Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator

By A. DOUCET

Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, U.K.
doucet@stats.ox.ac.uk

M. K. PITT

Department of Economics, University of Warwick, Coventry CV4 7AL, U.K.
m.pitt@warwick.ac.uk

G. DELIGIANNIDIS

Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, U.K.
deligiannidis@stats.ox.ac.uk

AND R. KOHN

Australian School of Business, University of New South Wales, Sydney 2052, Australia
r.kohn@unsw.edu.au

SUMMARY

When an unbiased estimator of the likelihood is used within a Metropolis–Hastings chain, it is necessary to trade off the number of Monte Carlo samples used to construct this estimator against the asymptotic variances of the averages computed under this chain. Using many Monte Carlo samples will typically result in Metropolis–Hastings averages with lower asymptotic variances than the corresponding averages that use fewer samples; however, the computing time required to construct the likelihood estimator increases with the number of samples. Under the assumption that the distribution of the additive noise introduced by the loglikelihood estimator is Gaussian with variance inversely proportional to the number of samples and independent of the parameter value at which it is evaluated, we provide guidelines on the number of samples to select. We illustrate our results by considering a stochastic volatility model applied to stock index returns.

Some key words: Intractable likelihood; Metropolis–Hastings algorithm; Particle filter; Sequential Monte Carlo; State-space model.

1. INTRODUCTION

The use of unbiased estimators within the Metropolis–Hastings algorithm was initiated by Lin et al. (2000), with a surge of interest in these ideas since their introduction in Bayesian statistics by Beaumont (2003). In a Bayesian context, an unbiased likelihood estimator is commonly constructed using importance sampling as in Beaumont (2003) or particle filters as in Andrieu et al. (2010). Andrieu & Roberts (2009) call this method the pseudo-marginal algorithm and establish some of its theoretical properties.

Apart from the choice of proposals inherent to any Metropolis–Hastings algorithm, the main practical issue with the pseudo-marginal algorithm is the choice of the number, N, of Monte Carlo samples or particles used to estimate the likelihood. For any fixed N, the transition kernel of the pseudo-marginal algorithm leaves the posterior distribution of interest invariant. Using many Monte Carlo samples usually results in pseudo-marginal averages with lower asymptotic variances than the corresponding averages obtained from fewer samples, as established in a 2014 unpublished paper by C. Andrieu and M. Vihola (arXiv:1404.6909) for likelihood estimators based on importance sampling. Empirical evidence suggests that this result also holds when the likelihood is estimated by particle filters. However, the computing cost of constructing the likelihood estimator increases with N. We aim to select N so as to minimize the computational resources necessary to achieve a specified asymptotic variance for a particular pseudo-marginal average. This amount of computational resources, a quantity referred to as the computing time, is typically proportional to N times the asymptotic variance of the particular average, this variance being itself a function of N. Assuming that the distribution of the additive noise introduced by the loglikelihood estimator is Gaussian with a variance inversely proportional to N and independent of the parameter value at which it is evaluated, this minimization was carried out by Pitt et al. (2012) and Sherlock et al. (2015). However, Pitt et al. (2012) assumed that the Metropolis–Hastings proposal is the posterior density, whereas Sherlock et al. (2015) relaxed the Gaussian noise assumption but restricted themselves to an isotropic normal random walk proposal and assumed that the posterior density factorizes into d independent and identically distributed components, where $d \rightarrow \infty$.

Our article addresses a similar problem but considers general proposal and target densities and relaxes the Gaussian noise assumption. In this more general setting, we cannot minimize the computing time itself, and instead minimize explicit upper bounds on it. Quantitative results are presented under a Gaussian assumption. In this scenario, our guidelines are that N should be chosen such that the standard deviation of the loglikelihood estimator is around $1{\cdot}0$ when the Metropolis–Hastings algorithm using the exact likelihood is efficient and around $1{\cdot}7$ when it is inefficient. In most practical scenarios, the efficiency of the Metropolis–Hastings algorithm using the exact likelihood is unknown, as the algorithm cannot be implemented. In such cases, our results suggest selecting a standard deviation around $1{\cdot}2$.

2. Metropolis–Hastings method using an estimated likelihood

We briefly review how an unbiased likelihood estimator can be used within a Metropolis–Hastings scheme in a Bayesian context. Let $y \in \mathsf{Y}$ be the observations and $\theta \in \Theta \subseteq \mathbb{R}^d$ the parameters of interest. The likelihood is denoted by $p(y \mid \theta)$, and the prior for θ admits a density $p(\theta)$ with respect to Lebesgue measure, so the posterior density of interest is $\pi(\theta) \propto p(y \mid \theta)p(\theta)$. We slightly abuse notation by using the same symbols for distributions and densities.

The Metropolis–Hastings scheme to sample from π simulates a Markov chain according to the transition kernel

$$Q_{\mathrm{EX}}(\theta, \mathrm{d}\vartheta) = q(\theta, \vartheta)\alpha_{\mathrm{EX}}(\theta, \vartheta)\,\mathrm{d}\vartheta + \{1 - \varrho_{\mathrm{EX}}(\theta)\}\delta_\theta(\mathrm{d}\vartheta),$$

where

$$\alpha_{\mathrm{EX}}(\theta, \vartheta) = \min\{1, r_{\mathrm{EX}}(\theta, \vartheta)\}, \quad \varrho_{\mathrm{EX}}(\theta) = \int q(\theta, \vartheta)\alpha_{\mathrm{EX}}(\theta, \vartheta)\,\mathrm{d}\vartheta,$$

with $r_{\mathrm{EX}}(\theta, \vartheta) = \pi(\vartheta)q(\vartheta, \theta)/\{\pi(\theta)q(\theta, \vartheta)\}$. This Markov chain cannot be simulated if $p(y \mid \theta)$ is intractable.

Assume that $p(y \mid \theta)$ is intractable but we have access to a nonnegative unbiased estimator $\hat{p}(y \mid \theta, U)$ of $p(y \mid \theta)$, where $U \sim m(\cdot)$ represents all the auxiliary random variables used to obtain this estimator. In this case, we introduce the joint density $\bar{\pi}(\theta, u)$ on $\Theta \times \mathcal{U}$, where

$$\bar{\pi}(\theta, u) = \pi(\theta)m(u)\hat{p}(y \mid \theta, u)/p(y \mid \theta). \tag{1}$$

This joint density admits the correct marginal density $\pi(\theta)$, because $\hat{p}(y \mid \theta, U)$ is unbiased. The pseudo-marginal algorithm is a Metropolis–Hastings scheme targeting (1) with proposal density $q(\theta, \cdot)m(\cdot)$, yielding the acceptance probability

$$\min\left\{1, \frac{\hat{p}(y \mid \vartheta, v)p(\vartheta)q(\vartheta, \theta)}{\hat{p}(y \mid \theta, u)p(\theta)q(\theta, \vartheta)}\right\} = \min\left\{1, \frac{\hat{p}(y \mid \vartheta, v)/p(y \mid \vartheta)}{\hat{p}(y \mid \theta, u)/p(y \mid \theta)} r_{\mathrm{EX}}(\theta, \vartheta)\right\} \tag{2}$$

for a proposal (ϑ, v). In practice, we record only $\{\theta, \log \hat{p}(y \mid \theta, u)\}$ instead of (θ, u). We follow Andrieu & Roberts (2009) and Pitt et al. (2012) and analyse this scheme using additive noise, $Z = \log \hat{p}(y \mid \theta, U) - \log p(y \mid \theta) = \psi(\theta, U)$, in the loglikelihood estimator, rather than U. In this parameterization, the target density on $\Theta \times \mathbb{R}$ becomes

$$\bar{\pi}(\theta, z) = \pi(\theta) \exp(z)g(z \mid \theta), \tag{3}$$

where $g(z \mid \theta)$ is the density of Z when $U \sim m(\cdot)$ and the transformation $Z = \psi(\theta, U)$ is applied.

To sample from $\bar{\pi}(\theta, z)$, we could use the scheme previously described to sample from $\bar{\pi}(\theta, u)$ and then set $z = \psi(\theta, u)$. Equivalently, we can use the transition kernel

$$Q\{(\theta, z), (\mathrm{d}\vartheta, \mathrm{d}w)\} = q(\theta, \vartheta)g(w \mid \vartheta)\alpha_Q\{(\theta, z), (\vartheta, w)\} \mathrm{d}\vartheta \, \mathrm{d}w$$
$$+ \{1 - \varrho_Q(\theta, z)\}\delta_{(\theta, z)}(\mathrm{d}\vartheta, \mathrm{d}w), \tag{4}$$

where

$$\alpha_Q\{(\theta, z), (\vartheta, w)\} = \min\{1, \exp(w - z) r_{\mathrm{EX}}(\theta, \vartheta)\} \tag{5}$$

is (2) expressed in the new parameterization. Henceforth, we make the following assumption.

Assumption 1. The noise density $g(z)$ is independent of θ.

Under this assumption, the target density (3) factorizes as $\pi(\theta)\pi_Z(z)$, where

$$\pi_Z(z) = \exp(z)g(z). \tag{6}$$

Assumption 1 allows us to analyse the performance of the pseudo-marginal algorithm in detail, but it is not satisfied in practical settings. However, in the stationary regime, we are concerned with the noise density at values of the parameter which arise from the target density $\pi(\theta)$ and the marginal density of the proposals at stationarity, $\int \pi(\mathrm{d}\vartheta)q(\vartheta, \theta)$. If the noise density does not vary significantly in regions of high probability mass of these densities, then this assumption is approximately satisfied. In §4, we examine experimentally how the noise density varies with respect to draws from $\pi(\theta)$ and $\int \pi(\mathrm{d}\vartheta)q(\vartheta, \theta)$.

3. MAIN RESULTS

3·1. *Outline*

This section presents the main results of the paper. The proofs are given in Appendix 1 and the Supplementary Material. We minimize upper bounds on the computing time of the pseudo-marginal algorithm, as discussed in § 1. This requires us to establish upper bounds on the asymptotic variance of an ergodic average under the kernel Q given in (4). To obtain these bounds, we introduce a new Markov kernel Q^*, defined by

$$
Q^*\{(\theta, z), (\mathrm{d}\vartheta, \mathrm{d}w)\} = q(\theta, \vartheta) g(w) \alpha_{Q^*}\{(\theta, z), (\vartheta, w)\} \, \mathrm{d}\vartheta \, \mathrm{d}w
$$
$$
+ \{1 - \varrho_{\mathrm{EX}}(\theta) \varrho_Z(z)\} \delta_{(\theta, z)}(\mathrm{d}\vartheta, \mathrm{d}w),
$$

where

$$
\alpha_{Q^*}\{(\theta, z), (\vartheta, w)\} = \alpha_{\mathrm{EX}}(\theta, \vartheta) \alpha_Z(z, w), \quad \alpha_Z(z, w) = \min\{1, \exp(w - z)\}, \tag{7}
$$

$$
\varrho_Z(z) = \int g(w) \alpha_Z(z, w) \, \mathrm{d}w.
$$

As Q and Q^* are reversible with respect to $\bar{\pi}$ and the acceptance probability in (7) is always smaller than (5), an application of the theorem in Peskun (1973) ensures that the variance of an ergodic average under Q^* is greater than or equal to the variance under Q. We obtain an exact expression for the variance under the bounding kernel Q^*, as well as simpler upper bounds, by exploiting a nonstandard representation of this variance, the factor form of the acceptance probability in (7), and the spectral properties of an auxiliary Markov kernel.

3·2. *Inefficiency of Metropolis–Hastings-type chains*

In this subsection we recall and establish various results on the integrated autocorrelation time of Markov chains, henceforth referred to as the inefficiency. In particular, we give a novel representation of the inefficiency of Metropolis–Hastings-type chains, which will be the basic ingredient of the proof of our main result.

Consider a Markov kernel Π on the measurable space $(\mathsf{X}, \mathcal{X}) = \{\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\}$, where $\mathcal{B}(\mathbb{R}^n)$ is the Borel σ-algebra on $\mathbb{R}^n$. For any measurable real-valued function f, measurable set A and probability measure μ, we write $\mu(f) = \int_{\mathsf{X}} \mu(\mathrm{d}x) f(x)$, $\bar{f} = f - \mu(f)$, $\mu(A) = \mu\{\mathbb{I}_A(\cdot)\}$, $\Pi f(x) = \int_{\mathsf{X}} \Pi(x, \mathrm{d}y) f(y)$ and, for $n \geqslant 2$, $\Pi^n(x, \mathrm{d}y) = \int_{\mathsf{X}} \Pi^{n-1}(x, \mathrm{d}z) \Pi(z, \mathrm{d}y)$, with $\Pi^1 = \Pi$. We introduce the Hilbert spaces

$$
L^2(\mathsf{X}, \mu) = \{f : \mathsf{X} \to \mathbb{R} : \mu(f^2) < \infty\}, \quad L_0^2(\mathsf{X}, \mu) = \{f : \mathsf{X} \to \mathbb{R} : \mu(f) = 0, \ \mu(f^2) < \infty\}
$$

equipped with the inner product $\langle f, g \rangle_\mu = \int f(x) g(x) \mu(\mathrm{d}x)$. We write $\phi_n(f, \Pi) = \langle \bar{f}, \Pi^n \bar{f} \rangle_\mu / \mu(\bar{f}^2)$ for the autocorrelation at lag $n \geqslant 0$ and $\mathrm{IF}(f, \Pi) = 1 + 2 \sum_{n=1}^\infty \phi_n(f, \Pi)$ for the associated inefficiency. A μ-invariant and ψ-irreducible Markov chain is said to be ergodic; see Tierney (1994) for the definition of ψ-irreducibility. The next result follows directly from Kipnis & Varadhan (1986) and Häggström & Rosenthal (2007, Theorem 4 and Corollary 6).

PROPOSITION 1. *Consider a μ-reversible and ergodic Markov kernel Π. Let $(X_i)_{i \geqslant 1}$ be a stationary Markov chain that evolves according to Π, and let $h \in L^2(\mathsf{X}, \mu)$ be such that $\mu(\bar{h}^2) > 0$. Then the following properties hold.*

(i) *There exists a probability measure $e(h, \Pi)$ on $[-1, 1)$, called the spectral measure, such that*

$$\phi_n(h, \Pi) = \int_{-1}^{1} \lambda^n e(h, \Pi)(\mathrm{d}\lambda), \quad \mathrm{IF}(h, \Pi) = \int_{-1}^{1} (1 + \lambda)(1 - \lambda)^{-1} e(h, \Pi)(\mathrm{d}\lambda). \quad (8)$$

(ii) *If $\mathrm{IF}(h, \Pi) < \infty$, then as $n \to \infty$,*

$$n^{-1/2} \sum_{i=1}^{n} \{h(X_i) - \mu(h)\} \to N\{0, \, \mu(\bar{h}^2)\mathrm{IF}(h, \Pi)\} \quad (9)$$

in distribution, where $N(a, b^2)$ denotes the normal distribution with mean a and variance b^2.

When estimating $\mu(h)$, (9) implies that we need approximately $n\mathrm{IF}(h, \Pi)$ samples from the Markov chain $(X_i)_{i \geqslant 1}$ to obtain an estimator of the same precision as an average of n independent draws from μ.

Henceforth we consider a μ-reversible kernel

$$P(x, \mathrm{d}y) = q(x, \mathrm{d}y)\alpha(x, y) + \{1 - \varrho(x)\}\delta_x(\mathrm{d}y), \quad \varrho(x) = \int q(x, \mathrm{d}y)\alpha(x, y),$$

where the proposal kernel is selected such that $q(x, \{x\}) = 0$, $\alpha(x, y)$ is the acceptance probability, and we assume that there does not exist an x such that $\mu(\{x\}) = 1$. We refer to P as a Metropolis–Hastings-type kernel because it is structurally similar to the Metropolis–Hastings kernel but we do not require $\alpha(x, y)$ to be the Metropolis–Hastings acceptance probability. This generalization is required when studying the kernel Q^*, as the acceptance probability $\alpha_{Q^*}\{(\theta, z), (\vartheta, w)\}$ in (7) is not the Metropolis–Hastings acceptance probability.

Let $(X_i)_{i \geqslant 1}$ be a Markov chain that evolves according to P. We now establish a nonstandard expression for $\mathrm{IF}(h, P)$ derived from the associated jump chain representation $(\tilde{X}_i, \tau_i)_{i \geqslant 1}$ of $(X_i)_{i \geqslant 1}$. In this representation, $(\tilde{X}_i)_{i \geqslant 1}$ corresponds to the sequence of accepted proposals and $(\tau_i)_{i \geqslant 1}$ the associated sojourn times, i.e., $\tilde{X}_1 = X_1 = \cdots = X_{\tau_1}$, $\tilde{X}_2 = X_{\tau_1 + 1} = \cdots = X_{\tau_1 + \tau_2}$ and so on, with $\tilde{X}_{i+1} \neq \tilde{X}_i$. We now state some properties of this jump chain; see Lemma 1 in Douc & Robert (2011).

LEMMA 1. *Let P be ψ-irreducible. Then $\varrho(x) > 0$ for any $x \in \mathsf{X}$ and $(\tilde{X}_i, \tau_i)_{i \geqslant 1}$ is a Markov chain with a $\bar{\mu}$-reversible transition kernel $\bar{P}$, where*

$$\bar{P}\{(x, \tau), (\mathrm{d}y, \zeta)\} = \tilde{P}(x, \mathrm{d}y)G\{\zeta; \varrho(y)\}, \quad \bar{\mu}(\mathrm{d}x, \tau) = \tilde{\mu}(\mathrm{d}x)G\{\tau; \varrho(x)\}$$

with

$$\tilde{P}(x, \mathrm{d}y) = \frac{q(x, \mathrm{d}y)\alpha(x, y)}{\varrho(x)}, \quad \tilde{\mu}(\mathrm{d}x) = \frac{\mu(\mathrm{d}x)\varrho(x)}{\mu(\varrho)},$$

and $G(\cdot; \upsilon)$ denotes the geometric distribution with parameter υ.

The next proposition gives the relationship between $\mathrm{IF}(h, P)$ and $\mathrm{IF}(h/\varrho, \tilde{P})$.

PROPOSITION 2. *Assume that P and $\tilde{P}$ are ergodic, that $h \in L_0^2(\mathsf{X}, \mu)$ and that* IF$(h, P) < \infty$. *Then $h/\varrho \in L_0^2(\mathsf{X}, \tilde{\mu})$,*

$$\mu(h^2)\{1 + \text{IF}(h, P)\} = \mu(\varrho)\tilde{\mu}(h^2/\varrho^2)\{1 + \text{IF}(h/\varrho, \tilde{P})\}$$

and IF$(h/\varrho, \tilde{P}) \leqslant$ IF(h, P).

Lemma 1 and Proposition 2 are used in § 3·3 to establish a representation of the inefficiency for the kernel $P = Q^*$.

We conclude this section by establishing some results on the positivity of the Metropolis–Hastings kernel and its associated jump kernel. Recall that a μ-invariant Markov kernel Π is positive if $\langle \Pi h, h \rangle_\mu \geqslant 0$ for any $h \in L^2(\mathsf{X}, \mu)$. If Π is reversible, then positivity is equivalent to $e(h, \Pi)([0, 1)) = 1$ for all $h \in L^2(\mathsf{X}, \mu)$, where $e(h, \Pi)$ is the spectral measure, and positivity also implies that IF$(h, \Pi) \geqslant 1$; see, for example, Geyer (1992). The positivity of the jump kernel $\tilde{P}$ associated with a Metropolis–Hastings kernel P is useful here, as several bounds on the inefficiency established subsequently require the spectral measure $e(h, \tilde{P})$ to be supported on $[0, 1)$. We now give sufficient conditions ensuring this property by extending Lemma 3.1 of Baxendale (2005). This complements results of Rudolf & Ullrich (2013).

PROPOSITION 3. *Assume that $\alpha(x, y)$ is the Metropolis–Hastings acceptance probability and $\mu(\mathrm{d}x) = \mu(x)\,\mathrm{d}x$. If P is ψ-irreducible, then $\tilde{P}$ and P are both positive if one of the following two conditions is satisfied:*

(i) *$q(x, \mathrm{d}y) = q(x, y)\,\mathrm{d}y$ is a ν-reversible kernel with $\nu(\mathrm{d}x) = \nu(x)\,\mathrm{d}x$, μ is absolutely continuous with respect to ν, and there exists $r : \mathsf{X} \times \mathsf{Z} \to \mathbb{R}^+$ such that $\nu(x)q(x, y) = \int r(x, z)r(y, z)\chi(\mathrm{d}z)$, where χ is a measure on Z;*

(ii) *$q(x, \mathrm{d}y) = q(x, y)\,\mathrm{d}y$, and there exists $s : \mathsf{X} \times \mathsf{Z} \to \mathbb{R}^+$ such that $q(x, y) = \int s(x, z)s(y, z)\chi(\mathrm{d}z)$, where χ is a measure on Z.*

Remark 1. Condition (i) in the proposition is satisfied for an independent proposal $q(x, y) = \nu(y)$ by taking $\mathsf{Z} = \{1\}$, $\chi(\mathrm{d}z) = \delta_1(\mathrm{d}z)$ and $r(x, 1) = \nu(x)$. It is also satisfied for autoregressive positively correlated proposals with normal or Student-t innovations. Condition (ii) holds if $q(x, y)$ is a symmetric random walk proposal whose increments are multivariate normal or Student-t.

3·3. *Inefficiency of the bounding chain*

In this subsection we apply the results of § 3·2 to establish an exact expression for IF(h, Q^*). The next lemma shows that IF(h, Q^*) is an upper bound on IF(h, Q).

