SVERIGES RIKSBANK
WORKING PAPER SERIES

307

SPEEDING UP MCMC BY DELAYED ACCEPTANCE AND DATA SUBSAMPLING

MATIAS QUIROZ

August 2015

WORKING PAPERS ARE OBTAINABLE FROM

Sveriges Riksbank • Information Riksbank • SE-103 37 Stockholm
Fax international: +46 8 787 05 26
Telephone international: +46 8 787 01 00
E-mail: info@riksbank.se

The Working Paper series presents reports on matters in the sphere of activities of the Riksbank that are considered to be of interest to a wider public.
The papers are to be regarded as reports on ongoing studies and the authors will be pleased to receive comments.

SPEEDING UP MCMC BY DELAYED ACCEPTANCE AND DATA SUBSAMPLING

MATIAS QUIROZ*

SVERIGES RIKSBANK WORKING PAPER SERIES

NO. 307

AUGUST 2015

Abstract. The complexity of Markov Chain Monte Carlo (MCMC) algorithms arises from the requirement of a likelihood evaluation for the full data set in each iteration. Payne and Mallick (2014) propose to speed up the Metropolis-Hastings algorithm by a delayed acceptance approach where the acceptance decision proceeds in two stages. In the first stage, an estimate of the likelihood based on a random subsample determines if it is likely that the draw will be accepted and, if so, the second stage uses the full data likelihood to decide upon final acceptance. Evaluating the full data likelihood is thus avoided for draws that are unlikely to be accepted. We propose a more precise likelihood estimator which incorporates auxiliary information about the full data likelihood while only operating on a sparse set of the data. It is proved that the resulting delayed acceptance MCMC is asymptotically more efficient compared to that of Payne and Mallick (2014). Furthermore, we adapt the method to handle data sets that are too large to fit in Random-Access Memory (RAM). This adaptation results in an algorithm that samples from an approximate posterior with well studied theoretical properties in the literature.

Keywords: Bayesian inference, Markov chain Monte Carlo, Delayed acceptance MCMC, Large data, Survey sampling

JEL Classification: C11, C13, C15, C55, C83

*Research Division, Sveriges Riksbank, SE-103 37 Stockholm, Sweden, Department of Statistics, Stockholm University and Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University. E-mail: quiroz.matias@gmail.com. The author was partially supported by VINNOVA grant 2010-02635. The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

1. Introduction

Markov Chain Monte Carlo (MCMC) methods have been the workhorse for sampling from nonstandard posterior distributions in Bayesian statistics for nearly three decades. Recently, with increasingly more complex models and/or larger data sets, there has been a surge of interest in improving the $O(n)$ complexity emerging from the necessity of a complete data scan in each iteration of the algorithm.

There are a number of approaches proposed in the literature to speed up MCMC. Some authors divide the data into different partitions and carry out MCMC for the partitions in a parallel and distributed manner. The draws from each partition's subposterior are subsequently combined to represent the full posterior distribution. This line of work includes Scott et al. (2013); Neiswanger et al. (2013); Wang and Dunson (2013); Minsker et al. (2014), among others. Other authors use a subsample of the data in each MCMC iteration to speed up the algorithm, see e.g. Korattikara et al. (2013), Bardenet et al. (2014), Maclaurin and Adams (2014), Maire et al. (2015), Bardenet et al. (2015) and Quiroz et al. (2015a, 2015b). Finally, delayed acceptance MCMC has been used to speed up computations (Banterle et al., 2014; Payne and Mallick, 2014). The main idea behind this approach is to avoid computations if there is an indication that the draw will ultimately be rejected.

This paper extends the delayed acceptance algorithms presented in Payne and Mallick (2014) merging with ideas developed in Quiroz et al. (2015a, 2015b). This combination provides an interesting alternative to the Pseudo-marginal MCMC (PMCMC) approach in Quiroz et al. (2015a, 2015b) if exact inference is of importance. Their algorithm targets a (slightly) perturbed posterior, whereas the delayed acceptance MCMC has the true posterior as invariant distribution.

The delayed acceptance algorithm in Payne and Mallick (2014) uses a random sample of the data in the first stage to obtain a computationally cheap estimate of the likelihood, which is used to compute a first Metropolis-Hastings (M-H) acceptance ratio. If accepted, the second stage computes the true (based on all data) M-H ratio. The algorithm speeds up the standard MCMC because it avoids evaluation of the full data likelihood for proposals that are unlikely to be accepted. However, the log-likelihood estimate in Payne and Mallick (2014) is based on a random sample obtained by simple random sampling (SI). In SI all observations are equally probable to be included in the sample. Quiroz et al. (2015a) conclude that an estimate based on this sampling scheme is inefficient because the contribution to the log-likelihood function varies substantially across observations. Ideally, observations who contribute more should be included in the sample with a higher probability. They propose to use Probability Proportional-to-Size sampling (PPS) to achieve this. In a related setting, Quiroz et al. (2015b) propose to use SI combined with the difference estimator from the survey sampling literature to estimate the log-likelihood unbiasedly. Broadly speaking, this estimator subtracts an approximation of each log-likelihood contribution from each log-likelihood contribution to obtain a new population with elements that are roughly of the same size, thereby avoiding the need for PPS sampling. We propose to use the difference estimator in a delayed acceptance MCMC setting. The variance of the resulting likelihood estimate is much smaller compared to the estimator used by Payne and Mallick (2014). Consequently, our method is more effective in filtering out proposals with a low acceptance probability and promoting good proposals to the second stage.

