Submitted to the Annals of Statistics

INFERENCE FOR STOCHASTIC VOLATILITY MODELS USING TIME CHANGE TRANSFORMATIONS

By Konstantinos Kalogeropoulos, Gareth O. Roberts and Petros Dellaportas

London School of Economics and Political Sciences, University of Warwick and Athens University of Economics and Business

We address the problem of parameter estimation for diffusion driven stochastic volatility models through Markov chain Monte Carlo (MCMC). To avoid degeneracy issues we introduce an innovative reparametrisation defined through transformations that operate on the time scale of the diffusion. A novel MCMC scheme which overcomes the inherent difficulties of time change transformations is also presented. The algorithm is fast to implement and applies to models with stochastic volatility. The methodology is tested through simulation based experiments and illustrated on data consisting of US treasury bill rates.

1. Introduction. Diffusion processes provide natural models for continuous time phenomena. They are used extensively in diverse areas such as finance, biology and physics. A diffusion process is defined through a stochastic differential equation (SDE)

$$dX_t = \mu(t, X_t, \theta)dt + \sigma(t, X_t, \theta)dW_t,\ 0 \leq t \leq T, \tag{1.1}$$

where W is standard Brownian motion. The drift $\mu(.)$ and volatility $\sigma(.)$ reflect the instantaneous mean and standard deviation respectively. In this paper we assume the existence of a unique weak solution to (1.1), which translates into some regularity conditions (locally Lipschitz with a linear growth bound) on $\mu(.)$ and $\sigma(.)$; see chapter 5 of [31] for more details.

The task of inference for diffusion processes is particularly challenging and has received remarkable attention in the recent literature; see [32] for an extensive review. The main difficulty is inherent in the nature of diffusions which are infinite dimensional objects. However, only a finite number of points may be observed and the marginal likelihood of these observations is generally unavailable in closed form. This has stimulated the development of various non-likelihood approaches which use indirect inference [18], estimating functions [6], or the efficient method of moments [14]; see also [13].

Most likelihood based methods approach the likelihood function through the transition density of (1.1). Denote the observations by Y_k, $k = 0, \ldots, n$, and with t_k their corresponding times. If the dimension of Y_k equals that of X (for each k) we can use the Markov property to write the likelihood, given the initial point Y_0, as:

$$\mathcal{L}(Y, \theta | Y_0) = \prod_{k=1}^{n} p_k(Y_k | Y_{k-1}; \theta, \Delta), \quad \Delta = t_k - t_{k-1} \tag{1.2}$$

The transition densities $p_k(.)$ are not available in closed form but several approximations are available. They may be analytical, see [1], [2], or simulation based, see [28], [9]. They usually

AMS 2000 subject classifications: Primary, 65C60, 60J60, 91B84; secondary, 62M99, 62F15

Keywords and phrases: Imputation, Markov Chain Monte Carlo, Diffusion Process

approximate the likelihood in a way so that the discretisation error can become arbitrarily small, although the methodology developed in [4] succeeds exact inference in the sense that it allows only for Monte Carlo error. A potential downside of these methods may be their dependence on the Markov property. In many interesting multidimensional diffusion models the observation regime is different and some of the diffusion components are not observed at all.

A famous such example is provided by stochastic volatility models, used extensively to model financial time series such as equity prices [19, 20, 33], or interest rates [3, 8, 15]. A stochastic volatility model is usually represented by a 2-dimensional diffusion

$$\begin{pmatrix} dX_t \\ d\alpha_t \end{pmatrix} = \begin{pmatrix} \mu_x(X_t, \alpha_t, \theta) \\ \mu_\alpha(\alpha_t, \theta) \end{pmatrix} dt + \begin{pmatrix} \sigma_x(\alpha_t, \theta) & 0 \\ 0 & \sigma_\alpha(\alpha_t, \theta) \end{pmatrix} \begin{pmatrix} dB_t \\ dW_t \end{pmatrix}, \tag{1.3}$$

where X denotes the observed equity (stock) log-price or the short term interest rate with volatility $\sigma_x(.)$, which is a function of a latent diffusion α. Note that the observed process X in (1.3) is not Markov; the distribution of a future stock price depends (besides the current price) on the current volatility which in turn depends on the entire price history. Nevertheless, since the 2-dimensional diffusion is Markov, state space approaches are possible and can be potentially be implemented using sequential Monte Carlo techniques [29]. While such approaches allow for online estimation of the volatility, various implications arise regarding inference for the parameters [25], [34].

An alternative approach to the problem adopts Bayesian inference utilizing Markov chain Monte Carlo (MCMC) methods. Adhering to the Bayesian framework, a prior $p(\theta)$ is first assigned on the parameter vector θ. Then, given the observations Y, the posterior $p(\theta|Y)$ can be explored through data augmentation [35], treating the unobserved paths of X (paths between observations) as missing data. Note that the augmented diffusion satisfies the Markov property irrespectively of the observation regime. Hence data augmentation approaches are more general.

Initial MCMC schemes following this programme were introduced by [21]; see also [22], [11] and [10]. However, as noted in the simulation based experiment of [10] and established theoretically by [30], any such algorithm's convergence properties will degenerate as the number of imputed points increases. The problem may be overcome with the reparametrisation of [30], and this scheme may be applied in all one-dimensional and some multi-dimensional contexts. However this framework does not cover general multidimensional diffusion models. [7] and [23] offer appropriate reparametrisations but only for a class of stochastic volatility models. Alternative reparametrisations were introduced in [17]; see also [16] for a sequential approach.

In this paper we introduce a novel reparametrisation that, unlike previous MCMC approaches, operates on the time scale of the observed diffusion rather than its path. This facilitates the construction of irreducible and efficient MCMC schemes, designed appropriately to accommodate the time change of the diffusion path. Being a data augmentation procedure our approach does not rely on the Markov property and can be applied to a much larger class of diffusions than those considered in [4] and [1]. Moreover it may be coupled with the existing approaches of [30] or [7] to handle more general models, that is almost every stochastic volatility model used in practice. The paper is organized as follows: Section 2 elaborates on the need for a transformation of the diffusion to avoid problematic MCMC algorithms. In Section 3 we introduce time change transformations whereas Section 4 provides the details for the corresponding non-trivial MCMC implementation. The proposed methodology of the paper is tested and illustrated through numerical experiments in section

5, and on US treasury bill rates in section 6. Finally, section 7 concludes and provides some relevant discussion.

2. The necessity of reparametrisation. A Bayesian data augmentation scheme bypasses a problematic sampling from the posterior $\pi(\theta|Y)$ by introducing a latent variable $\mathcal{X}$ that simplifies the likelihood $\mathcal{L}(Y;\mathcal{X},\theta)$. It usually involves the following two steps:

1. `Simulate` $\mathcal{X}$ `conditional on` Y `and` θ.
2. `Simulate` θ `from the augmented conditional posterior which is proportional to` $\mathcal{L}(Y;\mathcal{X},\theta)\pi(\theta)$.

It is not hard to adapt our problem to this setting. Y represents the observations of the price process X. The latent variables $\mathcal{X}$ introduced to simplify the likelihood evaluations are discrete skeletons of diffusion paths between observations or entirely unobserved diffusions. In other words, $\mathcal{X}$ is a fine partition of multidimensional diffusion with drift $\mu_X(t,X_t,\theta)$ and diffusion matrix

$$\Sigma_X(t,X_t,\theta) \;=\; \sigma(t,X_t,\theta) \;\times\; \sigma(t,X_t,\theta)',$$

and the augmented dataset is $\mathcal{X}_{i\delta},\ i=0,\dots,T/\delta$, where δ specifies the amount of augmentation. The likelihood can be approximated via the Euler scheme

$$\mathcal{L}^E(Y;\mathcal{X},\theta) = \prod_{i=1}^{T/\delta} p(\mathcal{X}_{i\delta}|\mathcal{X}_{(i-1)\delta}),\ \mathcal{X}_{i\delta}|\mathcal{X}_{(i-1)\delta} \sim \mathcal{N}\left(\mathcal{X}_{(i-1)\delta}+\delta\mu_{\mathcal{X}}(.),\delta\Sigma_{\mathcal{X}}(.)\right),$$

which is known to converge to the true likelihood $\mathcal{L}(Y;\mathcal{X},\theta)$ for small δ [28].

Another property of diffusion processes relates $\Sigma_{\mathcal{X}}(.)$ to the quadratic variation process. Specifically we know that

$$\lim_{\delta\to 0}\sum_{i=1}^{T/\delta}\left(\mathcal{X}_{i\delta}-\mathcal{X}_{(i-1)\delta}\right)\left(\mathcal{X}_{i\delta}-\mathcal{X}_{(i-1)\delta}\right)^T = \int_0^T \Sigma_{\mathcal{X}}(s,\mathcal{X}_s,\theta)ds \ \ a.s.$$

The solution of the equation above determines the diffusion matrix parameters. Hence, there exists perfect correlation between these parameters and $\mathcal{X}$ as $\delta\to 0$. This has disastrous implications for the mixing and convergence of the MCMC chain as it translates into reducibility for $\delta\to 0$. This issue was first noted by [30] for scalar diffusions and also confirmed by the simulation experiment of [10]. Nevertheless, it is not an MCMC specific problem. It turns out that the convergence of its deterministic analogue, EM algorithm, is problematic when the amount of information in the augmented data $\mathcal{X}$ strongly exceeds that of the observations. In our case $\mathcal{X}$ contains an infinite amount of information for $\delta\to 0$.