LEMMA 2. *The kernel Q^* is $\bar{\pi}$-reversible and* IF$(h, Q) \leqslant$ IF(h, Q^*) *for any $h \in L^2(\Theta \times \mathbb{R}, \bar{\pi})$.*

In practice, we are only interested in functions $h \in L^2(\Theta, \pi)$. To simplify notation, we write IF(h, Q) in this case, instead of introducing the function $\tilde{h} \in L^2(\Theta \times \mathbb{R}, \bar{\pi})$ which satisfies $\tilde{h}(\theta, z) = h(\theta)$ for all $z \in \mathbb{R}$ and writing IF$(\tilde{h}, Q)$. Proposition 2 shows that it is possible to express IF(h, Q^*) as a function of the inefficiency of its jump kernel $\tilde{Q}^*$, which is particularly useful as $\tilde{Q}^*$ admits a simple structure.

LEMMA 3. *Assume that Q^* is $\bar{\pi}$-irreducible. The jump kernel $\tilde{Q}^*$ associated with Q^* is*

$$\tilde{Q}^*\{(\theta, z), (\mathrm{d}\vartheta, \mathrm{d}w)\} = \tilde{Q}_{\mathrm{EX}}(\theta, \mathrm{d}\vartheta)\tilde{Q}_{\mathrm{Z}}(z, \mathrm{d}w), \tag{10}$$

where

$$\tilde{Q}_{\mathrm{EX}}(\theta, \mathrm{d}\vartheta) = \frac{q(\theta, \vartheta)\alpha_{\mathrm{EX}}(\theta, \vartheta)\,\mathrm{d}\vartheta}{\varrho_{\mathrm{EX}}(\theta)}, \quad \tilde{Q}_{\mathrm{Z}}(z, \mathrm{d}w) = \frac{g(w)\alpha_{\mathrm{Z}}(z, w)\,\mathrm{d}w}{\varrho_{\mathrm{Z}}(z)}.$$

The kernel $\tilde{Q}_{\mathrm{EX}}$ is reversible with respect to $\tilde{\pi}(\mathrm{d}\theta)$, and the kernel $\tilde{Q}_{\mathrm{Z}}$ is positive and reversible with respect to $\tilde{\pi}_{\mathrm{Z}}(\mathrm{d}z)$, where

$$\tilde{\pi}(\mathrm{d}\theta) = \frac{\pi(\mathrm{d}\theta)\varrho_{\mathrm{EX}}(\theta)}{\pi(\varrho_{\mathrm{EX}})}, \quad \tilde{\pi}_{\mathrm{Z}}(\mathrm{d}z) = \frac{\pi_{\mathrm{Z}}(\mathrm{d}z)\varrho_{\mathrm{Z}}(z)}{\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})}.$$

Moreover, $\tilde{Q}_{\mathrm{Z}}$ is uniformly ergodic and if $\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) < \infty$, then IF$(1/\varrho_{\mathrm{Z}}, \tilde{Q}_{\mathrm{Z}}) < \infty$.

In addition, if Q^* is ergodic, $h \in L_0^2(\Theta, \pi)$, IF$(h, Q^*) < \infty$ *and $\tilde{Q}^*$ is ergodic, then $h/\varrho_{\mathrm{EX}} \in L_0^2(\Theta, \tilde{\pi})$, $\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) < \infty$,* IF$\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}), \tilde{Q}^*\} < \infty$ *and*

$$\pi(h^2)\{1 + \text{IF}(h, Q^*)\} = \pi(\varrho_{\mathrm{EX}})\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}})\tilde{\pi}(h^2/\varrho_{\mathrm{EX}}^2)\left[1 + \text{IF}\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}), \tilde{Q}^*\}\right]. \tag{11}$$

The following theorem provides an expression for IF(h, Q^*) which decouples the contributions of the parameter and the noise components. The proof exploits the relationships between IF(h, Q_{EX}) and IF$(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$ and between IF(h, Q^*) and IF$\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}), \tilde{Q}^*\}$, as well as the spectral representation (8) of IF$\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathrm{Z}}), \tilde{Q}^*\}$. This spectral representation admits a simple structure due to the product form (10) of $\tilde{Q}^*$.

THEOREM 1. *Let $h \in L^2(\Theta, \pi)$. Assume that Q_{EX}, Q^*, $\tilde{Q}_{\mathrm{EX}}$ and $\tilde{Q}^*$ are ergodic with* IF$(h, Q^*) < \infty$. *Then* IF$(h, Q) \leqslant$ IF(h, Q^*) *and* IF(h, Q^*) *equals*

$$\frac{1 + \text{IF}(h, Q_{\mathrm{EX}})}{\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})} - 1$$

$$+ \frac{2\{1 + \text{IF}(h, Q_{\mathrm{EX}})\}}{1 + \text{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})}\left\{\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) - \frac{1}{\pi_{\mathrm{Z}}(\varrho_{\mathrm{Z}})}\right\}\sum_{n=0}^{\infty}\phi_n(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})\phi_n(1/\varrho_{\mathrm{Z}}, \tilde{Q}_{\mathrm{Z}}).$$

Remark 2. If $q(\theta, \vartheta) = \pi(\vartheta)$, then IF$(h, Q_{\mathrm{EX}}) =$ IF$(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) = 1$ and $\phi_n(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) = 0$ for $n \geqslant 1$. It follows from Theorem 1 that IF$(h, Q^*) = 2\pi_{\mathrm{Z}}(1/\varrho_{\mathrm{Z}}) - 1$. This result was established in Lemma 4 of Pitt et al. (2012).

Theorem 1 requires Q_{EX}, Q^*, $\tilde{Q}_{\mathrm{EX}}$ and $\tilde{Q}^*$ to be ergodic. The following proposition, which generalizes Theorem 2.2 of Roberts & Tweedie (1996), provides sufficient conditions ensuring this.

PROPOSITION 4. *Suppose that $\pi(\theta)$ is bounded away from 0 and ∞ on compact sets, and that there exist $\delta > 0$ and $\varepsilon > 0$ such that $|\theta - \vartheta| \leqslant \delta$ implies $q(\theta, \vartheta) \geqslant \varepsilon$. Then Q_{EX}, Q^*, $\tilde{Q}_{\mathrm{EX}}$ and $\tilde{Q}^*$ are ergodic.*

3·4. *Bounds on the relative inefficiency of the pseudo-marginal chain*

For any kernel Π, we define the relative inefficiency RIF$(h, \Pi) =$ IF$(h, \Pi)/$IF(h, Q_{EX}), which measures the inefficiency of Π compared to that of Q_{EX}. This section provides tractable upper bounds for RIF(h, Q). From Lemma 2, RIF$(h, Q) \leqslant$ RIF(h, Q^*), but the expression of

$\text{RIF}(h, Q^*)$ that follows from Theorem 1 is intricate and depends on the autocorrelation sequence $\{\phi_n(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}_{n \geqslant 1}$ as well as on other terms. The next corollary provides upper bounds on $\text{RIF}(h, Q)$ that depend only on $\text{IF}(h, Q_{\text{EX}})$ and $\phi_1(1/\varrho_z, \tilde{Q}_z)$. The latter can be evaluated numerically and, to simplify the notation, we refer to it as ϕ_z.

COROLLARY 1. *Under the assumptions of Theorem 1, the following hold:*

(i) $\text{RIF}(h, Q) \leqslant \text{URIF}_1(h)$*, where*

$$\text{URIF}_1(h) = \{1 + 1/\text{IF}(h, Q_{\text{EX}})\}[\pi_z(1/\varrho_z) + (1 - \phi_z)\{\pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z)\}]$$
$$- 1/\text{IF}(h, Q_{\text{EX}});$$

(ii) *if, in addition,* $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \geqslant 1$*, then* $\text{RIF}(h, Q) \leqslant \text{URIF}_2(h) \leqslant \text{URIF}_1(h)$*, where*

$$\text{URIF}_2(h) = \{1 + 1/\text{IF}(h, Q_{\text{EX}})\}\pi_z(1/\varrho_z) - 1/\text{IF}(h, Q_{\text{EX}}).$$

Proposition 3 gives sufficient conditions for the condition $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \geqslant 1$ of Corollary 1(ii) to hold.

Remark 3. The bounds above are tight in two cases. First, if $\pi_z(1/\varrho_z) \to 1$, then $\text{RIF}(h, Q)$, $\text{URIF}_1(h)$, $\text{URIF}_2(h) \to 1$. Second, if $q(\theta, \vartheta) = \pi(\vartheta)$, then $\text{RIF}(h, Q) = \text{URIF}_2(h)$.

We now provide upper bounds on $\text{RIF}(h, Q)$ and lower bounds on $\text{RIF}(h, Q^*)$ in terms of $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$.

COROLLARY 2. *Under the assumptions of Theorem 1:*

(i) $\text{RIF}(h, Q) \leqslant \text{URIF}_3(h)$ *where*

$$\text{URIF}_3(h) = \left\{1 + \frac{1}{\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})}\right\}\left[\frac{1}{\pi_z(\varrho_z)} + \phi_z\left\{\pi_z(1/\varrho_z) - \frac{1}{\pi_z(\varrho_z)}\right\}\right]$$
$$+ 2\{\pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z)\}(1 - \phi_z)/\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - 1/\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}});$$

(ii) $\text{RIF}(h, Q) \leqslant \text{URIF}_4(h)$ *where*

$$\text{URIF}_4(h) = \frac{\{1 + 1/\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})}\{\pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z)\}\{1 + \text{IF}(1/\varrho_z, \tilde{Q}_z)\}$$
$$+ 1/\pi_z(\varrho_z) + \frac{1}{\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})}\left\{\frac{1}{\pi_z(\varrho_z)} - 1\right\};$$

(iii) *if* $\tilde{Q}_{\text{EX}}$ *is positive, then* $\text{RIF}(h, Q^*) \geqslant \text{LRIF}_1(h)$ *where*

$$\text{LRIF}_1(h) = \frac{1}{\pi_z(\varrho_z)} + \frac{2}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})}\{\pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z)\};$$

(iv) $\text{RIF}(h, Q^*) \geqslant \text{LRIF}_2$ *where*

$$\text{LRIF}_2 = 1/\pi_z(\varrho_z), \tag{12}$$

and $\text{RIF}(h, Q^*)$, $\text{URIF}_4(h) \to \text{LRIF}_2$ *as* $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \to \infty$.

Again, Proposition 3 gives sufficient conditions for $\tilde{Q}_{\mathrm{EX}}$ to be positive. In § 3·5 these bounds are discussed in more detail.

3·5. *Optimizing the computing time under a Gaussian assumption*

This subsection provides quantitative guidelines on how to select the standard deviation σ of the noise density, under the following assumption.

Assumption 2. The noise density is $g^{\sigma}(z) = \varphi(z; -\sigma^2/2, \sigma^2)$, where $\varphi(z; a, b^2)$ is a univariate normal density with mean a and variance b^2.

Assumption 2 ensures that $\int \exp(z) g^{\sigma}(z)\, \mathrm{d}z = 1$, as required by the unbiasedness of the likelihood estimator. Consider a time series $y_{1:T} = (y_1, \ldots, y_T)$, where the likelihood estimator $\hat{p}(y_{1:T} \mid \theta)$ of $p(y_{1:T} \mid \theta)$ is computed through a particle filter with N particles. Theorem 1 of Bérard et al. (2014) shows that, under regularity assumptions, the loglikelihood error is distributed according to a normal density with mean $-\delta \gamma^2/2$ and variance $\delta \gamma^2$ as $T \to \infty$, for $N = \delta^{-1} T$. Hence, in this important scenario, the noise distribution satisfies approximately the form specified in Assumption 2 for large T, and the variance is asymptotically inversely proportional to the number of samples. This assumption is also made in Pitt et al. (2012), where it is justified experimentally. Section 4 below provides additional experimental results.

The next result is Lemma 4 in Pitt et al. (2012) and follows from Assumption 2, equation (6) and Remark 2. We now make the dependence on σ explicit in our notation.

COROLLARY 3. *Under Assumption 2*, $\pi_{\mathrm{Z}}^{\sigma}(z) = \varphi(z; \sigma^2/2, \sigma^2)$,

$$\varrho_{\mathrm{Z}}^{\sigma}(z) = 1 - \Phi(z/\sigma + \sigma/2) + \exp(-z)\Phi(z/\sigma - \sigma/2),$$

$$\pi_{\mathrm{Z}}^{\sigma}(1/\varrho_{\mathrm{Z}}^{\sigma}) = \int \frac{\varphi(w; 0, 1)}{1 - \bar{\varrho}_{\mathrm{Z}}^{\sigma}(w)}\, \mathrm{d}w,$$

where $\bar{\varrho}_{\mathrm{Z}}^{\sigma}(w) = \Phi(w + \sigma) - \exp(-w\sigma - \sigma^2/2)\Phi(w)$ *and* $\Phi(\cdot)$ *is the standard Gaussian cumulative distribution function. Additionally*, $\pi_{\mathrm{Z}}^{\sigma}(\varrho_{\mathrm{Z}}^{\sigma}) = 2\Phi(-\sigma/\sqrt{2})$.

The terms $\pi_{\mathrm{Z}}^{\sigma}(1/\varrho_{\mathrm{Z}}^{\sigma})$, $\phi_{\mathrm{Z}}^{\sigma}$ and $\mathrm{IF}(1/\varrho_{\mathrm{Z}}^{\sigma}, \tilde{Q}_{\mathrm{Z}})$ appearing in the bounds of Corollaries 1 and 2 do not admit analytical expressions, but they can be computed numerically. We note that $\pi_{\mathrm{Z}}^{\sigma}(1/\varrho_{\mathrm{Z}}^{\sigma})$ is finite and thus, by Lemma 3, $\mathrm{IF}(1/\varrho_{\mathrm{Z}}^{\sigma}, \tilde{Q}_{\mathrm{Z}})$ is also finite. Consequently, for specific values of σ, $\mathrm{IF}(h, Q_{\mathrm{EX}})$ and $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$, these bounds can be calculated.

We now use these bounds to guide the choice of σ. The quantity we aim to minimize is the relative computing time for Q, defined as $\mathrm{RCT}(h, Q; \sigma) = \mathrm{RIF}(h, Q; \sigma)/\sigma^2$, because $1/\sigma^2$ is usually approximately proportional to the number of samples N used to estimate the likelihood and the computational cost at each iteration is typically proportional to N, at least in the particle filter scenario described previously. We define $\mathrm{RCT}(h, Q^*; \sigma)$ similarly. As $\mathrm{RIF}(h, Q; \sigma)$ is intractable, we instead minimize the upper bounds $\mathrm{URCT}_i(h; \sigma) = \mathrm{URIF}_i(h; \sigma)/\sigma^2$ for $i = 1, \ldots, 4$. We similarly define the quantities $\mathrm{LRCT}_1(h; \sigma) = \mathrm{LRIF}_1(h; \sigma)/\sigma^2$ and $\mathrm{LRCT}_2(\sigma) = \mathrm{LRIF}_2(\sigma)/\sigma^2$, which bound $\mathrm{RCT}(h, Q^*; \sigma)$ from below. Figure 1 plots these bounds, except $\mathrm{LRCT}_1(h; \sigma)$, for clarity, against σ for different values of $\mathrm{IF}(h, Q_{\mathrm{EX}})$ and $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$. It is clear, however, that $\mathrm{LRCT}_1(h; \sigma)$ provides a much tighter lower bound than $\mathrm{LRCT}_2(\sigma)$ for low values of $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$.
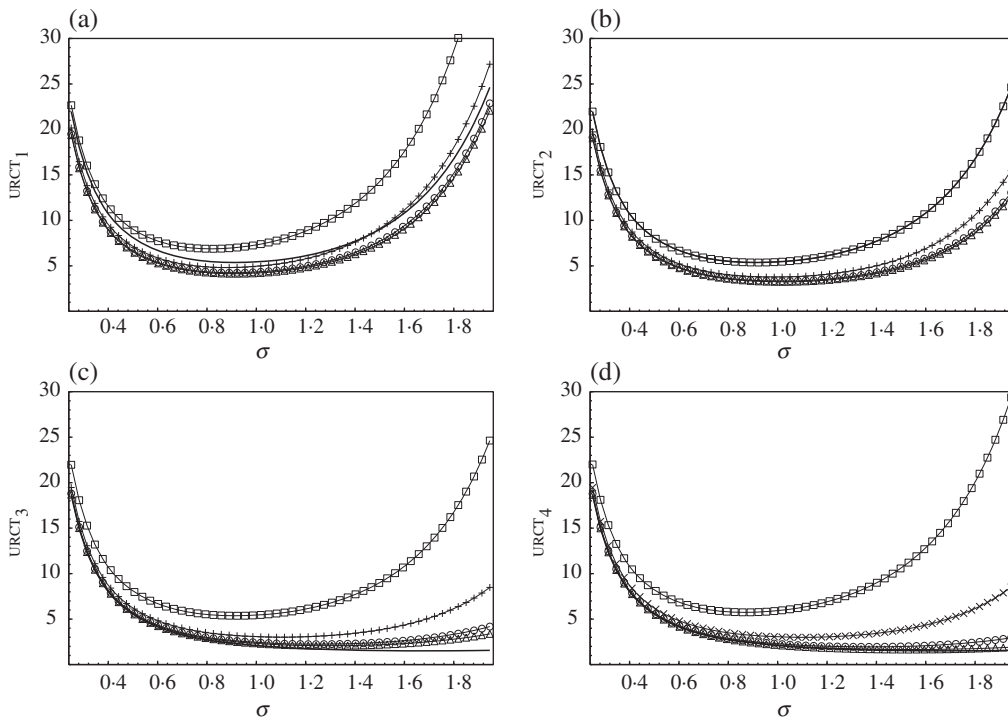
Fig. 1. Theoretical results for relative computing time as a function of σ: (a) $\mathrm{URCT}_1(h; \sigma)$ and (b) $\mathrm{URCT}_2(h; \sigma)$, with values of $\mathrm{IF}(h, Q_{\mathrm{EX}})$ taken to be 1 (squares), 4 (crosses), 20 (circles) and 80 (triangles), and where the solid line corresponds to the perfect proposal, as discussed in Remark 2; (c) $\mathrm{URCT}_3(h; \sigma)$ and (d) $\mathrm{URCT}_4(h; \sigma)$ plotted together with the lower bound $\mathrm{LRCT}_2(\sigma)$ (solid line), with values of $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$ taken to be 1 (squares), 4 (crosses), 20 (circles) and 80 (triangles).

Before discussing how these results guide the selection of σ, we outline some properties of the bounds. First, as the corresponding inefficiency increases, the upper bounds $\mathrm{URCT}_i(h; \sigma)$ displayed in Fig. 1 become flatter as functions of σ, and the corresponding minimizing argument σ_{opt} increases. This flattening effect suggests less sensitivity to the choice of σ for the pseudo-marginal algorithm. Second, for given σ, all the upper bounds are decreasing functions of the corresponding inefficiency, which suggests that the penalty from using the pseudo-marginal algorithm drops as the exact algorithm becomes more inefficient. Third, in the case discussed in Remark 2, where $q(\theta, \vartheta) = \pi(\vartheta)$ so that $\mathrm{IF}(h, Q_{\mathrm{EX}}) = \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) = 1$, we obtain $\mathrm{URCT}_2(h; \sigma) = \mathrm{URCT}_3(h; \sigma) = \mathrm{RCT}(h, Q^*; \sigma) = \mathrm{RCT}(h, Q; \sigma)$. Fourth, $\mathrm{URCT}_4(h; \sigma)$ agrees with the lower bound $\mathrm{LRCT}_2(\sigma)$ as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) \to \infty$, as indicated by Corollary 2(ii). In this case these two bounds, as well as $\mathrm{URCT}_1(h; \sigma)$, are sharp for $\mathrm{RCT}(h, Q^*; \sigma)$. Fifth, $\mathrm{URCT}_2(h; \sigma)$ is sharper than $\mathrm{URCT}_1(h; \sigma)$ for $\mathrm{RCT}(h, Q^*; \sigma)$, but requires a mild additional assumption.

As the likelihood is intractable, it is necessary to make a judgement on how to choose σ, because $\mathrm{IF}(h, Q_{\mathrm{EX}})$ and $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$ are unknown and cannot be easily estimated. Consider two extreme scenarios. The first is the perfect proposal $q(\theta, \vartheta) = \pi(\vartheta)$; in this case, by Corollary 3 and Remark 2, $\mathrm{RCT}(h, Q; \sigma) = \{2\pi_Z^\sigma(1/\varrho_Z^\sigma) - 1\}/\sigma^2$, which we denote by $\mathrm{RCT}(h, Q_\pi; \sigma)$, is minimized at $\sigma_{\mathrm{opt}} = 0\cdot 92$. The second scenario involves a very inefficient proposal corresponding to Corollary 2(iv), so that $\mathrm{RCT}(h, Q^*; \sigma) = \mathrm{LRCT}_2(\sigma)$, which is minimized at $\sigma_{\mathrm{opt}} = 1\cdot 68$. If we choose $\sigma_{\mathrm{opt}} = 1\cdot 68$ over $\sigma_{\mathrm{opt}} = 0\cdot 92$ in the first scenario, then $\mathrm{RCT}(h, Q_\pi; \sigma)$

rises from 5·36 to 12·73. Conversely, if we choose $\sigma_{\mathrm{opt}} = 0\cdot92$ over $\sigma_{\mathrm{opt}} = 1\cdot68$ in the second scenario, the relative computing time $\mathrm{RCT}(h, Q^*; \sigma)$ rises from 1·51 to 2·29. This suggests that the penalty in choosing the wrong value is much more severe if we incorrectly assume we are in the second scenario than if we incorrectly assume we are in the first scenario. This is because as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$ increases, $\mathrm{LRCT}_2(\sigma)$ is very flat relative to $\mathrm{RCT}(h, Q_\pi; \sigma)$ as a function of σ. In practice, choosing σ_{opt} slightly greater than 1·0 seems sensible. For example, a value of $\sigma = 1\cdot2$ leads to an increase in $\mathrm{RCT}(h, Q_\pi; \sigma)$ from the minimum value of 5·36 to 6·10 and an increase in $\mathrm{LRCT}_2(\sigma)$ from the minimum value of 1·51 to 1·75. In Appendix 2, we compute lower and upper bounds for the minimizing argument of $\mathrm{RCT}(h, Q^*; \sigma)$ for various values of $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$.