The delayed acceptance MCMC needs to compute the full data likelihood whenever the first stage is passed, which makes it unsuitable for data sets too large to fit in RAM. Payne and Mallick (2014) combine their algorithm with the consensus Monte Carlo in Scott et al. (2013) to overcome this issue. The consensus Monte Carlo samples from an approximate posterior and currently lacks any theoretical guarantees. To handle extremely large data sets we instead propose to replace the true likelihood evaluation in the second stage with an estimate. We call this method delayed acceptance PMCMC which, like the consensus Monte Carlo, samples from an approximate posterior. However, the theoretical framework developed in Quiroz et al. (2015a) can straightforwardly be applied to prove that the approximate posterior is within $O(m^{-1/2})$ of the true posterior, where m is the size of the subsample used for estimation in the second stage.

This paper is organized as follows. Section 2 outlines the methodology and its extension to the so called big data setting. Section 3 applies the method to a micro-economic data set containing nearly 5 million observations. Section 4 concludes and Appendix A proves Theorem 1.

2. Methodology

2.1. **Delayed acceptance MCMC.** The delayed acceptance MCMC was initially developed in Christen and Fox (2005) for inference in computationally expensive inverse problems. Payne and Mallick (2014) realize the potential of using this framework to speed up computations in the large data setting.

Let θ denote the vector of parameters. Let $p(y|\theta)$ and $p(\theta)$ denote the likelihood and prior, respectively, and we often suppress dependence on covariates for notational clarity. The aim is to design an MCMC algorithm which admits the posterior

$$\pi(\theta) = \frac{p(y|\theta)p(\theta)}{p(y)}, \quad \text{with} \quad p(y) = \int p(y|\theta)p(\theta)d\theta,$$

as invariant distribution. Moreover, the likelihood $p(y|\theta)$ should only be evaluated if there is a good chance of accepting the proposed θ.

Let $\hat{p}(y|\theta, v)$ be an approximation of $p(y|\theta)$ based on a subsample of the data represented by v. We discuss $\hat{p}(y|\theta, v)$ in detail in Section 2.2. The algorithm in Payne and Mallick (2014) proceeds as follows. Let θ_c denote the current state of the Markov chain. In the first stage, propose $\theta' \sim q_1(\cdot|\theta_c)$ and compute

$$\alpha_1(\theta_c, \theta') = \min\left\{1, \frac{\hat{p}(y|\theta', v)p(\theta')/q_1(\theta'|\theta_c)}{\hat{p}(y|\theta_c, v)p(\theta_c)/q_1(\theta_c|\theta')}\right\}. \tag{2.1}$$

Now, propose

$$\theta_p = \begin{cases} \theta' & \text{w.p.} \quad \alpha_1(\theta_c, \theta') \\ \theta_c & \text{w.p.} \quad 1 - \alpha_1(\theta_c, \theta'), \end{cases}$$

and accept to move the chain to the next state $\theta_i = \theta_p$ with probability

$$\alpha_2(\theta_c, \theta_p) \quad = \quad \min\left\{1, \frac{p(y|\theta_p)p(\theta_p)/q_2(\theta_p|\theta_c)}{p(y|\theta_c)p(\theta_c)/q_2(\theta_c|\theta_p)}\right\}, \tag{2.2}$$

where

$$q_2(\theta_p|\theta_c) \quad = \quad \alpha_1(\theta_c, \theta_p)q_1(\theta_p|\theta_c) + r(\theta_c)\delta_{\theta_c}(\theta_p), \quad r(\theta_c) = 1 - \int \alpha_1(\theta_c, \theta_p)q_1(\theta_p|\theta_c)d\theta_p,$$

and δ is the Dirac delta function. If rejected we set $\theta_i = \theta_c$.

The transition kernel of the Markov chain generated by this algorithm is

$$T(\theta_c \to d\theta_p) \quad = \quad T(\theta_c \to \theta_p)d\theta_p + \tilde{r}(\theta_c)\delta_{\theta_c}(d\theta_p)$$

where

$$T(\theta_c \to \theta_p) = q_2(\theta_p|\theta_c)\alpha_2(\theta_c, \theta_p), \quad \tilde{r}(\theta_c) = 1 - \int T(\theta_c \to \theta_p)d\theta_p,$$

and $\delta_{\theta_c}(d\theta_p) = 1$ if $\theta_c \in d\theta_p$, and zero otherwise.We now show that $T(\theta_c \to \theta_p)$ satisfies the detailed balance condition and therefore $\pi(\theta)$ is the invariant distribution (Chib and Greenberg, 1995). In fact, since

$$\alpha_2(\theta_c, \theta_p) \quad = \quad \frac{\pi(\theta_p)q_1(\theta_c|\theta_p)\alpha_1(\theta_p, \theta_c)}{\pi(\theta_c)q_1(\theta_p|\theta_c)\alpha_1(\theta_c, \theta_p)}\alpha_2(\theta_p, \theta_c),$$

we get

$$\begin{aligned}
\pi(\theta_c)T(\theta_c \to \theta_p) \quad &= \quad \pi(\theta_c)\alpha_1(\theta_c, \theta_p)q_1(\theta_p|\theta_c)\alpha_2(\theta_c, \theta_p) \\
&= \quad \pi(\theta_p)\alpha_1(\theta_p, \theta_c)q_1(\theta_c|\theta_p)\alpha_2(\theta_p, \theta_c) \\
&= \quad \pi(\theta_p)T(\theta_p \to \theta_c).
\end{aligned}$$

