$ using $\theta_{0:T}$ and (V, W).
4. Draw $V^{(k+1)}$ from $p(V|\gamma_{0:T}, W, y_{1:T})$.
5. Draw $W^{(k+1)}$ from $p(W|\gamma_{0:T}, V^{(k+1)}, y_{1:T})$, i.e. the same difficult density from before.

The State-Dist alternating sampler would replace step 3 with a draw from $p(\gamma_{0:T}|V, W, y_{1:T})$.

GIS and Alternating Algorithms for the Local Level Model

There are four possible GIS algorithms and four corresponding alternating algorithms:

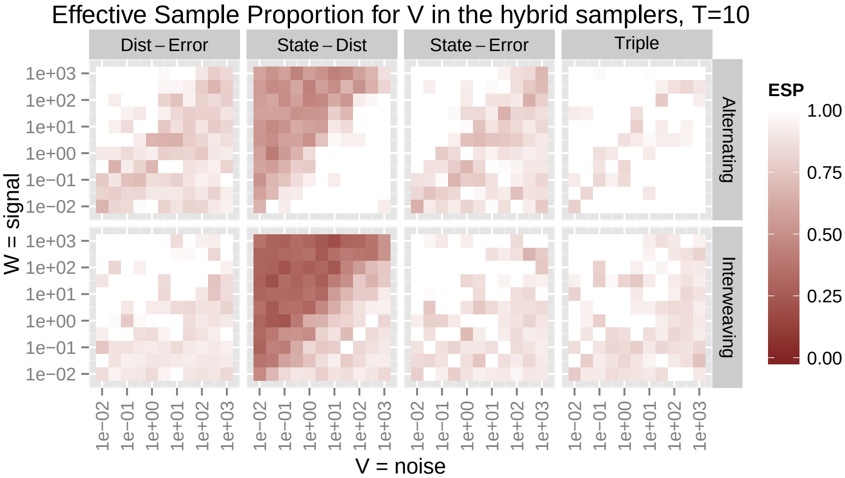
Alternating:	State + Dist	State + Error	Dist + Error	Triple
GIS:	State + Dist	State + Error	Dist + Error	Triple

The State-Dist interweaving sampler, for example:

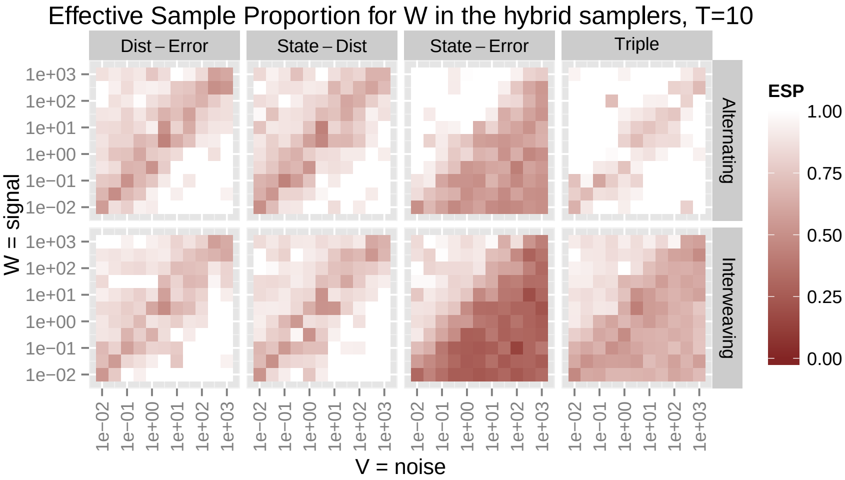
1. Draw $\theta_{0:T}$ from $p(\theta_{0:T}|V^{(k)}, W^{(k)}, y_{1:T})$ using FFBS.
2. Draw (V, W) from $p(V, W|\theta_{0:T}, y_{1:T})$.
3. Update $\gamma_{0:T}$ using $\theta_{0:T}$ and (V, W).
4. Draw $V^{(k+1)}$ from $p(V|\gamma_{0:T}, W, y_{1:T})$.
5. Draw $W^{(k+1)}$ from $p(W|\gamma_{0:T}, V^{(k+1)}, y_{1:T})$, i.e. the same difficult density from before.

The State-Dist alternating sampler would replace step 3 with a draw from $p(\gamma_{0:T}|V, W, y_{1:T})$.