The problem may be resolved if we apply a transformation so that the algorithm based on the transformed diffusion is no longer reducible as $\delta\to 0$. [30] provide appropriate diffusion transformations for scalar diffusions. In a multivariate context this requires a transformation to a diffusion with unit volatility matrix; see for instance [24]. [2] terms such diffusions as reducible and proves the non-reducibility of stochastic volatility models that obey (1.3). The transformations introduced in this paper follow a slightly different route and target the time scale of the diffusion. One of the appealing features of such a reparametrisation is the generalisation to stochastic volatility models.

3. Time change transformations. For ease of illustration we first provide the time change transformation and the relevant likelihood function for scalar diffusion models with constant volatility. Nevertheless, one of the main advantages of this technique is the applicability to general stochastic volatility models.

3.1. *Scalar diffusions.* Consider a diffusion X defined through the following SDE:

$$dX_t = \mu(t, X_t, \theta)dt + \sigma dW_t^X, \quad 0 < t < 1 \;\; X_0 = y_0, \quad \sigma > 0. \tag{3.1}$$

Without loss of generality, we assume a pair of observations $X_0 = y_0$ and $X_1 = y_1$. For more data, note that the same operations are possible for every pair of successive observations that are linked together through the Markov property. We introduce the latent 'missing' path of X for $0 \le t \le 1$, denoted by X^{mis}, so that $X = (y_0, X^{mis}, y_1)$. In the spirit of [30], the goal is to write the likelihood for θ, σ with respect to a parameter-free dominating measure. Using Girsanov's theorem we can get the Radon-Nikodym derivative between the law of the diffusion X, denoted by $\mathbb{P}^X$, and that of the driftless diffusion $dM_t = \sigma dW_t^X$ which represents Wiener measure and is denoted by $\mathbb{W}^X$. We can write

$$\frac{d\mathbb{P}(X)}{d\mathbb{W}^X} = G(t, X, \theta, \sigma) = \exp\left\{ \int_0^1 \frac{\mu(t, X_t, \theta)}{\sigma^2} dX_t - \frac{1}{2}\int_0^1 \frac{\mu(t, X_t, \theta)^2}{\sigma^2} dt \right\}.$$

Consider the factorisation $\mathbb{W}^X = \mathbb{W}_y^X \times Leb(y_1) \times f(y_1|\sigma)$, where $y_1 \sim \mathcal{N}(y_0, \sigma^2)$ and $Leb(.)$ denotes Lebesgue measure. This naturally factorises the measure of X as the Lebesgue density of y_1 under the dominating measure, multiplied by the conditional dominating measure $\mathbb{W}_y^X$. We can now write

$$\frac{d\mathbb{P}(X^{mis}, y_0, y_1)}{d\left\{\mathbb{W}_y^X \times Leb(y)\right\}} = G(t, X, \theta, \sigma) f(y_1|\sigma),$$

where clearly the dominating measure depends on σ, since it reflects a Brownian bridge with volatility σ.

Now consider the time change transformation which introduces a new time scale $s = \eta(t, \sigma)$

$$s = \eta(t, \sigma) = \int_0^t \sigma^2 ds = t\sigma^2, \tag{3.2}$$

and then defines the new transformed diffusion U as

$$U_s = \begin{cases} X_{\eta^{-1}(s,\sigma)}, & 0 \le s \le \sigma^2, \\ M_{\eta^{-1}(s,\sigma)}, & s > \sigma^2. \end{cases}$$

The definition for $s > \sigma^2$ is needed to ensure that U is well defined for different values of $\sigma^2 > 0$ which is essential in the context of a MCMC algorithm. Using standard time change properties, see for example [26], the SDE for U is

$$dU_s = \begin{cases} \frac{1}{\sigma^2}\mu(s, U_s, \theta)ds + dW_s^U & 0 \le s \le \sigma^2, \\ dW_s^U, & s > \sigma^2, \end{cases}$$

where W^U is another Brownian motion at the time scale of U. By using Girsanov's theorem again, the law of U, denoted by $\mathbb{P}$, is given through its Radon-Nikodym derivative with respect to the law $\mathbb{W}^U$ of the Brownian motion W^U at the $U-$time scale s

$$\begin{aligned} \frac{d\mathbb{P}}{d\mathbb{W}^U} = G(s, U, \theta, \sigma) &= \exp\left\{ \int_0^{+\infty} \frac{\mu(s, U_s, \theta)}{\sigma^2} dU_s - \frac{1}{2}\int_0^{+\infty} \frac{\mu(s, U_s, \theta)^2}{\sigma^4} ds \right\} \\ &= \exp\left\{ \int_0^{\sigma^2} \frac{\mu(s, U_s, \theta)}{\sigma^2} dU_s - \frac{1}{2}\int_0^{\sigma^2} \frac{\mu(s, U_s, \theta)^2}{\sigma^4} ds \right\}. \end{aligned} \tag{3.3}$$

If we condition the Wiener measure on y_1 at the time scale s, the likelihood can be written with respect to a Brownian bridge measure $\mathbb{W}_y^U$ as

$$d\mathbb{P}(U, y_0, y_1) = G(s, U, \theta, \sigma) f(y_1|\sigma) d\left\{\mathbb{W}_y^U \times Leb(y)\right\}.$$

However, this Brownian bridge is conditioned on the event $U_{\sigma^2} = y_1$ and therefore contains the parameter σ. For this reason we introduce a second transformation which applies to both the diffusion's time scale and its path. Define

$$U_s = (\sigma^2 - s) Z_{s/\{\sigma^2(\sigma^2-s)\}} + (1 - \frac{s}{\sigma^2}) y_0 + \frac{s}{\sigma^2} y_1, \;\; 0 \le s < \sigma^2. \tag{3.4}$$

Note that this transformation is a bijection. In the case $y_0 = y_1 = 0$, its inverse is given by

$$Z_v = \frac{1 + v\sigma^2}{\sigma^2} U_{v\sigma^4/(1+v\sigma^2)}, \;\; 0 \le v < \infty.$$

Applying Ito's formula and using time change properties we can also obtain the SDE of Z based on another driving Brownian motion W^Z operating at the $Z-$time denoted by v

$$dZ_v = \frac{\mu\left(t, \frac{\sigma^2}{1+v\sigma^2}\nu(Z_v, \sigma), \theta\right) + \nu(Z_v, \sigma)\sigma^2}{1 + v\sigma^2} dv + dW_v^Z, \;\; 0 \le v < \infty, \tag{3.5}$$

where $\nu(Z_v, \sigma) = U_s$. This operation essentially transforms to a diffusion that runs from 0 to infinity but in a way so that the unit volatility is preserved. By defining the likelihood in the v time scale through Girsanov and conditioning the dominating measure on y_1, we obtain $\mathbb{W}_y^Z$,

$$\frac{d\mathbb{P}(Z, y_0, y_1)}{d\left\{\mathbb{W}_y^Z \times Leb(y)\right\}} = G(v, Z, \theta, \sigma) f(y_1|\sigma). \tag{3.6}$$

The integrals in $G(v, Z, \theta, \sigma)$ run up to ∞, however the expression is finite being a bijection of the Radon-Nikodym derivative between $\mathbb{P}(U)$ and $\mathbb{W}^U$ given by (3.3). Using the following lemma, we prove that $\mathbb{W}_y^Z$ is the law of the standard Brownian motion and hence the likelihood is written with respect to a dominating measure that does not depend on any parameters.

LEMMA 3.1. *Let W be a standard Brownian motion in $[0, +\infty)$. Consider the process defined for $0 \le t \le T$*

$$B_t = (T - t) W_{t/\{T(T-t)\}} + (1 - \frac{t}{T}) y_0 + \frac{t}{T} y_1, \; 0 \le t < T$$

Then B is a Brownian bridge from y_0 at time 0 to y_1 at time T.

Proof: See [31, IV.40.1] for the case $y_0 = 0$, $T = 1$. The extension for general y_0 and T is trivial.

COROLLARY 3.1. *The process Z is standard Brownian motion under the dominating measure. In other words $\mathbb{W}_y^Z$ is standard Wiener measure.*

Proof: Note that $\mathbb{W}_y^U$ reflects a Brownian bridge from y_0 at time 0 to y_1 at time T and we obtained $\mathbb{W}_y^Z$ by using the transformation of Lemma 3.1. Since this transformation is a bijection, U is a Brownian bridge (under the dominating measure) if and only if Z is standard Brownian motion.

