7

Efficient Bayesian parameter estimation

Sylvia Frühwirth-Schnatter

Department of Applied Statistics (IFAS), Johannes Kepler University

Abstract

The paper contributes to practical Bayesian estimation of state space models using Markov chain Monte Carlo (MCMC) and data augmentation methods. The main topic is the relation between parameterisation and computational efficiency of the resulting MCMC sampler. Standard data augmentation based directly on the dynamic process of the model, also called centred parameterisation, is computationally efficient as long as the variability of the dynamic process is important to explain the variability of the marginal process. This centred parameterisation, however, is shown to be computationally inefficient for models where the unobserved component is hardly changing over time. For such a state space model various reparameterisation techniques are considered, among them noncentring the location as in Pitt and Shephard (1999a) and noncentring the variance as suggested in Meng and van Dyk (1998) for random effects models. A new reparameterisation technique based on the standardised disturbances is suggested that is computationally efficient over a wide range of overidentified models.

State Space and Unobserved Component Models: Theory and Applications, eds. Andrew C. Harvey, Siem Jan Koopman and Neil Shephard. Published by Cambridge University Press.

7.1 Introduction

State space models are a rather intuitive tool for modelling empirical time series through hidden processes. The state space model is usually formulated in a natural way, where the model parameters governing the dynamic part of the process appear only in the transition equation. Consider, as a typical example, the following time varying parameter model:

$$\beta_t = \phi\beta_{t-1} + (1-\phi)\mu + w_t, \qquad w_t \sim N\left(0, \sigma_w^2\right), \tag{7.1}$$

$$y_t = Z_t\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \tag{7.2}$$

$t = 1, \ldots, N$, where β_0 is an unknown starting value and ε_t and w_t are independent processes. Model (7.1) and (7.2) is a state space model with state process β_t. Among the unknown model parameters θ, where $\theta = (\mu, \phi, \sigma_w^2, \sigma_\varepsilon^2)$, all parameters that determine the stochastic law for the hidden process, namely (μ, ϕ, σ_w^2), appear only in the transition equation (7.1).

For unknown model parameters θ, Bayesian inference usually relies on data augmentation by choosing missing data $\tilde{X}$ and sampling from the joint posterior $\pi(\tilde{X}, \theta|y)$, where $y = (y_1, \ldots, y_N)$ is the observed time series. A natural candidate for the missing data is the state process: $\tilde{X} = (\beta_0, \ldots, \beta_N)$. Whereas the joint posterior $\pi(\tilde{X}, \theta|y)$ is rather complex, the conditional posterior densities $\pi(\tilde{X}|\theta, y)$ and $\pi(\theta|\tilde{X}, y)$ are often of closed form, especially for a Gaussian state space model. It has been recognised by various authors (Carlin, Polson and Stoffer 1992a, Carter and Kohn 1994, Frühwirth-Schnatter 1994c) that due to this specific structure MCMC estimation based on data augmentation could be implemented for joint estimation of $\tilde{X}$ and θ through a two-block sampler where one samples the missing data conditional on the model parameters and the model parameters conditional on the missing data.

As an example, consider again model (7.1) and (7.2). Conditional on a known value of θ, the model is a Gaussian linear state space model and the moments of the joint posterior $\pi(\tilde{X}|\theta, y)$ may be obtained with the Kalman filter and the associated smoothing formulae. Conditional on a known state process $\tilde{X}$, the model factors in two independent components, the autoregressive state equation (7.1) and the regression model (7.2) for which parameter estimation is pretty standard. Therefore a straightforward MCMC sampler for this model consists of two main blocks:

(a) Sample a path of $\tilde{X}$ conditional on known model parameters

$$\theta = (\mu, \sigma_w^2, \sigma_\varepsilon^2, \phi)$$

using one of the multimove samplers discussed in Carter and Kohn (1994), Frühwirth-Schnatter (1994c), de Jong and Shephard (1995) or Durbin and Koopman (2002).

(b) Sample the model parameters $(\mu, \sigma_w^2, \sigma_\varepsilon^2, \phi)$ conditional on $\tilde{X}$. Conditional on knowing $\tilde{X}$, the parameters μ, ϕ and σ_w^2 appearing in the state equation (7.1) are independent from the parameter σ_ε^2 appearing in the observation equation (7.2).

Bayesian parameter estimation for state space models using an MCMC sampler, which is alternating between the updates of $\tilde{X}$ and θ, is very convenient and often leads to sensible results. Under certain circumstances, however, it turns out that the convergence properties of this sampler may be extremely poor leading to a painfully slow and computationally inefficient estimation method. For the time varying parameter model (7.1) and (7.2), for instance, this sampler will be mixing poorly, if β_t is changing only slowly over time (σ_w^2 close to 0), see also Figure 7.1. The same problem is also encountered in many other state space models of practical relevance for financial econometrics such as dynamic factor models or stochastic volatility models, see e.g. Kim, Shephard and Chib (1998), Frühwirth-Schnatter and Geyer (1996), Shephard (1996), and Roberts, Papaspiliopoulos and Dellaportas (2004).

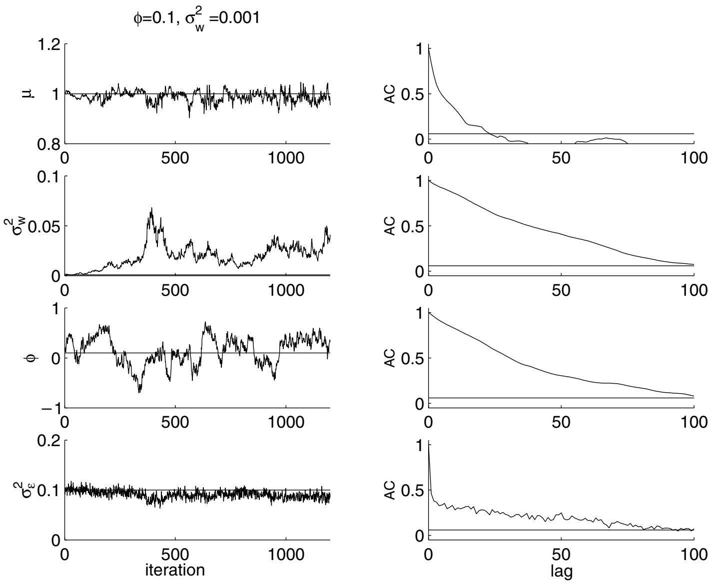

Fig. 7.1. Data simulated from model (7.1) and (7.2) with $Z_t = 1$, $\phi = 0.1$ and $\sigma_w^2 = 0.001$; MCMC draws with empirical autocorrelation obtained for μ, σ_w^2, σ_ε^2 and ϕ for parameterisation (7.1) and (7.2).

Somewhat surprisingly, considerable improvement is often possible by simple reparameterisation techniques. The relation between the convergence properties of the alternating MCMC sampler and the parameterisation is best understood for the random-effects model which is that special case of model (7.1) and (7.2) where $\phi = 0$. Gelfand, Sahu and Carlin (1995) discuss the effect of recentring the location μ of β_t, whereas Meng and van Dyk (1998) discussed the effect of recentring the scale σ_w of β_t. It has been recognised, that there exists a whole continuum of reparameterisations both of the locations (Papaspiliopoulos, Roberts and Sköld 2003) and the scale (Meng and van Dyk 1998). Far less is known about optimal parameterisation for time series models where $\phi \neq 0$. Recentring the location of a time series has been studied by Pitt and Shephard (1999a). Reparameterisation techniques for stochastic volatility models are discussed in Shephard (1996), Roberts, Papaspiliopoulos and Dellaportas (2004) and Frühwirth-Schnatter and Sögner (2002).

This contribution reviews some of the existing reparameterisation techniques and introduces some new techniques, especially designed for time series models. The discussion will be confined entirely to the time varying parameter model (7.1) and (7.2). Section 7.2 reviews reparametrisation techniques for an unknown location parameter μ for the case where $\sigma_w^2, \sigma_\varepsilon^2$ and ϕ are known. Section 7.3 explores the case where additionally the variance parameters σ_w^2 and σ_ε^2 are unknown, whereas ϕ is still known. A new parameterisation for time series models is suggested that is based on recentring the scale of the hidden process. Section 7.4 considers the most interesting case where all parameters are unknown. Here an MCMC sampler is introduced that is based on the noncentred disturbances of the hidden process. Section 7.5 presents more details on implementing MCMC estimation for all parameterisations. Section 7.6 concludes by discussing the results of this paper in the light of more general state space models.

7.2 Reparameterisation of the location

This section reviews the reparametrisation technique for an unknown location parameter μ of the time varying parameter model (7.1) and (7.2) for the case where all other model parameters, namely $\sigma_w^2, \sigma_\varepsilon^2$ and ϕ are known. The material is based on Gelfand, Sahu and Carlin (1995), Papaspiliopoulos, Roberts and Sköld (2003) and Pitt and Shephard (1999a).