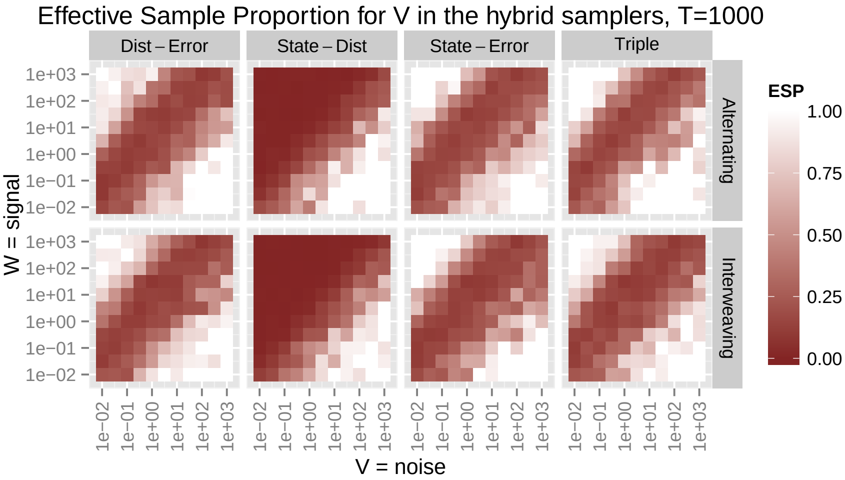
Simulation Results for the GIS Algorithms, $T = 10$



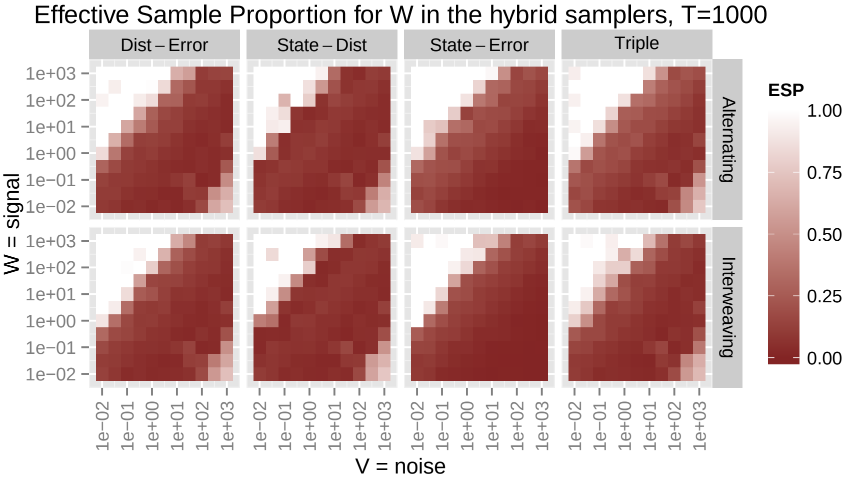
Simulation Results for the GIS Algorithms, $T = 10$



Simulation Results for the GIS Algorithms, $T = 1000$



Simulation Results for the GIS Algorithms, $T = 1000$



Takeaways from GIS Algorithm Simulations

For T large enough, alternating or interweaving doesn't make a difference for ESP (but interweaving is less computationally expensive).

The Dist-Error interweaving algorithm and the Triple interweaving algorithm appear to have identical ESPs.

The "beauty and the beast" intuition works for any given parameter, i.e. for W, the Dist-Error and State-Dist algorithms have the best ESP, but the State-Error algorithms has worse ESP in some regions of the parameter space.

For the reasonable areas of the parameter space (W/V not too small or large), there are still major mixing problems.

Takeaways from GIS Algorithm Simulations

For T large enough, alternating or interweaving doesn't make a difference for ESP (but interweaving is less computationally expensive).

The Dist-Error interweaving algorithm and the Triple interweaving algorithm appear to have identical ESPs.

The "beauty and the beast" intuition works for any given parameter, i.e. for W, the Dist-Error and State-Dist algorithms have the best ESP, but the State-Error algorithms has worse ESP in some regions of the parameter space.

For the reasonable areas of the parameter space (W/V not too small or large), there are still major mixing problems.

Partial CIS for the Local Level Model

Recall $\theta_{0:T}$ is SA for $W|V$ and both $\gamma_{0:T}$ and $\psi_{0:T}$ are AA for (V, W). So all we need is a SA for $V|W$ for a CIS algorithm.