Some caution should be exercised in interpreting these results, as the lower bounds apply to $\mathrm{RCT}(h, Q^*; \sigma)$ but not, in general, to $\mathrm{RCT}(h, Q; \sigma)$. Similarly, while $\mathrm{URCT}_4(h; \sigma)$ and the lower bounds become exact for $\mathrm{RCT}(h, Q^*; \sigma)$ as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) \to \infty$, they only provide upper bounds for $\mathrm{RCT}(h, Q; \sigma)$. However, there are scenarios in which $\mathrm{RCT}(h, Q^*; \sigma) \approx \mathrm{RCT}(h, Q; \sigma)$. This is the case when the acceptance probability α_{EX} of Q_{EX} is close to unity so that $\alpha_{Q^*} \approx \alpha_Q$, which occurs, for instance, when the proposal is reversible with respect to a distribution approximating π or when the proposal is a random walk with a very small step size.

The numerical results in this section are based on Assumption 2. However, the bounds on the relative inefficiences of Q and Q^* presented in Corollaries 1 and 2 can be calculated for any other noise distribution $g(z)$, subject to $\int \exp(z)g(z)\,\mathrm{d}z = 1$. For example, one could consider a Laplace distribution with mean $\log(1 - \sigma^2)$ and scale parameter $\sigma/\sqrt{2}$ as in Sherlock et al. (2015). These bounds can in turn be used to construct corresponding bounds on the relative computing times of Q and Q^*, provided that an appropriate penalization term can be found to relate the parameterization of the noise distribution to the computational effort of obtaining the likelihood estimator.

3·6. *Discussion*

We now compare informally the bound $\mathrm{LRIF}_2(\sigma) = 1/\{2\Phi(-\sigma/\sqrt{2})\}$ in Corollary 2(iv) with the results in Sherlock et al. (2015). These authors make Assumption 1 and also assume that the target factorizes into d independent and identically distributed components and that the proposal is an isotropic Gaussian random walk of jump size $d^{-1/2}l$. In the Gaussian noise case, for $h(\theta) = \theta_1$ where $\theta = (\theta_1, \ldots, \theta_d)$, their results and a standard calculation with their diffusion limit suggest that as $d \to \infty$,

$$\frac{\mathrm{IF}(h, Q; \sigma, l)}{\mathrm{IF}(h, Q_{\mathrm{EX}}; l)} = \mathrm{RIF}(h, Q; \sigma, l) \to \mathrm{ARIF}(\sigma, l) = \frac{J_{\sigma^2=0}(l)}{J_{\sigma^2}(l)} = \frac{\Phi(-l/2)}{\Phi\{-(2\sigma^2 + l^2)^{1/2}/2\}},$$

where the expression for $J_{\sigma^2}(l)$ is given by equations (3.2) and (3.3) in Sherlock et al. (2015). We observe that $\mathrm{ARIF}(\sigma, l)$ converges to $\mathrm{LRIF}_2(\sigma)$ as $l \to 0$. This is unsurprising. As $d \to \infty$, we conjecture that in this scenario the conditions of Corollary 2(iv) hold; in particular, $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) \to \infty$ for any $l > 0$. Therefore, in this case, $\mathrm{RIF}(h, Q^*; \sigma, l) \to \mathrm{LRIF}_2(\sigma)$. As $l \to 0$, we have informally that $\varrho_{\mathrm{EX}}(\theta) \to 1$, so it is reasonable to conjecture that $\mathrm{RIF}(h, Q; \sigma, l)/\mathrm{RIF}(h, Q^*; \sigma, l) \to 1$. If one of these limits holds uniformly, then $\mathrm{ARIF}(\sigma, l) \to \mathrm{LRIF}_2(\sigma)$.

4. APPLICATION

4·1. *Stochastic volatility model*

In this section we describe a multivariate partially observed diffusion model, which was introduced by Chernov et al. (2003) and discussed in Huang & Tauchen (2005). The logarithm of

regularly observed price $P(t)$ evolves according to

$$d \log P(t) = \mu_y \, dt + \sigma(t) \, dB(t),$$
$$dv_1(t) = -k_1 \{v_1(t) - \mu_1\} \, dt + \sigma_1 \, dW_1(t),$$
$$dv_2(t) = -k_2 v_2(t) \, dt + \{1 + \beta_{12} v_2(t)\} \, dW_2(t),$$

with $\sigma(t) = \text{s-exp}[\{v_1(t) + \beta_2 v_2(t)\}/2]$, where the function $\text{s-exp}(\cdot)$ is a spliced exponential function to ensure nonexplosive growth. The leverage parameters corresponding to the correlations between the driving Brownian motions are $\phi_1 = \text{corr}\{B(t), W_1(t)\}$ and $\phi_2 = \text{corr}\{B(t), W_2(t)\}$. The two components for volatility allow for quite sudden changes in the log price while retaining long memory in volatility. We note that the Brownian motion of the price process may be expressed as $dB(t) = a_1 dW_1(t) + a_2 dW_2(t) + \sqrt{b} \, d\bar{B}(t)$, where $a_1 = \phi_1(1 - \phi_2^2)/(1 - \phi_1^2\phi_2^2)$, $a_2 = \phi_2(1 - \phi_1^2)/(1 - \phi_1^2\phi_2^2)$ and $b = (1 - \phi_1^2)(1 - \phi_2^2)/(1 - \phi_1^2\phi_2^2)$. Here $\bar{B}(t)$ is an independent Brownian motion. Suppose that the log prices are observed at equally spaced times $\tau_1 < \cdots < \tau_T < \tau_{T+1}$, and let $\Delta = \tau_{s+1} - \tau_s$ for any s, which gives returns $Y_s = \log P(\tau_{s+1}) - \log P(\tau_s)$ for $s = 1, \ldots, T$. The distribution of these returns conditional on the volatility paths and the driving processes $W_1(t)$ and $W_2(t)$ can be expressed in closed form as $Y_s \sim N(\mu_y \Delta + a_1 Z_{1,s} + a_2 Z_{2,s}, b \sigma_s^{2*})$ where

$$Z_{1,s} = \int_{\tau_s}^{\tau_{s+1}} \sigma(u) \, dW_1(u), \quad Z_{2,s} = \int_{\tau_s}^{\tau_{s+1}} \sigma(u) \, dW_2(u), \quad \sigma_s^{2*} = \int_{\tau_s}^{\tau_{s+1}} \sigma^2(u) \, du. \quad (13)$$

An Euler scheme is used to approximate the evolution of the volatilities $v_1(t)$ and $v_2(t)$ by placing $M - 1$ latent points between τ_s and τ_{s+1}. The volatility components are denoted by $v_{1,1}^s, \ldots, v_{1,M-1}^s$ and $v_{2,1}^s, \ldots, v_{2,M-1}^s$. For notational convenience, the start and end points are set to $v_{1,0}^s = v_1(\tau_s)$ and $v_{1,M}^s = v_1(\tau_{s+1})$, and similarly for $v_2(t)$. These latent points are evenly spaced in time by $\delta = \Delta/M$. The equation for the Euler evolution, starting at $v_{1,0}^s = v_{1,M}^{s-1}$ and $v_{2,0}^s = v_{2,M}^{s-1}$, is

$$v_{1,m+1}^s = v_{1,m}^s - k_1(v_{1,m}^s - \mu_1)\delta + \sigma_1\sqrt{\delta} u_{1,m},$$
$$v_{2,m+1}^s = v_{2,m}^s - k_2 v_{2,m}^s \delta + (1 + \beta_{12} v_{2,m}^s)\sqrt{\delta} u_{2,m} \quad (m = 0, \ldots, M-1),$$

where $u_{1,m} \sim N(0, 1)$ and $u_{2,m} \sim N(0, 1)$. Conditional on these trajectories and the innovations, the distribution of the returns has a closed form, so that $Y_s \sim N(\mu_y \Delta + a_1 \hat{Z}_{1,s} + a_2 \hat{Z}_{2,s}, b \hat{\sigma}_s^{2*})$, where $\hat{Z}_{1,s}$, $\hat{Z}_{2,s}$ and $\hat{\sigma}_s^{2*}$ are the Euler approximations to the corresponding expressions in (13).

4·2. *Empirical results for the error of the loglikelihood estimator*

We consider T daily returns, $y = (y_1, \ldots, y_T)$, for the S&P 500 index from three different datasets: $T = 40$, 29 April 1983 to 27 June 1983; $T = 300$, 27 April 1982 to 27 June 1983 or $T = 2700$, 7 March 1975 to 7 November 1985. In this subsection we investigate Assumptions 1 and 2 by examining the behaviour of the error of the loglikelihood estimator $Z = \log \hat{p}_N(y \mid \theta) - \log p(y \mid \theta)$ for the nine-dimensional parameter vector $\theta = (k_1, \mu_1, \sigma_1, k_2, \beta_{12}, \beta_2, \mu_y, \phi_1, \phi_2)$ of the stochastic volatility model introduced in the previous subsection. The likelihood is estimated using the bootstrap particle filter with N particles. We select values of N to ensure that, for each dataset, the variance of Z evaluated at $\bar{\theta}$ is approximately unity, where $\bar{\theta}$ is the posterior mean corresponding to a vague prior. We take $\delta = 0.5$ for the Euler scheme.
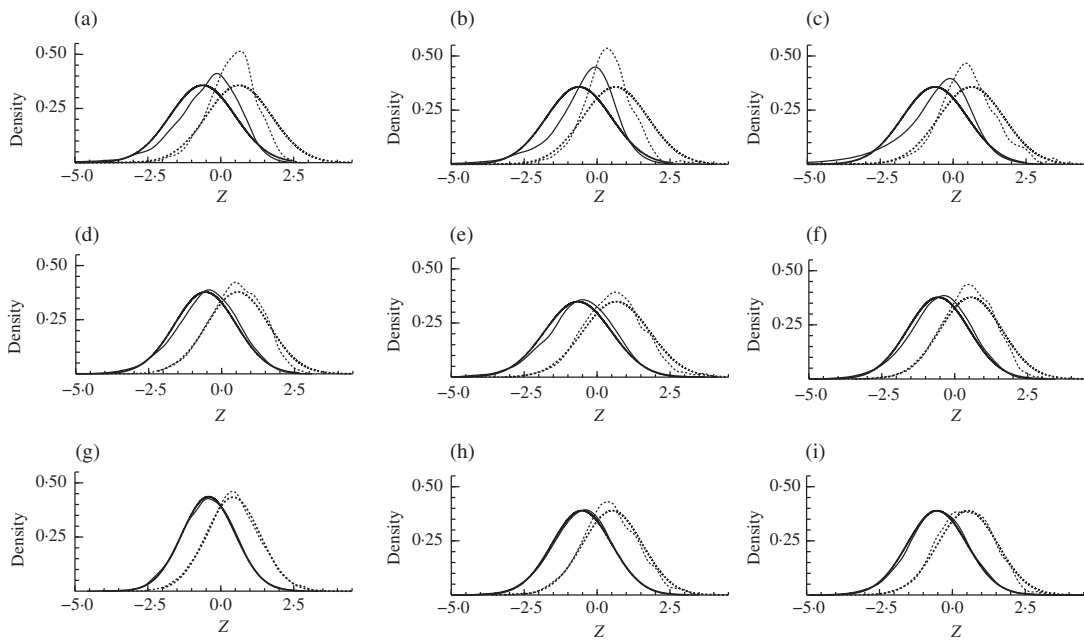
Fig. 2. Estimated and theoretical densities for Z, in the three cases where $T = 40$ and $N = 4$ (panels (a)–(c)), $T = 300$ and $N = 80$ (panels (d)–(f)), and $T = 2700$ and $N = 700$ (panels (g)–(i)); the kernel density estimates associated with $g_N(z \mid \theta)$ (thin solid) and $\pi_N(z \mid \theta)$ (thin dotted) evaluated at the posterior mean $\bar{\theta}$ are plotted in panels (a), (d) and (g), the estimates evaluated over values from the posterior $\pi(\theta)$ are plotted in panels (b), (e) and (h), while those evaluated over values from $\int \pi(\mathrm{d}\vartheta) q(\vartheta, \theta)$ are plotted in panels (c), (f) and (i); the theoretical densities $g_Z^\sigma(z)$ (thick solid) and $\pi_Z^\sigma(z)$ (thick dotted) are overlaid.

Panels (a), (d) and (g) of Fig. 2 display the kernel density estimates corresponding to the density of Z for $\theta = \bar{\theta}$, denoted by $g_N(z \mid \bar{\theta})$, which is obtained by running $S = 6000$ particle filters at this value. As $p(y \mid \bar{\theta})$ is unknown, it is estimated by averaging these estimates. The Metropolis–Hastings algorithm is then used to obtain samples from $\pi_N(z \mid \bar{\theta}) = \exp(z) g_N(z \mid \bar{\theta})$. On each kernel density estimate we overlay the corresponding assumed density, $g_Z^\sigma(z)$ or $\pi_Z^\sigma(z)$, where σ^2 is the sample variance of Z over the S particle filters. For $T = 40$, there is a discrepancy between the assumed Gaussian densities and the kernel density estimates representing $g_N(z \mid \bar{\theta})$ and $\pi_N(z \mid \bar{\theta})$. In particular, although $g_N(z \mid \bar{\theta})$ is well approximated over most of its support, it is slightly lighter-tailed than the assumed Gaussian density in the right tail and much heavier-tailed in the left tail. This translates into a smaller discrepancy between $g_N(z \mid \theta)$ and $\pi_N(z \mid \theta)$ and a higher acceptance rate for the pseudo-marginal algorithm than the Gaussian assumption suggests. For $T = 300$ and $T = 2700$, the assumed Gaussian densities are very accurate.

We also examine Z when θ is distributed according to $\pi(\theta)$. We record 200 samples from $\pi(\theta)$, for $T = 40$, 300 and 2700. These samples are obtained using the pseudo-marginal algorithm. We use a multivariate Student-t random walk proposal $q(\vartheta, \theta)$ with step size proportional to $T^{-1/2}$ on the parameter components transformed to the real line. For each of these samples, we run the particle filter 300 times in order to estimate the true likelihood at these values. The resulting kernel density estimates, corresponding to the densities $\int \pi(\mathrm{d}\theta) g_N(z \mid \theta)$ and $\int \pi(\mathrm{d}\theta) \pi_N(z \mid \theta)$, are displayed in panels (b), (e) and (h) of Fig. 2. We similarly examine the density of Z when θ is distributed according to the marginal proposal density in the stationary regime $\int \pi(\mathrm{d}\vartheta) q(\vartheta, \theta)$. Panels (c), (f) and (i) of Fig. 2 show the resulting density estimates. In both scenarios, Assumptions 1 and 2 are problematic for $T = 40$, as $g_N(z \mid \bar{\theta})$ is not close to Gaussian since
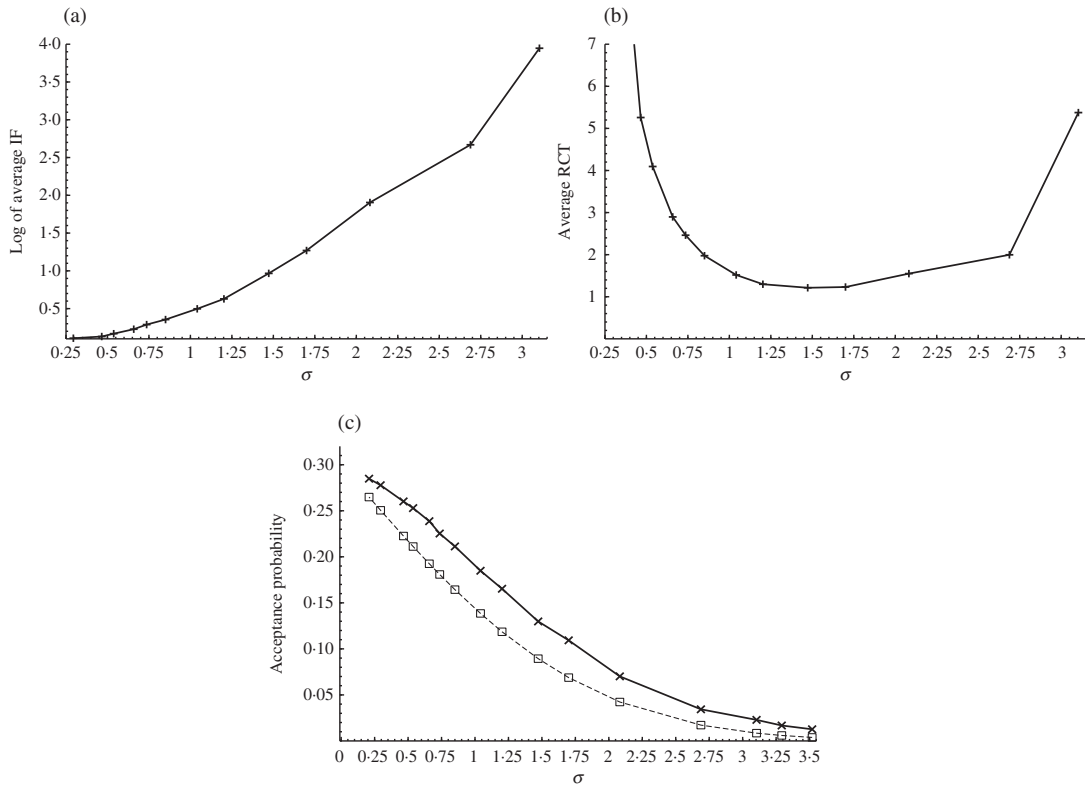
Fig. 3. Performance summary for the pseudo-marginal algorithm applied to the S&P 500 daily returns data for the Huang & Tauchen (2005) two-factor model, with $T = 300$: (a) inefficiencies IF, averaged over the nine parameter components, plotted against σ; (b) relative computing times RCT, averaged over the nine parameter components, plotted against σ; (c) marginal acceptance probability $\bar{\pi}(\varrho_Q)$ (crosses) together with the lower bound $2\Phi(-\sigma/\sqrt{2})\pi(\varrho_{EX})$ (squares) plotted against σ.

the central limit theorem provides a poor approximation. Moreover, since T is small, $\pi(\theta)$ and $\int \pi(\mathrm{d}\vartheta)q(\vartheta, \theta)$ are relatively diffuse. Consequently, $g_N(z \mid \bar{\theta})$ is not close to $g_N(z \mid \theta)$ marginalized over $\pi(\theta)$ or $\int \pi(\mathrm{d}\vartheta)q(\vartheta, \theta)$. For $T = 300$ and $T = 2700$, the assumed densities $g_Z^\sigma(z)$ and $\pi_Z^\sigma(z)$ are close to the corresponding kernel estimates and Assumptions 1 and 2 appear to capture reasonably well the salient features of the densities associated with Z. In particular, the approximation suggested by the central limit theorem becomes very good. Additionally, $\pi(\theta)$ and $\int \pi(\mathrm{d}\vartheta)q(\vartheta, \theta)$ are sufficiently concentrated to ensure that the variance of Z as a function of θ exhibits little variability.

4·3. *Empirical results for the pseudo-marginal algorithm*

We examine the behaviour of the pseudo-marginal algorithm when sampling from the posterior density $\pi(\theta)$ with $\delta = 0.05$ and $T = 300$ for various values of N. We use the same pseudo-marginal algorithm as described in the previous subsection.

The standard deviation $\sigma(\bar{\theta}; N)$ of $\log \hat{p}_N(y \mid \bar{\theta})$, where $\bar{\theta}$ is the posterior mean, is evaluated by Monte Carlo simulations. For each value of N, we compute the inefficiencies, denoted by IF, and the corresponding approximate relative computing times, denoted by RCT, of all parameter components. The quantity RCT is computed as IF$/\sigma^2(\bar{\theta}; N)$ divided by the inefficiency of Q when $N = 2000$, the latter being an approximation of the inefficiency of Q_{EX}. The results are very similar for all parameter components and so, for clarity, Fig. 3 shows the average quantities

over the nine components. For most parameters, the optimal value for $\sigma(\bar{\theta}; N)$ is between $1 \cdot 2$ and $1 \cdot 5$, corresponding to $N = 40$ and 60. The results agree with the bound $\text{URCT}_4(h; \sigma)$ in § 3·5. This can be partly explained by the inefficiencies associated with $\tilde{Q}$ for $N = 2000$ being large, suggesting that the inefficiencies associated with $\tilde{Q}_{\text{EX}}$ are large.