7.2.1 Noncentring the location for the random-effects model

7.2.1.1 Recentring of the location

The relation between the convergence properties of the alternating MCMC sampler reviewed in Section 7.1 and the parameterisation chosen for the location is best understood for a random-effects model under the assumption of known variances σ_w^2 and σ_ε^2. A random-effects model is that special case of model (7.1) and (7.2) where $\phi = 0$:

$$\beta_t = \mu + w_t, \qquad w_t \sim N\left(0, \sigma_w^2\right), \tag{7.3}$$

$$y_t = Z_t\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right). \tag{7.4}$$

A very clear discussion of how to parameterise the location for a random-effects models appears in Gelfand, Sahu and Carlin (1995). They called parameterisation (7.3) and (7.4) the centred parameterisation, as the random effect β_t is centred around the *a priori* expected value: $\mathrm{E}(\beta_t) = \mu$. An alternative parameterisation is based on the transformation $\tilde{\beta}_t = \beta_t - \mu$:

$$\tilde{\beta}_t = w_t, \qquad w_t \sim N\left(0, \sigma_w^2\right), \tag{7.5}$$

$$y_t = Z_t\mu + Z_t\tilde{\beta}_t + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right). \tag{7.6}$$

Gelfand, Sahu and Carlin (1995) called parameterisation (7.5) and (7.6) the noncentred parameterisation, as the corresponding random effect $\tilde{\beta}_t$ is no longer centred around the *a priori* expected value, but around 0: $\mathrm{E}(\tilde{\beta}_t) = 0$. An interesting effect of this reparameterisation has been to move the location parameter μ from the transition equation (7.3) to the observation equation (7.6) where it now appears as a fixed effect.

7.2.1.2 To centre or not to centre the location?

For known variances σ_w^2 and σ_ε^2, Gelfand, Sahu and Carlin (1995) gave general conditions, under which the centred parameterisation yields MCMC algorithms with faster convergence than the noncentred one. In general, the noncentred parameterisation should be preferred, if only a small fraction of the (marginal) variability $\mathrm{V}(y_t)$ of y_t is caused by heterogeneity in β_t, i.e. $\mathrm{V}(y_t|\beta_t)$ is only slightly smaller than $\mathrm{V}(y_t)$. For a random design Z_t this is equivalent to the condition that the coefficient of determination defined by

$$D = 1 - \frac{V(y_t|\beta_t)}{V(y_t)} = \frac{E(Z_t^2)\sigma_w^2}{\sigma_\varepsilon^2 + E(Z_t^2)\sigma_w^2} = \frac{E(Z_t^2)\sigma_w^2/\sigma_\varepsilon^2}{1 + E(Z_t^2)\sigma_w^2/\sigma_\varepsilon^2} \tag{7.7}$$

is close to 0; $\sigma_w^2/\sigma_\varepsilon^2$ is known as the signal-to-noise ratio.

Roberts and Sahu (1997) showed that for a model with known variances the coefficient D determines the convergence rate of the alternating MCMC sampler under the two different parameterisation. Whereas the convergence rate is $1 - D$ for the centred parameterisation, it is equal to D under the noncentred parameterisation. Therefore, for the random effects model the parameterisation that is centred in the location is preferred if D is bigger than $1/2$, whereas the noncentred parameterization is preferred if D is smaller than $1/2$.

An important (and disturbing) conclusion from Roberts and Sahu (1997) is the following. If the signal-to-noise ratio $\sigma_w^2/\sigma_\varepsilon^2$ goes to infinity, the noncentred parameterisation will not converge geometrically, whereas the centred parameterisation will not converge geometrically if the signal-to-noise ratio $\sigma_w^2/\sigma_\varepsilon^2$ goes to 0. Each of the two parameterisations may be arbitrarily bad under unfortunate parameter constellations.

7.2.2 Noncentring the location for a time series model

7.2.2.1 Noncentring the location

Far less is known about recentring the location of a time series models where $\phi \neq 0$, a notable exception being Pitt and Shephard (1999a). Following Gelfand, Sahu and Carlin (1995), Pitt and Shephard (1999a) called parameterisation (7.1) and (7.2) the centred one also for a time series model, as the long-run mean of β_t is equal to μ. As an alternative, Pitt and Shephard (1999a) considered a model that is noncentred in the location:

$$\tilde{\beta}_t = \phi\tilde{\beta}_{t-1} + w_t, \qquad w_t \sim N\left(0, \sigma_w^2\right), \tag{7.8}$$

$$y_t = Z_t\mu + Z_t\tilde{\beta}_t + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right). \tag{7.9}$$

For $\phi \to 0$, this corresponds to the noncentred random-effects model (7.3) and (7.4).

7.2.2.2 To centre or not to centre the location?

For a model where $Z_t = 1$ and the variances are assumed to be known, the following results are proven in Pitt and Shephard (1999a, p. 71). For $\phi \to 1$, the convergence rate of the centred parameterisation goes to 0, whereas the convergence rate of the noncentred parameterisation goes to 1. Thus for the limiting random walk model the noncentred parameterisation does not converge geometrically regardless of the signal-to-noise ratio $\sigma_w^2/\sigma_\varepsilon^2$.

For $\phi < 1$, the variances matter, and the centred parameterisation is better than the noncentred parameterisation if

$$\frac{\sigma_w^2}{(1-\phi)^2} > \sigma_\varepsilon^2. \tag{7.10}$$

For $\phi = 0$, this result coincides with the previous result, that the centred parameterisation is better if $\sigma_w^2 > \sigma_\varepsilon^2$. Also for $\phi > 0$, the centred parameterisation is better in the case of high heterogeneity, whereas the noncentred parameterisation is better in the case of low heterogeneity. With ϕ approaching 1, however, the range of variance parameters σ_w^2 for which the noncentred parameterisation is better than the centred one becomes smaller and may be just a very small region close to 0 for highly persistent processes.

7.2.2.3 Illustration for simulated data

For illustration, we simulated time series of length $N = 500$ from a model where $\mu = 1$ and $\sigma_\varepsilon^2 = 0.1$ under three different levels of heterogeneity: a low level of heterogeneity with $\sigma_w^2 = 0.001$, a medium level of heterogeneity with $\sigma_w^2 = 0.05$ and a high level of heterogeneity with $\sigma_w^2 = 1$. Here, Z_t takes randomly the values -1, 0, 1, a choice that guarantees identifiability also for ϕ approaching 0.

MCMC sampling of $\tilde{X}$ and μ, with all other model parameters fixed, is carried out under the two parameterisations for all settings of heterogeneity for $\phi = 0.1$ and $\phi = 0.95$. Figure 7.2 illustrates the following results already obtained in Section 7.2.2.2 for a time series model, where the variances are known:

- For time series with low autocorrelation ($\phi = 0.1$ in the example) we have a result similar to that of the random-effects model, namely that the centred parameterisation is better than the noncentred one for a high level of heterogeneity.
- For highly persistent time series ($\phi = 0.95$ in the example), the centred parameterisation is more efficient even for a low level of heterogeneity.

7.2.3 Partially noncentring of the location

Papaspiliopoulos, Roberts and Sköld (2003) showed for a random-effects model with known variances that there always exists a reparameterisation which has the optimal convergence rate 0. They call this parameterisation partially noncentred. The partially noncentred parameterisation is defined as a continuum between the centred parameterisation (7.3) and (7.4) and

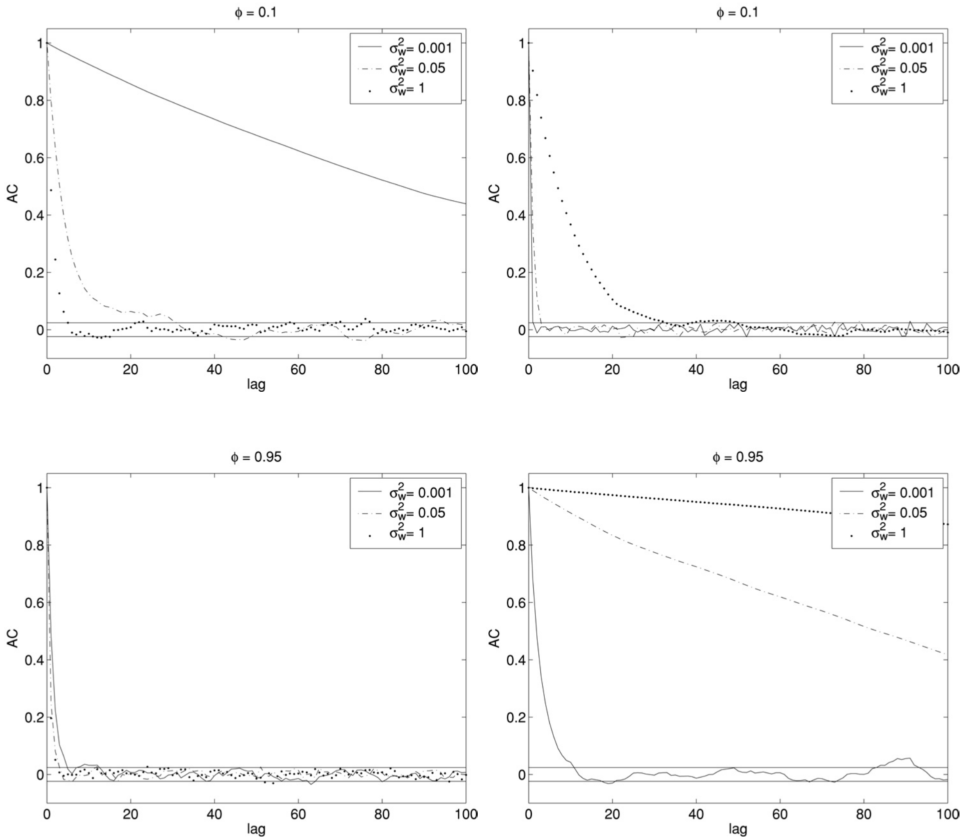

Fig. 7.2. Data simulated from model (7.1) and (7.2) with $\phi = 0.1$ (top) and $\phi = 0.95$ (bottom) for various values of σ_w^2; empirical autocorrelation of the MCMC draws of μ (ϕ, σ_w^2 and σ_ε^2 fixed at their true values) for the centred parameterisation (left hand side) and the parameterisation noncentred in the location (right hand side).