Note that when $\theta_p = \theta_c$ then $\alpha_2(\theta_c, \theta_p) = 1$. Otherwise, it can be shown that (Result 1 in Payne and Mallick, 2014)

$$\alpha_2(\theta_c, \theta_p) \quad = \quad \min\left\{1, R_m = \frac{\hat{p}(y|\theta_c, v)/p(y|\theta_c)}{\hat{p}(y|\theta_p, v)/p(y|\theta_p)}\right\}, \tag{2.3}$$

where we introduce the dependence on the sample size m. We note from Equation (2.3) that if $\hat{p}(y|\theta_c, v)$ and $\hat{p}(y|\theta_p, v)$ are good approximations of $p(y|\theta_c)$ and $p(y|\theta_p)$, respectively, then $\alpha_2(\theta_c, \theta_p)$ will be close to 1 and the algorithm is efficient (it evaluates the full data set only for good proposals). The likelihood estimators are discussed in the next subsection but we already now state the following theorem, which implies that an algorithm with a more accurate estimator of R_m will (on average) result in a higher $\alpha_2(\theta_c, \theta_p)$.

Theorem 1. *Suppose that we have two delayed acceptance algorithms with the ratios in* (2.3) *denoted by*

$$R_m^{(1)} = \frac{\hat{p}^{(1)}(y|\theta_c, v)/p(y|\theta_c)}{\hat{p}^{(1)}(y|\theta_p, v)/p(y|\theta_p)} \quad \textit{and} \quad R_m^{(2)} = \frac{\hat{p}^{(2)}(y|\theta_c, v)/p(y|\theta_c)}{\hat{p}^{(2)}(y|\theta_p, v)/p(y|\theta_p)},$$

where $\hat{p}^{(i)}$ is the likelihood estimator for the ith algorithm. Let

$$\sigma_1^2 = V_v[\log(R_m^{(1)})], \quad \sigma_2^2 = V_v[\log(R_m^{(2)})], \quad \textit{and assume} \quad \sigma_1^2 < \sigma_2^2.$$

Then, asymptotically in m,

$$E_v[\alpha_2^{(1)}(\theta_c, \theta_p)] \quad > \quad E_v[\alpha_2^{(2)}(\theta_c, \theta_p)],$$

where $\alpha_2^{(i)}(\theta_c, \theta_p)$ denotes the acceptance probability in the second stage for the algorithm with ratio $R_m^{(i)}$.

Proof. See Appendix A. □

Lemma 1 in Appendix A derives σ^2 for the difference estimator and the estimator in Payne and Mallick (2014). We illustrate in our application that, for a given sample size m, σ^2 is lower for the difference estimator, and hence it is a more efficient algorithm (with respect to α_2) by Theorem 1.

2.2. **Likelihood approximators.** Consider a model parametrized by $p(y_k|\theta, x_k)$, where y_k is a potentially multivariate response vector and x_k is a vector of covariates for the kth observation. Let $l_k(\theta) = \log p(y_k|\theta, x_k)$ denote the kth observation's log-density, $k = 1, \ldots, n$.

Given conditionally independent observations, the likelihood function can be written

$$p(y|\theta) = \exp\left[l(\theta)\right], \tag{2.4}$$

where $l(\theta) = \sum_{k=1}^{n} l_k(\theta)$ is the log-likelihood function. This setting is more general than iid. observations, although we require that the log-likelihood can be written as a sum of terms where each term depends on a unique piece of data information.

To approximate $p(y|\theta)$ we estimate $l(\theta)$ based on a subsample by methods from survey sampling and use Equation (2.4). See Särndal et al. (2003) for an introduction to survey sampling. Let $F = (1, \ldots, n)$ denote the indices of the full population and define $v = (v_1, \ldots, v_n)$, where $v_k = 1$ if observation k is included in the subsample and $v_k = 0$ otherwise. Let S denote the set of indices in the subsample with $|S| = \sum_{k=1}^{n} v_k = m$. With simple random sampling without replacement (SI) we have

$$\Pr(v_k = 1) = \frac{m}{n} \quad \text{for} \quad k \in F.$$

We can obtain an unbiased estimate of $l(\theta)$ by the Horvitz-Thompson (H-T) estimator (Horvitz and Thompson, 1952), which under SI is

$$\hat{l}_m(\theta) = \frac{n}{m} \sum_{k \in S} l_k(\theta), \quad \text{with} \quad V[\hat{l}_m(\theta)] = n^2 \frac{(1-f)}{m} s_F^2, \tag{2.5}$$

where $f = m/n$ is the sampling fraction and $s_F^2 = \frac{1}{n-1} \sum_{k \in F} (l_k(\theta) - \bar{l}_F(\theta))^2$ with obvious notation. The likelihood approximator in Payne and Mallick (2014) is

$$\hat{p}_{pm}(y|\theta) \;=\; \exp(\hat{l}_m(\theta)), \quad \text{with} \quad \hat{l}_m(\theta) \text{ as in (2.5).} \tag{2.6}$$