As all the bounds in the paper are based on Q^*, it is useful to assess the discrepancy between Q and Q^*. One approach is to examine the marginal acceptance probability $\bar{\pi}(\varrho_Q)$ under Q against $\sigma = \sigma(\bar{\theta}, N)$ as N varies. Using the acceptance criterion (7) of Q^*, we obtain under Assumptions 1 and 2 that $\bar{\pi}(\varrho_Q) \geqslant 2\Phi(-\sigma/\sqrt{2})\pi(\varrho_{\text{EX}})$. If Q and Q^* are close in the sense of having similar marginal acceptance probabilities, then we expect $\bar{\pi}(\varrho_Q)$ to have a similar shape to its lower bound where $\pi(\varrho_{\text{EX}})$ is approximated using $\bar{\pi}(\varrho_Q)$ with $N = 2000$. For this model, the two functions on either side of the inequality, displayed in Fig. 3, are similar.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes the proofs of Propositions 2–4 and Lemmas 1 and 3, as well as an additional example.

APPENDIX 1

Proof of Lemma 2. It is straightforward to establish that Q^* is $\bar{\pi}$-reversible. Moreover, for any $a, b \geqslant 0$, $\min(1, a)\min(1, b) \leqslant \min(1, ab)$, so $\alpha_{Q^*}\{(\theta, z), (\vartheta, w)\} \leqslant \alpha_Q\{(\theta, z), (\vartheta, w)\}$ for any θ, z, ϑ and w. Hence, Theorem 4 in Tierney (1998), which is a general state-space extension of Peskun (1973), applies and yields the result. □

Proof of Theorem 1. Without loss of generality, let $h \in L_0^2(\Theta, \pi)$. By Theorem 7 of Andrieu & Vihola (2015), $\text{IF}(h, Q_{\text{EX}}) \leqslant \text{IF}(h, Q)$, and by Lemma 2, $\text{IF}(h, Q) \leqslant \text{IF}(h, Q^*)$, where $\text{IF}(h, Q^*) < \infty$ by assumption. Hence $\text{IF}(h, Q_{\text{EX}}) < \infty$, and Proposition 2 applied to Q_{EX} yields that $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}), \tilde{\pi}(h^2/\varrho_{\text{EX}}^2) < \infty$ and

$$\pi(h^2)\{1 + \text{IF}(h, Q_{\text{EX}})\} = \pi(\varrho_{\text{EX}})\tilde{\pi}(h^2/\varrho_{\text{EX}}^2)\{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}. \tag{A1}$$

Since the assumptions of Lemma 3 are satisfied, we can substitute (A1) into (11) to obtain

$$1 + \text{IF}(h, Q^*) = \pi_z(1/\varrho_z)\frac{\{1 + \text{IF}(h, Q_{\text{EX}})\}}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})}\left[1 + \text{IF}\{h/(\varrho_{\text{EX}}\varrho_z), \tilde{Q}^*\}\right]. \tag{A2}$$

We now provide a spectral representation for $\text{IF}\{h/(\varrho_{\text{EX}}\varrho_z), \tilde{Q}^*\}$. With $\tilde{\pi} \otimes \tilde{\pi}_z(\mathrm{d}\theta, \mathrm{d}z) = \tilde{\pi}(\mathrm{d}\theta)\tilde{\pi}_z(\mathrm{d}z)$,

$$\begin{aligned} \text{IF}\{h/(\varrho_{\text{EX}}\varrho_z), \tilde{Q}^*\} &= 1 + 2\sum_{n=1}^{\infty} \frac{\langle \varrho_{\text{EX}}^{-1}\varrho_z^{-1}h, (\tilde{Q}^*)^n \varrho_{\text{EX}}^{-1}\varrho_z^{-1}h \rangle_{\tilde{\pi} \otimes \tilde{\pi}_z}}{\tilde{\pi} \otimes \tilde{\pi}_z(\varrho_z^{-2}\varrho_{\text{EX}}^{-2}h^2)} \\ &= 1 + 2\sum_{n=1}^{\infty} \frac{\langle \varrho_z^{-1}, (\tilde{Q}_z)^n \varrho_z^{-1}\rangle_{\tilde{\pi}_z}\langle \varrho_{\text{EX}}^{-1}h, (\tilde{Q}_{\text{EX}})^n \varrho_{\text{EX}}^{-1}h \rangle_{\tilde{\pi}}}{\tilde{\pi}_z(\varrho_z^{-2})\tilde{\pi}(\varrho_{\text{EX}}^{-2}h^2)} \end{aligned} \tag{A3}$$

and, as $\tilde{Q}_z$ and $\tilde{Q}_{EX}$ are reversible, the following spectral representations, as in (8), hold:

$$\phi_n(1/\varrho_z, \tilde{Q}_z) = \frac{\langle \varrho_z^{-1}, (\tilde{Q}_z)^n \varrho_z^{-1} \rangle_{\tilde{\pi}_z} - \{\tilde{\pi}_z(\varrho_z^{-1})\}^2}{\mathbb{V}_{\tilde{\pi}_z}(\varrho_z^{-1})} = \int_{-1}^{1} \lambda^n \tilde{e}_z(d\lambda),$$

$$\phi_n(h/\varrho_{EX}, \tilde{Q}_{EX}) = \frac{\langle \varrho_{EX}^{-1} h, (\tilde{Q}_{EX})^n \varrho_{EX}^{-1} h \rangle_{\tilde{\pi}}}{\tilde{\pi}(\varrho_{EX}^{-2} h^2)} = \int_{-1}^{1} \omega^n \tilde{e}_{EX}(d\omega),$$

where, to simplify notation, we define

$$\mathbb{V}_{\tilde{\pi}_z}(\varrho_z^{-1}) = \tilde{\pi}_z\big[\{\varrho_z^{-1} - \tilde{\pi}_z(\varrho_z^{-1})\}^2\big], \quad \tilde{e}_z(d\lambda) = e(\varrho_z^{-1}, \tilde{Q}_z)(d\lambda), \quad \tilde{e}_{EX}(d\omega) = e(\varrho_{EX}^{-1} h, \tilde{Q}_{EX})(d\omega).$$

Using $\tilde{\pi}_z(\varrho_z^{-1}) = 1/\pi_z(\varrho_z)$, we can rewrite $\text{IF}\{h/(\varrho_{EX}\varrho_z), \tilde{Q}^*\}$ in (A3) as

$$1 + 2\sum_{n=1}^{\infty} \frac{1}{\tilde{\pi}_z(\varrho_z^{-2})} \left\{ \mathbb{V}_{\tilde{\pi}_z}(\varrho_z^{-1}) \int \lambda^n \tilde{e}_z(d\lambda) + \frac{1}{\pi_z(\varrho_z)^2} \right\} \int \omega^n \tilde{e}_{EX}(d\omega)$$

$$= 1 + 2(1-\gamma) \int \frac{\omega}{1-\omega} \tilde{e}_{EX}(d\omega) + 2\gamma \iint \frac{\lambda\omega}{1-\lambda\omega} \tilde{e}_z(d\lambda)\tilde{e}_{EX}(d\omega)$$

$$= -1 + 2(1-\gamma) \int \left(1 + \frac{\omega}{1-\omega}\right) \tilde{e}_{EX}(d\omega) + 2\gamma \iint \left(1 + \frac{\omega\lambda}{1-\lambda\omega}\right) \tilde{e}_z(d\lambda)\tilde{e}_{EX}(d\omega), \qquad \text{(A4)}$$

where the second expression is finite since $\int (1+\omega)(1-\omega)^{-1}\tilde{e}_{EX}(d\omega) = \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX}) < \infty$ and

$$\gamma = \mathbb{V}_{\tilde{\pi}_z}(\varrho_z^{-1})/\tilde{\pi}_z(\varrho_z^{-2}) = \big\{\pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z)\big\}/\pi_z(\varrho_z^{-1}). \qquad \text{(A5)}$$

Upon rearranging (A4), we have

$$1 + \text{IF}\{h/(\varrho_{EX}\varrho_z), \tilde{Q}^*\} = \{1 + \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX})\}(1-\gamma) + \gamma\beta$$

$$= \frac{1 + \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX})}{\pi_z(\varrho_z)\pi_z(\varrho_z^{-1})} + \left\{\frac{\pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z)}{\pi_z(\varrho_z^{-1})}\right\}\beta, \qquad \text{(A6)}$$

with

$$\frac{\beta}{2} = \iint \frac{\tilde{e}_z(d\lambda)\tilde{e}_{EX}(d\omega)}{1-\omega\lambda} = \sum_{n=0}^{\infty} \phi_n(h/\varrho_{EX}, \tilde{Q}_{EX})\phi_n(1/\varrho_z, \tilde{Q}_z). \qquad \text{(A7)}$$

By substituting (A6) into (A2) and using that

$$\text{IF}(h, Q^*) = \pi_z(1/\varrho_z)\frac{1 + \text{IF}(h, Q^{EX})}{1 + \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX})}\left[1 + \text{IF}\{h/(\varrho_{EX}\varrho_z), \tilde{Q}^*\}\right] - 1$$

$$= \frac{1 + \text{IF}(h, Q_{EX})}{1 + \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX})}\left\{\pi_z(\varrho_z^{-1}) - \frac{1}{\pi_z(\varrho_z)}\right\}\beta + \frac{1 + \text{IF}(h, Q_{EX})}{\pi_z(\varrho_z)} - 1, \qquad \text{(A8)}$$

we obtain the result. □

Proof of Corollary 1. Upon dividing (A8) by $\text{IF}(h, Q_{EX})$, we obtain

$$\text{RIF}(h, Q^*) = \frac{\pi_z(\varrho_z^{-1})\{1 + \text{IF}(h, Q_{EX})\}}{\text{IF}(h, Q_{EX})\{1 + \text{IF}(h/\varrho_{EX}, \tilde{Q}_{EX})\}}A - \frac{1}{\text{IF}(h, Q_{EX})}, \qquad \text{(A9)}$$

where A is the quantity in (A6) and can be expressed in terms of γ, defined in (A5), as

$$A = 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - 2\gamma \iint \left\{ \frac{1}{(1-\omega)} - \frac{1}{(1-\lambda\omega)} \right\} \tilde{e}_z(d\lambda)\tilde{e}_{\text{EX}}(d\omega)$$

$$= 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - 2\gamma \iint \frac{\omega(1-\lambda)}{(1-\omega)(1-\omega\lambda)} \tilde{e}_z(d\lambda)\tilde{e}_{\text{EX}}(d\omega).$$

Lemma 3 ensures that the kernel $\tilde{Q}_z$ is positive, implying that $\tilde{e}_z\{[0,1)\} = 1$. Hence,

$$\iint \left\{ \frac{\omega(1-\lambda)}{(1-\omega)(1-\omega\lambda)} - \frac{\omega(1-\lambda)}{(1-\omega)} \right\} \tilde{e}_z(d\lambda)\tilde{e}_{\text{EX}}(d\omega) = \iint \frac{\omega^2(1-\lambda)\lambda}{(1-\omega)(1-\omega\lambda)} \tilde{e}_z(d\lambda)\tilde{e}_{\text{EX}}(d\omega) \geqslant 0.$$

We can now bound A from above as follows:

$$A \leqslant 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - 2\gamma \iint \frac{\omega(1-\lambda)}{(1-\omega)} \tilde{e}_z(d\lambda)\tilde{e}_{\text{EX}}(d\omega)$$

$$= 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - \gamma \left\{ 1 - \int \lambda\tilde{e}_z(d\lambda) \right\} \int \frac{2\omega}{(1-\omega)} \tilde{e}_{\text{EX}}(d\omega)$$

$$= 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - \gamma(1-\phi_z)\{\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) - 1\}$$

$$= \{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}\{\phi_z + (1-\phi_z)(1-\gamma)\} + 2(1-\phi_z)\gamma$$

$$\leqslant \{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}\{\phi_z + (1-\phi_z)(1-\gamma) + 2(1-\phi_z)\gamma\}$$

$$= \{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\} \left\{ 2(1-\phi_z/2) - \frac{(1-\phi_z)}{\pi_z(1/\varrho_z)\pi_z(\varrho_z)} \right\}, \tag{A10}$$

where we have used the identity $\phi_z = \int \lambda\tilde{e}_z(d\lambda)$. The last inequality is established by noting that $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$ and γ are nonnegative. Substituting the expression into (A9) establishes statement (i) of the corollary. By observing that if $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}) \geqslant 1$, then (A10) is bounded above by

$$\{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\} \left[\phi_z + \frac{(1-\phi_z)}{\pi_z(1/\varrho_z)\pi_z(\varrho_z)} + (1-\phi_z) \left\{ 1 - \frac{1}{\pi_z(1/\varrho_z)\pi_z(\varrho_z)} \right\} \right]$$

$$= 1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}}),$$

we prove the inequality in (ii). $\qquad\square$

Proof of Corollary 2. We establish the upper bound $\text{URIF}_3(h)$ in (i) by noting that (A10) implies

$$A \leqslant \{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}\{\phi_z + (1-\phi_z)(1-\gamma)\} + 2(1-\phi_z)\gamma,$$

where A is the quantity in (A6) and γ is given by (A5). Upon substituting into (A9), we obtain

$$\text{RIF}(h, Q^*) + \frac{1}{\text{IF}(h, Q_{\text{EX}})} \leqslant \frac{\{1 + \text{IF}(h, Q_{\text{EX}})\}}{\text{IF}(h, Q_{\text{EX}})} \left\{ \phi_z \pi_z(\varrho_z^{-1}) + \frac{(1-\phi_z)}{\pi_z(\varrho_z)} \right\}$$

$$+ \frac{2(1-\phi_z)\{1 + \text{IF}(h, Q_{\text{EX}})\}}{\text{IF}(h, Q_{\text{EX}})\{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})\}} \left\{ \pi_z(\varrho_z^{-1}) - \frac{1}{\pi_z(\varrho_z)} \right\},$$

and, after further manipulations, we have

$$\text{RIF}(h, Q^*) \leqslant \phi_z\{\pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z)\} + 1/\pi_z(\varrho_z)$$

$$+ \frac{1}{\text{IF}(h, Q_{\text{EX}})} \left[\phi_z\{\pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z)\} + \frac{1}{\pi_z(\varrho_z)} - 1 \right]$$

$$+ 2\frac{\{1 + 1/\text{IF}(h, Q_{\text{EX}})\}}{1 + \text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})}\{\pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z)\}(1-\phi_z)$$

$$\leqslant \phi_z\{\pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z)\} + 1/\pi_z(\varrho_z)$$
$$+ \frac{1}{\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})}\left[\phi_z\{\pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z)\} + \frac{1}{\pi_z(\varrho_z)} - 1\right]$$
$$+ \frac{2}{\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})}\{\pi_z(1/\varrho_z) - 1/\pi_z(\varrho_z)\}(1 - \phi_z),$$

as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) \leqslant \mathrm{IF}(h, Q_{\mathrm{EX}})$ from Proposition 2.

To establish the upper bound $\mathrm{URIF}_4(h)$ in (ii), we use that, in the right-hand side of the equality (A9), the term β defined in (A7) and appearing in A satisfies the inequality

$$\beta = \iint \frac{2}{(1 - \lambda\omega)}\,\tilde{e}_z(\mathrm{d}\lambda)\tilde{e}_{\mathrm{EX}}(\mathrm{d}\omega) \leqslant \int \frac{2}{(1 - \lambda)}\,\tilde{e}_z(\mathrm{d}\lambda) = 1 + \mathrm{IF}(1/\varrho_z, \tilde{Q}_z), \qquad (\mathrm{A}11)$$

where $\mathrm{IF}(1/\varrho_z, \tilde{Q}_z) = \int(1 + \lambda)/(1 - \lambda)\tilde{e}_z(\mathrm{d}\lambda) < \infty$ by assumption. Therefore, upon substituting into (A9), we obtain

$$\mathrm{RIF}(h, Q^*) \leqslant \frac{\pi_z(\varrho_z^{-1})\{1 + \mathrm{IF}(h, Q_{\mathrm{EX}})\}}{\mathrm{IF}(h, Q_{\mathrm{EX}})\{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})\}}\left\{1 - \frac{1}{\pi_z(\varrho_z^{-1})\pi_z(\varrho_z)}\right\}\{1 + \mathrm{IF}(1/\varrho_z, \tilde{Q}_z)\}$$
$$+ \frac{\{1 + \mathrm{IF}(h, Q_{\mathrm{EX}})\}}{\mathrm{IF}(h, Q_{\mathrm{EX}})}\frac{1}{\pi_z(\varrho_z)} - \frac{1}{\mathrm{IF}(h, Q_{\mathrm{EX}})}$$
$$= \frac{1}{\pi_z(\varrho_z)} + \frac{\{1 + 1/\mathrm{IF}(h, Q_{\mathrm{EX}})\}}{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})}\{\pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z)\}\{1 + \mathrm{IF}(1/\varrho_z, \tilde{Q}_z)\}$$
$$+ \frac{1}{\mathrm{IF}(h, Q_{\mathrm{EX}})}\left\{\frac{1}{\pi_z(\varrho_z)} - 1\right\}$$
$$\leqslant \frac{1}{\pi_z(\varrho_z)} + \frac{\{1 + 1/\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})\}}{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})}\{\pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z)\}\{1 + \mathrm{IF}(1/\varrho_z, \tilde{Q}_z)\}$$
$$+ \frac{1}{\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})}\left\{\frac{1}{\pi_z(\varrho_z)} - 1\right\},$$

since $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) \leqslant \mathrm{IF}(h, Q_{\mathrm{EX}})$.

To establish the inequality in (iii), we combine (A6) and (A9) to obtain

$$\mathrm{RIF}(h, Q^*) = \frac{\pi_z(\varrho_z^{-1})\{1 + 1/\mathrm{IF}(h, Q_{\mathrm{EX}})\}}{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})}\gamma\beta + (1 - \gamma)\pi_z(\varrho_z^{-1})\{1 + 1/\mathrm{IF}(h, Q_{\mathrm{EX}})\} - \frac{1}{\mathrm{IF}(h, Q_{\mathrm{EX}})}$$
$$= \frac{1}{\pi_z(\varrho_z)} + \frac{\{1 + 1/\mathrm{IF}(h, Q_{\mathrm{EX}})\}}{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})}\{\pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z)\}\beta + \frac{\{1/\pi_z(\varrho_z) - 1\}}{\mathrm{IF}(h, Q_{\mathrm{EX}})}$$
$$\geqslant \frac{1}{\pi_z(\varrho_z)} + \frac{2\{1 + 1/\mathrm{IF}(h, Q_{\mathrm{EX}})\}}{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})}\{\pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z)\} + \frac{\{1/\pi_z(\varrho_z) - 1\}}{\mathrm{IF}(h, Q_{\mathrm{EX}})} \qquad (\mathrm{A}12)$$
$$\geqslant \frac{1}{\pi_z(\varrho_z)} + \frac{2}{1 + \mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})}\{\pi_z(\varrho_z^{-1}) - 1/\pi_z(\varrho_z)\}.$$

The first inequality follows because the identity for β given in (A11) shows that $\beta \geqslant 2$ when $\tilde{Q}_{\mathrm{EX}}$ is positive. The second inequality follows from $\mathrm{IF}(h, Q_{\mathrm{EX}}) \geqslant 0$.

From (A9), we have that $\mathrm{RIF}(h, Q^*) \geqslant 1/\pi_z(\varrho_z)$, as the second and third terms on the left-hand side of the inequality (A12) are both positive. This establishes the inequality of (iv). We examine the limit of $\mathrm{RIF}(h, Q^*)$ as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) \to \infty$, again noting that $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) \leqslant \mathrm{IF}(h, Q_{\mathrm{EX}})$. Using the inequality for β in (A11) and the fact that $\mathrm{IF}(1/\varrho_z, \tilde{Q}_z) < \infty$ by Lemma 3, we obtain the limiting form, as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}}) \to \infty$, given by (12) for $\mathrm{RIF}(h, Q^*)$. □

Table A1. *Sandwiching results for different values of* $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$, *based on the upper bounds for* $\mathrm{RCT}(h, Q^*; \sigma)$ *given by* $\mathrm{URCT}_3(h; \sigma)$ *and* $\mathrm{URCT}_4(h; \sigma)$ *and on the lower bound* $\mathrm{LRCT}_1(h; \sigma)$

$\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$	1	10	25	100	1000
$\mathrm{RCT}(h, Q^*; \sigma_{\mathrm{opt}})$	(3·201, 5·327)	(2·020, 2·256)	(1·773, 1·876)	(1·595, 1·625)	(1·518, 1·522)
σ_{opt}	(0·548, 1·572)	(1·018, 1·598)	(1·205, 1·658)	(1·421, 1·730)	(1·607, 1·730)

APPENDIX 2

We exploit the two upper bounds $\mathrm{URCT}_3(h; \sigma)$ and $\mathrm{URCT}_4(h; \sigma)$, together with the lower bound $\mathrm{LRCT}_1(h; \sigma)$, in order to find an interval where the optimal value σ_{opt} for $\mathrm{RCT}(h, Q^*; \sigma)$ lies. We consider how this interval varies as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$ increases. To do this, we compute the interval where $\mathrm{LRCT}_1(h; \sigma)$ lies below the minimum of $\inf_\sigma \mathrm{URCT}_3(h; \sigma)$ and $\inf_\sigma \mathrm{URCT}_4(h; \sigma)$. Table A1 displays this interval together with the minimum of the two upper bounds and the minimum of the lower bound. It is straightforward to see that σ_{opt} is contained in this interval and $\mathrm{RCT}(h, Q^*; \sigma_{\mathrm{opt}})$ is contained in the corresponding interval in Table A1. It is apparent that the intervals tighten as $\mathrm{IF}(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$ increases. Similarly, the endpoints of the interval containing $\mathrm{RCT}(h, Q^*; \sigma_{\mathrm{opt}})$ both decrease, whereas the lower endpoint of the interval containing σ_{opt} increases. A degree of caution should be exercised, however, as this argument applies to Q^* rather than Q.