This appears hard to find at first glance, instead we can try partial CIS, i.e. an algorithm which only interweaves in one of the Gibbs steps:

1. Draw $\theta_{0:T}$ from $p(\theta_{0:T}|V^{(k)}, W^{(k)}, y_{1:T})$ using FFBS.
2. Draw $V^{(k+1)}$ from $p(V|W^{(k)}, \theta_{0:T}, y_{1:T})$
3. Draw W from $p(W|V^{(k+.5)}, \theta_{0:T}, y_{1:T})$
4. Update $\gamma_{0:T}$ where $\gamma_0 = \theta_0$ and $\gamma_t = (\theta_t - \theta_{t-1})/\sqrt{W}$.
5. Draw $W^{(k+1)}$ from $p(W|V^{(k+1)}, \gamma_{0:T}, y_{1:T})$

Note that we have to use $\gamma_{0:T}$ in step 4, otherwise steps 5 and 3 would be identical draws because $p(W|V, \theta_{0:T}, y_{1:T}) = p(W|V, \psi_{0:T}, y_{1:T})$.

Partial CIS for the Local Level Model

Recall $\theta_{0:T}$ is SA for $W|V$ and both $\gamma_{0:T}$ and $\psi_{0:T}$ are AA for (V, W). So all we need is a SA for $V|W$ for a CIS algorithm.

This appears hard to find at first glance, instead we can try partial CIS, i.e. an algorithm which only interweaves in one of the Gibbs steps:

1. Draw $\theta_{0:T}$ from $p(\theta_{0:T}|V^{(k)}, W^{(k)}, y_{1:T})$ using FFBS.
2. Draw $V^{(k+1)}$ from $p(V|W^{(k)}, \theta_{0:T}, y_{1:T})$
3. Draw W from $p(W|V^{(k+.5)}, \theta_{0:T}, y_{1:T})$
4. Update $\gamma_{0:T}$ where $\gamma_0 = \theta_0$ and $\gamma_t = (\theta_t - \theta_{t-1})/\sqrt{W}$.
5. Draw $W^{(k+1)}$ from $p(W|V^{(k+1)}, \gamma_{0:T}, y_{1:T})$

Note that we have to use $\gamma_{0:T}$ in step 4, otherwise steps 5 and 3 would be identical draws because $p(W|V, \theta_{0:T}, y_{1:T}) = p(W|V, \psi_{0:T}, y_{1:T})$.

Full CIS for the Local Level Model

For $t = 1, 2, ..., T$ define

$$\tilde{\gamma}_t = \frac{\sqrt{W}}{\sqrt{V}}\gamma_t = \frac{\theta_t - \theta_{t-1}}{\sqrt{V}} = \frac{w_t}{\sqrt{V}}$$

with $\tilde{\gamma}_0 = \gamma_0 = \theta_0$.

The model written in terms of $\tilde{\gamma}_{0:T}$ is

$$y_t|\tilde{\gamma}_{0:T}, V, W \overset{ind}{\sim} N(\tilde{\gamma}_0 + \sqrt{V}\textstyle\sum_{s=1}^{t-1}\tilde{\gamma}_s, V)$$
$$\tilde{\gamma}_t|V, W \overset{iid}{\sim} N(0, W/V)$$

So $\gamma_{0:T}$ & $\tilde{\gamma}_{0:T}$ make a AA-SA pair for $W|V$

Full CIS for the Local Level Model

For $t = 1, 2, ..., T$ define

$$\tilde{\gamma}_t = \frac{\sqrt{W}}{\sqrt{V}}\gamma_t = \frac{\theta_t - \theta_{t-1}}{\sqrt{V}} = \frac{w_t}{\sqrt{V}}$$

with $\tilde{\gamma}_0 = \gamma_0 = \theta_0$.

The model written in terms of $\tilde{\gamma}_{0:T}$ is

$$y_t|\tilde{\gamma}_{0:T}, V, W \overset{ind}{\sim} N(\tilde{\gamma}_0 + \sqrt{V}\textstyle\sum_{s=1}^{t-1}\tilde{\gamma}_s, V)$$
$$\tilde{\gamma}_t|V, W \overset{iid}{\sim} N(0, W/V)$$

So $\gamma_{0:T}$ & $\tilde{\gamma}_{0:T}$ make a AA-SA pair for $W|V$

Full CIS for the Local Level Model

For $t = 1, 2, ..., T$ define

$$\tilde{\psi}_t = \frac{\sqrt{V}}{\sqrt{W}}\psi_t = \frac{y_t - \theta_t}{\sqrt{W}} = \frac{v_t}{\sqrt{W}}$$

with $\tilde{\psi}_0 = \psi_0 = \theta_0$.