the noncentred parameterisation (7.5) and (7.6). The state vector is defined as a weighted mean of β_t and $\tilde{\beta}_t$:

$$\beta_t^w = W_t \tilde{\beta}_t + (1 - W_t)\beta_t, \tag{7.11}$$

with $W_t = 0$ corresponding to the centred and $W_t = 1$ corresponding to the noncentred parameterisation, respectively. The state vector of the partially centred parameterisation is related to the state vector β_t of the centred parameterisation through

$$\beta_t^w = \beta_t - W_t \mu, \tag{7.12}$$

whereas the relation to the state vector $\tilde{\beta}$ of the noncentred parameterisation is given by

$$\beta_t^w = \tilde{\beta}_t + (1 - W_t)\mu. \tag{7.13}$$

The partially noncentred model takes the following form:

$$\beta_t^w = (1 - W_t)\mu + w_t, \qquad w_t \sim N\left(0, \sigma_w^2\right), \tag{7.14}$$

$$y_t = Z_t \beta_t^w + Z_t W_t \mu + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right). \tag{7.15}$$

Thus for a partially noncentred model the location parameter μ appears both in the transition and the observation equation.

For a model with time invariant predictor $Z_t \equiv Z$, Papaspiliopoulos, Roberts and Sköld (2003) showed that iid samples may be produced by choosing

$$W_t \equiv 1 - D = \frac{1}{1 + Z^2 \sigma_w^2 / \sigma_\varepsilon^2}, \tag{7.16}$$

where D is defined in (7.7) and $\sigma_w^2/\sigma_\varepsilon^2$ is the signal-to-noise ratio. Again we find that the centred parameterisation results under a high signal-to-noise ratio whereas the noncentred parametrisation results under a small signal-to-noise ratio. For more general models with time varying Z_t, Papaspiliopoulos, Roberts and Sköld (2003) showed that iid samples may be produced by choosing data dependent weights:

$$W_t = \frac{1}{1 + Z_t^2 \sigma_w^2 / \sigma_\varepsilon^2}. \tag{7.17}$$

The partially noncentred parameterisation introduced in Papaspiliopoulos, Roberts and Sköld (2003) is limited to the case of a random-effects model and no results were given for time series models. Nevertheless as above, one may define β_t^w as in (7.12) also for a time series model:

$$\beta_t^w = \beta_t - W_t \mu, \tag{7.18}$$

to obtain the following partially centred parameterisation:

$$\beta_t^w = \phi \beta_{t-1}^w + \left[(1 - \phi) - (W_t - \phi W_{t-1})\right] \mu + w_t, \quad w_t \sim N\left(0, \sigma_w^2\right), \tag{7.19}$$

$$y_t = Z_t \beta_t^w + Z_t W_t \mu + \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right). \tag{7.20}$$

For $W_t = 0$ we obtain the centred parameterisation and for $W_t = 1$ the noncentred one. It is, however, still unclear how to select W_t for a time series model in such a way that the corresponding two block MCMC sampler produces (nearly) iid draws.

7.3 Reparameterisation of the scale

7.3.1 The effect of unknown variances

In the previous section we showed that under high levels of heterogeneity MCMC estimation of $(\tilde{X}, \mu)$ may be safely based on the centred parameterisation (7.1) and (7.2). With σ_w^2 approaching zero, however, the sampler deteriorates and noncentring the location may lead to a dramatic improvement of the convergence properties of the MCMC sampler. These conclusions, however, were drawn under the assumption of *known* variances, and little is known about the effect of also considering the variances σ_w^2 and σ_ε^2 as unknown parameters especially for time series models with $\phi \neq 0$.

For a random-effects model with unknown variances it has been noted by Meng and van Dyk (1998) that for models with low but unknown levels of heterogeneity noncentring the location, although improving convergence in μ, leads to a sampler that is rather inefficient in sampling σ_w^2. For random-effects models with low (but unknown) heterogeneity, Meng and van Dyk (1998) and Van Dyk and Meng (2001) suggested recentring the model both in location and scale and demonstrated considerable improvement over the parameterisation that is noncentred only in the location. This material will be reviewed in Section 7.3.2. The extension to time series models where ϕ is a known nonzero parameter will be considered in Section 7.3.3.

7.3.2 Noncentring both location and scale for a random effects model

7.3.2.1 Recentring of location and scale

Meng and van Dyk (1998) showed for a random-effects model which is noncentred in the location that sampling both the location and the variances leads to a poor sampler, when the coefficient of determination D defined in (7.7) is small. Note that although the parameterisation is noncentred in the location, which is preferable to the centred one for D small, the parameterisation is obviously not optimal for the scale parameters. As a remedy, Meng and van Dyk (1998) suggested using the following reparameterisation which is based on rescaling the state vector to a variable with unit variance:

$$\beta_t^\star = \frac{\tilde{\beta}_t}{\sigma_w}, \tag{7.21}$$

leading to the following random-effects model:

$$\beta_t^\star = w_t, \qquad w_t \sim N(0,1), \tag{7.22}$$
$$y_t = Z_t\mu + Z_t\sigma_w\beta_t^\star + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right). \tag{7.23}$$

One effect of recentring the scale is that the unknown variance σ_w^2, or rather the standard deviation σ_w, is moved from the transition equation (7.22) to the observation equation (7.23). The model is now parameterised in such a way that the square root of σ_w^2 appears in the observation equation as an unknown regression coefficient with latent, unobserved regressor $\beta_t^\star$. Therefore an MCMC sampler under this parameterisation is easily derived. We will present further details in Section 7.5.

As in (7.22) and (7.23) the scale of the state variable $\beta_t^\star$, measured by $\beta_t^{\star 2}$, is centred around 1, $E(\beta_t^{\star 2}) = 1$, we call this parameterisation noncentred in the scale. In contrast to that parameterisation (7.1) and (7.2) as well as (7.5) and (7.6) are centred in the scale: $\mathrm{E}(\tilde{\beta}_t^2) = \sigma_w^2$. Note that parameterisation (7.22) and (7.23) is noncentred both in the location and in the scale.

7.3.2.2 To centre or not to centre?

As for the location parameter, there exists a whole continuum of reparameterisations of the scale. Meng and van Dyk (1997, 1999) were the first to discuss partially recentring of the scale of a t-distribution. These results were extended to the univariate random-effects model in Meng and van Dyk (1998, p.574). Starting from a model that is noncentred in the location, they defined

$$\beta_t^a = \frac{\tilde{\beta}_t}{(\sigma_w)^A}, \tag{7.24}$$

leading to the following state space model:

$$\beta_t^a = (\sigma_w)^{1-A} w_t, \qquad w_t \sim N(0,1), \tag{7.25}$$
$$y_t = Z_t\mu + Z_t(\sigma_w)^A\beta_t^a + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right). \tag{7.26}$$

Note that $(\beta_t^a)^2$ is a weighted geometric mean between the state vector $\tilde{\beta}_t^2$, that is centred in the scale, and $(\beta_t^\star)^2$ which is noncentred in the scale:

$$(\beta_t^a)^2 = (\tilde{\beta}_t^2)^{1-A}(\beta_t^{\star,2})^A. \tag{7.27}$$

The effect of the reparameterisation is that the unknown variance σ_w^2 appears both in the transition equation (7.25) as well as in the observation equation (7.26). For $A = 0$, we obtain a model that is centred in the scale, whereas for $A = 1$ we obtain a model that is noncentred in the scale.

Meng and van Dyk (1998, equation (4.4)), suggested the following choice for A:

$$A = \frac{2(1-D)}{2-D} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \overline{Z_t^2}\sigma_w^2/2}, \tag{7.28}$$

where D is defined in (7.7). From (7.28) we obtain that for the small heterogeneity case A will be close to 1, and noncentring in the scale is preferred, whereas under high heterogeneity, A is close to 0 and centring in the variance is preferred.

7.3.3 Noncentring both location and scale for a time series model

7.3.3.1 Recentring of location and scale

Practically nothing is known about the gain of recentring a time series model in the scale parameter. A model that is noncentred both in location and scale is given by

$$\beta_t^\star = \phi\beta_{t-1}^\star + w_t, \qquad w_t \sim N(0,1), \tag{7.29}$$

$$y_t = Z_t\mu + Z_t\sigma_w\beta_t^\star + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right). \tag{7.30}$$

As above we will assume that the parameter ϕ is known. As for a random-effects model, the magnitude of the unknown variances plays a crucial role. If the true value of the unknown variance σ_w^2 is not too small, the conclusions drawn in Section 7.2.2.2 about the circumstances under which the centred parameterisation should be preferred remain valid. With σ_w^2 approaching zero, however, these conclusions are no longer valid. Noncentring the location, although improving convergence in μ, leads to a sampler that is rather inefficient in sampling σ_w^2. Noncentring both location and scale improves the sampler considerably, as will be demonstrated below for the simulated data.