REFERENCES

ANDRIEU, C., DOUCET, A. & HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods (with Discussion). *J. R. Statist. Soc.* B **72**, 269–342.

ANDRIEU, C. & ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37**, 697–725.

ANDRIEU, C. & VIHOLA, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Prob.* **25**, 1030–77.

BAXENDALE, P. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Prob.* **15**, 700–38.

BEAUMONT, M. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164**, 1139–60.

BÉRARD, J., DEL MORAL, P. & DOUCET, A. (2014). A lognormal central limit theorem for particle approximations of normalizing constants. *Electron. J. Prob.* **19**, 1–28.

CHERNOV, M., GALLANT, A. R., GHYSELS, E. & TAUCHEN, G. (2003). Alternative models of stock price dynamics. *J. Economet.* **116**, 225–57.

DOUC, R. & ROBERT, C. P. (2011). A vanilla Rao–Blackwellization of Metropolis–Hastings algorithms. *Ann. Statist.* **39**, 261–77.

GEYER, C. J. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* **7**, 473–83.

HÄGGSTRÖM, O. & ROSENTHAL, J. S. (2007). On variance conditions for Markov chain central limit theorems. *Electron. Commun. Prob.* **12**, 454–64.

HUANG, X. & TAUCHEN, G. (2005). The relative contribution of jumps to total price variation. *J. Finan. Economet.* **3**, 456–99.

KIPNIS, C. & VARADHAN, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Commun. Math. Phys.* **104**, 1–19.

LIN, L., LIU, K. F. & SLOAN, J. (2000). A noisy Monte Carlo algorithm. *Phys. Rev.* D **61**, article no. 074505.

PESKUN, P. H. (1973). Optimum Monte–Carlo sampling using Markov chains. *Biometrika* **60**, 607–12.

PITT, M. K., SILVA, R., GIORDANI, P. & KOHN, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *J. Economet.* **171**, 134–51.

ROBERTS, G. O. & TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83**, 95–110.

RUDOLF, D. & ULLRICH, M. (2013). Positivity of hit-and-run and related algorithms. *Electron. Commun. Prob.* **18**, 1–8.

SHERLOCK, C., THIERY, A. H., ROBERTS, G. O. & ROSENTHAL, J. S. (2015). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* **43**, 238–75.

TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with Discussion). *Ann. Statist.* **21**, 1701–62.

TIERNEY, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Prob.* **8**, 1–9.

Supplementary material for
'Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator'

BY A. DOUCET

Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, U.K.
doucet@stats.ox.ac.uk

M. K. PITT

Department of Economics, University of Warwick, Coventry CV4 7AL, U.K.
m.pitt@warwick.ac.uk

G. DELIGIANNIDIS

Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, U.K.
deligiannidis@stats.ox.ac.uk

AND R. KOHN

Australian School of Business, University of New South Wales, Sydney 2052, Australia
r.kohn@unsw.edu.au

SUMMARY

This supplement provides some technical proofs and an additional example. Section S1 presents the proof of Proposition 2. Section S2 presents the proofs of Propositions 3 and 4 and Lemmas 1 and 3. Section S3 contains some auxiliary technical results. Section S4 compares the two lower bounds, $\mathrm{LRCT}_1(h; \sigma)$ and $\mathrm{LRCT}_2(\sigma)$, defined in § 3·5 of the main article. Section S5 illustrates the upper bound on the inefficiency in Corollary 2(iv) and compares it with the results in Sherlock et al. (2014). In § S6 we apply the pseudo-marginal algorithm to a linear Gaussian state-space model and present additional simulation results for the stochastic volatility model discussed in the main paper. Section S7 explains how the bounds on the inefficiency introduced in § 3·5 of the main paper are computed.

All code was implemented in the $\mathtt{Ox}$ language with pre-compiled C code for computationally intensive routines.

S1. PROOF OF PROPOSITION 2

The proof of Proposition 2 relies on Lemmas 5–8 given below. Lemmas 5–7 establish that $h/\varrho \in L^2(\mathsf{X}, \tilde{\mu})$ and $\mathrm{IF}(h/\varrho, \tilde{P}) < \infty$ whenever $\mathrm{IF}(h, P) < \infty$. To prove this result, we define the map that sends the functional h to h/ϱ as a linear operator between two Hilbert spaces, $\mathcal{H}$ and $\tilde{\mathcal{H}}$ defined below. The space $\mathcal{H}$ (respectively, $\tilde{\mathcal{H}}$) corresponds to the set of functions that have finite inefficiencies under P (respectively, under $\tilde{P}$). We then exploit the structure of the Metropolis–Hastings-type kernel P to prove that this linear operator is bounded on a dense subspace $\mathcal{H}_P \subset \mathcal{H}$, which allows us to extend the operator to $\mathcal{H}$. The proof is then completed by

checking that the unique extension constructed in this way is the one required. Lemma 8 is a general result on the central limit theorem for reversible and ergodic Markov chains which are not started in their stationary regime. The proof of Proposition 2 uses these preliminary results to establish the identity of interest.

In the notation of Proposition 2, we write $\| \cdot \|_\mu$ and $\langle \cdot, \cdot \rangle_\mu$ for the norm and inner product of $L^2(\mathsf{X}, \mu)$, respectively, with a similar notation for $L^2(\mathsf{X}, \tilde{\mu})$. By reversibility of P and $\tilde{P}$ with respect to μ and $\tilde{\mu}$, respectively, it is easy to check that $I - P$ and $I - \tilde{P}$ are positive self-adjoint operators on $L^2(\mathsf{X}, \mu)$ and $L^2(\mathsf{X}, \tilde{\mu})$, respectively. By Theorem 13.11 in Rudin (1991), the inverses $(I - P)^{-1}$ and $(I - \tilde{P})^{-1}$ are densely defined and self-adjoint. They are also positive, because for any $f \in \mathrm{Domain}\{(I - P)^{-1}\}$ there exists a function g such that $f = (I - P)g$, and thus

$$\langle (I - P)^{-1} f, f \rangle_\mu = \langle (I - P)^{-1}(I - P)g, (I - P)g \rangle_\mu = \langle g, (I - P)g \rangle_\mu \geq 0$$

since $I - P$ is positive. Therefore, by Theorem 13.31 in Rudin (1991), there exists a unique self-adjoint positive operator $(I - P)^{-1/2}$ such that $(I - P)^{-1} = (I - P)^{-1/2}(I - P)^{-1/2}$. Finally, since $(I - P)^{-1}$ is densely defined, so is $(I - P)^{-1/2}$. Similar considerations show the existence and uniqueness of the positive self-adjoint operator $(I - \tilde{P})^{-1/2}$, which is densely defined on $L^2(\mathsf{X}, \tilde{\mu})$.

We now introduce the inner product spaces $(\mathcal{H}, \langle \cdot, \cdot \rangle_\mathcal{H})$ and $(\tilde{\mathcal{H}}, \langle \cdot, \cdot \rangle_{\tilde{\mathcal{H}}})$, where

$$\mathcal{H} = \{f \in L_0^2(\mathsf{X}, \mu) : \|f\|_\mu^2 + \|(I - P)^{-1/2}f\|_\mu^2 < \infty\},$$

$$\langle f, g \rangle_\mathcal{H} = \langle f, g \rangle_\mu + \langle (I - P)^{-1/2}f, (I - P)^{-1/2}g \rangle_\mu,$$

$$\tilde{\mathcal{H}} = \{f \in L_0^2(\mathsf{X}, \tilde{\mu}) : \|f\|_{\tilde{\mu}}^2 + \|(I - \tilde{P})^{-1/2}f\|_{\tilde{\mu}}^2 < \infty\},$$

$$\langle f, g \rangle_{\tilde{\mathcal{H}}} = \langle f, g \rangle_{\tilde{\mu}} + \langle (I - \tilde{P})^{-1/2}f, (I - \tilde{P})^{-1/2}g \rangle_{\tilde{\mu}}.$$

Clearly, the space $\mathcal{H}$ corresponds to the set of functions having finite inefficiencies under P, and $\tilde{\mathcal{H}}$ corresponds to the set of functions having finite inefficiencies under $\tilde{P}$.

LEMMA 5. *Let P and $\tilde{P}$ be ergodic. Then $(\mathcal{H}, \langle \cdot, \cdot \rangle_\mathcal{H})$ and $(\tilde{\mathcal{H}}, \langle \cdot, \cdot \rangle_{\tilde{\mathcal{H}}})$ are Hilbert spaces.*

Proof. Since P and $\tilde{P}$ are ergodic, the only solutions in $L^2(\mathsf{X}, \mu)$ and $L^2(\mathsf{X}, \tilde{\mu})$ of $h = Ph$ and $g = \tilde{P}g$, respectively, are almost surely constant with respect to μ and $\tilde{\mu}$. If $f = Pf$ μ-almost surely, then

$$0 = \|f - Pf\|_\mu^2 = \int_{-1}^1 (1 - \lambda)^2 e(f, P)(\mathrm{d}\lambda),$$

where $e(f, P)$ is the spectral measure of P with respect to the function f. Therefore, $e(f, P)$ must be an atom at 1, which is impossible as P is ergodic; see the proof of Lemma 17 in Häggström & Rosenthal (2007) and Proposition 17.4.1 in Meyn & Tweedie (2009). Since $I - P$ and $I - \tilde{P}$ are injective in $L_0^2(\mu)$ and $L_0^2(\tilde{\mu})$, respectively, $(I - P)^{1/2}$ and $(I - \tilde{P})^{1/2}$ must also be injective on the corresponding spaces, because $(I - P)^{1/2}h = 0$ implies $(I - P)h = 0$. In addition, as mentioned above, these operators are self-adjoint and so their inverses, $(I - P)^{-1/2}$ and $(I - \tilde{P})^{-1/2}$, are densely defined and self-adjoint by Theorem 13.11 in Rudin (1991).

By Theorem 13.9 in Rudin (1991), $(I - P)^{-1/2}$ and $(I - \tilde{P})^{-1/2}$ are closed operators on $L_0^2(\mathsf{X}, \mu)$ and $L_0^2(\mathsf{X}, \tilde{\mu})$, respectively, because they are self-adjoint. By Rudin (1991, § 13.1), a possibly unbounded operator T on a Hilbert space $\mathcal{F}$ is said to be closed if and only if its graph

$$\mathfrak{G}(T) = \{(x, Tx) : x \in \mathcal{F}\}$$

is a closed subset of $\mathcal{F} \times \mathcal{F}$; equivalently, T is closed if $x_n \to x$ and $Tx_n \to y$ together imply $Tx = y$. In particular, x is in the domain of T. It follows that $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ and $(\tilde{\mathcal{H}}, \langle \cdot, \cdot \rangle_{\tilde{\mathcal{H}}})$ are Hilbert spaces by Proposition 1.4 in Schmüdgen (2012). □

LEMMA 6. *The linear space*

$$\mathcal{H}_P = \text{Range}\{(I - P)\} = \{h \in L_0^2(\mathsf{X}, \mu) : h = (I - P)g, \ g \in L^2(\mathsf{X}, \mu)\}$$

is dense in $\mathcal{H}$ with respect to the norm induced by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.

Proof. For $h \in \mathcal{H}$, we have

$$\|(I - P)^{-1/2}h\|_\mu = \int_{-1}^1 \frac{e(h, P)(\mathrm{d}\lambda)}{1 - \lambda} < \infty,$$

where $e(h, P)$ is the spectral measure associated with h and P. For $\epsilon > 0$, define

$$h_\epsilon = (I - P)\{(1 + \epsilon)I - P\}^{-1}h \in \mathcal{H}_P.$$

Then

$$\left\|(I - P)^{-1/2}(h_\epsilon - h)\right\|_\mu^2 = \left\|(I - P)^{-1/2}\big[(I - P)\{(1 + \epsilon)I - P\}^{-1} - I\big]h\right\|_\mu^2$$

$$= \int_{-1}^1 \frac{1}{1 - \lambda}\left(\frac{1 - \lambda}{1 + \epsilon - \lambda} - 1\right)^2 e(h, P)(\mathrm{d}\lambda)$$

$$= \int_{-1}^1 (1 - \lambda)\left(\frac{1}{1 + \epsilon - \lambda} - \frac{1}{1 - \lambda}\right)^2 e(h, P)(\mathrm{d}\lambda)$$

$$= \int_{-1}^1 \frac{\epsilon^2 e(h, P)(\mathrm{d}\lambda)}{(1 + \epsilon - \lambda)^2(1 - \lambda)}.$$

The integrand is bounded above by $1/(1 - \lambda)$, since $|\lambda| \le 1$ implies $\epsilon^2/(1 + \epsilon - \lambda)^2 \le 1$; thus, by the dominated convergence theorem, the integral vanishes as $\epsilon \to 0$. Since $I - P$ is bounded, $\|h_\epsilon - h\|_\mu$ also vanishes. Therefore, $h_\epsilon \to h$ in $\mathcal{H}$. In particular, $\mathcal{H}_P$ is dense in $\mathcal{H}$. □

LEMMA 7. *If $\text{IF}(h, P) < \infty$, then $h/\varrho \in L^2(\mathsf{X}, \tilde{\mu})$ and $\text{IF}(h/\varrho, \tilde{P}) < \infty$.*

Proof. For $h \in \mathcal{H}_P$, there exists $g \in L^2(\mathsf{X}, \mu)$ such that

$$h(x) = (I - P)g(x) = \varrho(x)(I - \tilde{P})g(x).$$

Therefore, $h(x)/\varrho(x) = (I - \tilde{P})g(x) \in \tilde{\mathcal{H}}$ since $\|g\|_{\tilde{\mu}}^2 \le \|g\|_\mu^2/\mu(\varrho)$. Thus, we can define the multiplication operator $T : \mathcal{H}_P \to \tilde{\mathcal{H}}$ by $T : h \to h/\varrho$.

Let $h(x) = (I - P)g(x)$. Then

$$\|h\|_{\mathcal{H}}^2 = \|h\|_\mu^2 + \langle h, (I - P)^{-1}(I - P)g\rangle_\mu \ge \langle h, g\rangle_\mu,$$

because $I - P$ is self-adjoint. Similarly,

$$\|Th\|_{\tilde{\mathcal{H}}}^2 = \|h/\varrho\|_{\tilde{\mu}}^2 + \|(I - \tilde{P})^{-1/2}(h/\varrho)\|_{\tilde{\mu}}^2$$

$$= \|(I - \tilde{P})g\|_{\tilde{\mu}}^2 + \|(I - \tilde{P})^{1/2}g\|_{\tilde{\mu}}^2 \le K\|(I - \tilde{P})^{1/2}g\|_{\tilde{\mu}}^2,$$

where $K = 1 + \|(I - \tilde{P})^{1/2}\|$ with $\|(I - \tilde{P})^{1/2}\|$ being the finite norm of the operator $(I - \tilde{P})^{1/2}$. Recalling that $h(x) = (I - P)g(x) = \varrho(x)(I - \tilde{P})g(x)$, we obtain

$$\|(I - \tilde{P})^{1/2}g\|_{\tilde{\mu}}^2 = \int g(x)(I - \tilde{P})g(x)\frac{\varrho(x)\mu(\mathrm{d}x)}{\mu(\varrho)} = \frac{\langle g, h\rangle_\mu}{\mu(\varrho)}.$$

It follows that $T : \mathcal{H}_P \to \tilde{\mathcal{H}}$ is bounded, as

$$\sup_{h \in \mathcal{H}_P} \frac{\|Th\|_{\tilde{\mathcal{H}}}^2}{\|h\|_{\mathcal{H}}^2} \leq \frac{K\|(I - \tilde{P})^{1/2}g\|_{\tilde{\mu}}^2}{\|h\|_{\mu}^2 + \langle h, g \rangle_{\mu}} = \frac{K}{\mu(\varrho)} \frac{\langle g, h \rangle_{\mu}}{\|h\|_{\mu}^2 + \langle g, h \rangle_{\mu}} \leq \frac{K}{\mu(\varrho)}.$$

Since $\mathcal{H}_P$ is dense, given $h \in \mathcal{H}$, there is a sequence $h_n \in \mathcal{H}_P$ such that $\|h_n - h\|_{\mathcal{H}} \to 0$ as $n \to \infty$. This implies that h_n is a Cauchy sequence in $\mathcal{H}$, i.e.,

$$\|h_n - h_m\|_{\mathcal{H}} \to 0$$

as $n \geq m \to \infty$. Since h_n and $h_n - h_m$ are in $\mathcal{H}_P$, we have that $Th_n, T(h_n - h_m) \in \tilde{\mathcal{H}}$ and, from the above calculation,

$$\|Th_n - Th_m\|_{\tilde{\mathcal{H}}} \leq \frac{K}{\mu(\varrho)} \|h_n - h_m\|_{\mathcal{H}} \to 0$$

as $m, n \to \infty$. Therefore, Th_n forms a Cauchy sequence in $\tilde{\mathcal{H}}$; in particular, h_n and $(I - \tilde{P})^{-1/2}h_n$ are Cauchy in $L^2(\mathsf{X}, \tilde{\mu})$. Since $L^2(\mathsf{X}, \tilde{\mu})$ is complete, we have that $h_n \to g \in L^2(\mathsf{X}, \tilde{\mu})$ and $(I - \tilde{P})^{-1/2}h_n \to f \in L^2(\mathsf{X}, \tilde{\mu})$. Since $Q = (I - \tilde{P})^{-1/2}$ is a closed operator, we can conclude that

$$g \in \text{Domain}\{Q\}, \quad Qg = f$$

and, in particular, $g \in \tilde{\mathcal{H}}$.

To complete the proof, we need to show that $g = h/\varrho$. Recall that $h_n \to h$ in $\mathcal{H}$ implies $\|h_n - h\|_{\mu} \to 0$. We can then choose a subsequence $n(k)$ such that $h_{n(k)} \to h$ μ-almost surely. Since $\tilde{\mu}$ is absolutely continuous with respect to μ, we also have that $h_{n(k)}/\varrho \to h/\varrho$ $\tilde{\mu}$-almost surely.

In addition, we know that $Th_n = h_n/\varrho \to g$ in $\tilde{\mathcal{H}}$ and hence in $L^2(\mathsf{X}, \tilde{\mu})$. Therefore, $h_{n(k)}/\varrho \to g$ in $L^2(\mathsf{X}, \tilde{\mu})$. We can now choose a further subsequence $n'(k)$ such that $h_{n'(k)}/\varrho \to g$ $\tilde{\mu}$-almost surely. Since $h_{n(k)}/\varrho$ also converges to h/ϱ $\tilde{\mu}$-almost surely, and $n'(k)$ is a subsequence of $n(k)$, we conclude that $g = h/\varrho$ $\tilde{\mu}$-almost surely. $\qquad\square$

LEMMA 8. *Assume that Π is μ-reversible and ergodic, $h \in L_0^2(\mathsf{X}, \mu)$ and $\text{IF}(h, \Pi) < \infty$. Let $(X_i)_{i \geq 1}$ be a Markov chain that evolves according to Π. If $X_1 \sim \nu$, where ν is absolutely continuous with respect to μ, then as $n \to \infty$,*

$$n^{-1/2} \sum_{i=1}^{n} h(X_i) \to N\{0, \mu(h^2)\text{IF}(h, \Pi)\}.$$

Proof of Lemma 8. Let $e(h, \Pi)(\mathrm{d}\lambda)$ be the associated spectral measure and define $S_n = \sum_{i=1}^{n} h(X_i)$. Then

$$\frac{1}{n} E_{\mu}\{E(S_n \mid X_1)^2\} = \frac{1}{n} \int \mu(\mathrm{d}x) \left\{ \sum_{i=0}^{n-1} \Pi^i h(x) \right\}^2 = \frac{1}{n} \int_{-1}^{1} \left(\sum_{i=0}^{n-1} \lambda^i \right)^2 e(h, \Pi)(\mathrm{d}\lambda)$$

$$= \int_{-1}^{1} \left(\frac{1}{n} \sum_{i=0}^{n-1} \lambda^i \right) \frac{1 - \lambda^n}{1 - \lambda} e(h, \Pi)(\mathrm{d}\lambda) \to 0$$

as $n \to \infty$ by dominated convergence, since $\int (1 - \lambda)^{-1} e(h, \Pi)(\mathrm{d}\lambda) < \infty$ by assumption. Hence, equation (4) in Wu & Woodroofe (2004) holds with $\sigma_n^2 = E_{\mu}(S_n^2) \sim \sigma^2 n$, where $\sigma^2 = \mu(h^2)\text{IF}(h, \Pi)$. It is straightforward to check, by calculations similar to those above, that the solution to the approximate Poisson equation given in the proof of Theorem 1.3 in Kipnis &

Varadhan (1986),

$$
h_n(x) = \left\{ \left(1 + \frac{1}{n} \right) I - \Pi \right\}^{-1} h(x),
$$

satisfies equation (5) in Wu & Woodroofe (2004), while equation (1.10) in Kipnis & Varadhan (1986) shows that $H_n(x_0, x_1) = h_n(x_1) - \Pi h_n(x_0)$ converges in $L^2(\mathsf{X} \times \mathsf{X}, \mu \otimes \Pi)$. Therefore, the conditions of Corollary 2 in Wu & Woodroofe (2004) are satisfied, and so the lemma follows from their equation (10); see their comments after this equation. $\square$