The model written in terms of $\tilde{\psi}_{0:T}$ is

$$y_t|\tilde{\psi}_{0:T}, y_{0:t-1}, V, W \sim N(y_{t-1} + \sqrt{W}(\tilde{\psi}_t - \tilde{\psi}_{t-1}), W)$$
$$\tilde{\psi}_t|V, W \overset{iid}{\sim} N(0, V/W)$$

except for $t = 1$, the mean of the system equation is $\sqrt{W}\tilde{\psi}_1 - \tilde{\psi}_0$.

So $\psi_{0:T}$ & $\tilde{\psi}_{0:T}$ make a AA-SA pair for $V|W$

Full CIS for the Local Level Model

For $t = 1, 2, ..., T$ define

$$\tilde{\psi}_t = \frac{\sqrt{V}}{\sqrt{W}}\psi_t = \frac{y_t - \theta_t}{\sqrt{W}} = \frac{v_t}{\sqrt{W}}$$

with $\tilde{\psi}_0 = \psi_0 = \theta_0$.

The model written in terms of $\tilde{\psi}_{0:T}$ is

$$y_t|\tilde{\psi}_{0:T}, y_{0:t-1}, V, W \sim N(y_{t-1} + \sqrt{W}(\tilde{\psi}_t - \tilde{\psi}_{t-1}), W)$$
$$\tilde{\psi}_t|V, W \overset{iid}{\sim} N(0, V/W)$$

except for $t = 1$, the mean of the system equation is $\sqrt{W}\tilde{\psi}_1 - \tilde{\psi}_0$.

So $\psi_{0:T}$ & $\tilde{\psi}_{0:T}$ make a AA-SA pair for $V|W$

Full CIS for the Local Level Model

It turns out that

$$p(W|\tilde{\gamma}_{0:T}, V, y_{1:T}) = p(W|\theta_{0:T}, V, y_{1:T})$$
$$p(V|\tilde{\psi}_{0:T}, W, y_{1:T}) = p(V|\theta_{0:T}, W, y_{1:T})$$

which makes full CIS look like an extention of partial CIS:

1. Draw $\theta_{0:T}$ from $p(\theta_{0:T}|V^{(k)}, W^{(k)}, y_{1:T})$.
2. Draw V from $p(V|W^{(k)}, \theta_{0:T}, y_{1:T})$.
3. Update $\psi_{0:T}$ from V and $\theta_{0:T}$.
4. Draw $V^{(k+1)}$ from $p(V|W^{(k)}, \psi_{0:T}, y_{1:T})$.
5. Update $\theta_{0:T}$ from $V^{(k+1)}$ and $\psi_{0:T}$.
6. Draw W from $p(W|V^{(k+1)}, \theta_{0:T}, y_{1:T})$.
7. Update $\gamma_{0:T}$ from W and $\theta_{0:T}$.
8. Draw $W^{(k+1)}$ from $p(W|V^{(k+1)}, \gamma_{0:T}, y_{1:T})$.