7.3.3.2 Illustration simulated for data

We reconsider the simulated data introduced in Section 7.2.2.3. For nonpersistent time series with $\phi = 0.1$, Figure 7.3 compares, for three different levels of heterogeneity, the preferred parameterisation for the location, with

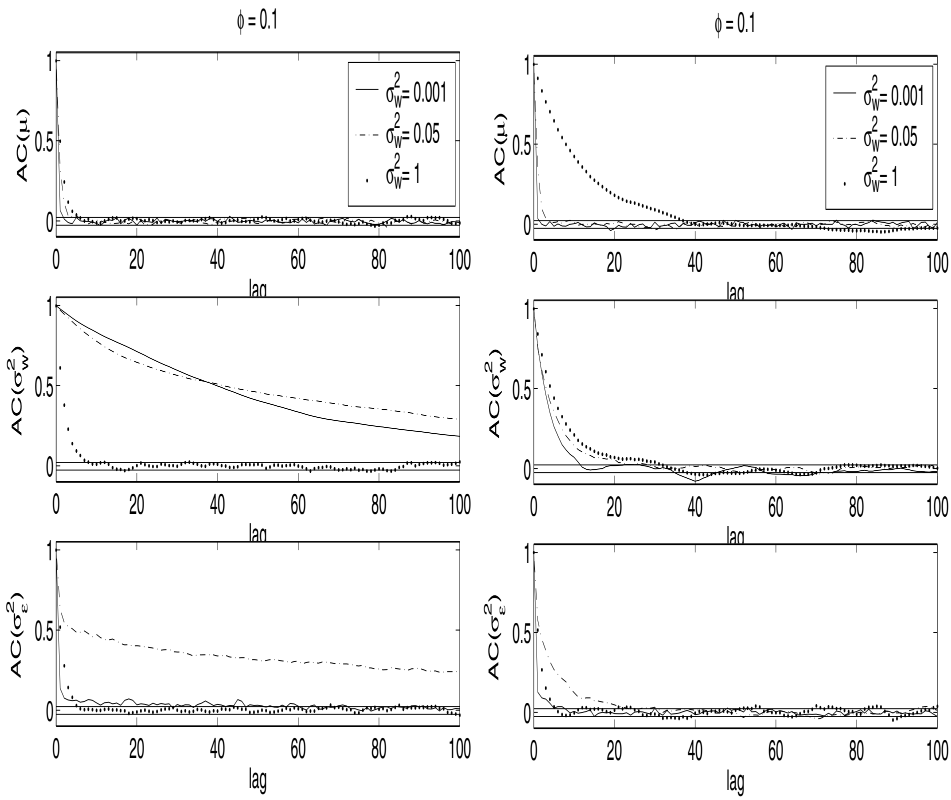

Fig. 7.3. Data simulated from model (7.1) and (7.2) with $\phi = 0.1$ for various values of σ_w^2; empirical autocorrelation of the MCMC draws of μ (top), σ_w^2 (middle) and σ_ε^2 (bottom) (ϕ fixed at the true value) for the preferred parameterisation for the location (left hand side, centred only for $\sigma_w^2 = 1$) and the parameterisation which is noncentred both in location and scale (right hand side)

the variances assumed to be unknown, with the parameterisation that is noncentred both in location and scale. For nonpersistent time series models with unknown small variances noncentring only in the location, although improving convergence in μ, leads to a sampler that is rather inefficient in sampling σ_w^2. Noncentring both location and scale improves the sampler considerably, see especially the example in Figure 7.3 where $\phi = 0.1$ and $\sigma_w^2 = 0.001$.

For persistent time series with $\phi = 0.95$, this comparison is carried out in Figure 7.4. Also for persistent time series model (ϕ close to 1) with unknown small variances the centred parameterisation, although efficient for sampling

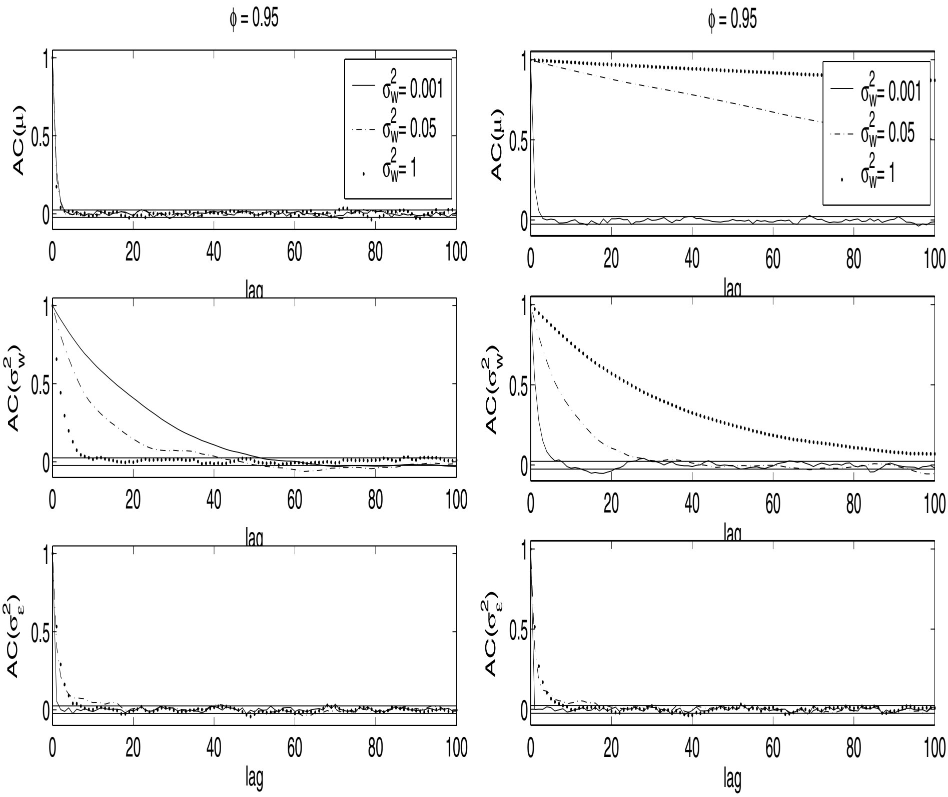

Fig. 7.4. Data simulated from model (7.1) and (7.2) with $\phi = 0.95$ for various values of σ_w^2; empirical autocorrelation of the MCMC draws of μ (top), σ_w^2 (middle) and σ_ε^2 (bottom) (ϕ fixed at the true value) for the centred parameterisation (left hand side) and the parameterisation which is noncentred both in location and scale (right hand side)

μ, leads to a sampler that is rather inefficient for sampling σ_w^2. Again, noncentring both location and scale improves the sampler considerably, see especially the example in Figure 7.4 where $\phi = 0.95$ and $\sigma_w^2 = 0.001$. Compared with nonpersistent time series, however, the range where the noncentred parameterisation is sensible is much smaller for persistent time series; compare Figure 7.3 and Figure 7.4. For highly persistent time series MCMC sampling based on the noncentred parameterisation quickly leads to rather poor results, when σ_w^2 moves away from 0. This result, however, is not surprising in the light of the results of Pitt and Shephard (1999a) discussed in Section 7.2.2.2.

7.4 Reparameterising time series models based on disturbances

7.4.1 The effect of an unknown autocorrelation structure

The results of the previous section hold under the assumption that the parameter ϕ, which determines the autocorrelation structure, is a known quantity. We now discuss what happens if ϕ is an unknown parameter. We will discuss only the case where $\phi > 0$.

The addition of ϕ as an extra unknown parameter does not affect the conclusions drawn so far if the variance σ_w^2 is not too small. However, problems arise when σ_w^2 is too close to 0. In this case the state space model is nearly oversised, making MCMC estimation of the model parameters difficult. Consider for instance, the case where $\sigma_w^2 = 0.001$. For ϕ fixed, the parameterisation (7.29) and (7.30) that is noncentred both in location and scale leads to a very efficient sampler, see again Figures 7.3 and 7.4. When adding ϕ as an unknown parameter, however, the resulting sampler is rather inefficient, see Figures 7.5 and 7.6. Mixing is slow mainly for ϕ, but rather fast for the other parameters. Remember that for this parameterisation ϕ still remains in the transition equation, whereas μ and σ_w^2 are moved to the observation equation. This suggests considering a parameterisation where ϕ is moved from the transition equation to the observation equation as well.