Proof of Proposition 2. Let $(X_i)_{i \geq 1}$ be a Markov chain that evolves according to P, and let $(\tilde{X}_i, \tau_i)_{i \geq 1}$ be the associated jump chain representation that evolves according to $\bar{P}$, as defined in Lemma 1. We denote by $\mathrm{pr}_{\nu, \Pi}$ the law of a Markov chain with initial distribution ν and transition kernel Π. By Theorem 1.3 in Kipnis & Varadhan (1986), we have that under $\mathrm{pr}_{\mu, P}$,

$$
S_n = \sum_{i=1}^{n} h(X_i) = M_n + \xi_n, \tag{S1}
$$

where M_n is a square-integrable martingale with respect to the natural filtration of $(X_i)_{i \geq 1}$. We also have the following convergence in probability:

$$
n^{-1/2} \sup_{1 \leq i \leq n} |\xi_i| \xrightarrow{\mathrm{pr}_{\mu, P}} 0. \tag{S2}
$$

Define $T_n = \tau_1 + \cdots + \tau_n$. The kernel $\bar{P}$ is ergodic because $\tilde{P}$ is ergodic. Hence we have that $\mathrm{pr}_{\tilde{\mu}, \tilde{P}}$-almost surely,

$$
\frac{T_n}{n} \to \tilde{\mu}(1/\varrho) = \frac{1}{\mu(\varrho)}. \tag{S3}
$$

The above limit also holds $\mathrm{pr}_{\mu, \tilde{P}}$-almost surely, since $\mathrm{pr}_{\mu, \tilde{P}}$ is absolutely continuous with respect to $\mathrm{pr}_{\tilde{\mu}, \tilde{P}}$. We first show that

$$
\{n/\mu(\varrho)\}^{-1/2}(M_{T_n} - M_{\lfloor n/\mu(\varrho) \rfloor}) \xrightarrow{\mathrm{pr}_{\mu, P}} 0. \tag{S4}
$$

Let $\epsilon > 0$ be arbitrary and define the event

$$
A_n = \left\{ (1 - \epsilon)\frac{n}{\mu(\varrho)} \leq T_n < (1 + \epsilon)\frac{n}{\mu(\varrho)} \right\}.
$$

By (S3), we have $\mathrm{pr}_{\mu, P}(A_n) \to 1$. The following inequality holds on the event A_n:

$$
\left| M_{T_n} - M_{\lfloor n/\mu(\varrho) \rfloor} \right| \leq \left| M_{T_n} - M_{\lfloor (1-\epsilon)n/\mu(\varrho) \rfloor} \right| + \left| M_{\lfloor n/\mu(\varrho) \rfloor} - M_{\lfloor (1-\epsilon)n/\mu(\varrho) \rfloor} \right|
$$
$$
\leq 2 \sup_{1 \leq i \leq 2\lfloor \epsilon n/\mu(\varrho) \rfloor + 1} |\tilde{M}_i|,
$$

where $\tilde{M}_i := M_{\lfloor (1-\epsilon)n/\mu(\varrho) \rfloor + i} - M_{\lfloor (1-\epsilon)n/\mu(\varrho) \rfloor}$ is a square-integrable martingale with stationary increments. Thus, for any $\delta > 0$,

$$
\mathrm{pr}_{\mu, P}\left[\left| M_{T_n} - M_{\lfloor n/\mu(\varrho) \rfloor} \right| > \delta\{n/\mu(\varrho)\}^{1/2} \right]
$$
$$
\leq \mathrm{pr}_{\mu, P}\left(\left[\left| M_{T_n} - M_{\lfloor n/\mu(\varrho) \rfloor} \right| > \delta\{n/\mu(\varrho)\}^{1/2} \right] \cap A_n \right) + \mathrm{pr}_{\mu, P}(A_n^c)
$$

$$\leq \mathrm{pr}_{\mu,P}\left[2\sup_{1\leq i\leq 2\lfloor \epsilon n/\mu(\varrho)\rfloor+1}|\tilde{M}_i| > \delta\{n/\mu(\varrho)\}^{1/2}\right] + o(1)$$

$$\leq C_1 \frac{E_{\mu,P}\{\tilde{M}^2_{2\lfloor \epsilon n/\mu(\varrho)\rfloor+1}\}}{\delta^2 n/\mu(\varrho)} + o(1)$$

$$\leq C_2 \frac{\epsilon n/\mu(\varrho)}{\delta^2 n/\mu(\varrho)} + o(1) \leq C_3\epsilon/\delta^2 + o(1),$$

where $C_1, C_2, C_3 < \infty$. The third inequality follows from Doob's maximal inequality. The last inequality holds because, for any square-integrable martingale $(N_i)_{i\geq 1}$ with stationary increments, $E_{\mu,P}(N_n^2) = E_{\mu,P}(N_1^2)\,n$. This bound is uniform in n, and therefore

$$\limsup_{n\to\infty}\mathrm{pr}_{\mu,P}\left[\left|M_{T_n} - M_{\lfloor n/\mu(\varrho)\rfloor}\right| > \delta\{n/\mu(\varrho)\}^{1/2}\right] \leq \epsilon/\delta^2.$$

As $\epsilon > 0$ is arbitrary,

$$\lim_{n\to\infty}\mathrm{pr}_{\mu,P}\left[\left|M_{T_n} - M_{\lfloor n/\mu(\varrho)\rfloor}\right| > \delta\{n/\mu(\varrho)\}^{1/2}\right] = 0$$

for any $\delta > 0$, and therefore (S4) holds. Now, by Proposition 1, $n^{-1/2}S_n \to N\{0, \mu(h^2)\mathrm{IF}(h,P)\}$. By the asymptotic negligibility (S2) of ξ_n and (S4), we have by Slutsky's theorem that $\{n/\mu(\varrho)\}^{-1/2}M_{T_n} \to N\{0, \mu(h^2)\mathrm{IF}(h,P)\}$ or, equivalently, $n^{-1/2}M_{T_n} \to N\{0, \mu(h^2)\,\mathrm{IF}(h,P)/\mu(\varrho)\}$. Finally, note that for any $\delta > 0$,

$$\mathrm{pr}_{\mu,P}\left(|\xi_{T_n}| > \delta n^{1/2}\right) \leq \mathrm{pr}_{\mu,P}\left[\{|\xi_{T_n}| > \delta n^{1/2}\} \cap A_n\right] + \mathrm{pr}_{\mu,P}(A_n^{\mathrm{c}})$$

$$\leq \mathrm{pr}_{\mu,P}\left\{\sup_{1\leq i\leq \lfloor(1+\epsilon)n/\mu(\varrho)\rfloor}|\xi_i| > \delta n^{1/2}\right\} + o(1) \to 0$$

by (S2). Therefore, using (S1) and Slutsky's theorem, we have that $n^{-1/2}S_{T_n} \to N\{0, \mu(h^2)\mathrm{IF}(h,P)/\mu(\varrho)\}$ when $\tilde{X}_1 = X_1 \sim \mu$. However, this result also holds when $\tilde{X}_1 \sim \tilde{\mu}$, as established in Lemma 8. In particular, the asymptotic variance is the same. Moreover, $(\tilde{X}_i, \tau_i)_{i\geq 1}$ is reversible and ergodic, while Lemma 7 guarantees that $h/\varrho \in L_0^2(\mathsf{X}, \tilde{\mu})$ and $\mathrm{IF}(h/\varrho, \tilde{P}) < \infty$. Hence, Proposition 1 applied to $(\tilde{X}_i, \tau_i)_{i\geq 1}$ ensures that the asymptotic variance is also given by the integrated autocovariance time. Equating the two expressions, we obtain

$$\mu(h^2)\mathrm{IF}(h,P)/\mu(\varrho) = \bar{\mu}(\tau^2 h^2) + 2\sum_{n\geq 1}\langle \tau h, \bar{P}^n \tau h\rangle_{\bar{\mu}}$$

$$= \tilde{\mu}\left(\frac{2-\varrho}{\varrho^2}h^2\right) + 2\sum_{n\geq 1}\left\langle \frac{h}{\varrho}, \tilde{P}^n\frac{h}{\varrho}\right\rangle_{\tilde{\mu}}$$

$$= \tilde{\mu}(h^2/\varrho^2) + \tilde{\mu}(h^2/\varrho^2)\mathrm{IF}(h/\varrho, \tilde{P}) - \mu(h^2)/\mu(\varrho),$$

where the equality in the second line follows from the expressions of $\bar{\mu}$ and $\bar{P}$ given in Lemma 1 and the properties of the geometric distribution. This yields the equality in Proposition 2, which can also be written as

$$\frac{1 + \mathrm{IF}(h/\varrho, \tilde{P})}{1 + \mathrm{IF}(h,P)} = \frac{\mu(h^2)}{\mu(\varrho)\tilde{\mu}(h^2/\varrho^2)} = \frac{\mu(h^2)}{\mu(h^2/\varrho)} \leq 1$$

as $0 < \varrho \leq 1$, implying that $\mathrm{IF}(h/\varrho, \tilde{P}) \leq \mathrm{IF}(h,P)$. $\qquad\square$

S2. PROOFS OF SOME OTHER TECHNICAL RESULTS IN THE MAIN PAPER

Proof of Lemma 1. As P is ψ-irreducible, it is also μ-irreducible because it is μ-invariant; see, for example, Tierney (1994, p. 1759). Hence, for any $x \in \mathsf{X}$ and $A \in \mathcal{X}$ with $\mu(A) > 0$, there exists an $n \geq 1$ such that $P^n(x, A) > 0$. As μ is not concentrated on a single point by assumption, this implies that $\varrho(x) > 0$ for any $x \in \mathsf{X}$. The rest of the proposition follows directly from Lemma 1 in Douc & Robert (2011). □

Proof of Lemma 3. Equation (10) and the expressions for the associated invariant distributions follow from a direct application of Lemma 1. The positivity of $\tilde{Q}_{\mathsf{z}}$ follows directly from Proposition 3; see Remark 1. We write $\tilde{\pi} \otimes \tilde{\pi}_{\mathsf{z}}(\mathrm{d}\theta, \mathrm{d}z) = \tilde{\pi}(\mathrm{d}\theta)\tilde{\pi}_{\mathsf{z}}(\mathrm{d}z)$. By applying Proposition 2 to Q^*, we obtain that for any $h \in L_0^2(\Theta, \pi)$, $h/(\varrho_{\mathrm{EX}}\varrho_{\mathsf{z}}) \in L_0^2(\Theta \times \mathbb{R}, \tilde{\pi} \otimes \tilde{\pi}_{\mathsf{z}})$, $\mathrm{IF}\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathsf{z}}), \tilde{Q}^*\} < \infty$ and

$$
\pi(h^2)\{1 + \mathrm{IF}(h, Q^*)\} = \bar{\pi}(\varrho_{\mathrm{EX}}\varrho_{\mathsf{z}})\tilde{\pi} \otimes \tilde{\pi}_{\mathsf{z}}\{h^2/(\varrho_{\mathrm{EX}}^2\varrho_{\mathsf{z}}^2)\}\big[1 + \mathrm{IF}\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathsf{z}}), \tilde{Q}^*\}\big]
$$
$$
= \pi(\varrho_{\mathrm{EX}})\pi_{\mathsf{z}}(\varrho_{\mathsf{z}})\tilde{\pi}(h^2/\varrho_{\mathrm{EX}}^2)\tilde{\pi}_{\mathsf{z}}(1/\varrho_{\mathsf{z}}^2)\big[1 + \mathrm{IF}\{h/(\varrho_{\mathrm{EX}}\varrho_{\mathsf{z}}), \tilde{Q}^*\}\big].
$$

The identity follows easily, as $\tilde{\pi}_{\mathsf{z}}(1/\varrho_{\mathsf{z}}^2) = \pi_{\mathsf{z}}(1/\varrho_{\mathsf{z}})/\pi_{\mathsf{z}}(\varrho_{\mathsf{z}})$ and $\pi_{\mathsf{z}}(1/\varrho_{\mathsf{z}}) < \infty$.

We now establish the uniform ergodicity of $\tilde{Q}_{\mathsf{z}}$. We have

$$
\tilde{Q}_{\mathsf{z}}(z, \mathrm{d}w) = \frac{g(\mathrm{d}w)\alpha_{\mathsf{z}}(z, w)}{\varrho_{\mathsf{z}}(z)}
$$

where

$$
\varrho_{\mathsf{z}}(z) = \int g(\mathrm{d}w) \min\{1, \exp(w - z)\}
$$
$$
\leq \min\left\{\int g(\mathrm{d}w), \exp(-z)\int \exp(w)g(\mathrm{d}w)\right\} = \min\{1, \exp(-z)\}
$$

by Jensen's inequality and the fact that $\int \exp(w)g(\mathrm{d}w) = 1$. Now, using the inequality $\alpha_{\mathsf{z}}(z, w) \geq \min\{1, \exp(w)\}\min\{1, \exp(-z)\}$, it follows that there exists $0 < \varepsilon \leq 1$ and a probability measure $\upsilon(\mathrm{d}w)$ such that

$$
\tilde{Q}_{\mathsf{z}}(z, \mathrm{d}w) \geq \varepsilon\upsilon(\mathrm{d}w),
$$

where

$$
\varepsilon = \int g(\mathrm{d}w)\min\{1, \exp(w)\},
$$
$$
\upsilon(\mathrm{d}w) = \varepsilon^{-1}g(\mathrm{d}w)\min\{1, \exp(w)\}.
$$

If $\pi_{\mathsf{z}}(1/\varrho_{\mathsf{z}}) < \infty$, then $\tilde{\pi}_{\mathsf{z}}(1/\varrho_{\mathsf{z}}^2) < \infty$ and so the inequality $\mathrm{IF}(1/\varrho_{\mathsf{z}}, \tilde{Q}_{\mathsf{z}}) < \infty$ follows directly from Theorem 1 in Roberts & Rosenthal (2008). □

Proof of Proposition 3. If $\langle f, \tilde{P}f\rangle_{\tilde{\mu}} \geq 0$ for any $f \in L^2(\mathsf{X}, \tilde{\mu})$, then $\tilde{P}$ is positive by definition, implying the positivity of P, since $L^2(\mathsf{X}, \mu) \subseteq L^2(\mathsf{X}, \tilde{\mu})$ and

$$
\langle f, Pf\rangle_\mu = \mu(\varrho)\langle f, \tilde{P}f\rangle_{\tilde{\mu}} + \mu\{(1 - \varrho)f^2\}.
$$

For a proposal of the form $q(x, y) = \int s(x, z)s(y, z)\chi(\mathrm{d}z)$, Lemma 3.1 in Baxendale (2005) establishes that $\langle f, \tilde{P}f\rangle_{\tilde{\mu}} \geq 0$ for any $f \in L^2(\mathsf{X}, \tilde{\mu})$. For a ν-reversible proposal such that $\nu(x)q(x, y) = \int r(x, z)r(y, z)\chi(\mathrm{d}z)$, we have that for any $f \in L^2(\mathsf{X}, \tilde{\mu})$,

$$
\mu(\varrho)\langle f, \tilde{P}f\rangle_{\tilde{\mu}}
$$

$$
= \iint f(x)f(y)\nu(x)q(x,y)\min\left\{\frac{\mu(x)}{\nu(x)},\frac{\mu(y)}{\nu(y)}\right\}\mathrm{d}x\,\mathrm{d}y
$$

$$
= \iint f(x)f(y)\nu(x)q(x,y)\left[\int_0^\infty \mathbb{I}_{\{0,\mu(x)/\nu(x)\}}(t)\mathbb{I}_{\{0,\mu(y)/\nu(y)\}}(t)\,\mathrm{d}t\right]\mathrm{d}x\,\mathrm{d}y
$$

$$
= \int_0^\infty\left[\iiint f(x)r(x,z)f(y)r(y,z)\mathbb{I}_{\{0,\mu(x)/\nu(x)\}}(t)\mathbb{I}_{\{0,\mu(y)/\nu(y)\}}(t)\,\mathrm{d}x\,\mathrm{d}y\,\chi(\mathrm{d}z)\right]\mathrm{d}t
$$

$$
= \int_0^\infty\left(\int\left[\int f(x)r(x,z)\mathbb{I}_{\{0,\mu(x)/\nu(x)\}}\mathrm{d}x\right]^2\chi(\mathrm{d}z)\right)\mathrm{d}t \geq 0,
$$

by repeated application of Fubini's theorem. □

Proof of Proposition 4. Theorem 2.2 in Roberts & Tweedie (1996) establishes the ergodicity of Q_{EX}. We extend their argument to prove the ergodicity of Q^*. For the ball $B(\theta,L)$ centred at θ of radius L, we define

$$
\eta(\theta,L) = \left\{\sup_{\vartheta\in B(\theta,L)}\pi(\vartheta)\right\}^{-1}\inf_{\vartheta\in B(\theta,L)}\pi(\vartheta),
$$

which, by assumption, satisfies $0 < \eta(\theta,L) < \infty$. Then, for any $(\theta,z)\in\Theta\times\mathbb{R}$, $\vartheta\in B(\theta,\delta)$ and $w\in\mathbb{R}$,

$$
Q^*\{(\theta,z),(\mathrm{d}\vartheta,\mathrm{d}w)\} \geq q(\theta,\vartheta)\alpha_{\mathrm{ex}}(\theta,\vartheta)g(w)\alpha_{\mathrm{z}}(z,w)\,\mathrm{d}\vartheta\,\mathrm{d}w
$$

$$
\geq \varepsilon\eta(\theta,\delta)\min\{g(w),\exp(-z)\pi_{\mathrm{z}}(w)\}\,\mathrm{d}\vartheta\,\mathrm{d}w, \tag{S5}
$$

which is strictly positive on $S = \{w : g(w) > 0\} = \{w : \pi_{\mathrm{z}}(w) > 0\}$. Hence, the n-step density part of $(Q^*)^n$ is strictly positive for all $(\vartheta,z)\in B(\theta,n\delta)\times S$. This establishes the $\mathrm{d}\vartheta\times\pi_{\mathrm{z}}(\mathrm{d}z)$ irreducibility of Q^* and hence its ergodicity, as it is $\bar{\pi}$-invariant. For $\tilde{Q}^*$, we have that for any $(\theta,z)\in\Theta\times\mathbb{R}$, $\vartheta\in B(\theta,\delta)$ and $w\in\mathbb{R}$,

$$
\tilde{Q}^*\{(\theta,z),(\mathrm{d}\vartheta,\mathrm{d}w)\} = \frac{q(\theta,\vartheta)\alpha_{\mathrm{EX}}(\theta,\vartheta)g(w)\alpha_{\mathrm{z}}(z,w)}{\varrho_{\mathrm{EX}}(\theta)\varrho_{\mathrm{z}}(z)}\,\mathrm{d}\vartheta\,\mathrm{d}w
$$

$$
\geq \varepsilon\eta(\theta,\delta)\min\{g(w),\exp(-z)\pi_{\mathrm{z}}(w)\}\,\mathrm{d}\vartheta\,\mathrm{d}w,
$$

using calculations as in (S5) and the fact that $0 < \varrho_{\mathrm{EX}}(\theta)\varrho_{\mathrm{z}}(z) \leq 1$ for any $(\theta,z)\in\Theta\times\mathbb{R}$, because Q^* is irreducible. Finally, the ergodicity of $\tilde{Q}_{\mathrm{EX}}$ follows, using similar arguments, from the ergodicity of Q_{EX}. □

S3. STATEMENTS AND PROOFS OF AUXILIARY TECHNICAL RESULTS

PROPOSITION 5. *Define the relative computing time* $\mathrm{uRCT}_2(h;\sigma)$ *by*

$$
\mathrm{uRCT}_2(h;\sigma) = \frac{\mathrm{uRIF}_2(h;\sigma)}{\sigma^2},
$$

where $\mathrm{uRIF}_2(h;\sigma)$ *is the relative inefficiency. Under the same assumptions as in Theorem* 1, *the following properties hold.*

(i) *If* $\mathrm{IF}(h,Q_{\mathrm{EX}}) = 1$, *then* $\mathrm{uRCT}_2(h;\sigma)$ *is minimized at* $\sigma_{\mathrm{opt}} = 0.92$, *and we have* $\mathrm{RIF}(h,Q;\sigma_{\mathrm{opt}}) = \mathrm{IF}(h,Q_\pi;\sigma_{\mathrm{opt}}) = 4.54$ *and* $\pi_{\mathrm{z}}^{\sigma_{\mathrm{opt}}}(\varrho_{\mathrm{z}}^{\sigma_{\mathrm{opt}}}) = 0.51$.

(ii) *If* $\mathrm{IF}(h/\varrho_{\mathrm{EX}},\tilde{Q}_{\mathrm{EX}}) \geq 1$, *then* σ_{opt} *increases to* 1.02 *as* $\mathrm{IF}(h,Q_{\mathrm{EX}}) \to \infty$.

(iii) $\mathrm{uRIF}_2(h;\sigma)$ *and* $\mathrm{uRCT}_2(h;\sigma)$ *are decreasing functions of* $\mathrm{IF}(h,Q_{\mathrm{EX}})$.