Full CIS for the Local Level Model

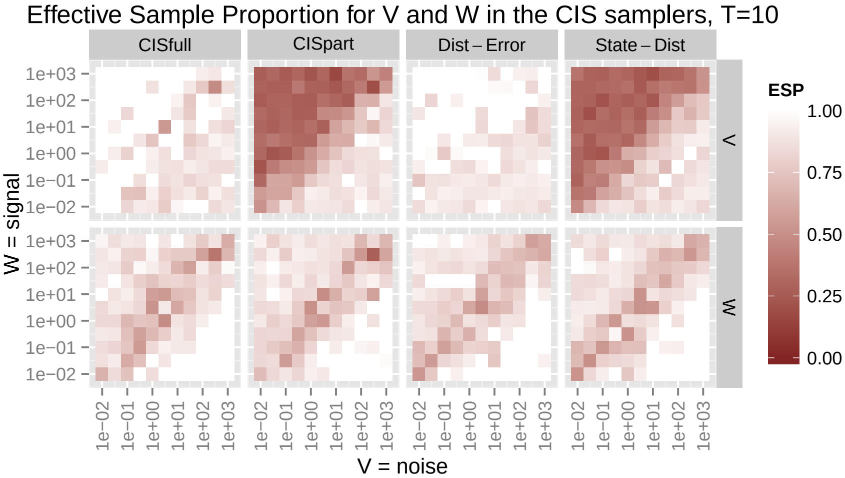
It turns out that

$$
\begin{aligned}
p(W|\tilde{\gamma}_{0:T}, V, y_{1:T}) &= p(W|\theta_{0:T}, V, y_{1:T}) \\
p(V|\tilde{\psi}_{0:T}, W, y_{1:T}) &= p(V|\theta_{0:T}, W, y_{1:T})
\end{aligned}
$$

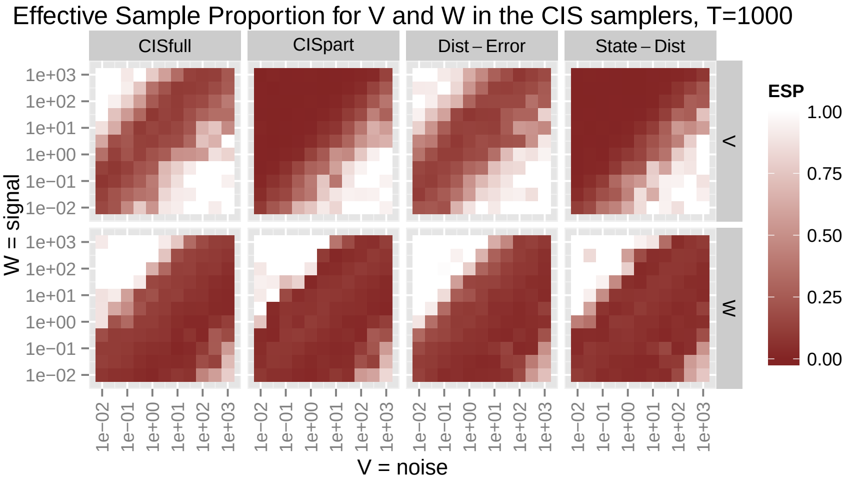
which makes full CIS look like an extention of partial CIS:

1. Draw $\theta_{0:T}$ from $p(\theta_{0:T}|V^{(k)}, W^{(k)}, y_{1:T})$.
2. Draw V from $p(V|W^{(k)}, \theta_{0:T}, y_{1:T})$.
3. Update $\psi_{0:T}$ from V and $\theta_{0:T}$.
4. Draw $V^{(k+1)}$ from $p(V|W^{(k)}, \psi_{0:T}, y_{1:T})$.
5. Update $\theta_{0:T}$ from $V^{(k+1)}$ and $\psi_{0:T}$.
6. Draw W from $p(W|V^{(k+1)}, \theta_{0:T}, y_{1:T})$.
7. Update $\gamma_{0:T}$ from W and $\theta_{0:T}$.
8. Draw $W^{(k+1)}$ from $p(W|V^{(k+1)}, \gamma_{0:T}, y_{1:T})$.

Simulation Results for CIS Algorithms, $T = 10$



Simulation Results for CIS Algorithms, $T = 1000$



Main Takeaways from CIS Simulations

Full CIS looks identical to Dist-Error GIS and Partial CIS looks identical to State-Dist GIS.

This suggests that another Partial CIS algorithm exists that implicitly uses $\tilde{\psi}_{0:T}$ instead of $\tilde{\gamma}_{0:T}$ and behaves identically to the State-Error GIS algorithm.

Upshot: there's no good reason to use CIS instead of GIS in the local level model since CIS requires more computation.

Main Takeaways from CIS Simulations

Full CIS looks identical to Dist-Error GIS and Partial CIS looks identical to State-Dist GIS.

This suggests that another Partial CIS algorithm exists that implicitly uses $\tilde{\psi}_{0:T}$ instead of $\tilde{\gamma}_{0:T}$ and behaves identically to the State-Error GIS algorithm.

Upshot: there's no good reason to use CIS instead of GIS in the local level model since CIS requires more computation.

Recommendations in the Local Level Model

If T is small (< 100), use the state sampler.