7.4.2 Parameterisation based on noncentred disturbances

A parameterisation in which all unknown parameters are moved into the observation equation and where we use the standardised disturbances w_t in (7.29) rather than $\beta_t^\star$ as missing data is.

$$w_t \sim N(0,1), \tag{7.31}$$

$$y_t = Z_t\mu + Z_t\sigma_w\beta_t^\star + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right), \tag{7.32}$$

where $\beta_t^\star = \phi\beta_{t-1}^\star + w_t$. The missing data are defined as

$$\tilde{X} = (\beta_0^\star, w_1, \ldots, w_N).$$

The transition equation (7.31) and consequently the prior of $w_1, \ldots, w_N$ is independent of any model parameter. It is no longer possible to sample ϕ through a Gibbs sampler, instead we use a random walk Metropolis Hastings algorithm of the Gibbs step, see Section 7.5 for more details. Note that due

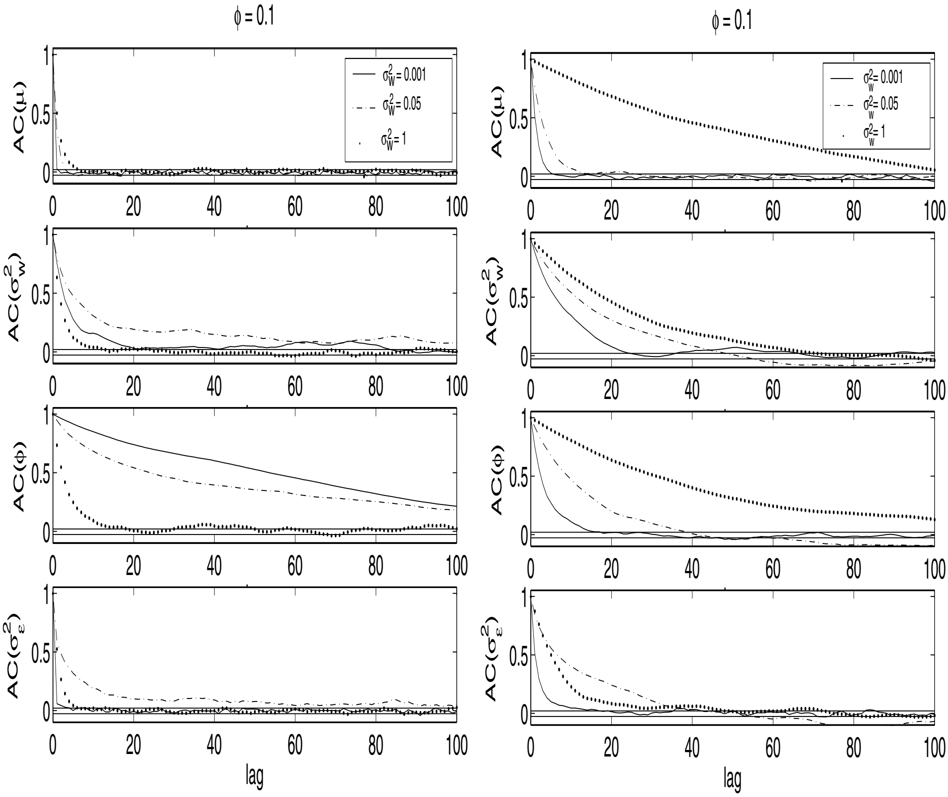

Fig. 7.5. Data simulated from model (7.1) and (7.2) with $\phi = 0.1$ for various values of σ_w^2; empirical autocorrelation of the MCMC draws of μ (top), σ_w^2, ϕ and σ_ε^2 (bottom) for the preferred parameterisation with ϕ known (left hand side, centred in the location and scale for $\sigma_w^2 = 1$, noncentred otherwise) and the parameterisation based on the noncentred disturbances (right hand side)

to the specific choice of $\tilde{X}$, the state process $\beta_1^\star, \ldots, \beta_N^\star$ changes whenever we change ϕ.

The examples in Figures 7.5 and 7.6 show that for σ_w^2 small the parameterisation based on the noncentred disturbances is considerable better than the parameterisation based on the noncentred, but correlated state process $\beta_t^\star$.

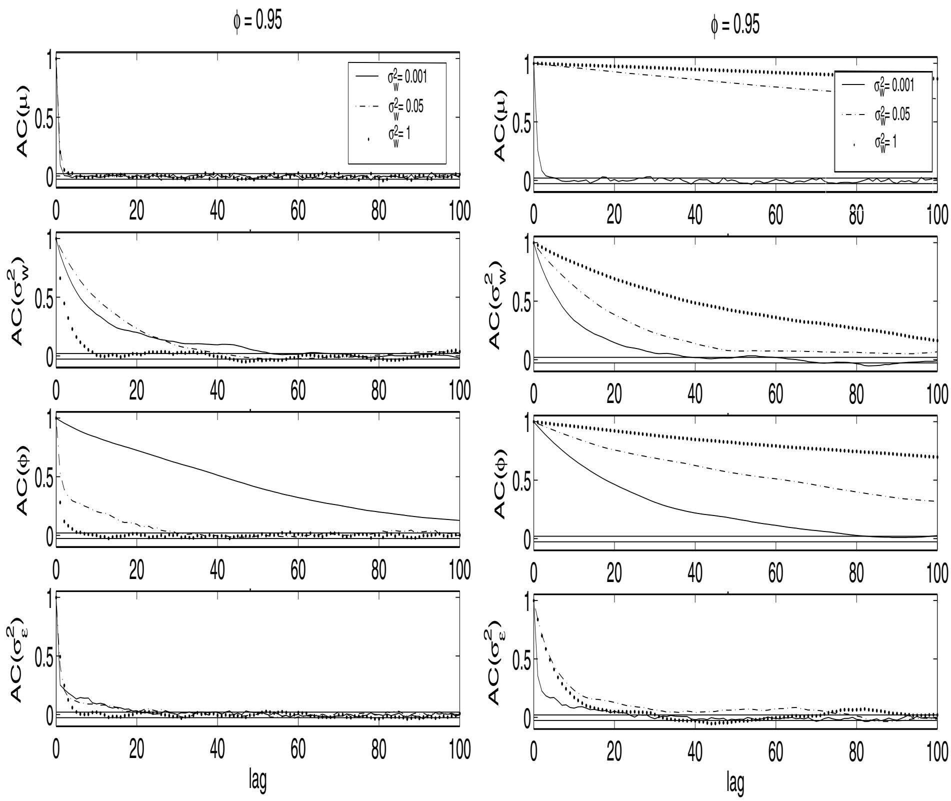

Fig. 7.6. Data simulated from model (7.1) and (7.2) with $\phi = 0.95$ for various values of σ_w^2; empirical autocorrelation of the MCMC draws of μ (top), σ_w^2, ϕ and σ_ε^2 (bottom) for the preferred parameterisation with ϕ known (left hand side, noncentred in the location and scale for $\sigma_w^2 = 0.001$, centered otherwise) and the parameterisation based on the noncentred disturbances (right hand side)

7.4.3 Illustration for simulated data

We reconsider the simulated data introduced in Section 7.2.2.3. First, we consider that parameterisation for the location and scale, which was the preferred one in Section 7.3.3.2 for ϕ fixed, however, this time with ϕ assumed to be unknown. Results are reported in Figure 7.5 for the non-persistent time series with $\phi = 0.1$ and in Figure 7.6 for the persistent time series with $\phi = 0.95$. In both cases, adding ϕ as an unknown parameter leads to a sampler that is inefficient in sampling ϕ for the small heterogeneity case $\sigma_w^2 = 0.001$. For larger values of σ_w^2, however, adding ϕ does not impose any convergence problems. For highly persistent time series the centred parameterisation is valid even for ϕ unknown, as long as σ_w^2 is not too close to 0.

Second, we consider the parameterisation that is based on the noncentred disturbances. A comparison of these two parameterisations is carried out in Figures 7.5 and 7.6 for $\phi = 0.1$ and $\phi = 0.95$, respectively. For a low level of heterogeneity with $\sigma_w^2 = 0.001$, choosing this parameterisation leads to a dramatic improvement of the convergence properties of the sampler compared with the parameterisation discussed above.

This improvement, however, is valid only for nearly oversized models. Figures 7.5 and 7.6 clearly show that the parameterisation based on the noncentred disturbances may be much worse than the centred parameterisation, when the true value of σ_w^2 moves away from 0. For the highly persistent time series in Figures 7.6, the parameterisation based on the noncentred disturbances is worse for all values but $\sigma_w^2 = 0.001$. For the nonpersistent time series in Figure 7.5 the range for which the parameterisation based on the noncentred disturbances is better than or comparable to the other parameterisations is much larger. Nevertheless, none of the parameterisations dominates the other one.

7.4.4 An alternative parameterisation based on disturbances

Various alternative parameterisations based on the disturbances are possible, an example being a parameterisation which is based on choosing the centred shocks u_t in

$$\beta_t = \phi\beta_{t-1} + u_t, \qquad u_t \sim N\left(\gamma, \sigma_w^2\right),$$

where $\gamma = (1-\phi)\mu$, as missing data. This parameterisation reads

$$u_t \sim N\left(\gamma, \sigma_w^2\right), \tag{7.33}$$

$$y_t = Z_t\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \tag{7.34}$$

where $\beta_t = \phi\beta_{t-1} + u_t$. The missing data are defined as

$$\tilde{X} = (\beta_0, u_1, \ldots, u_N).$$

The transition equation (7.33) and consequently the prior of $u_1, \ldots, u_N$ is independent of ϕ, whereas the model parameters μ and σ_w^2 remain in the transition equation. A comparable parameterisation has been applied in Roberts, Papaspiliopoulos and Dellaportas (2001) for the stochastic volatility model of Barndorff-Nielsen and Shephard (2001).

For the time-varying parameter model we could not identify any parameter combinations where a parameterisation based on the centred shocks outperformed either the centred parameterisation or the parameterisation based on the noncentred disturbances. For nearly oversised models with $\sigma_w^2 = 0.001$ the resulting sampler is poor, as μ and σ_w^2 still remain in the transition equation. For models that are best estimated with the centered parameterisation (e.g. for models with $\sigma_w^2 = 1$), the effect of moving ϕ from the transition to the observation equation is a most undesirable one. Moving ϕ introduces conditional independence between ϕ and γ, whereas marginally ϕ and γ are highly correlated, see also Section 7.5.4.

7.4.5 Random selection of the parameterisation

From the investigations of the previous sections we identified only two sensible parameterisations, one centred in both location and scale and one based on the noncentred disturbances, with the noncentred parameterisation being the preferred one for nearly oversized models.