Note that $\mathbb{W}_y^Z$ is the probability law of the driftless version of the conditional diffusion Z, whereas the SDE in (3.5) corresponds to the unconditional version of Z itself. The conditional SDE of Z is generally not available but this does not create a problem. For the path updates we may use the fact that

$$\frac{d\mathbb{P}}{d\mathbb{W}_y^Z}(Z|y_0,y_1) = G(v,Z,\theta,\sigma)\frac{f(y_1|\sigma)}{f^P(y_1|\theta,\sigma)} \propto G(v,Z,\theta,\sigma), \tag{3.7}$$

where $\mathbb{P}_y$ is the law of the conditional version of Z and $f^P(.)$ is the density of y_1 under $\mathbb{P}$. Both $\mathbb{P}_y$ and $f^P(.)$ are generally unknown but $G(.)$ and $f(.)$, which appear in (3.6) and (3.7), are available.

3.2. *Stochastic volatility models.* Consider the general class of stochastic volatility models with SDE given by (1.3) for $0 \le t \le t_1$. Without loss of generality, we may assume a pair of observations,$X_0 = y_0$, $X_{t_1} = y_1$, due to the Markov property of the 2-dimensional diffusion (X, α). The likelihood can then be divided into two parts: The first contains the marginal likelihood of the diffusion α and the remaining part corresponds to the diffusion X conditioned on the path of α

$$\mathbb{P}_\theta(X,\alpha) = \mathbb{P}_\theta(\alpha)\mathbb{P}_\theta(X|\alpha).$$

Denote the marginal likelihood for α by $\mathcal{L}_\alpha(\alpha,\theta)$. To overcome reducibility issues arising from the paths of α one may use the reparametrisations of [7] or [23]. The relevant transformations of the latter are

$$\beta_t = h(\alpha_t,\theta),\ \frac{\partial h(\alpha_t,\theta)}{\partial \alpha_t} = \{\sigma_\alpha(\alpha_t,\theta)\}^{-1},$$
$$\gamma_t = \beta_t - \beta_0,\ \ \beta_t = \eta(\gamma_t),$$

and the marginal likelihood for the transformed latent diffusion γ becomes

$$\mathcal{L}_\gamma(\gamma,\theta) = \frac{d\mathbb{P}}{d\mathbb{W}}(\gamma) = G_\gamma\{\eta(\gamma),\theta\}, \tag{3.8}$$

where $\mathbb{W}$ is standard Wiener measure. By letting $\alpha_t = g_t^\gamma = h^{-1}(\eta(\gamma_t),\theta)$, the SDE of X conditional on γ becomes:

$$dX_t = \mu_x(X_t, g_t^\gamma,\theta)dt + \sigma_x(g_t^\gamma,\theta)dB_t,\ \ 0\le t\le t_1.$$

Given the paths of the diffusion α, the volatility function $\sigma_x(g_t^\gamma,\theta)$ may be viewed as a deterministic function of time. The situation is similar to that of the previous section. We can introduce a new time scale s

$$s = \eta(t,\gamma,\theta) = \int_0^t \sigma_x^2(g_t^\gamma,\theta)ds,$$

$$T = \eta(t_1, \gamma, \theta),$$

and define U with the new time scale as before (M is a Brownian motion on the $U-$time scale)

$$U_s = \begin{cases} X_{\eta^{-1}(s)}, & 0 \le s \le T, \\ M_{\eta^{-1}(s)}, & s > T. \end{cases} \tag{3.9}$$

The SDE for U now becomes

$$dU_s = \begin{cases} \frac{\mu_x\left(U_s, \gamma_{\eta^{-1}(s,\gamma,\theta)}, \theta\right)}{\sigma_x^2(\gamma_{\eta^{-1}(s,\gamma,\theta)}, \theta)} ds + dW_s^U, & 0 \le s \le T, \\ dW_s^U, & s > T, \end{cases}$$

We obtain the Radon Nikodym derivative between the distribution of U with respect to that of the Brownian motion W^U,

$$\frac{d\mathbb{P}}{d\mathbb{W}^U} = G(U, \gamma, \theta),$$

and introduce $\mathbb{W}_y^U$ as before. The density of y_1 under $\mathbb{W}^U$, denoted by $f(y_1, \gamma, \theta)$, is just

$$f(y_1|\gamma, \theta) \equiv N(y_0, T).$$

The dominating measure $\mathbb{W}_y^U$ reflects a Brownian motion conditioned to equal y_1 at a parameter depended time $T = \eta(t_1, \gamma, \theta)$. To remove this dependency we introduce a second time change

$$U_s = (T - s) Z_{s/\{T(T-s)\}} + (1 - \frac{s}{T}) y_0 + \frac{s}{T} y_1, \; 0 \le s < T. \tag{3.10}$$

or else (for $y_0 = y_1 = 0$)

$$Z_v = \frac{1 + vT}{T} U_{vT^2/(1+vT)}, \; 0 \le v < \infty.$$

Therefore, the SDE for Z is now given by

$$dZ_v = \frac{T}{1 + vT} \left\{ \frac{\mu\left(\frac{T}{1+vT}\nu(Z_v), \gamma_{k(v,\gamma,\theta)}, \theta\right)}{\sigma_x^2(\gamma_{k(v,\gamma,\theta)}, \theta)} + \nu(Z_v) \right\} dv + dW_v^Z, \;\; 0 \le v < \infty,$$

where $k(v, \gamma, \theta)$ is equal with the initial time scale of X, t, and $\nu(Z_v) = U_s$.
Conditional on γ, the likelihood can be written in a similar manner as in (3.6):

$$\frac{d\mathbb{P}}{d\left\{\mathbb{W}_y^Z \times Leb(y)\right\}}(Z|y_0, y_1, \gamma) = G(Z, \gamma, \theta) f(y_1|\gamma, \theta) \tag{3.11}$$

It is not hard to see that $\mathbb{W}_y^Z$ reflects a standard Wiener measure and therefore the dominating measure is independent of parameters. To obtain the full likelihood we need to multiply the two parts given by (3.8) and (3.11).

3.3. *Incorporating leverage effect.* In the previous section we made the assumption that the increments of X and γ are independent, in other words we assumed no leverage effect. This assumption can be relaxed in the following way: In the presence of a leverage effect ρ, the SDE of X conditional on γ can be written as

$$dX_t = \mu_x(X_t, g_t^\gamma, \theta)dt + \rho\sigma_x(g_t^\gamma, \theta)dW_t + \sqrt{1-\rho^2}\sigma_x(g_t^\gamma, \theta)dB_t, \;\; 0 \le t \le t_1,$$

where W is the driving Brownian motion of γ). Note that given γ, W can be regarded as a function of γ and its parameters θ. Therefore, the term $\rho\sigma_x(g_t^\gamma, \theta)dW_t$ can be viewed as a deterministic function of time, and it can be treated as part of the drift of X_t. However, this operation introduces additional problems as the assumptions ensuring a weakly unique solution to the SDE of X are violated. To avoid this issue we introduce the infinitesimal transformation

$$X_t = \mathcal{H}(H_t, \rho, \gamma, \theta) = H_t + \int_0^t \rho\sigma_x(g_s^\gamma, \theta)dW_s,$$

which leads us to the following SDE for H:

$$dH_t = \mu_x\left\{\mathcal{H}(X_t, \rho, \gamma, \theta), g_t^\gamma, \theta\right\}dt + \sqrt{1-\rho^2}\sigma_x(g_t^\gamma, \theta)dB_t, \;\; 0 \le t \le t_1.$$

We can now proceed as before, defining U and Z based on the SDE of H in a similar manner as in (3.9) and (3.10) respectively.

3.4. *State dependent volatility.* Consider the family of state dependent stochastic volatility models where conditional on γ, the SDE of X may be written as:

$$dX_t = \mu_x(X_t, g_t^\gamma, \theta)dt + \sigma_1(g_t^\gamma, \theta)\sigma_2(X_t, \theta)dB_t, \;\; 0 \le t \le t_1.$$

This class contains among others, the models of [3], [15], [8], [11]. In order to apply the time change transformations of section 3.2, we should first transform X to $\dot{X}_t$, through $\dot{X}_t = h(X_t, \theta)$, so that it takes the form of (1.3). Such a transformation, which may be viewed as the first transformation in [30], but only for X, should satisfy the following differential equation

$$\frac{\partial h(X_t, \theta)}{\partial X_t} = \frac{1}{\sigma_2(X_t, \theta)}.$$

The time change transformations for U and Z may then be defined on the basis of $\dot{X}$ that will now have volatility $\sigma_1(g_t^\gamma, \theta)$. Note that the observations $(\dot{y}_0, \dot{y}_1)$ may now be functions of the $\sigma_2(X_t, \theta)$ parameters. These parameters enter the reparametrised likelihood in two ways: first through the $f(y_1; \gamma, \theta)$ which now should include the relevant Jacobian term, and second through $\nu(Z_t)$ in the drift functio of Z which depends on the transformed endpoints.