We now turn to the difference estimator in Quiroz et al. (2015b) which we propose to use in the likelihood approximator. Let $w_k(\theta)$ denote an approximation of $l_k(\theta)$ and decompose

$$\begin{aligned} l(\theta) &= \sum_{k \in F} w_k(\theta) + \sum_{k \in F} \left[l_k(\theta) - w_k(\theta)\right] \\ &= w + d, \end{aligned}$$

where

$$w = \sum_{k \in F} w_k(\theta), \quad d = \sum_{k \in F} d_k(\theta), \quad \text{and} \quad d_k(\theta) = l_k(\theta) - w_k(\theta).$$

Here w is known prior to sampling and we estimate d with the H-T estimator

$$\hat{d}_m(\theta) = \frac{n}{m} \sum_{k \in S} d_k(\theta), \quad \text{with} \quad V[\hat{d}_m(\theta)] = n^2 \frac{(1-f)}{m} s_F^2 \tag{2.7}$$

and $s_F^2 = \frac{1}{n-1} \sum_{k \in F} (d_k(\theta) - \bar{d}_F(\theta))^2$. The difference estimator is

$$\hat{l}_m(\theta) \;\; = \;\; w + \hat{d}_m(\theta) \tag{2.8}$$

and the likelihood approximator becomes

$$\hat{p}_{de}(y|\theta) \;\; = \;\; \exp(w + \hat{d}_m(\theta)), \quad \text{with} \quad \hat{d}_m(\theta) \text{ as in (2.7).} \tag{2.9}$$

SI usually gives a huge variance of H-T and many other estimators. The difference estimator omits this problem because, since $w_k(\theta)$ is an approximation of $l_k(\theta)$, $l_k(\theta) - w_k(\theta)$ should be roughly of the same size for all $k \in F$. We follow Quiroz et al. (2015b) and set $w_k(\theta)$ to a Taylor series approximation of $l_k(\theta)$. Moreover, to overcome the $O(n)$ complexity of computing w, we obtain a sparse set of the data through local data clusters, see Quiroz et al. (2015b) for details.

The approximators in (2.6) and (2.9) differ from the class of estimators considered in Quiroz et al. (2015a) on two aspects. First, they are not bias-corrected. This correction is not needed in the delayed acceptance setting because the final acceptance decision is based on the true likelihood. Second, the sampling is without replacement. The reason Quiroz et al. (2015a) use with replacement is to facilitate the derivation of explicit upper bounds of the error in the approximation. The delayed acceptance method is exact and we choose without replacement because it gives a smaller variance of the estimator (Särndal et al., 2003). However, Theorem 1 is proved under the assumption of with replacement sampling. When $m << n$ this provides a good approximation of the corresponding without replacement sampling.

2.3. **Delayed acceptance PMCMC.** For data sets to large to fit in RAM, Payne and Mallick (2014) suggest to combine the delayed acceptance algorithm with the consensus Monte Carlo (Scott et al., 2013). At present, there are no theoretical results to assess the errors in the consensus method.

As an alternative we propose to combine the delayed acceptance algorithm with the pseudo-marginal framework for data subsampling initially developed in Quiroz et al. (2015a), which we call delayed acceptance PMCMC. The method replaces the true likelihood evaluation in the second stage of the delayed acceptance with an estimator based on $\tilde{m}$ observations. In the first stage an approximation of this estimator is computed using $m < \tilde{m}$ observations to determine if the proposed draw is likely to pass the second stage.

Let u be a vector of auxiliary variables corresponding to the subset of observations to include when estimating $p(y|\theta)$ in the second stage. Let $\hat{p}_{\tilde{m}}(y|\theta, u)$ denote a biased estimator of $p(y|\theta)$ with expectation

$$p_{\tilde{m}}(y|\theta) \quad = \quad \int \hat{p}_{\tilde{m}}(y|\theta, u)p(u)du. \tag{2.10}$$

The sampling is now on the augmented space (θ, u), targeting the posterior

$$\tilde{\pi}_{\tilde{m}}(\theta, u) = \hat{p}_{\tilde{m}}(y|\theta, u)p(u)p(\theta)/p_{\tilde{m}}(y), \text{ with } p_{\tilde{m}}(y) = \int p_{\tilde{m}}(y|\theta)p(\theta)d\theta. \tag{2.11}$$

We follow Quiroz et al. (2015b) and use the estimator

$$\hat{p}_{\tilde{m}}(y|\theta, u) \quad = \quad \exp\left(\hat{l}_{\tilde{m}} - \hat{\sigma}_z^2/2\right), \tag{2.12}$$

where $\hat{l}_{\tilde{m}}$ is similar as in (2.8) but using with replacement sampling of $\tilde{m}$ observations, $z = \hat{l} - l$ is the estimation error and $\hat{\sigma}_z^2$ is an unbiased estimate of $\hat{\sigma}_z^2 = V[z]$. Quiroz et al. (2015b) outline in detail how to sample from (2.11) which here constitutes the second stage in the delayed acceptance PMCMC. In the first stage, we use an approximation of $\hat{p}_{\tilde{m}}(y|\theta, u)$ which we denote $\hat{p}(y|\theta, u, v)$ and corresponds to $\hat{p}(y|\theta, v)$ in Section 2.2. Applying the same computations as in Section 2.1, but for the transition kernel