Proof of Proposition 5. We consider minimizing $\text{uRCT}_2(h;\sigma)$ with respect to σ. Then

$$\text{uRCT}_2(h;\sigma) = \frac{\{1 + \text{IF}(h, Q_{\text{EX}})\}}{2\,\text{IF}(h, Q_{\text{EX}})} \frac{\text{IF}(h, Q_\pi; \sigma)}{\sigma^2} + \frac{\text{IF}(h, Q_{\text{EX}}) - 1}{2\sigma^2 \text{IF}(h, Q_{\text{EX}})}.$$

To obtain statement (i), note that $\text{uRCT}_2(h;\sigma) = \text{IF}(h, Q_\pi; \sigma)/\sigma^2$ when $\text{IF}(h, Q_{\text{EX}}) = 1$. We define $H(\sigma) = \text{IF}(h, Q_\pi; \sigma)/\sigma^2$. Using Lemma 5 in Pitt et al. (2012), one can verify that $H(\sigma)$ is minimized at $\sigma_{\text{opt}} = 0{\cdot}92$ and that $\partial^2 H(\sigma)/\partial \sigma^2 > 0$. The numerical values in (i) at $\sigma_{\text{opt}} = 0{\cdot}92$ can be found in Pitt et al. (2012). To obtain statement (ii), we note that

$$\partial \text{uRCT}_2(h;\sigma)/\partial\sigma = \{1 + \text{IF}(h, Q_{\text{EX}})\}/\{2\,\text{IF}(h, Q_{\text{EX}})\} \times \{\partial H(\sigma)/\partial\sigma\}$$
$$- \{\text{IF}(h, Q_{\text{EX}}) - 1\}/\{\sigma^3 \text{IF}(h, Q_{\text{EX}})\},$$
$$\partial^2 \text{uRCT}_2(h;\sigma)/\partial\sigma^2 = \{1 + \text{IF}(h, Q_{\text{EX}})\}/\{2\,\text{IF}(h, Q_{\text{EX}})\} \times \{\partial^2 H(\sigma)/\partial\sigma^2\}$$
$$+ 3\{\text{IF}(h, Q_{\text{EX}}) - 1\}/\{\sigma^4 \text{IF}(h, Q_{\text{EX}})\},$$

so that $\partial^2 \text{uRCT}_2(h;\sigma)/\partial\sigma^2 > 0$ if $\text{IF}(h, Q_{\text{EX}}) \geq 1$. For the limiting case in (ii),

$$\lim_{\text{IF}_{\text{EX}} \uparrow \infty} \partial \text{uRCT}_2(h;\sigma)/\partial\sigma = \{\partial H(\sigma)/\partial\sigma\}/2 - 1/\sigma^3,$$

which we can verify numerically is equal to 0 at $\sigma_{\text{opt}} = 1{\cdot}02$. For general values of $\text{IF}(h, Q_{\text{EX}})$,

$$\partial\big\{\partial \text{uRCT}_2(h;\sigma)/\partial\sigma\big|_{\sigma=\sigma_{\text{opt}}}\big\}/\partial \text{IF}(h, Q_{\text{EX}})$$
$$= -1/\{\text{IF}(h, Q_{\text{EX}})\}^2\big\{\partial H(\sigma)/\partial\sigma\big|_{\sigma=\sigma_{\text{opt}}}/2 + 1/\sigma_{\text{opt}}^3\big\} < 0,$$

where $\partial H(\sigma)/\partial\sigma > 0$ for $\sigma > 0{\cdot}92$. Hence, σ_{opt} increases with $\text{IF}(h, Q_{\text{EX}})$, which verifies assertion (ii). Finally, to obtain (iii), it is straightforward to show that

$$\text{uRIF}_2(h;\sigma) = \frac{\text{IF}(h, Q_\pi; \sigma) + 1}{2} + \frac{\text{IF}(h, Q_\pi; \sigma) - 1}{2\,\text{IF}(h, Q_{\text{EX}})},$$

so that $\text{uRIF}_2(h;\sigma)$ and $\text{uRCT}_2(h;\sigma) = \text{uRIF}_2(h;\sigma)/\sigma^2$ are decreasing functions of $\text{IF}(h, Q_{\text{EX}})$, holding σ constant. $\square$

S4. LOWER BOUNDS ON THE RELATIVE COMPUTING TIME OF Q^*

In this section we compare the two lower bounds, $\text{lRCT}_1(h;\sigma)$ and $\text{lRCT}_2(\sigma)$, defined in § 3·5 of the main article. It can be seen that $\text{lRCT}_2(\sigma)$ provides a universal lower bound which does not depend on the function h and is also independent of the inefficiency associated with the exact jump chain. It is clear from Fig. S1 that $\text{lRCT}_1(h;\sigma)$ provides a tighter lower bound for low values of $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$ but requires $\tilde{Q}_{\text{EX}}$ to be positive. The two bounds become similar for very large values of $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$. It is also clear that the optimal value of σ associated with $\text{lRCT}_1(h;\sigma)$ increases with $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$ towards the optimal argument $\sigma = 1{\cdot}68$ of lRCT_2.

S5. ASYMPTOTIC UPPER BOUND

In this section we illustrate, in the Gaussian noise case, the lower bound on the inefficiency $\text{lRIF}_2(\sigma) = 1/\{2\Phi(-\sigma/\sqrt{2})\}$ and the exact relative inefficiency $\text{aRIF}(\sigma, l) = \Phi(-l/2)/\Phi\{-(2\sigma^2 + l^2)^{1/2}/2\}$ obtained by Sherlock et al. (2014) and discussed in § 3·6 of the main paper. Recall that $\text{aRIF}(\sigma, l) \to \text{lRIF}_2(\sigma)$ as $l \to 0$ and note that $\text{aRIF}(\sigma, l) \to \Psi(\sigma) = \exp(\sigma^2/4)/\sigma^2$ as $l \to \infty$. Figure S2 displays the corresponding relative computing times
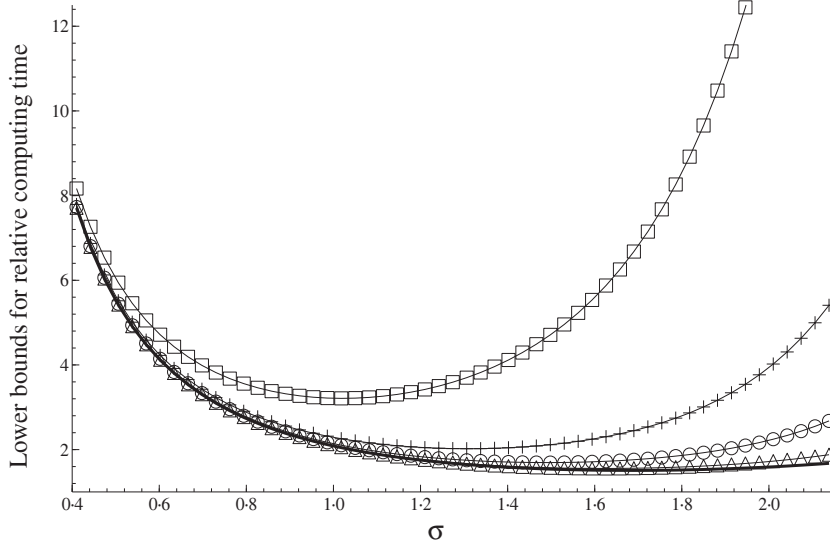
Fig. S1: Theoretical results for the lower bounds $\text{LRCT}_1(h;\sigma)$ and $\text{LRCT}_2(\sigma)$ as functions of σ: $\text{LRCT}_1(h;\sigma)$ is plotted for the inefficiency $\text{IF}(h/\varrho_{\text{EX}}, \tilde{Q}_{\text{EX}})$ associated with the exact jump chain taking the values 1 (squares), 10 (crosses), 40 (circles) and 200 (triangles); $\text{LRCT}_2(\sigma)$ is shown as the solid black curve.

$\text{LRCT}_2(\sigma) = \text{LRIF}_2(\sigma)/\sigma^2$ and $\text{ARCT}(\sigma; l) = \text{ARIF}(\sigma, l)/\sigma^2$. They are very similar in shape as functions of σ, regardless of l, and are also minimized at similar values: $\text{LRCT}_2(\sigma)$ is minimized at $\sigma_1 = 1\cdot68$ and $\Psi(\sigma)$ is minimized at $\sigma_2 = 2\cdot00$; $\text{LRCT}_2(\sigma_1) = 1\cdot51$ and $\text{LRCT}_2(\sigma_2) = 1\cdot59$, while $\Psi(\sigma_1) = 0\cdot72$ and $\Psi(\sigma_2) = 0\cdot68$.

S6. SIMULATION RESULTS

In this section we apply the pseudo-marginal algorithm to a linear Gaussian state-space model, and present additional simulation results for the stochastic volatility model discussed in the main paper. The linear Gaussian state-space model we consider is a first-order autoregression $\text{AR}(1)$ observed with noise. In this case, $Y_t = X_t + \sigma_\varepsilon \varepsilon_t$, and the state evolution is $X_{t+1} = \mu_x(1 - \phi) + \phi X_t + \sigma_\eta \eta_t$, where ε_t and η_t are standard normal and independent. We take $\phi = 0\cdot8$, $\mu_x = 0\cdot5$ and $\sigma_\eta^2 = 1 - \phi^2$, so that the marginal variance σ_x^2 of the state X_t is 1. We consider a series of length T, where $\sigma_\varepsilon^2 = 0\cdot5$ is assumed to be known. The parameters of interest are therefore $\theta = (\phi, \mu_x, \sigma_x)$. The analysis is very similar to that in §4 of the main paper. However, for this state-space model, the likelihood can be calculated by using the Kalman filter. This facilitates the analysis in §§ S6·1 and S6·2 in two ways. First, in the calculation of the loglikelihood error $Z = \log \hat{p}_N(y \mid \theta) - \log p(y \mid \theta)$, the true likelihood term is known rather than estimated. Second, because the likelihood is known, we can directly examine the exact chain Q_{EX} and estimate the inefficiency $\text{IF}(h, Q_{\text{EX}})$.

S6·1. *Empirical results for the error of the loglikelihood estimator*

The analysis in this subsection mirrors that in §4·2 of the main paper. We investigate empirically Assumptions 1 and 2 by examining the behaviour of $Z = \log \hat{p}_N(y \mid \theta) - \log p(y \mid \theta)$ for $T = 40$, 300 and 2700. Corresponding values of N are selected to ensure that the variance
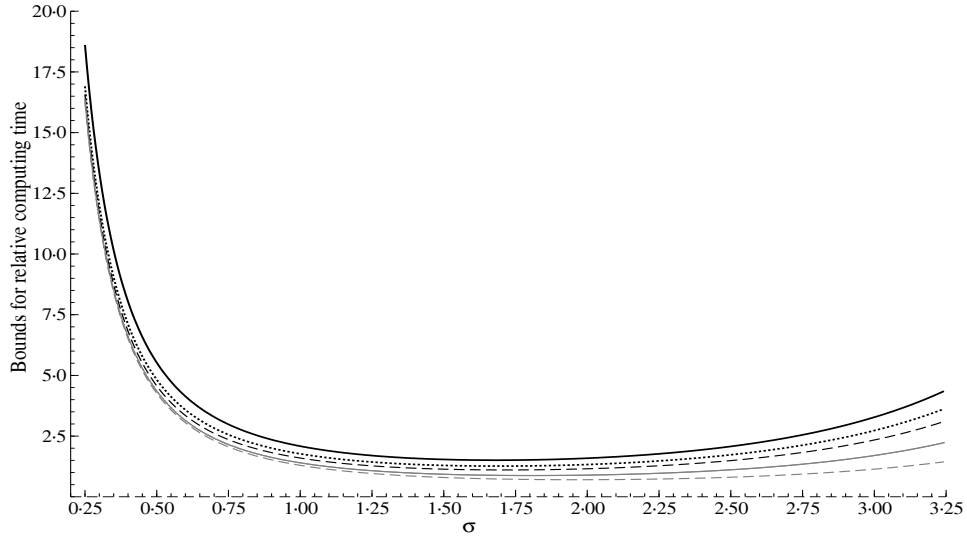
Fig. S2: Theoretical results for the relative computing time: LRCT$_2(\sigma)$ (solid black) is displayed together with ARCT(σ, l) as functions of σ; ARCT(σ, l), the relative computing time for the limiting case of a random walk proposal, is evaluated for $l = 0.5$ (dotted black), 1 (dashed black), 2·5 (solid grey) and 10 (dashed grey), where l is the scaling factor in the proposal.

of Z evaluated at the posterior mean $\bar{\theta}$ is approximately unity in each case. The left panels of Fig. S3 display the histograms corresponding to the density $g_N(z \mid \bar{\theta})$ of Z for $\theta = \bar{\theta}$, obtained by running $S = 6000$ particle filters at this value. We overlay on each histogram a kernel density estimate together with the corresponding assumed density, $g_Z^{\sigma}(z)$ of Assumption 2, where σ^2 is the sample variance of Z over the S particle filters. For $T = 40$, there is a slight discrepancy between the assumed Gaussian densities and the true histograms representing $g_N(z \mid \bar{\theta})$. In particular, although $g_N(z \mid \bar{\theta})$ is well approximated over most of its support, it is heavier-tailed in the left tail. For $T = 300$ and $T = 2700$, the assumed Gaussian densities are very accurate.

We also examine Z when θ is distributed according to $\pi(\theta)$. We record 100 samples from $\pi(\theta)$ for $T = 40$, 300 and 2700. For each of these samples, we run the particle filter 300 times to estimate the true likelihood at these values. The resulting histograms, corresponding to the density $\int \pi(d\theta) g_N(z \mid \theta)$, are displayed in the right panels of Fig. S3. For $T = 300$ and $T = 2700$, the assumed densities $g_Z^{\sigma}(z)$ are close to the corresponding histograms, and Assumptions 1 and 2 again appear to capture reasonably well the salient features of the densities associated with Z.

It is important, in examining departures from Assumption 1, to consider the heterogeneity of the conditional density $g_N(z \mid \theta)$ as θ varies over $\pi(\theta)$. In Fig. S4, the conditional moments associated with the density $g_N(z \mid \theta)$ are estimated, based on running the particle filter independently $S = 300$ times for each of 100 values of θ from $\pi(\theta)$. We record the estimates of the mean, variance, and third and fourth central moments at each value of θ, for $T = 300$ and $T = 2700$. There is a small degree of variability for $T = 300$ around the values that we would expect, which are -0.5, 1, 0 and 3, corresponding to $g_Z^{\sigma}(z)$ where $\sigma = 1$. This variability reduces as T rises to 2700. A small degree of variability is expected as these are moments estimated from $S = 300$ samples. This lack of heterogeneity explains why the values of Z, marginalized over $\pi(\theta)$, in the
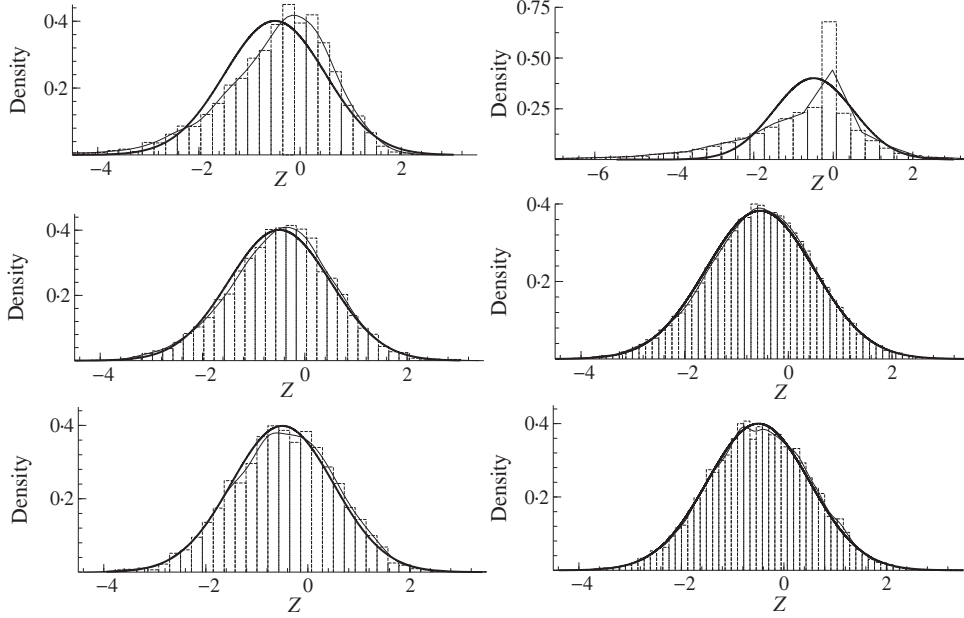
Fig. S3: Results for Z arising from the AR(1)-plus-noise model experiment. From top to bottom, the panels show plots for $T = 40$ and $N = 4$ (top), $T = 300$ and $N = 50$ (middle), and $T = 2700$ and $N = 500$ (bottom). The three panels on the left show histograms, kernel density estimates and theoretical densities associated with $g_N(z \mid \theta)$ evaluated at the posterior mean $\bar{\theta}$, and the panels on the right show the results for $g_N(z \mid \theta)$ evaluated over values from the posterior $\pi(\theta)$; the densities $g_Z^\sigma(z)$ are overlaid (thick solid curves).

right panels of Fig. S3 are close to $g_Z^\sigma(z)$ for time series of moderate and large lengths. Figure S5 presents results of a similar experiment for the stochastic volatility model and data considered in § 4 of the main paper. There is rather more variability, as the true value of the likelihood in this case is unknown and has to be estimated. However, the results are similar, and the variability again reduces as T rises to 2700.

S6·2. *Empirical results for the pseudo-marginal algorithm*

The pseudo-marginal algorithm is applied to data with $T = 300$. The true likelihood of the data is computed by the Kalman filter, as the model is a linear Gaussian state-space model. This allows the exact Metropolis–Hastings scheme Q_{EX} to be implemented, so that the corresponding inefficiency $\text{IF}(h, Q_{\text{EX}})$ can be estimated easily. We consider varying N so that the standard deviation $\sigma(\bar{\theta}; N)$ of the loglikelihood estimator varies. The grid of values that we consider for N is $\{11, 16, 22, 31, 43, 60, 83, 116, 161, 224, 312\}$; see Table S1. The value $N = 60$ results in $\sigma(\bar{\theta}; N) = 0.92$.

We transform each of the parameters to the real line so that $\Psi = k(\theta)$, where both θ and Ψ are three-dimensional vectors, and place a Gaussian prior on Ψ centred at zero with a large variance. We use the autoregressive Metropolis proposal $q(\Psi, \Psi^*)$, with

$$\Psi^* = (1 - \rho)\hat{\Psi} + \rho\Psi + (1 - \rho^2)^{1/2}\{(\nu - 2)/\nu\}^{1/2}\Sigma^{1/2}t_\nu,$$

for both the pseudo-marginal algorithm and the exact likelihood schemes, where $\hat{\Psi}$ is the mode of the loglikelihood obtained from the Kalman filter and the covariance Σ is the negative inverse of
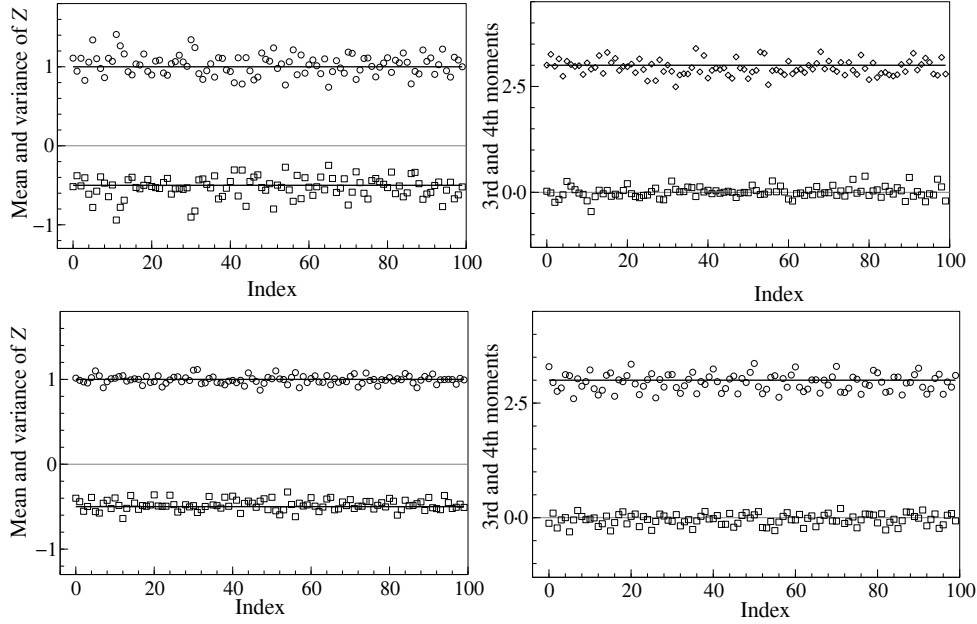
Fig. S4: Estimated conditional moments for Z arising from the AR(1)-plus-noise model experiment. The upper panels show results for $T = 300$ and $N = 50$, and the lower panels show results for $T = 2700$ and $N = 500$. The left panels plot the mean (squares) and variance (circles) associated with $g_N(z \mid \theta)$ for 100 different values of θ from $\pi(\theta)$; the right panels display the corresponding estimates of the third (squares) and fourth (circles) moments.

the second derivative of the loglikelihood at the mode. Here t_ν denotes a standard multivariate t-distributed random variable with ν degrees of freedom. We set $\nu = 5$. We use this autoregressive proposal with the scalar autoregressive parameter ρ chosen from $\{0, 0\cdot4, 0\cdot6, 0\cdot9\}$. We first apply this proposal for these four values of ρ using the known likelihood in the Metropolis–Hastings scheme, and then estimate the inefficiency for each of the parameters $\theta = (\mu_x, \phi, \sigma_\eta)$.