3.5. *Multivariate stochastic volatility models.* We may use the techniques of section 3.3 to define time change transformations for multidimensional diffusions. Consider a $d-$dimensional version of the SDE in (3.1) where σ now is a 2×2 matrix ($[\sigma]_{ij} = \sigma_{ij}$). As noted in [24], the mapping between σ and the volatility matrix $\sigma\sigma^T$ should be a bijection in order to ensure identifiability of the σ parameters. A way to achieve this, is by allowing σ to be the lower triangular matrix that produces the Cholesky decomposition of $\sigma\sigma^T$. For $d = 2$, the SDE of such a diffusion is given by

$$dX_t^{\{1\}} = \mu(X_t^{\{1\}}, X_t^{\{2\}}, \theta)dt + \sigma_{11}dB_t,$$

$$dX_t^{\{2\}} = \mu(X_t^{\{1\}}, X_t^{\{2\}}, \theta)dt + \sigma_{21}dB_t + \sigma_{22}dW_t.$$

The time change transformations for $X^{\{1\}}$ will be exactly as in section 3.1. For $X^{\{2\}}$ note that given $X^{\{1\}}$ the term $\sigma_{21}dB_t$ is now a deterministic function of time and may be treated as part of the drift. Thus, we may proceed following the route of the section 3.3.

Similar transformations can be applied for diffusions that have, or may be transformed to have, more general volatility functions. For example we may assume two correlated price processes with correlation ρ_x:

$$[\sigma]_{11} = \sigma_x^{\{1\}}(g_t^\gamma, \theta),$$
$$[\sigma]_{21} = \rho_x\sigma_x^{\{2\}}(g_t^\gamma, \theta),$$
$$[\sigma]_{22} = \sqrt{1-\rho_x^2}\sigma_x^{\{2\}}(g_t^\gamma, \theta).$$

We may proceed in a similar manner for multivariate stochastic volatility models of general dimension d.

4. MCMC implementation. The construction of an appropriate data augmentation algorithm involves several issues. The time change transformations introduce three interesting features to the MCMC algorithm which we address separately: the presence of three time scales; the need to update diffusion paths that run from 0 to $+\infty$; and the fact that time scales depend on parameters. For ease of illustration we will assume the simple case of a univariate diffusion with constant volatility and a pair observations ($X_0 = y_0$ and $X_1 = y_1$) with $0 \leq t \leq 1$. For the case of stochastic volatility models the time change transformation need only be applied for the diffusion paths of Z. The updates of transformed diffusion γ_t that governs the volatility, may be carried out using overlapping blocks as in [23]. Details are given below.

4.1. *Three time scales.* We introduce m intermediate points of X at equidistant times between 0 and 1, to give $X = \{X_{i/(m+1)},\ i = 0, 1, \ldots, m+1\}$. In addition, we make the assumption that m is large enough for accurate likelihood approximations and any error induced by the time discretisation is negligible for the purposes of our analysis.

Given a value of the time scale parameter σ, we can get the $U-$time points by applying (3.2) to each one of the existing points X, so that

$$U_{\sigma^2 i/(m+1)} = X_{i/(m+1)},\ i = 0, 1, \ldots, m+1.$$

Note that it is only the times that change, the values of the diffusion remain intact. In a stochastic volatility model we would use the quantities

$$\int_{\frac{i}{m+1}}^{\frac{i+1}{m+1}} \sigma_x^2(.)ds$$

for each pair of consecutive imputed points.

The points of Z are multiplied by a time factor which corrects the deviations from unit volatility. The $Z-$time points may be obtained by

$$t_i^Z = \frac{\sigma^2 i/(m+1)}{\sigma^2(\sigma^2 - \sigma^2 i/(m+1))},\ i = 0, 1, \ldots, m.$$

Clearly this does not apply to the last point which occurs at time $+\infty$. The paths of X, or U, are thus more convenient for likelihood evaluations and maybe used instead exploiting the fact that the relevant transformations are bijections. However, the component of the relevant Gibbs sampling scheme is the diffusion Z.

4.2. *Updating the paths of Z.* The paths of Z may be updated using an independence sampler with the reference measure as a proposal. Here $\mathbb{W}^Z$ reflects a Brownian motion at the $Z-$time which is fixed given the current values of the time-scale parameters and the paths of γ_t in the case of stochastic volatility models. An appropriate algorithm is given by the following steps.

- `Step 1: Propose a Brownian motion on the` $Z-$`time, say` Z^*`. The value at the endpoint (time` $+\infty$`) is not needed.`
- `Step 2: Transform back to` X^*`, using (3.4).`
- `Step 3: Accept with probability:` $\min\left\{1, \frac{G(X^*,\theta,\sigma)}{G(X,\theta,\sigma)}\right\}$.

4.3. *Updating time scale parameters and volatility paths.* The updates of parameters that define the time scale, such as σ, are of particular interest. In almost all cases, their conditional posterior density is not available in closed form, and Metropolis steps are inevitable. However, the proposed values of these parameters will imply different $Z-$ time scales. In other words, for each potential proposed value for σ there exists a different set of $Z-$ points needed for accurate approximations of the likelihood and the Metropolis accept-reject probabilities. In theory, this would pose no issues had we been able to store an infinitely thin partition of Z, but of course this is not possible.

We use retrospective sampling ideas; see [27] and [5] for applications in different contexts. Under the assumption of a sufficiently fine partition of Z, all the non-recorded intermediate points contribute nothing to the likelihood and they are irrelevant in that respect; the set of recorded points is sufficient for likelihood approximation purposes. In other words, their distribution is given by the likelihood reference measure which reflects a Brownian motion. Thus, they can be drawn after the proposal of the candidate value of the time scale parameter. To ensure compatibility with the recorded partition of Z, it suffices to condition on their neighboring points. This is easily done using standard Brownian bridge properties: Suppose that we want to simulate the value of Z at time t_b which fall between the recorded values at times t_a and t_c, so that $t_a \leq t_b \leq t_c$. Denote by Z_{t_a} and Z_{t_c} the corresponding Z values. Under the assumption that Z is distributed according to $\mathbb{W}^Z_y$ between t_a and t_c we have that

$$Z_{t_b}|\, Z_{t_a}, Z_{t_c} \sim N\left\{\frac{(t_b-t_a)Z_{t_c}+(t_c-t_b)Z_{t_a}}{t_c-t_a},\ \ \frac{(t_b-t_a)(t_c-t_b)}{t_c-t_a}\right\}. \tag{4.1}$$

We describe the algorithm for the general case for more than two observations $X_0 = y_0$, $X_{t_1} = y_1$, ..., $X_{t_n} = y_n$. The time change transformations are applied for each pair of successive observations for which we get a separate Z process. However, as mentioned earlier, we may use X to evaluate the likelihood, since X is a function of the n separate Z diffusions. A suitable algorithm for the $\sigma-$updates may be summarized through the following steps:

- `Step 1: Propose a candidate value for` σ`, say` σ^*`.`
- `Step 2: Repeat for each pair of successive points:`
  - `Use (3.2) and (3.4) to get the new times associated with it.`
  - `Draw the values of` Z `at the new times using (4.1).`
  - `Transform back to` X^* `(that corresponds to the time between the pair of successive points), using (3.4).`

- Step 3: Form the entire path X^* by appropriately joining the bits between successive observations.
- Step 4: Accept with probability: $\min\left\{1, \frac{G(X^*,\theta,\sigma^*)f(y_1|\sigma^*)}{G(X,\theta,\sigma)f(y_1|\sigma)}\right\}$.

In stochastic volatility models the updates paths of the transformed diffusion γ_t may be implemented using overlapping blocks. Note that these paths are associated with the time scale of the process Z and therefore a similar algorithm as above should be embedded in their updates. For simplicity consider blocks of γ_t paths that correspond to times between non-successive observations. Each block is further split into intervals between successive observations, thus providing a separate Z diffusion for each interval whose time scale is affected by the relevant $gamma_t$ path. A suitable algorithm for updates of such blocks is presented below:

- Step 1: Propose γ_t^*, between the times of two non-successive observations y_A and y_B, by a Brownian bridge connecting the relevant endpoints of γ_t.
- Step 2: Repeat for each pair of successive observed points between (and including) y_A and y_B
  - Use (3.9) and (3.10) to get the new times associated with the proposed path.
  - Draw the values of Z at the new times using (4.1).
  - Transform back to the path of X^* (between the pair of successive observations), using (3.10).
- Step 3: Join the bits of X^* to form its path between y_A and y_B.
- Step 4: Accept with probability: $\min\left\{1, \frac{G_\gamma(\eta(\gamma^*),\theta)G(X^*,\theta,\sigma^*)f(y|\gamma^*)}{G_\gamma(\eta(\gamma),\theta)G(X,\theta,\sigma)f(y|\gamma)}\right\}$.