$$T\left\{(\theta_c, u_c) \to (d\theta_p, du_p)\right\}$$

on the augmented space, it is straightforward to show that the detailed balance condition is fulfilled for $\tilde{\pi}_{\tilde{m}}(\theta, u)$. Thus, $\tilde{\pi}_{\tilde{m}}(\theta, u)$ is the invariant distribution with the perturbed posterior

$$\pi_{\tilde{m}}(\theta) \quad = \quad \int \tilde{\pi}_{\tilde{m}}(\theta, u) du$$

as marginal distribution. Quiroz et al. (2015a) prove that, for a particular class of estimators, $\pi_{\tilde{m}}(\theta)$ is within $O(\tilde{m}^{-1/2})$ of $\pi(\theta)$ and derive expressions for the upper bound of the error in the approximation. The difference estimator in (2.12) belongs to this class (Quiroz et al., 2015b) and therefore the delayed acceptance PMCMC is theoretically justified by Theorem 1 in Quiroz et al. (2015a).

3. Application

3.1. **Data and model.** We model the probability of bankruptcy conditional on a set of covariates using a data set of $534,717$ Swedish firms for the time period 1991-2008. We have in total $n = 4,748,089$ firm-year observations. The variables included are: earnings before interest and taxes, total liabilities, cash and liquid assets, tangible assets, logarithm of deflated total sales and logarithm of firm age in years. We also include the macroeconomic variables GDP-growth rate (yearly) and the interest rate set by the Swedish central bank. See Giordani et al. (2014) for a detailed description of the data set.

We consider the logistic regression model

$$p(y_k|x_k, \beta) \quad = \quad \left(\frac{1}{1+\exp(x_k^T\beta)}\right)^{y_k} \left(\frac{1}{1+\exp(-x_k^T\beta)}\right)^{1-y_k},$$

where x_k includes the variables above plus an intercept term. We set $p(\beta) \sim N(0, 10I)$ for simplicity.

3.2. **Performance evaluation.** The Inefficiency Factor (IF), or the integrated autocorrelation time, is defined as

$$IF = 1 + 2\sum_{l=1}^{\infty} \rho_l, \tag{3.1}$$

where ρ_l is the autocorrelation at the lth lag of the chain. We estimate IF using the CODA package in R (Plummer et al., 2006). IF measures the number of draws required to obtain the equivalent of a single independent draw.

We evaluate the performance using the Effective Draws (ED)

$$ED = \frac{N}{IF \times t}, \tag{3.2}$$

where N is the number of MCMC iterations and t is the computing time. The measure of interest is the effective draws of delayed acceptance (DMCMC) relative to that of standard MCMC, i.e.

$$RED = \frac{ED^{DMCMC}}{ED^{MCMC}}. \tag{3.3}$$

Our method and also Payne and Mallick (2014) require some additional computations compared with the standard M-H algorithm (e.g. draw a subsample, construct the approximations used by the difference estimator). These computations should ideally be implemented in a low-level language such as C, as opposed to our current implementation in Python. We therefore also provide a measure that is independent of the implementation, where t in (3.2) is replaced by the average number of density evaluations. This measure provides an estimate of the potential speedup gain in an ideal programming environment.

3.3. **Implementation details.** The model is estimated with the delayed MCMC algorithm using the difference estimator and the estimator in Payne and Mallick (2014). Both methods are compared to the standard M-H algorithm.

In correspondence with the authors we found that an alternative implementation of the algorithm in Section 2.1 is used in Payne and Mallick (2014), where the denominator $\hat{p}(y|\theta_c, v)$

in (2.1) is replaced by $p(y|\theta_c)$. The detailed balance is still satisfied with this implementation. Furthermore, Payne and Mallick (2014) generate a new v in each iteration to estimate the numerator in (2.1). When both numerator and denominator are estimated as in our implementation of their method, it is important that the estimates (in a given iteration) use the same subset of observations (i.e. same v) as the variance of the ratio becomes much smaller in this case. We find that our implementation of the algorithm in Payne and Mallick (2014) is more efficient for our application (not reported here). We will therefore use this implementation for comparison.

Since the bankruptcy observations ($y_k = 1$) are sparse in the data we follow Payne and Mallick (2014) and estimate the likelihood only for the $y_k = 0$ observations. That is, we decompose

$$l(\beta) = \sum_{\{k;y_k=1\}} l_k(\beta) + \sum_{\{k;y_k=0\}} l_k(\beta),$$

and evaluate the first term whereas a random sample is only taken to estimate the second term.

We consider a Random walk M-H proposal for β. The proposal covariance is obtained as follows. We optimize on a subsample of $n_{sub} = 10,000$ observations and compute the inverse Hessian at the optimum β^*. Our experience is that if the off-diagonal elements of the covariance of the posterior based on the subset is not in agreement to that of the full data posterior then the proposal distribution can be very poor. Therefore we set the off-diagonal elements to zero and the diagonal elements are scaled with n_{sub}/n so that the proposal has the same scale as the full data posterior. Finally, the proposal covariance is multiplied with $2.38/\sqrt{d}$ (Roberts et al., 1997) where d is the number of parameters. All algorithms use the same proposal distribution and starting value β^*.