As the choice between the two parameterisation depends on the unknown parameters, it is in general not possible to decide *a priori* which parameterisation to choose. A rather pragmatic solution is to try both of them and to select the one with the lower autocorrelations in the MCMC draws. Alternatively, we could randomly select one of these two parameterisations. First experiences with this hybrid sampler for the simulated time series of this paper are rather promising, see also Figure 7.7. For this hybrid sampler we randomly select the parameterisation before each draw. In particular this gives a 0.5 probability that the parameterisation remains unchained and a 0.5 probability that the model is recentred.

7.5 Bayesian estimation and MCMC implementation

7.5.1 Prior distributions

For Bayesian estimation we assume the following prior distributions. First, we assume a flat prior for μ, an inverted Gamma for σ_ε^2 and a Beta distribution for ϕ. The initial state β_0 is assumed to arise from the marginal distribution of the AR(1) process: $\beta_0|\phi, \mu \sim N(\mu, \sigma_w^2/(1-\phi^2))$, see Schotman (1994) for a motivation.

The prior on σ_w^2 depends on the parameterisation. For a parameterisation that is centred in the scale or is based on the centred shocks, we assume

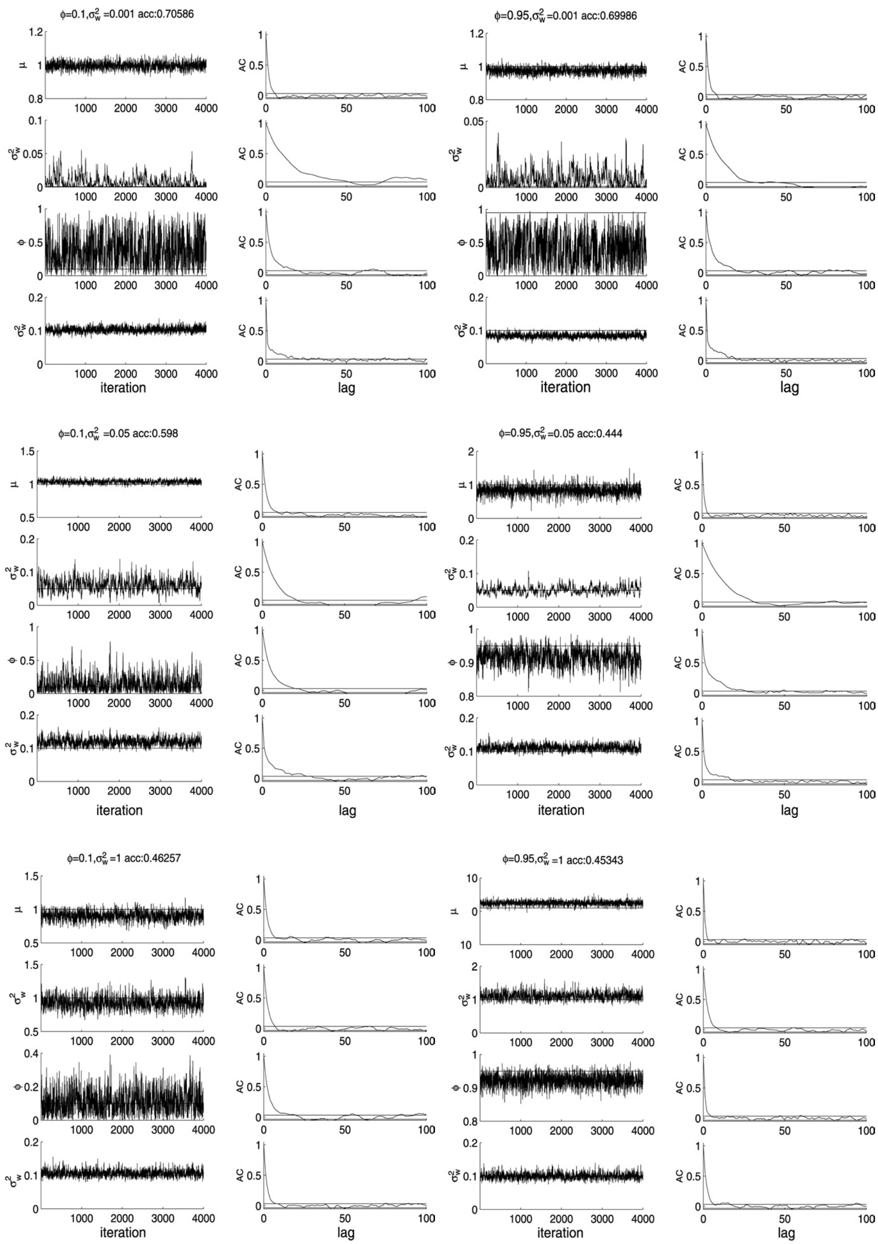

Fig. 7.7. Data simulated from model (7.1) and (7.2) with $\phi = 0.1$ (left hand side) and $\phi = 0.95$ (right hand side); $\sigma_w^2 = 0.001$ (top), $\sigma_w^2 = 0.05$ (middle) and $\sigma_w^2 = 1$ (bottom); MCMC draws with empirical autocorrelation obtained for μ, σ_w^2, ϕ and σ_ε^2 by choosing randomly the centred parameterisation or the parameterisation based on noncentred disturbances and keeping every second draw

that σ_w^2 is inverted Gamma, whereas for a parameterisation that is noncentred in the scale or based on noncentred disturbances we assume that σ_w is normal with prior mean equals 0 and prior variance proportional to σ_ε^2.

7.5.2 MCMC implementation

7.5.2.1 MCMC estimation for parameterisation centred in location and scale

MCMC estimation for a parameterisation that is centred in both location and scale is pretty straightforward and has been already outlined in the introduction. We discuss here more details of step (b). Conditional on $\tilde{X}$, (μ, ϕ, σ_w^2) is independent from σ_ε^2. However, rather than sampling (μ, ϕ, σ_w^2) directly, we sample the transformed parameter $(\gamma, \phi, \sigma_w^2)$ from the 'regression model' model (7.1), where $\gamma = (1 - \phi)\mu$. From (7.1) we obtain a marginal inverted Gamma proposal for σ_w^2, and a conditional normal proposal for $\gamma, \phi|\sigma_w^2$. To include the prior on β_0, which depends on $(\gamma, \phi, \sigma_w^2)$ in a nonconjugate way, we use a Metropolis–Hastings algorithm as in Chib and Greenberg (1994). Finally, from regression model (7.2), σ_ε^2 follows an inverted Gamma posterior.

7.5.2.2 MCMC estimation for parameterisations noncentred in location and centred in scale

For a model that is noncentred in the location, the following modification of step (b) is necessary. Conditional on $\tilde{X}$, the parameters ϕ and σ_w^2 appearing in the state equation (7.8) are independent of the parameters μ and σ_ε^2 appearing in the observation equation (7.9). We sample σ_w^2 and ϕ jointly from the autoregressive model (7.8) using an obvious modification of the Metropolis–Hastings algorithm discussed in Section 7.5.2.1. μ and σ_ε^2 are sampled jointly from the regression model (7.9).

7.5.2.3 MCMC estimation for parameterisations noncentred in location and scale

MCMC estimation for parameterisations that are noncentred in the scale is less straightforward. Van Dyk and Meng (2001) discussed MCMC estimation for random-effects model that are noncentred in the scale. Here we will present further details and discuss the extension to time series models. A straightforward MCMC sampler for parameterisation (7.29) and (7.30) which is noncentred in location and scale, but with ϕ remaining in the transition equation, is the following:

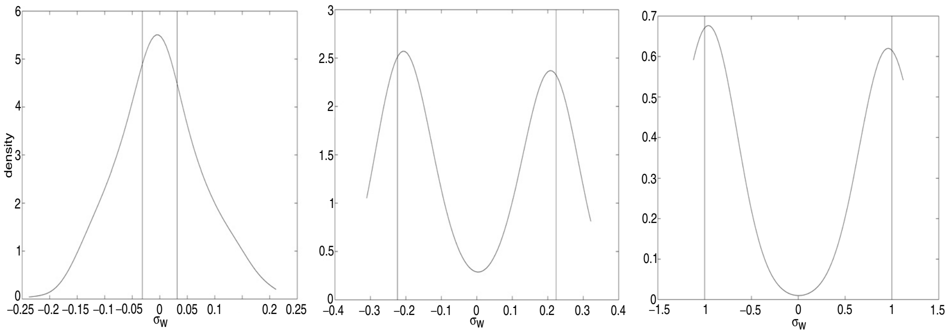

Fig. 7.8. Data simulated from model (7.1) and (7.2) with $\phi = 0.1$ and $\sigma_w^2 = 0.001$ (left), $\sigma_w^2 = 0.05$ (middle) and $\sigma_w^2 = 1$ (right); density of the marginal posterior distribution of σ_w (all model parameters considered unknown) under the parameterisation based noncentred disturbances; each MCMC draw has been concluded by a random sign switch; vertical lines indicate the true values

(a) Sample a path of $\tilde{X} = (\beta_0^\star, \dots, \beta_N^\star)$ conditional on known parameters using any multimove sampler.
(b) Conditional on $\tilde{X}$, the parameters $(\mu, \sigma_w, \sigma_\varepsilon^2)$ are independent of ϕ. From model (7.29) obtain a normal proposal for ϕ that can be used within a Metropolis–Hastings algorithm to sample ϕ from the conditional posterior $\pi(\phi|\tilde{X}, y)$ which included the prior on $\beta_0^\star$. Finally, sample the parameters $(\mu, \sigma_w, \sigma_\varepsilon^2)$ jointly conditional on $\tilde{X}$ from 'regression model' (7.30).