Note that in each block the endpoints of the transformed diffusion γ_t are not updated. The use of overlapping blocks is therefore crucial to preserve the irreducibility of the chain.

5. Simulations. As discussed in section 2, appropriate reparametrisations are necessary to avoid issues regarding the mixing and convergence of the MCMC algorithm. In fact, the chain becomes reducible as the level of augmentation increases. This is also verified by the numerical examples performed in [23] even in very simple stochastic volatility models. In this section we perform a simulation based experiment to check the immunity of MCMC schemes to increasing levels of augmentation, as well as the ability of our estimation procedure to retrieve the correct values of the diffusion parameters despite the fact that the series is partially observed at only a finite number of points. We simulate data from the following stochastic volatility model

$$
\begin{aligned}
dX_t &= \kappa_x(\mu_x - X_t)dt + \rho\exp(\alpha_t/2)dW_t + \sqrt{1-\rho^2}\exp(\alpha_t/2)dB_t,\\
d\alpha_t &= \kappa_\alpha(\mu_\alpha - \alpha_t)dt + \sigma dW_t,
\end{aligned}
$$

where B and W are independent Brownian motions, and ρ reflects the correlation between the increments of X and α, also term as leverage effect. A high frequency Euler approximating scheme with a step of 0.001 was used for the simulation of the diffusion paths. Specifically, 500,001 points were drawn and one value of X for every 1000 was recorded, thus forming a

TABLE 1

Summaries of the posterior draws for the simulation example of Section 5 for $m = 50$.

Parameter	True value	Post. mean	Post. SD	Post 2.5%	Post median	Post 97.5%
κ_x	0.2	0.244	0.038	0.173	0.243	0.321
μ_x	0.1	0.313	0.174	-0.046	0.317	0.641
κ_α	0.3	0.304	0.148	0.110	0.277	0.672
μ_α	-0.2	-0.268	0.107	-0.484	-0.267	-0.059
σ	0.4	0.406	0.130	0.202	0.390	0.705
ρ	-0.5	0.477	0.138	-0.657	-0.491	-0.066

dataset of 501 observations of X at $0 \leq t \leq 500$. The parameter values were set to $\rho = -0.5$, $\sigma = 0.4$, $\kappa_x = 0.2$, $\mu_x = 0.1$, $\kappa_\alpha = 0.3$ and $\mu_\alpha = -0.2$

The transformations required to construct an irreducible data augmentation scheme are listed below. First we transform α to γ through

$$\gamma_t = \frac{\alpha_t - \alpha_0}{\sigma}, \; 0 \leq t \leq 500,$$
$$\alpha_t = \nu(\gamma_t, \sigma, \alpha_0) = \alpha_0 + \sigma\gamma_t.$$

Given γ, and for each pair of consecutive observation times t_{k-1} and t_k ($k = 1, 2, \ldots, 500$) on X, we transform as follows: First, we remove the term introduced from the leverage effect

$$H_t = X_t - \int_{t_{k-1}}^{t} \rho \exp\{\nu(\gamma_s, \sigma, \alpha_0)/2\}\, dW_s, \; t_{k-1} \leq t \leq t_k,$$

and consequently we set

$$s = \eta(t) = \int_{t_{k-1}}^{t} (1-\rho)^2 \exp\{\nu(\gamma_t, \sigma, \alpha_0\}\, dt.$$

Then, U and Z may be defined again from 3.9 and 3.10 respectively, but based on H rather on X. The elements of the MCMC scheme are Z,γ, α_0 and the parameters $(\kappa_x, \mu_x, \kappa_\alpha, \mu_\alpha, \rho, \sigma)$.

Vague priors were assigned to all the parameters, restricting κ_x, κ_α, σ to be positive and ρ to be in $(-1, 1)$. The chain was run several times for 50,000 iterations on different levels of augmentation, by setting the number of imputed points was set to 2, 30, 40 and 50, to assess the approximation error. As in [23], it was noted that the choice of the length of the overlapping blocks, needed for the updates of γ, may improve substantially the mixing of the chain. We present results with blocks of length corresponding to 8 observations; the acceptance rate for each block was around 75%, whereas the acceptance rate for each path of Z was around 95%. The running time needed for such a chain with $m = 40$ was roughly 4 hours in a mid-specification PC. We also noted a linear relationship between the running times and m that confirms the fact that our algorithm is $O(m)$ (see discussion). Figure 1 shows autocorrelation plots for all the parameters. There is no sign of any increase in the autocorrelation to raise suspicions against the irreducibility of the chain. Figure 2 shows density plots for all parameters and on all values of m. In most cases the densities do not differ across augmentation levels, except for κ_x and μ_α. These plots may be used to monitor the deterioration of the discretisation error. In this example a choice of $m = 2$ may have been suboptimal, but a choice of m over 30 seems to perform well. Also the plots reveal good agreement with true values of the parameters, which is also supported by Table 1.

6. Application: US treasury bill rates. To illustrate the time change methodology we fit a stochastic volatility model to US treasury bill rates. The dataset consists of 1809

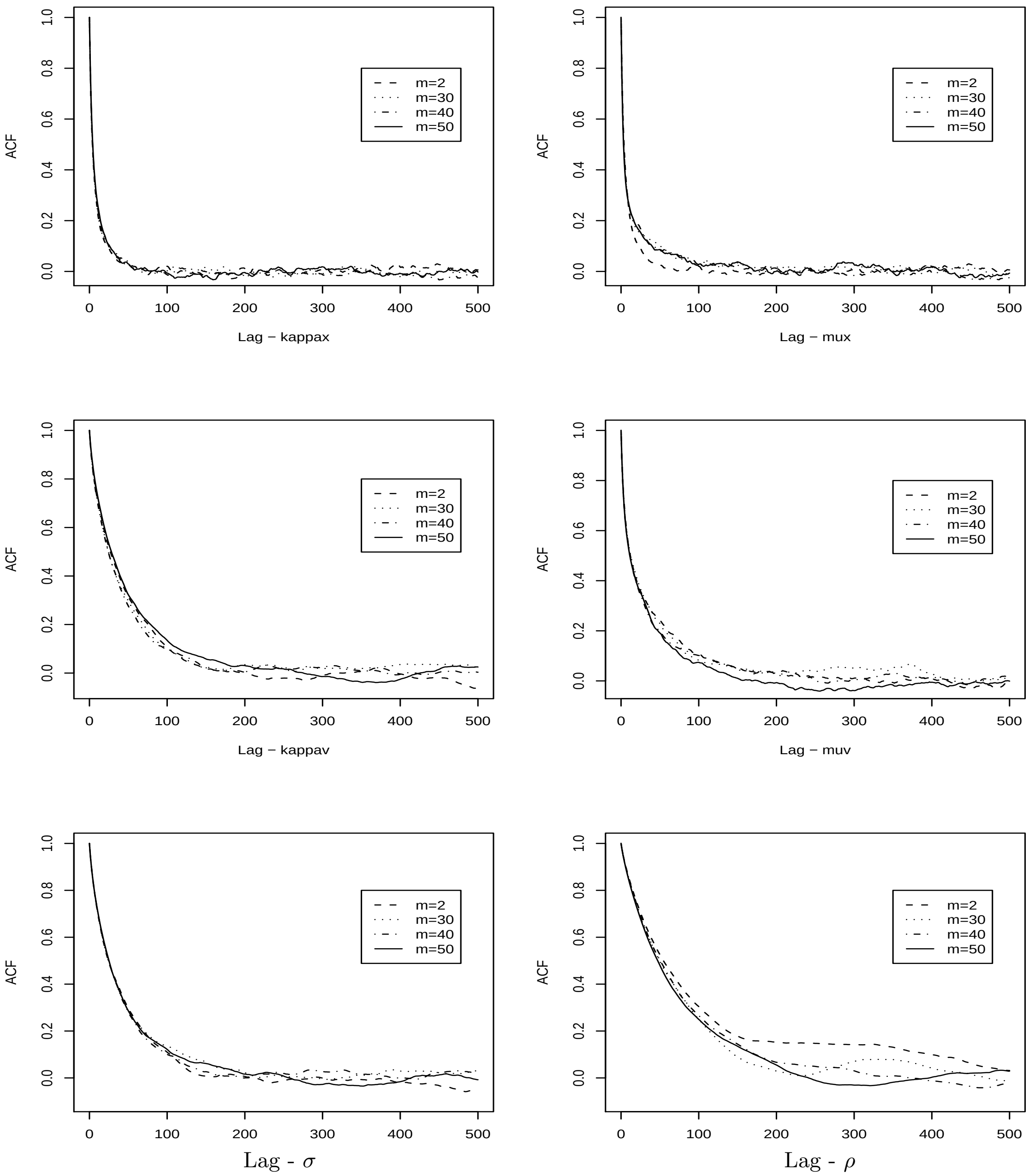

FIG 1. *Autocorrelation plots of the parameter posterior draws for different numbers of imputed points* $m = 10, 30, 40, 50$. *Simulation example of Section 5.*