Two main implementations of the difference estimator are considered. The first computes w_k with the second order term evaluated at β, which we call *dynamic*. The second, which we call *static*, fixes the second order term at the optimum β^*. The dynamic approach clearly provides a better approximation but is more expensive to compute. For both the dynamic and static approaches we compare four different sparse representations of the data for computing w in (2.8), each with a different number of clusters. The clusters are obtained using Algorithm 1 in Quiroz et al. (2015b) on the observations for which $y = 0$ ($4,706,523$ observations). We note that, as more clusters are used to represent the data, the approximation of the likelihood is more accurate but also more expensive to compute.

For all algorithms we sample $N = 205,000$ draws from the posterior and discard $5,000$ as burn-in. The delayed acceptance algorithms are implemented with an update of v every 100th iteration.

3.4. **Results.** Table 1 and 2 summarize the results, respectively, for the difference estimator and the estimator in Payne and Mallick (2014). It is evident that the difference estimator has a larger second stage acceptance probability α_2 (for a given sample size), which is a consequence of Theorem 1 because it has a lower $\sigma^2 = V[\log(R_m)]$. We also note from Table 2 that for some sample sizes Payne and Mallick (2014) performs more poorly than the standard Metropolis-Hastings algorithm. One possible explanation is that the applications in Payne and Mallick (2014) have a small number of continuous covariates (one in the first application and three in the second) and the rest are binary. It is clear that the continuous covariate case results in more variation among the log-likelihood contributions which is detrimental for SI. In this application we have eight continuous covariates which explains why SI performs poorly for small sampling fractions.

To facilitate comparison between the methods, Figure 1 shows the fraction of relative effective draws between the difference estimator (with a particular approximation; see the caption) and the estimator in Payne and Mallick (2014). In terms of the average number of density evaluations our method is superior for all cases. If execution time is considered instead, for relatively large sample sizes the improvement is not so pronounced which is mostly attributed to the implementation in a high-level language (Python). However, accurate posterior estimators are achieved with small sample sizes for the difference estimator (see the next paragraph), and for these cases our method is superior with respect to execution time as well.

TABLE 1. *Delayed acceptance MCMC with the difference estimator.* The table shows some quantities for the static and dynamic implementation with different sparse representations of the data represented by K, which is the number of clusters (expressed as % of n). For each approximation different sample sizes ($0.1, 1, 5$ in % of n) are considered. The quantities are the mean RED_1 and RED_2 in (3.3) measured with respect to computing time and average density evaluations, respectively. Furthermore, $\bar{\sigma}$ is the mean (over MCMC iterations) standard deviation of $\log(R_m)$ (see σ_1^2 in part (i) of Lemma 1). Finally, α_1 and α_2 are the acceptance probabilities in (2.1) and (2.3) (expressed in %), where the latter is computed conditional on acceptance in the first stage. The results of the most efficient algorithms with respect to RED_1 (black boldface) and RED_2 (red italic) are highlighted. The standard M-H algorithm has an acceptance rate of 14%.

	Static					**Dynamic**				
	RED_1	RED_2	$\bar{\sigma}$	α_1	α_2	RED_1	RED_2	$\bar{\sigma}$	α_1	α_2
$K = 0.03$										
0.1	0.79	0.80	6.63	28	12	1.81	2.25	2.81	18	36
1	2.26	2.90	2.14	17	46	1.34	4.77	0.90	14	73
5	1.54	3.80	0.96	14	73	0.37	4.56	0.40	14	88
$K = 0.21$										
0.1	1.60	1.69	3.75	21	26	**2.80**	**4.45**	**1.14**	**15**	**67**
1	3.01	4.08	1.20	15	66	1.55	5.80	0.37	14	89
5	1.74	4.39	0.54	14	84	0.39	5.01	0.16	14	95
$K = 0.71$										
0.1	2.25	2.67	2.35	17	44	2.25	5.43	0.57	14	83
1	**3.24**	**4.78**	**0.74**	**14**	**78**	*1.27*	*5.92*	*0.18*	*14*	*95*
5	1.77	4.64	0.33	14	90	0.38	5.01	0.08	14	98
$K = 3.68$										
0.1	2.15	4.01	1.02	14	70	0.73	5.33	0.18	14	95
1	*2.28*	*4.79*	*0.33*	*14*	*90*	0.59	5.23	0.06	14	98
5	1.39	4.26	0.15	14	95	0.28	4.47	0.03	14	99

Figure 2 shows the kernel density estimates of the marginal posterior distributions for four parameters (to save space). The posteriors are estimated using different sample sizes for the following cases: (i) The difference estimator implemented with the dynamic and static second order approximation using $K = 0.71$ (see Table 1). (ii) The algorithm in Payne and Mallick (2014). All the panels include the standard M-H algorithm for comparison. Recall that the delayed acceptance is exact regardless of the sample size for estimating the likelihood.

Table 2. *Delayed acceptance MCMC with the Payne and Mallick (2014) estimator.* The table shows some quantities for different sample sizes $(0.1, 1, 5, 50, 80,$ in % of $n)$ to approximate the likelihood. The quantities RED_1 and RED_2 in (3.3) measured with respect to computing time and average density evaluations, respectively. Furthermore, $\bar{\sigma}$ is the mean (over MCMC iterations) standard deviation of $\log(R_m)$ (see σ_2^2 in part (ii) of Lemma 1). Finally, α_1 and α_2 are the acceptance probabilities in (2.1) and (2.3) (expressed in %), where the latter is computed conditional on acceptance in the first stage. The results of the most efficient algorithm with respect to RED_1 and RED_2 are marked in boldface (they coincide). The standard M-H algorithm has an acceptance rate of 14%.