If T is large, use the Dist-Error GIS sampler, but spend some time obtaining efficient draws from the complex densities — $p(W|\gamma_{0:T}, V, y_{1:T}$ and $p(V|\psi_{0:T}, W, y_{1:T})$.

Possibly use metropolis steps for (V, W) jointly in the Dist-Error GIS sampler — untested, but likely has decent mixing properties and avoids the expensive sampling steps.

Recommendations in the Local Level Model

If T is small (< 100), use the state sampler.

If T is large, use the Dist-Error GIS sampler, but spend some time obtaining efficient draws from the complex densities — $p(W|\gamma_{0:T}, V, y_{1:T}$ and $p(V|\psi_{0:T}, W, y_{1:T})$.

Possibly use metropolis steps for (V, W) jointly in the Dist-Error GIS sampler — untested, but likely has decent mixing properties and avoids the expensive sampling steps.

Generalizations to other DLMS

Suppose y_t and θ_t are now vectors and

$$y_t|\theta_{0:T} \overset{ind}{\sim} N(F_t\theta_t, V)$$
$$\theta_t|\theta_{0:t-1} \sim N(G_t\theta_{t-1}, W)$$

with F_t, G_t known matricies for $t = 1, 2, ..., T$ and V, W unknown covariance matricies. Then for $t = 1, 2, ..., T$:

$$\gamma_t = W^{-\frac{1}{2}}(\theta_t - G_t\theta_{t-1})$$
$$\psi_t = V^{-\frac{1}{2}}(y_t - F_t\theta_t)$$

Note $dim(\gamma_t) = dim(\theta_t)$ and $dim(\psi_t) = dim(y_t)$ while in general $dim(\theta_t) \neq dim(y_t)$.

Multivariate analogue of W/V: $|W|/|V|$? Ratio of eigenvalues?

Generalizations to other DLMS

Suppose y_t and θ_t are now vectors and

$$y_t|\theta_{0:T} \overset{ind}{\sim} N(F_t\theta_t, V)$$
$$\theta_t|\theta_{0:t-1} \sim N(G_t\theta_{t-1}, W)$$

with F_t, G_t known matricies for $t = 1, 2, ..., T$ and V, W unknown covariance matricies.Then for $t = 1, 2, ..., T$:

$$\gamma_t = W^{-\frac{1}{2}}(\theta_t - G_t\theta_{t-1})$$
$$\psi_t = V^{-\frac{1}{2}}(y_t - F_t\theta_t)$$

Note $dim(\gamma_t) = dim(\theta_t)$ and $dim(\psi_t) = dim(y_t)$ while in general $dim(\theta_t) \neq dim(y_t)$.

Multivariate analogue of W/V: $|W|/|V|$? Ratio of eigenvalues?

Generalizations to other DLMS

Suppose y_t and θ_t are now vectors and

$$
\begin{aligned}
y_t|\theta_{0:T} &\overset{ind}{\sim} N(F_t\theta_t, V)\\
\theta_t|\theta_{0:t-1} &\sim N(G_t\theta_{t-1}, W)
\end{aligned}
$$

with F_t, G_t known matricies for $t = 1, 2, ..., T$ and V, W unknown covariance matricies.Then for $t = 1, 2, ..., T$:

$$
\begin{aligned}
\gamma_t &= W^{-\frac{1}{2}}(\theta_t - G_t\theta_{t-1})\\
\psi_t &= V^{-\frac{1}{2}}(y_t - F_t\theta_t)
\end{aligned}
$$

Note $dim(\gamma_t) = dim(\theta_t)$ and $dim(\psi_t) = dim(y_t)$ while in general $dim(\theta_t) \neq dim(y_t)$.

Multivariate analogue of W/V: $|W|/|V|$? Ratio of eigenvalues?

References I

[1] Sylvia Frühwirth-Schnatter. Efficient bayesian parameter estimation for state space models based on reparameterizations. *State Space and Unobserved Component Models: Theory and Applications*, pages 123–151, 2004.

[2] Jun S Liu and Ying Nian Wu. Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94 (448):1264–1274, 1999.

[3] Omiros Papaspiliopoulos, Gareth O Roberts, and Martin Sköld. A general framework for the parametrization of hierarchical models. *Statistical Science*, pages 59–73, 2007.

[4] David A Van Dyk and Xiao-Li Meng. The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1), 2001.

[5] Yaming Yu and Xiao-Li Meng. 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(3):531–570, 2011.