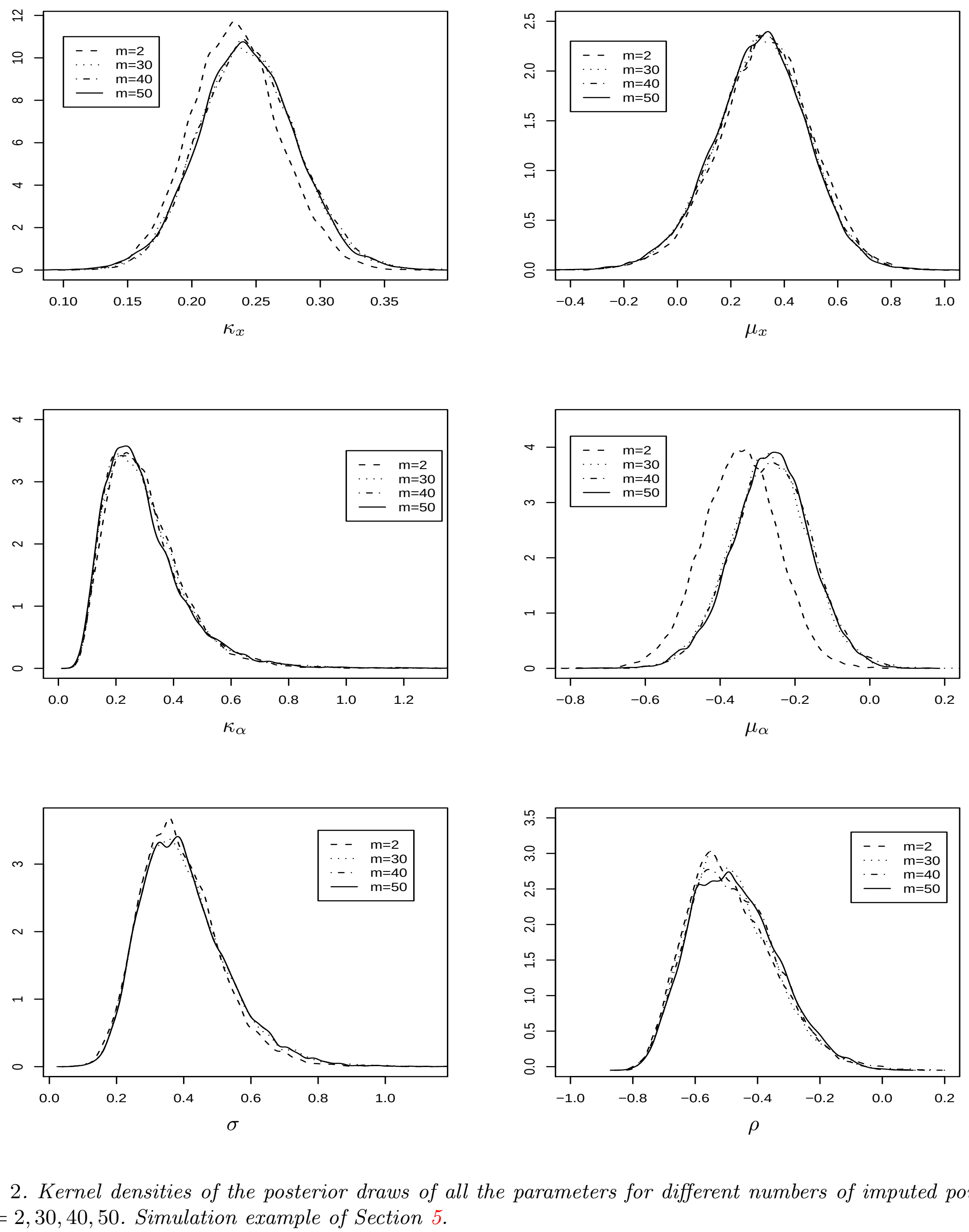

FIG 2. *Kernel densities of the posterior draws of all the parameters for different numbers of imputed points* $m = 2, 30, 40, 50$. *Simulation example of Section 5.*

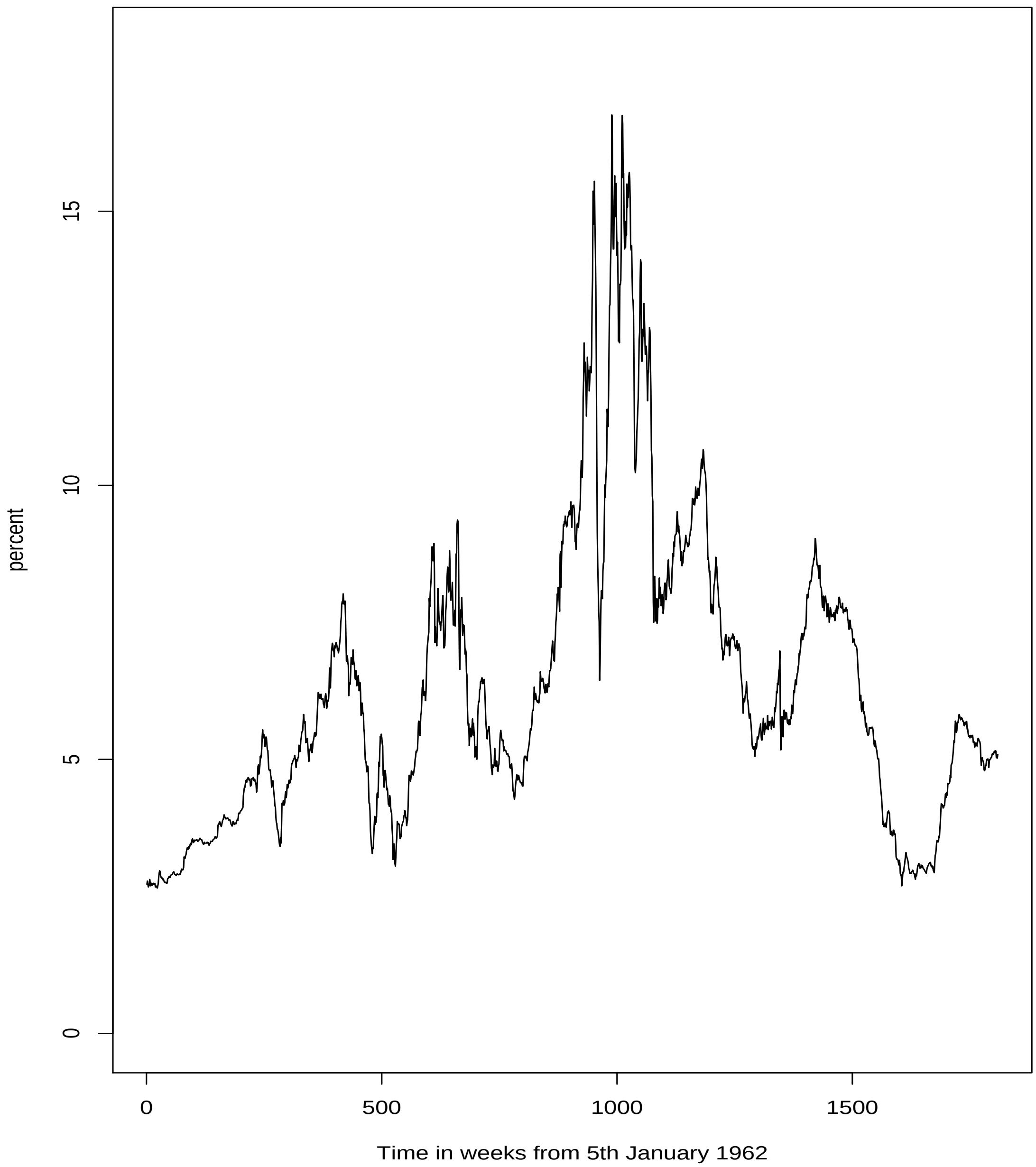

Fig 3. *Weekly* 3−*month US Treasury bill rate from the* 5*th of January* 1962 *up to the* 30*th of August* 1996.

weekly observations (Wednesday) of the 3−month US Treasury bill rate from the 5th of January 1962 up to the 30th of August 1996. The data are plotted in Figure 3.