	RED_1	RED_2	$\bar{\sigma}$	α_1	α_2
0.1	0.18	0.18	24.82	42	3
1	0.65	0.66	7.91	30	10
5	**1.61**	**1.64**	**3.52**	**20**	**30**
50	1.10	1.34	1.12	14	77
80	0.88	1.07	0.89	14	89

However, the sample size clearly affects the effective draws (see Tables 1 and 2) and this is also evident in the figure, in particular for the estimator in Payne and Mallick (2014) with small sample sizes.

4. Conclusions

We explore the use of the efficient and robust difference estimator in a delayed acceptance MCMC setting. The estimator incorporates auxiliary information about the contribution to the log-likelihood function while keeping the computational complexity low by operating on a sparse set of the data.

In an application to modeling of firm-bankruptcy, we find that the proposed delayed acceptance algorithm is more efficient than both the algorithm proposed by Payne and Mallick (2014) and the standard M-H algorithm. Moreover, we prove that our method is asymptotically better, as measured by the probability of accepting the second stage conditional that the first stage was accepted.

The inevitable step of scanning the complete data when deciding upon final acceptance makes any delayed MCMC algorithm infeasible when facing data sets to large to fit in RAM.

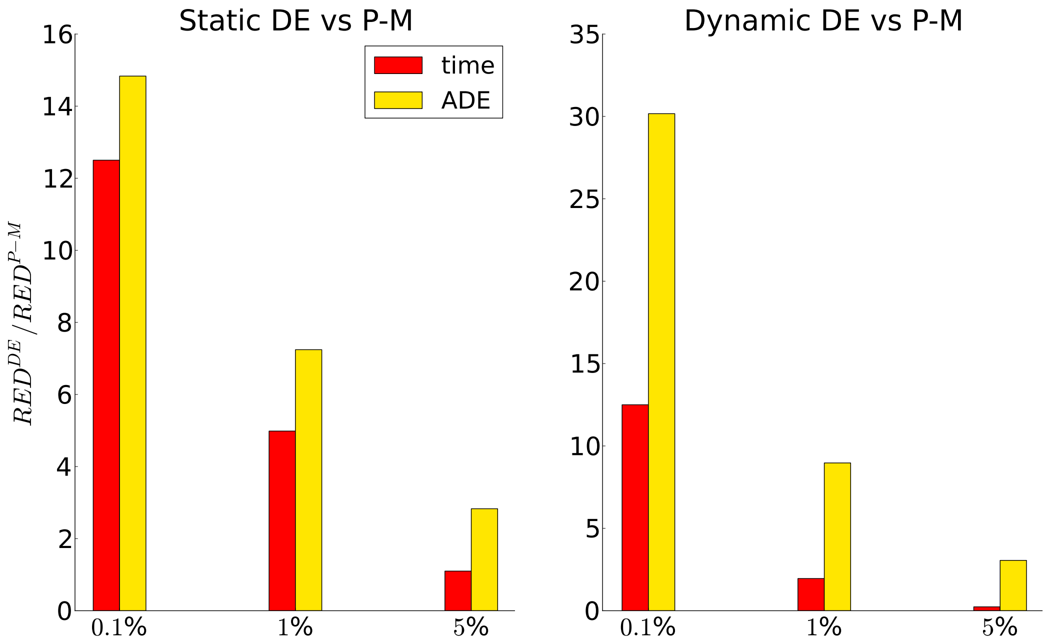

FIGURE 1. *Comparing relative effective draws.* The figure shows the fraction RED^{DE}/RED^{P-M} for sample sizes $m = 0.1, 1, 5$ (in % of n), where DE and P-M denotes the difference estimator and the estimator in and Payne and Mallick (2014), respectively. The fraction is computed with respect to the measures time (red bars) and Average Density Evaluations (ADE, yellow bars). The left panel shows the result for the static difference estimator, whereas the right panel shows the corresponding for the dynamic case. Both difference estimators use an approximation with $K = 0.71$ (% of n) number of clusters.

As an alternative to the solution of combining with the consensus Monte Carlo proposed by Payne and Mallick (2014), we propose delayed acceptance PMCMC which utilizes an estimated likelihood based on a subsample of size $\tilde{m}$ in the final acceptance step. We make the connection to previous literature transparent and it follows that the delayed PMCMC converges to the true posterior as $\tilde{m}$ increases. Moreover, the upper bound of the error can be addressed directly by results in previous work. This is an attractive feature that the consensus Monte Carlo approach currently lacks.

References

Banterle, M., Grazian, C., and Robert, C. P. (2014). Accelerating Metropolis-Hastings algorithms: Delayed acceptance with prefetching. *arXiv preprint arXiv:1406.2660*.