Figure S6 plots the acceptance probability for the pseudo-marginal algorithm against $\sigma(\bar{\theta}; N)$ for the four values of the proposal parameter ρ. The lower bound for the acceptance probabilities, as discussed at the end of § 4·3 of the main paper, is also displayed, and there is close correspondence in all cases. The histograms for the accepted and rejected values of Z, for $N = 60$ when $\sigma(\bar{\theta}; N) = 0\cdot92$, are also displayed. The approximating asymptotic Gaussian densities, with $\sigma = 0\cdot92$, are superimposed. This figure shows that the approximating densities correspond very closely to the two histograms. It should be noted that these are the marginal values for Z over the draws from the posterior $\pi(\theta)$ obtained by running the pseudo-marginal scheme, rather than being based on a fixed value of the parameters.

Tables S1 and S2 show the pseudo-marginal algorithm results for $\rho = 0$ and $\rho = 0\cdot9$. For the independent Metropolis–Hastings proposal, it is clear from Table S1 that the computing time is minimized around $N = 43$ or 60, depending on which parameter is examined. The corresponding values of $\sigma(\bar{\theta}; N)$ are $1\cdot11$ and $0\cdot92$, which supports the findings that when an efficient proposal is used, the optimal value of σ is close to unity. This is again supported by Fig. S7, for which the relative computing time ($\rho = 0$, top right plot) is plotted against $\sigma(\bar{\theta}; N)$. We note that the relative inefficiencies and computing times are straightforward to calculate, as the exact chain inefficiencies for the three parameters have been calculated and are given in the top row of
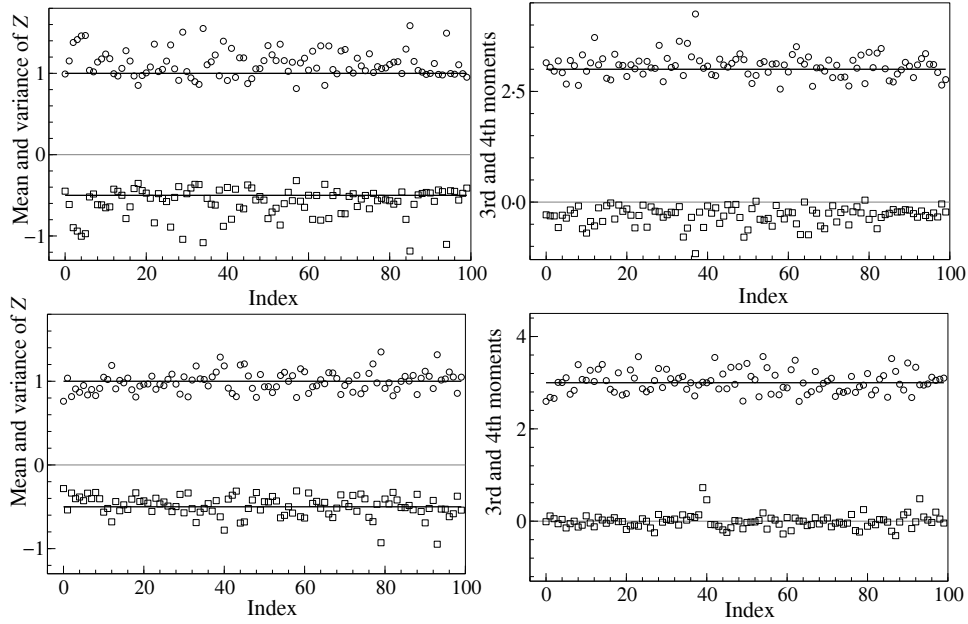
Fig. S5: Estimated conditional moments for Z arising from the two-factor model experiment for the S&P 500 data discussed in §4 of the main paper. The upper panels show results for $T = 300$ and $N = 80$, and the lower panels show results for $T = 2700$ and $N = 700$. The left panels plot the mean (squares) and variance (circles) associated with $g_N(z \mid \theta)$ for 100 different values of θ from $\pi(\theta)$; the right panels display the corresponding estimates of the third (squares) and fourth (circles) moments.

Table S1: *Results for the pseudo-marginal algorithm applied to the* AR(1)*-plus-noise example with proposal parameters* $\rho = 0$, $T = 300$, $\phi = 0.8$, $\mu = 0.5$, $\sigma_x^2 = 1$ *and* $\sigma_\varepsilon^2 = 0.5$; *inefficiencies* IF *and computing times* CT $= N \times$ IF *are reported for* (ϕ, μ, σ_x) *and marginal probabilities of acceptance* pr(Acc); *cf. Fig.* S6 *and Fig.* S7

| Q_{EX}		IF(ϕ)	IF(μ)	IF(σ_x)				pr(Acc)
		2·5845	2·5040	2·4163				0·7678
Q								
N	$\sigma(\bar{\theta}; N)$	IF(ϕ)	IF(μ)	IF(σ_x)	CT(ϕ)	CT(μ)	CT(σ_x)	pr(Acc)
11	2·2886	136·32	132·41	128·66	1499·5	1456·5	1415·3	0·11424
16	1·8692	61·403	63·756	66·609	982·45	1020·1	1065·7	0·19036
22	1·6063	37·256	40·486	37·367	819·63	890·68	822·07	0·25549
31	1·3412	15·880	18·099	19·135	492·29	561·08	593·20	0·32622
43	1·1096	11·320	9·7400	10·710	486·75	418·82	460·54	0·39347
60	0·9197	7·5040	8·0428	7·6168	450·24	482·57	457·01	0·45933
83	0·8058	5·7253	5·5841	5·9348	475·20	463·48	492·59	0·50885
116	0·6828	4·3756	4·7106	4·1693	507·57	546·43	483·63	0·56621
161	0·5828	3·8112	4·2379	3·6388	613·61	682·30	585·84	0·60160
224	0·4838	3·2711	3·1605	3·3134	732·73	707·94	742·19	0·63562
312	0·4096	3·0774	3·4768	2·8355	960·14	1084·8	884.67	0·65793
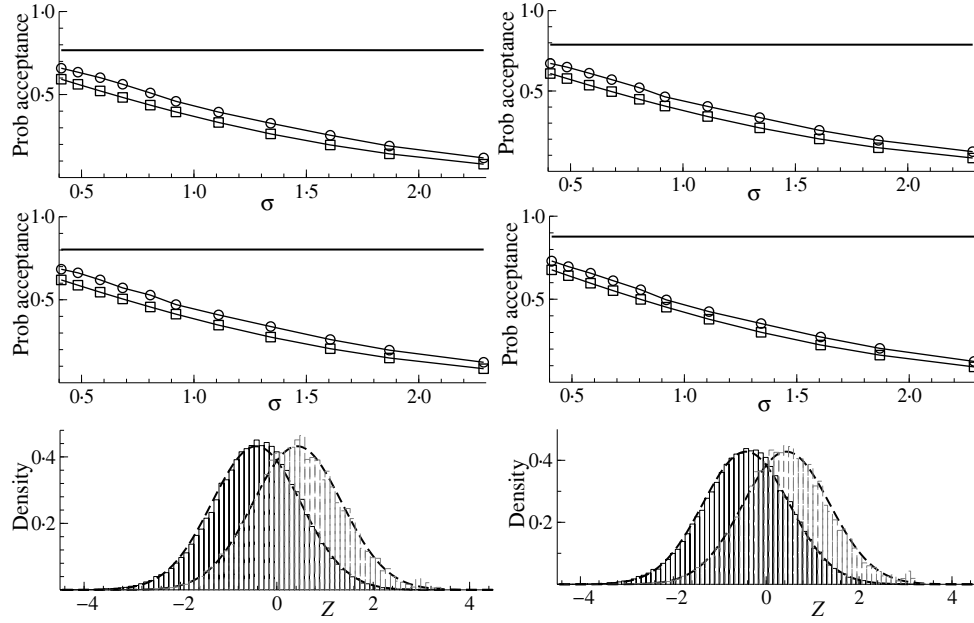
Fig. S6: Results for the pseudo-marginal algorithm applied to the AR(1)-plus-noise example with $T = 300, \phi = 0\cdot8, \mu = 0\cdot5, \sigma_x^2 = 1$ and σ_ε^2 fixed at $0\cdot5$. The upper four panels show the marginal acceptance probabilities plotted against $\sigma(\bar\theta; N)$, with autoregressive proposal parameters $\rho = 0$ (top left), $0\cdot4$ (top right), $0\cdot6$ (middle left) and $0\cdot9$ (middle right). The estimated (constant) marginal probabilities of acceptance for Q_{EX} (squares) are shown together with the estimated probabilities from Q (circles); the lower bound (squares) is given as the probability from the exact scheme times $2\Phi(-\sigma/\sqrt2)$. The bottom two panels display histograms and theoretical densities for the accepted and proposed values of Z, the loglikelihood error for $\rho = 0$ (left) and for $\rho = 0\cdot9$ (right).

Table S2: *Results for the pseudo-marginal algorithm applied to the AR(1)-plus-noise example with proposal parameter $\rho = 0\cdot9$; other settings are the same as in Table S1*

| Q_{EX}		IF(ϕ)	IF(μ)	IF(σ_x)				pr(Acc)
		25·59	22·21	24·44				0·87717
Q								
N	$\sigma(\bar\theta; N)$	IF(ϕ)	IF(μ)	IF(σ_x)	CT(ϕ)	CT(μ)	CT(σ_x)	pr(Acc)
11	2·2886	594·64	488·30	639·04	6541·1	5371·3	7029·5	0·12579
16	1·8692	157·49	183·78	182·07	2519·9	2940·4	2913·1	0·20410
22	1·6063	126·87	115·84	125·37	2791·2	2548·6	2758·2	0·27279
31	1·3412	69·541	67·421	71·982	2155·9	2089·9	2231·5	0·35385
43	1·1096	53·053	62·344	58·002	2281·1	2680·9	2494·0	0·42577
60	0·9197	49·351	47·476	45·194	2961·1	2848·6	2711·6	0·49610
83	0·8058	37·709	29·550	38·266	3129·8	2452·7	3176·1	0·55764
116	0·6828	29·360	36·943	34·892	3405·8	4285·4	4047·4	0·61174
161	0·5828	28·277	27·883	29·864	4552·6	4489·2	4808·1	0·65704
224	0·4838	27·770	29·471	30·533	6220·5	6601·5	6839·4	0·69674
312	0·4096	29·231	25·549	29·967	9120·2	7971·4	9349·8	0·73057
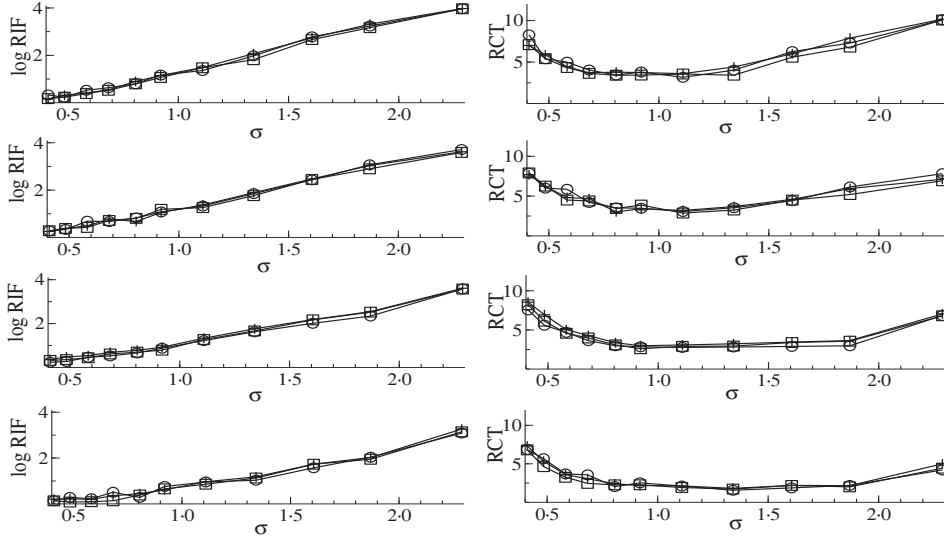
Fig. S7: Relative inefficiencies and computing times for the AR(1)-plus-noise example with $T = 300$, $\phi = 0.8$, $\mu = 0.5$, $\sigma_x^2 = 1$ and σ_ε^2 fixed at 0·5. The left panels plot the logarithm of RIF against $\sigma(\bar{\theta}; N)$, and the right panels show RCT = RIF$/\sigma^2(\bar{\theta}; N)$ plotted against $\sigma(\bar{\theta}; N)$. In each panel, the three graphs are for ϕ (squares), μ (circles) and σ_x (crosses). From top to bottom, the panels correspond to $\rho = 0$, 0·4, 0·6 and 0·9. Here $\sigma(\bar{\theta}; N)$ is the standard deviation of the loglikelihood estimator evaluated at the posterior mean $\bar{\theta}$; see Tables S1 and S2.

Table S1. Table S2 shows the results for the more persistent proposal where $\rho = 0.9$. In this case, for all three parameters the optimal value of N is around 31, at which $\sigma(\bar{\theta}; N) = 1.34$. The corresponding graph of the relative computing time is given in the bottom right panel of Fig. S7. The findings are consistent with the discussion in § 3·5 of the main paper. In particular, as ρ increases, we find that IF(h, Q_{EX}) and IF$(h/\varrho_{\mathrm{EX}}, \tilde{Q}_{\mathrm{EX}})$ increase. It is apparent from Fig. S7 that, as expected, the optimal value of $\sigma(\bar{\theta}; N)$ increases and the relative computing time decreases for any given $\sigma(\bar{\theta}; N)$. In addition, the relative computing time becomes flatter as a function of $\sigma(\bar{\theta}; N)$ as ρ increases.

As an additional check of the theoretical results in § 3·5 of the main paper, we implement the Q^* chain associated with the AR(1)-plus-noise example. This is straightforward as the exact likelihood is available, allowing the acceptance criterion to be evaluated exactly. The results are displayed in Fig. S8. In this case, a reasonably persistent proposal, $\rho = 0.6$, is chosen, and the acceptance probability associated with the exact chain Q_{EX} is high. As a consequence, the discrepancy between the two chains Q and Q^* is small. This is reflected in the graphs of the inefficiency, relative computing time and marginal acceptance probability for both chains as displayed in Fig. S8.

S7. NUMERICAL PROCEDURES

Under the Gaussian assumption, Corollary 3 specifies the function ϱ_Z^σ, and the term $\pi_Z^\sigma(1/\varrho_Z^\sigma)$ can be accurately evaluated using numerical quadrature. This section explains how we numerically evaluate the terms ϕ_Z^σ and IF$(1/\varrho_Z^\sigma, \tilde{Q}^z)$ which appear in the bounds of Corollaries 1 and 2. The inefficiency IF$(1/\varrho_Z^\sigma, \tilde{Q}^z)$ is finite by Lemma 3, because $\pi_Z^\sigma(1/\varrho_Z^\sigma)$ is finite. The autocorrelations quickly descend to zero as a function of n, for all σ. Hence, it is straightforward to estimate
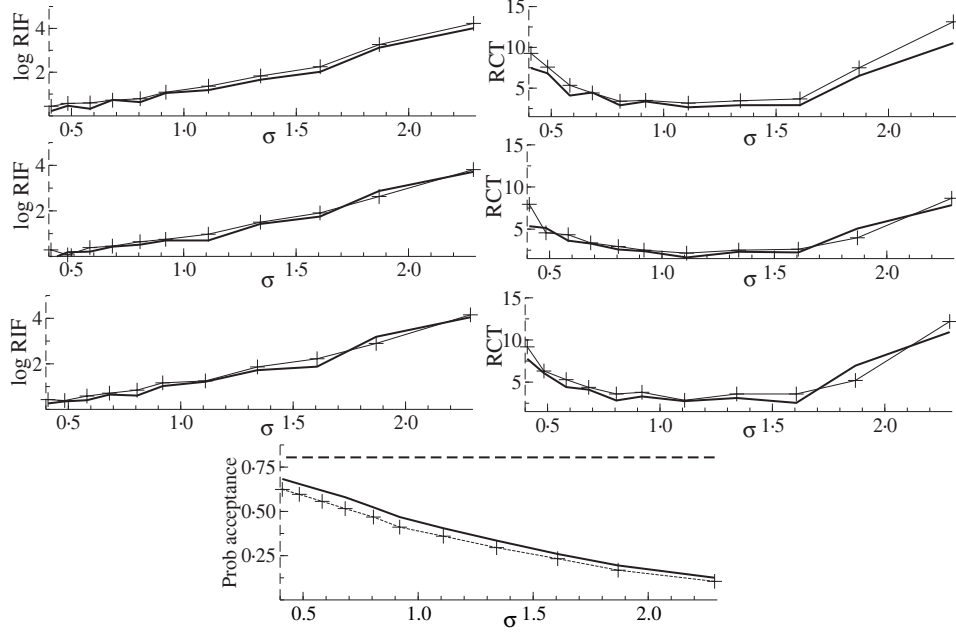
Fig. S8: Relative inefficiencies and computing times for the AR(1)-plus-noise example with $T = 300$, $\phi = 0 \cdot 8$, $\mu = 0 \cdot 5$, $\sigma_x^2 = 1$ and σ_ε^2 fixed at $0 \cdot 5$. The upper six panels show results for ϕ (first row), μ (second row) and σ_x^2 (third row), where the left panels plot the logarithm of RIF against $\sigma(\bar\theta; N)$ and the right panels plot RCT $=$ RIF$/\sigma^2(\bar\theta; N)$ against $\sigma(\bar\theta; N)$. In each panel, the output for Q is shown by a thick solid line and that for Q^* by a thin line with crosses. The bottom panel shows the acceptance probability for the exact scheme (dashed line), Q (solid line) and Q^* (crosses).

IF$(1/\varrho_Z^\sigma, \tilde{Q}^Z)$ by appropriate summation of the autocorrelations, and to tabulate it against σ for use in the bounds of Corollaries 1 and 2. The autocorrelation ϕ_Z^σ for $n = 1$ is similarly tabulated.

From Lemma 3,

$$\tilde{Q}^Z(z, \mathrm{d}w) = \frac{g(w)\min\{1, \exp(w-z)\}\,\mathrm{d}w}{\varrho_Z(z)}, \quad \tilde{\pi}_Z(\mathrm{d}z) = \frac{\pi_Z(\mathrm{d}z)\varrho_Z(z)}{\pi_Z(\varrho_Z)},$$

so the autocorrelation at lag n is

$$\phi_n(\varrho_Z^{-1}, \tilde{Q}^Z) = \frac{\langle \varrho_Z^{-1}, (\tilde{Q}^Z)^n \varrho_Z^{-1}\rangle_{\tilde{\pi}_Z} - \{\tilde{\pi}_Z(\varrho_Z^{-1})\}^2}{\mathbb{V}_{\tilde{\pi}_Z}(\varrho_Z^{-1})}$$

with

$$\langle \varrho_Z^{-1}, (\tilde{Q}^Z)^n \varrho_Z^{-1}\rangle_{\tilde{\pi}_Z} = \int \varrho_Z^{-1}(z_0)\varrho_Z^{-1}(z_n)\tilde{\pi}_Z(\mathrm{d}z_0)(\tilde{Q}^Z)^n(z_0, \mathrm{d}z_n)$$

$$= \pi_Z(\varrho_Z)^{-1}\int \varrho_Z^{-1}(z_n)\tilde{\pi}_Z(\mathrm{d}z_0)(\tilde{Q}^Z)^n(z_0, \mathrm{d}z_n). \tag{S6}$$

The term $\pi_Z(\varrho_Z)$ can be computed by quadrature. The term (S6) can also be accurately calculated using Monte Carlo integration, by simulating a large number M of independent and identically distributed samples $Z_0^i \sim \pi_Z$ and then propagating each sample through the transition kernel $\tilde{Q}^Z$ n times to obtain $Z_n^i \sim \pi_Z(\tilde{Q}^Z)^n$, yielding the estimate $M^{-1}\sum_{i=1}^M \varrho_Z^{-1}(Z_n^i)$.

REFERENCES

BAXENDALE, P. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Prob.* **15**, 700–38.

DOUC, R. & ROBERT, C. P. (2011). A vanilla Rao–Blackwellization of Metropolis–Hastings algorithms. *Ann. Statist.* **39**, 261–77.

HÄGGSTRÖM, O. & ROSENTHAL, J. S. (2007). On variance conditions for Markov chain central limit theorems. *Electron. Commun. Prob.* **12**, 454–64.

KIPNIS, C. & VARADHAN, S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Commun. Math. Phys.* **104**, 1–19.

MEYN, S. & TWEEDIE, R.L. (1991). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge: Cambridge University Press.

PITT, M. K., SILVA, R., GIORDANI, P. & KOHN, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *J. Economet.* **171**, 134–51.

ROBERTS, G. O. & TWEEDIE, R. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83**, 95–110.

ROBERTS, G. O. & ROSENTHAL, J. S. (2008). Variance bounding Markov chains. *Ann. Appl. Prob.* **18**, 1201–14.

RUDIN, W. (1991). *Functional Analysis.* New York: McGraw-Hill.

SHERLOCK, C., THIERY, A. H., ROBERTS, G. O. & ROSENTHAL, J. S. (2014). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* to appear.

SCHMÜDGEN, K. (2012). *Unbounded Self-adjoint Operators on Hilbert Space*. Berlin: Springer.

TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with Discussion). *Ann. Statist.* **21**, 1701–62.

WU, W. & WOODROOFE, M. (2004). Martingale approximations for sums of stationary processes. *Ann. Prob.* **32**, 1674–90.