A (small) variance σ_w now appears in (7.30) as a regression coefficient close to 0. For Bayesian estimation we do not constrain σ_w to be positive, making the whole model unidentified:

$$\psi = (\beta_1^\star, \dots, \beta_N^\star, \mu, \sigma_w, \phi, \sigma_\varepsilon^2)$$

will result in the same likelihood function $L(y|\psi)$ as $\tilde{\psi}$, defined by:

$$\tilde{\psi} = (-\beta_1^\star, \dots, -\beta_N^\star, \mu, -\sigma_w, \phi, \sigma_\varepsilon^2).$$

The advantage of this Bayesian unidentifiability is that by allowing σ_w to move freely in the parameter space, we avoid the boundary space problem for small variances. This seems to be the main reason why mixing improves a lot for models with small variances.

The symmetry of the likelihood around $\sigma_w = 0$ causes bimodality of the marginal posterior density $\pi(\sigma_w|y)$ for models where the true value of σ_w is actually different from 0. If the true value of σ_w is close to 0, a unimodal posterior which concentrates around 0, results. Thus the posterior distribution of σ_w may be used to explore how far, for the data at hand, the model is bounded away from a model where $\sigma_w^2 = 0$. Figure 7.8 shows the posterior distributions of σ_w for those data in Section 7.2.2.3 that were simulated from a model with $\phi = 0.1$ under the three different settings of heterogeneity. For MCMC estimation all parameters including ϕ are considered unknown. For the time series simulated under $\sigma_w^2 = 0.001$, the model is not bounded away from 0 and the mode of the posterior lies practically at 0. For the two other time series, however, the bimodality of the posterior is well pronounced and the posterior is bounded away from 0.

As expected theoretically, the posterior densities in Figure 7.8 are symmetric around 0. If the posterior density is estimated from MCMC sampling as described above, it need not be symmetric. The sampler may stay at one mode or may show occasional sign switching to the other mode. This makes it difficult to compare MCMC draws from different runs, to assess convergence and to interpret the posterior density. In order to explore the entire posterior distribution, we performed a random sign switching after each MCMC draw in order to produce Figure 7.8.

7.5.2.4 MCMC estimation for parameterisations based on the noncentred disturbances

Finally, we discuss MCMC estimation for the parameterisation that is based on the noncentred disturbances.

(a) Sample a path of $\tilde{X} = (\beta_0^\star, w_1, \ldots, w_N)$ conditional on known parameters using the disturbance simulation smoother of de Jong and Shephard (1995) and Durbin and Koopman (2002).
(b) Sample the parameters $(\mu, \sigma_w, \phi, \sigma_\varepsilon^2)$ jointly conditional on $\tilde{X}$.

Note that under this parameterisation ϕ is no longer independent of $(\mu, \sigma_w, \sigma_\varepsilon^2)$ conditional on $\tilde{X}$ as in the previous subsection. Hence this setup differs from that in the previous subsection. For joint sampling of $(\mu, \sigma_w, \phi, \sigma_\varepsilon^2)$ we use a Metropolis–Hastings algorithm. First, we use a proposal for ϕ that is either independent of all parameters or depends only on the old value of ϕ. For nearly oversized models ($\sigma_w^2 = 0.001$ in our examples) we prefer to sample from the prior on ϕ, whereas for all other models we use a random walk proposal. Conditional on ϕ, we propose $(\mu, \sigma_w, \sigma_\varepsilon^2)$ from 'regression model' (7.30) as in the previous subsection.

Again, σ_w acts as a regression coefficient in (7.30). As in the previous subsection, we do not constrain σ_w to be positive, leaving the whole model unidentified. To produce posterior densities of σ_w, we again conclude each MCMC draw with a random sign switch.

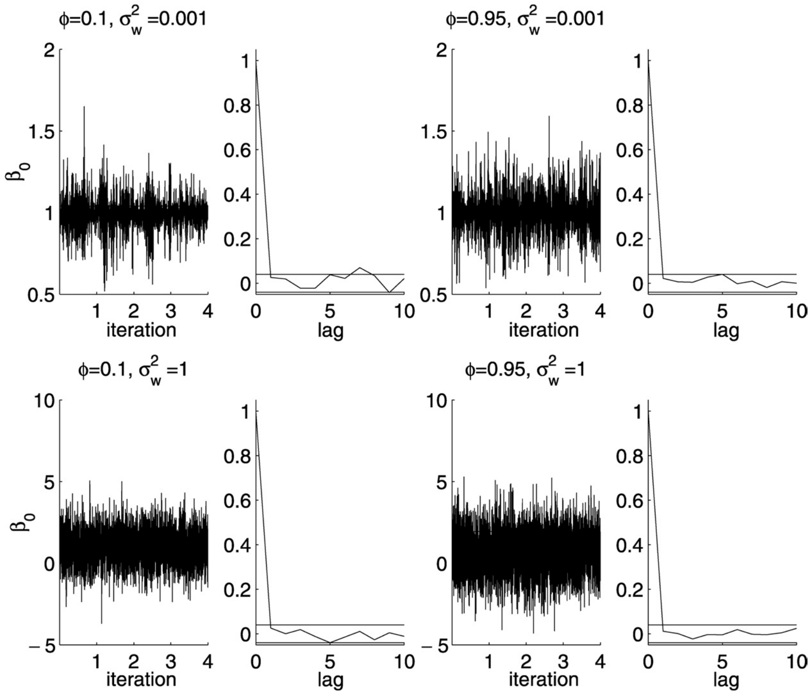

Fig. 7.9. Data simulated from model (7.1) and (7.2) with $\phi = 0.1$ (left hand side) and $\phi = 0.95$ (right hand side); $\sigma_w^2 = 0.001$ (top) and $\sigma_w^2 = 1$ (bottom); 4000 MCMC draws of β_0 with empirical autocorrelation obtained for the preferred parameterisation.

7.5.3 Sampling the unobserved initial state

All MCMC schemes discussed so far consider the unobserved initial parameter β_0 (or appropriate transformations of β_0) as part of the missing data $\tilde{X}$. Alternatively, we may view β_0 as part of the model parameter θ, and sample $\tilde{X}$ conditional on β_0. We found, however, that sampling β_0 (or appropriate transformations of β_0) as part of the missing data $\tilde{X}$ is the most efficient strategy over the whole range of parameters considered in this paper, see also Figure 7.9.

7.5.4 Blocking or transforming model parameters

Step (b) of the MCMC sampler discussed in Sections 7.5.2.1–7.5.2.4 is based on sampling the parameters $(\mu, \sigma_w^2, \phi, \sigma_\varepsilon^2)$ (or transformations of these parameters) jointly within one block rather than separately in two or more blocks. The reason for this is that we might have to deal with considerable or even extreme posterior correlation among the parameters, see for instance Figure 7.10 and Figure 7.11. In this case blocking the parameters is able to prevent slow mixing due to high correlation between parameters in different

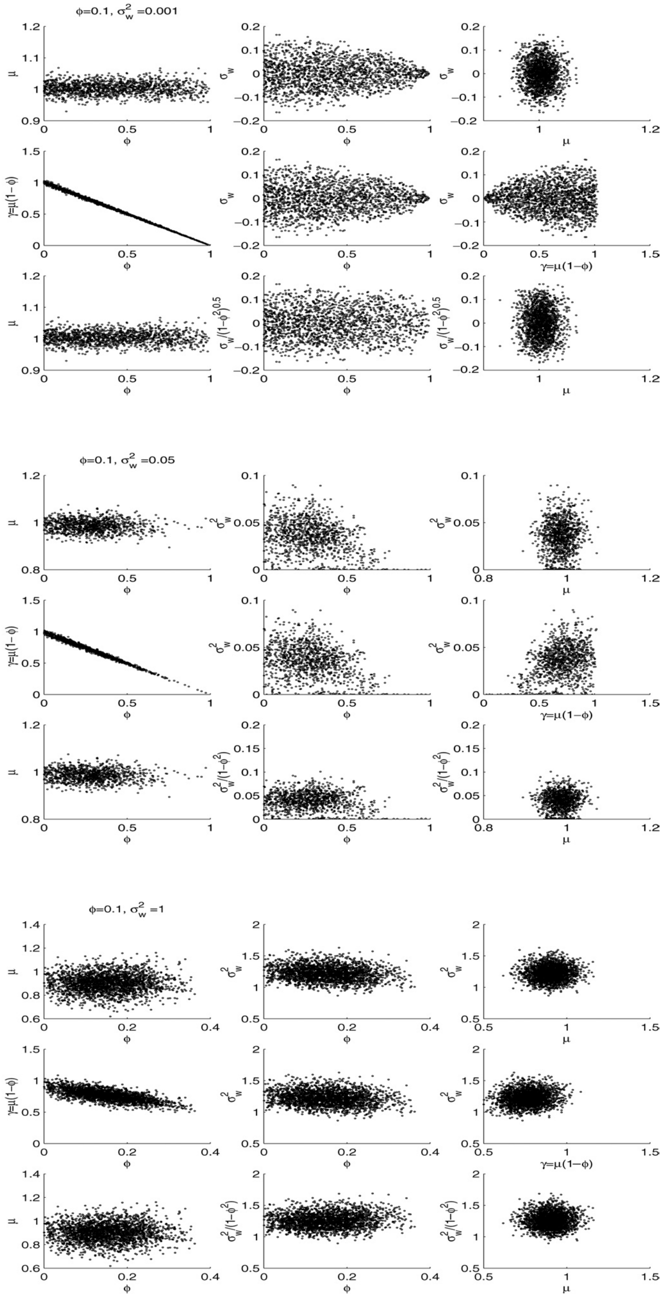

Fig. 7.10. Data simulated from model (7.1) and (7.2) with $\phi = 0.1$ and $\sigma_w^2 = 0.001$ (top), $\sigma_w^2 = 0.05$ (middle) and $\sigma_w^2 = 1$ (bottom); scatter plots of the MCMC draws under different definitions of model parameters