Previous analyses of these data include [3], [15], [8], [9], [11], and [16]. Apart from some slight deviations the adopted stochastic volatility models consisted of the following SDE.

$$\begin{aligned} dr_t &= (\theta_0 - \theta_1 r_t)dt + r_t^{\psi} \exp(\alpha_t/2) dB_t, \\ d\alpha_t &= \kappa(\mu - \alpha_t)dt + \sigma dW_t, \end{aligned} \tag{6.1}$$

with independent Brownian motions B and W. In some cases the following equivalent model was used:

$$\begin{aligned} dr_t &= (\theta_0 - \theta_1 r_t)dt + \sigma_r r_t^{\psi} \exp(\alpha_t/2) dB_t, \\ d\alpha_t &= -\kappa \alpha_t dt + \sigma dW_t. \end{aligned} \tag{6.2}$$

The model in (6.1) was chosen, as posterior draws of its parameters exhibit substantially less autocorrelation. In line with [15] and [16], we also set $\psi = 1$ for model parsimony reasons. [11], [8] and [9] assume general 'elasticity of variance' ψ but their estimates do not indicate a significant deviation from 1. By setting $X_t = \log(r_t)$, the volatility of X_t becomes $\exp(\alpha_t/2)$. Therefore the $U-$time for two consecutive observation times t_{k-1} and t_k is defined as

$$s = \eta(t) = \int_{t_{k-1}}^{t} \exp(\alpha_t) dt,$$

and U and Z are given by (3.9) and (3.10) respectively . As before α should also be transformed to γ

$$\gamma_t = \frac{\alpha_t - \alpha_0}{\sigma},$$
$$\alpha_t = \nu(\gamma_t, \sigma, \alpha_0) = \alpha_0 + \sigma\gamma_t,$$

to avoid degeneracy issues. We implemented MCMC algorithms based on Z and γ to sample from the posterior of the parameters θ_0, θ_1, κ, μ and σ. The time was measured in years setting the distance between successive Wednesdays to 5/252. Vague priors were assigned to all the parameters, restricting κ and σ to be positive to ensure identifiability and eliminate the possibility of explosion. The algorithm was run for 50,000 iterations and for different values of m equal to 2, 10 and 20. To optimize the efficiency of the chain we set the length of the overlapping blocks of γ to 10 observations which produced an acceptance rate of 51.9% for each block. The corresponding acceptance rate for each path of Z was 98.6% .

The kernel density plots of the posterior parameters and likelihood (Figure 4) indicate that a time discretisation corresponding to an m of 10 or 20 provides reasonable approximations. A choice of $m = 2$ produces similar parameter posterior draws but the log-likelihood plot (bottom right) seems to be slightly off. The corresponding autocorrelation plots of Figure 5 do not provide evidence of increasing autocorrelation in m. Finally, summaries of the posterior draws for all the parameters are provided in Table 2. The parameters κ, μ and σ are different from 0 verifying the existence of stochastic volatility. On the other hand, there is no evidence to support the existence of mean reversion on the rate, as θ_0 and θ_1 are not far from 0. The results are in line with those of [8], [9] and [16].

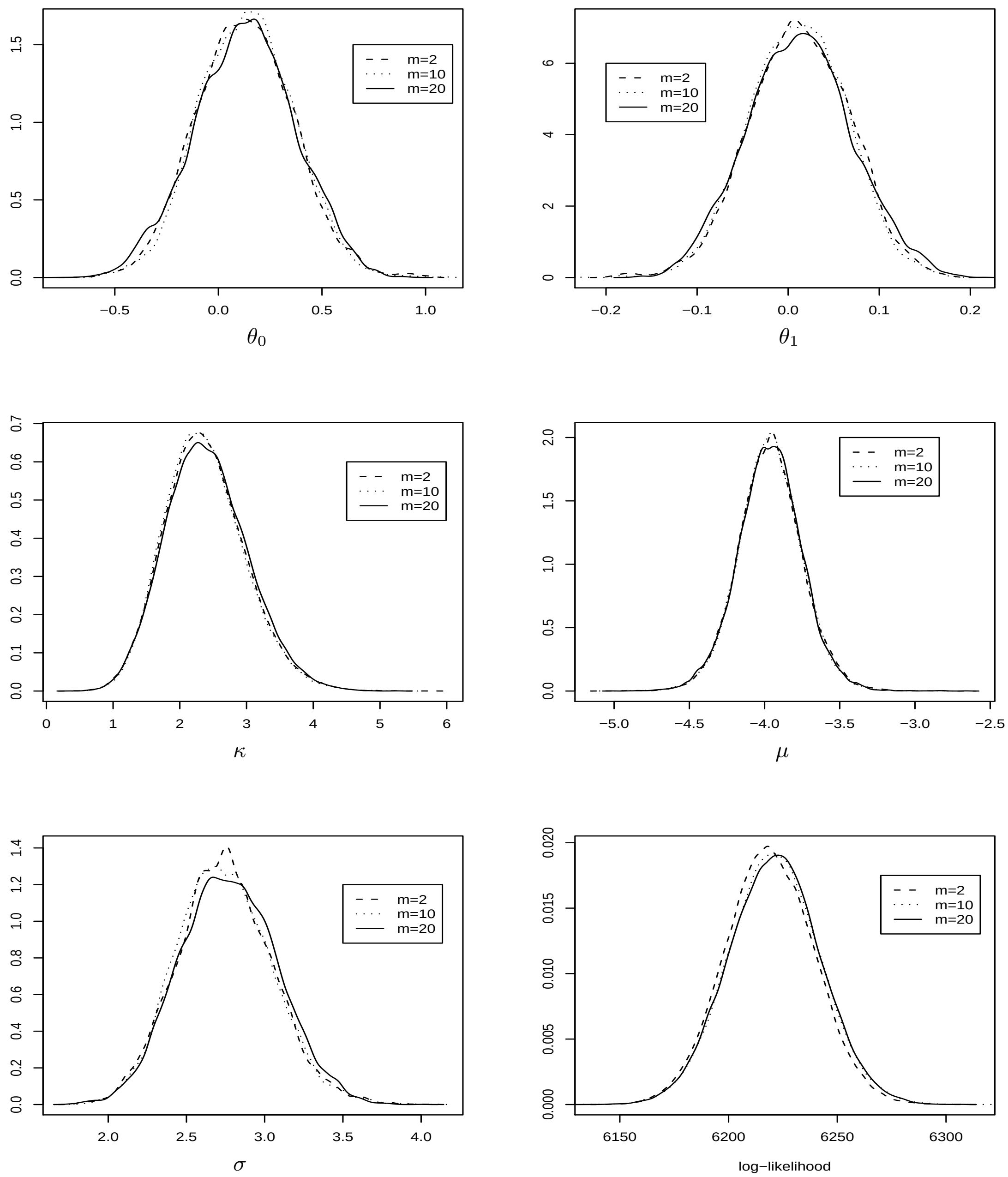

Fig 4. *Kernel densities of the posterior draws of all the parameters and the log-likelihood for different values of imputed points $m = 2, 10, 20$. Example on Weekly $3-$month US Treasury bill rates.*

Table 2

Summaries of the posterior draws for the stochastic volatility model of Weekly $3-$month US Treasury bill rates.

Parameter	Post. mean	Post. SD	Post 2.5%	Post median	Post 97.5%
θ_0	0.130	0.238	-0.347	0.132	0.589
θ_1	0.013	0.057	-0.096	0.013	0.125
κ	2.403	0.620	1.319	2.360	3.745
μ	-3.966	0.211	-4.384	-3.964	-3.547
σ	2.764	0.311	2.199	2.750	3.420

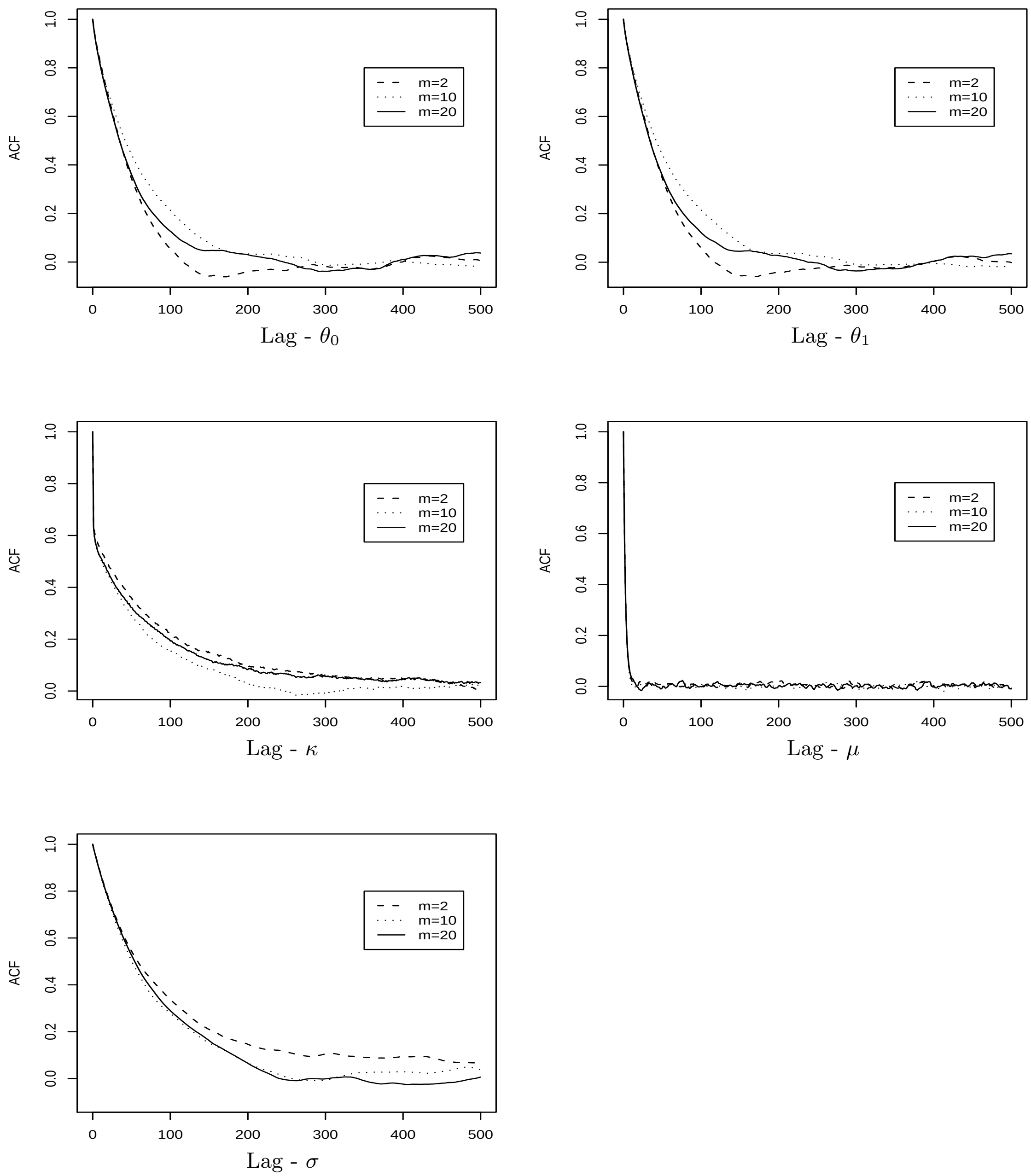

FIG 5. *Autocorrelation plots for the posterior draws of the model parameters for different numbers of imputed points $m = 2, 10, 20$ for the analysis of Weekly 3−month US Treasury bill rates.*

7. Discussion. Data augmentation MCMC schemes constitute a very useful tool for likelihood-based inference on diffusion models. They may not have the appealing properties of complete elimination of the time discretisation error [4], or the closed form approximate likelihood expressions of [1], but nevertheless they give a satisfactory and very general solution to the problem. However data augmentation schemes require careful construction to avoid the degeneracy issues described at the beginning of this paper.

Here, we introduce an innovative transformation which operates by altering the time axis of the diffusion. To accommodate the special features of time change transformations we also introduce a novel efficient MCMC scheme which mixes rapidly and is not prohibitively computationally expensive. Our method is also easy to implement and introduces no additional approximation error other than that included in methodologies based on a discretisation of the diffusion path. Moreover it has a broad range of applications which include general stochastic volatility models.

One clear advantage of the time change methodology is that in its pure form it produces algorithms whose mixing time is bounded as m goes to infinity, as in [30]. In addition the computing cost per iteration of our methods is $O(m)$ as with other competing methods. Thus the overall computing cost of our approach is $O(m)$ which compares favourably with competing methods are typically $O(m^2)$. In our experience mixing properties of the methods introduced in this paper are good in comparison with competing methods for these types of models and data. Furthermore we have found that implementation can routinely be carried out in a few hours on a mid-specification PC.

Further work will consider problems with state-dependent volatility and models which involve jump diffusions, to which the methodology introduced here can be easily applied. Fundamental to our approach here has been the introduction of a non-centered parametrisation to decouple dependence inherent in the model between missing data and volatility parameters. However non-centered constructions are not unique, as illustrated by the choice in the diffusion context between the state rescaling approaches of [17, 30] and the time-stretching strategy adopted here. Clearly, further work is required to investigate the relative merits of these approaches in different situations.

Acknowledgements. This work was supported by the EPSRC grant GR/S61577/01 and by the Marie Curie fellowships. The authors would also like to thank Omiros Papaspiliopoulos for helpful discussions.

References.

[1] Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed form approximation approach. *Econometrica*, 70:223–262.

[2] Aït-Sahalia, Y. (2008). Closed form likelihood expansions for multivariate diffusions. *Annals of Statistics*, 36:906–937.

[3] Andersen, T. G. and Lund, J. (1998). Estimating continuous-time stochastic volatility models of the short term interest rate. *Journal of Econometrics*, 77:343–377.

[4] Beskos, A., Papaspiliopoulos, O., Roberts, G., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382.

[5] Beskos, A. and Roberts, G. O. (2005). Exact simulation of diffusions. *Annals of Applied Probability*, 15(4):2422–2444.

[6] Bibby, B. and Sorensen, M. (1995). Martingale estimating functions for discretely observed diffusion processes. *Bernoulli*, 1:17–39.

[7] Chib, S., Pitt, M. K., and Shephard, N. (2006). Likelihood based inference for diffusion models. Submitted.

[8] Durham, G. B. (2003). Likelihood based specification analysis of continuous time models of the short term interest rate. *Journal of Financial Economics*, 70(3):463–487.

[9] Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.*, 20(3):297–316. With comments and a reply by the authors.

[10] Elerian, O. S., Chib, S., and Shephard, N. (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, 69:959–993.

[11] Eraker, B. (2001). Markov chain Monte Carlo analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics*, 19(2):177–191.

[12] Fearnhead, P., Papaspiliopoulos, O. and Roberts, Gareth O. (2008). Particle filter for partially observed diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):755–777.

[13] Gallant, A. R. and Long, J. R. (1997). Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika*, 84(1):125–141.

[14] Gallant, A. R. and Tauchen, G. (1996). Which moments to match? *Econometric Theory*, 12(4):657–681.

[15] Gallant, A. R. and Tauchen, G. (1998). Reprojecting partially observed systems with applications to interest rate diffusions. *Journal of American Statistical Association*, 93(441):10–24.

[16] Golightly, A. and Wilkinson, D. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Statistics and Computing*, 16:323–338.

[17] Golightly, A. and Wilkinson, D. (2007). Bayesian inference for nonlinear multivariate diffusions observed with error. *Computational Statistics and Data Analysis*, 52(3):1674–1693.

[18] Gouriéroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8:S85–S118.

[19] Heston, S. (1993). A closed-form solution for options with stochastic volatility. with applications to bonds and currency options. *Review of Financial Studies*, 6:327–343.

[20] Hull, J. C. and White, A. D. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42(2):281–300.

[21] Jones, C. S. (1999). Bayesian estimation of continuous-time finance models. Unpublished paper, Simon School of Business, University of Rochester.

[22] Jones, C. S. (2003). Nonlinear mean reversion in the short-term interest rate. *The review of financial studies*, 16:793–843.

[23] Kalogeropoulos, K. (2007). Likelihood based inference for a class of multidimensional diffusions with unobserved paths. *Journal of Statistical Planning and Inference*, 137:3092–3102.

[24] Kalogeropoulos, K., Dellaportas, P. and Roberts, G. (2007). Likelihood-based inference for correllated diffusions. Submitted.

[25] Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In Sequential Monte Carlo Methods in Practice, by A. Doucet, J. F. G. De Freitas and N. J. Gordon Springer-Verlag, New York.

[26] Oksendal, B. (2000). *Stochastic differential equations*. Springer, 5th edition.

[27] Papaspiliopoulos, O. and Roberts, G. (2005). Retrospective MCMC for Dirichlet process hierarchical models. *Biometrika*, 95:169–186.

[28] Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22(1):55–71.

[29] Pitt, M. K. and Shephard N. (1999). Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association*, 94(446):590–599.

[30] Roberts, G. and Stramer, O. (2001). On inference for partial observed nonlinear diffusion models using the metropolis-hastings algorithm. *Biometrika*, 88(3):603–621.

[31] Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov processes and martingales, 2, Ito calculus.* Wiley, Chicester.

[32] Sørensen, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: a survey. *International Statistical Review*, 72(3):337–354.

[33] Stein, E. M. and Stein, J. C. (1991). Stock proce distributions with stochastic volatility: an analytic approach. *Review of Financial Studies*, 4(4):727–752.

[34] Stroud, J. R., Polson, N. G. and Mller, P. (2004). Practical Filtering for Stochastic Volatility Models. *State Space and Unobserved Components Models (Harvey et al., eds.)*, Cambridge University Press, 236–247.

[35] Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Address of Konstantinos Kalogeropoulos
London School of Economics and Political Sciences
Department of Statistics
Houghton Street, WC2A 2AE, London, UK.
E-mail: K.Kalogeropoulos@lse.ac.uk
URL: http://stats.lse.ac.uk/Kalogeropoulos/

Address of Gareth O. Roberts
University of Warwick
Department of Statistics
CV4 7AL, Coventry, UK.
E-mail: Gareth.O.Roberts@warwick.ac.uk

Address of Petros Dellaportas
Athens University of Economics and Business
Department of Statistics
76 Patission Street, 10434, Athens, Greece.
E-mail: Petros@aueb.gr