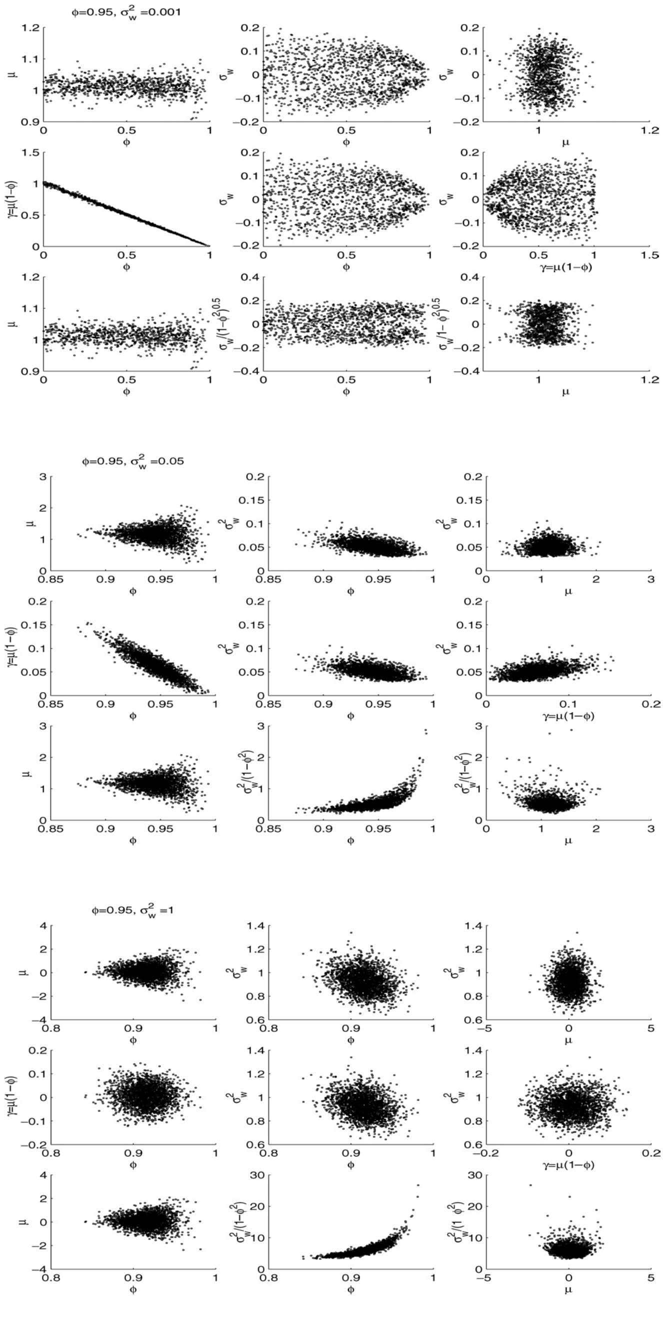

Fig. 7.11. Data simulated from model (7.1) and (7.2) with $\phi = 0.95$ and $\sigma_w^2 = 0.001$ (top), $\sigma_w^2 = 0.05$ (middle) and $\sigma_w^2 = 1$ (bottom); scatter plots of the MCMC draws under different definitions of model parameters

blocks. Under normal disturbances as in (7.1) and (7.2), tailor-made proposals are available that capture potentially strong posterior correlation as long as the model parameters are sampled jointly. Under nonnormal disturbances, however, it is much harder to find a suitable proposal density. It is therefore worthwhile to consider transformations of the parameters of the hidden model that prevent strong posterior correlations.

Starting from the 'natural' model parameter $\theta = (\mu, \sigma_w^2, \phi)$, we considered a transformed model parameter, called θ^C, that is based on the conditional moments of the shocks, $\theta^C = (\gamma, \sigma_w^2, \phi)$, where $\gamma = \mu(1-\phi)$, as well as another transformed model parameter, called θ^U, that is based on the unconditional moments of the state process, $\theta^U = (\mu, \sigma_\beta^2, \phi)$, where $\sigma_\beta^2 = \sigma_w^2/(1-\phi^2)$. Figures 7.10 and 7.11 show that transforming the model parameters may have a considerable influence on the shape of the posterior distribution.

For oversized models where the noncentred parameterisation is valid near independence is achieved by selecting the unconditional model parameter θ^U. Under the natural model parameter θ, the parameters ϕ and σ_w^2 suffer from posterior correlation, whereas under the conditional model parameter θ^C, the parameters ϕ and γ are highly correlated. This suggests combining the noncentred parameterisation with the unconditional model parameter θ^U.

For highly persistent time series, where the centred parameterisation is valid, posterior densities with elliptical contours are achieved by selecting the conditional model parameter θ^C. Single-move sampling should be avoided for ϕ and γ; however, the larger σ_w^2 the smaller the posterior correlation. Under the natural parameter θ, the posterior of μ strongly depends on how close ϕ is to the unit root, whereas under the unconditional model parameter θ^U the joint marginal posterior of ϕ and σ_β^2 is banana-shaped. This suggests combining the centred parameterisation with the conditional model parameter θ^C.

7.6 Concluding remarks

Although the discussion here has been confined entirely to the univariate time varying parameter model (7.1) and (7.2), the results seem to be of interest also for more general state space models.

Many other state space models, such as the dynamic factor model or stochastic volatility models, are based on hidden stationary processes and are formulated in such a way that the model parameters governing the hidden process appear only in the transition equation. It appears natural to call such a parameterisation centred also for more general state space model. MCMC estimation under the centred parameterisation considers the hidden process as missing data and uses a sampler that draws the missing data conditional on the model parameters and the model parameters conditional on the missing data.

Sampling a parameter that appears only in the transition equation of the state space model will not cause any problems as long as the randomness of the hidden process is actually of relevance for explaining the variance of the observed process. If the assumed randomness of the hidden process is of no relevance for explaining the variance of the observed process, for instance if the variance of the process is very close to 0, then selecting this process as missing data leads to a sampler that is poorly mixing for the parameters in the transition equation.

A very useful technique is then noncentring of the model in those parameters for which MCMC sampling based on the centred parameterisation is inefficient. Noncentring a model in a certain parameter means choosing the missing data $\tilde{X}$ in such a way that this model parameter no longer appears in the transition equation or equivalently in the prior distribution of the missing data, but is moved to the likelihood of the data given $\tilde{X}$. We have demonstrated in this paper that this may lead to considerable improvement of the resulting MCMC sampler in all cases where the centred parameterisation is performing in a poor way. Although noncentred parameterisations are rather useful, especially for nearly oversized models, the following aspects should be kept in mind.

First, model parameters that are not correlated in cases where the centred parameterisation is valid, may be highly correlated under the noncentred parameterisation especially if the model is oversized and one or more parameters are nearly unidentified. It is therefore often insufficient to substitute the Gibbs sampler that causes problems for a certain parameter under the centred parameterisation simply by a single-move Metropolis–Hastings algorithm under the noncentred parameterisation.

Second, if the randomness of the true underlying hidden process is of considerable relevance for explaining the variance of the observed process, the noncentred parameterisation will be rather poor and does not converge geometrically for the limiting case where all variance of the observed process stems from the hidden process. Furthermore, the range of variance parameters over which the noncentred parameterisation is better than the centred one depends very much on the autocorrelation of the hidden process. For a highly persistent hidden process which is close to a unit root, this range is restricted to processes with extremely small variance. The more one moves away from the unit root process, the bigger is the range of variance parameters for which the noncentred parameterisation is valid. Thus, in a setting with multivariate hidden factors, it might be necessary to use a hybrid parameterisation where a centred parameterisation for a highly persistent and/or highly random factor is combined with a noncentred parameterisation for a nearly constant factor with small variance.

Third, it is still unclear how to extend the results of this paper to state space models with nonnormal disturbances. Papaspiliopoulos, Roberts and Sköld (2003) showed for a random-effects model with all model parameters but μ being known that the centred parameterisation does not converge geometrically if the normal disturbances in the observation equation (7.4) are substituted by disturbances from a Cauchy distribution. On the other hand, the parameterisation which is noncentred in the location does not converge geometrically if this time the normal disturbances in the transition equation (7.4) are substituted by disturbances from a Cauchy distribution. This result would favor the centred parameterisation whenever the tails of the transition equation are fat compared with the tails of the observation distribution.

Finally, for nonstationary state space models such as the local level model or the basic structural model, the conditions (if any) under which the centred parameterisation fails and the way in which a noncentred parameterisation may be obtained still need to be explored.