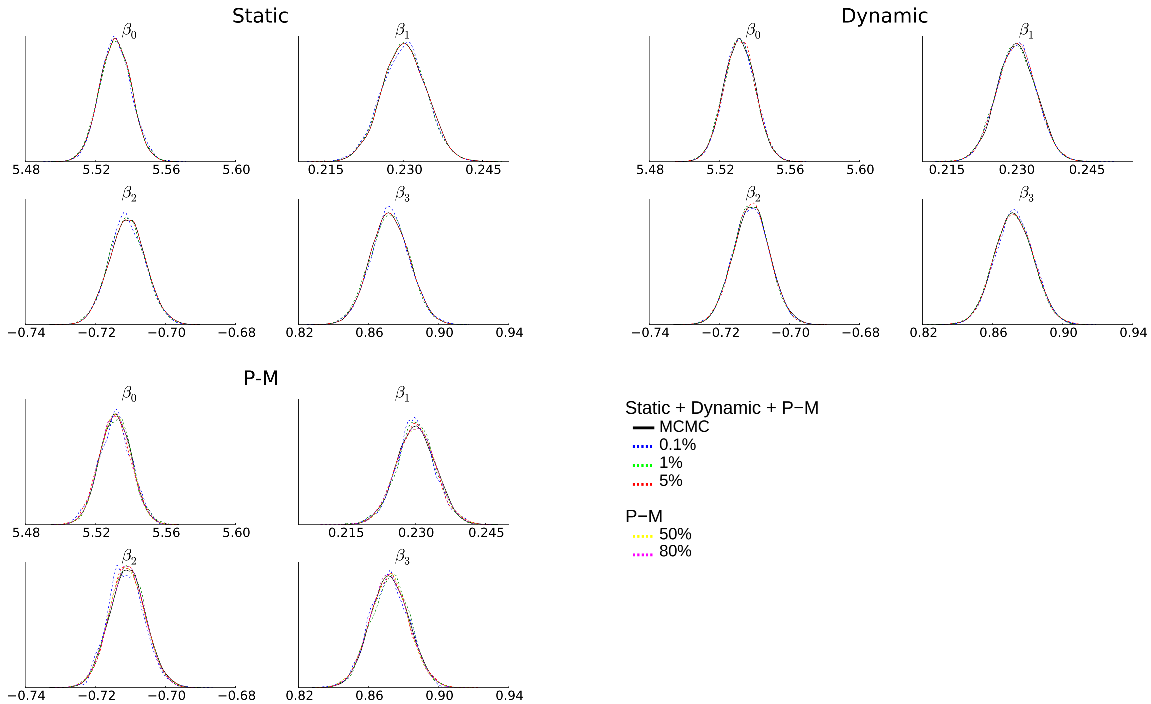

FIGURE 2. *Kernel density estimations of marginal posteriors.* The figure shows the marginal posteriors of four parameters obtained with different algorithms for some subsample sizes expressed as % of n (dashed colored lines) and the standard M-H (solid black line). The upper-left and upper-right panels show, respectively, the difference estimator with static and dynamic second order term in the approximation. The approximations are based on $K = 0.71$ (expressed as % of n) number of clusters. The lower-left panel shows the algorithm in Payne and Mallick (2014) (P-M).

Bardenet, R., Doucet, A., and Holmes, C. (2014). Towards scaling up Markov chain Monte Carlo: an adaptive subsampling approach. In *Proceedings of The 31st International Conference on Machine Learning*, pages 405–413.

Bardenet, R., Doucet, A., and Holmes, C. (2015). On Markov chain Monte Carlo methods for tall data. *arXiv preprint arXiv:1505.02827*.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335.

Christen, J. A. and Fox, C. (2005). MCMC using an approximation. *Journal of Computational and Graphical Statistics*, 14(4):795–810.

Giordani, P., Jacobson, T., Schedvin, E. v., and Villani, M. (2014). Taking the twists into account: Predicting firm bankruptcy risk with splines of financial ratios. *Journal of Financial and Quantitative Analysis*, 49(04):1071–1099.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

Korattikara, A., Chen, Y., and Welling, M. (2013). Austerity in MCMC land: Cutting the Metropolis-Hastings budget. *arXiv preprint arXiv:1304.5299*.

Maclaurin, D. and Adams, R. P. (2014). Firefly Monte Carlo: Exact MCMC with subsets of data. *arXiv preprint arXiv:1403.5693*.

Maire, F., Friel, N., and Alquier, P. (2015). Light and widely applicable MCMC: Approximate Bayesian inference for large datasets. *arXiv preprint arXiv:1503.04178*.

Minsker, S., Srivastava, S., Lin, L., and Dunson, D. (2014). Scalable and robust Bayesian inference via the median posterior. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1656–1664.

Neiswanger, W., Wang, C., and Xing, E. (2013). Asymptotically exact, embarrassingly parallel MCMC. *arXiv preprint arXiv:1311.4780*.

Payne, R. D. and Mallick, B. K. (2014). Bayesian big data classification: A review with complements. *arXiv preprint arXiv:1411.5653v2*.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1):7–11.

Quiroz, M., Villani, M., and Kohn, R. (2015a). Speeding up MCMC by efficient data subsampling. *arXiv preprint arXiv:1404.4178v2*.

Quiroz, M., Villani, M., and Kohn, R. (2015b). Fast and efficient MCMC for large data problems using data subsampling and the difference estimator. *Manuscript*.

Roberts, G. O., Gelman, A., Gilks, W. R., et al. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.

Särndal, C.-E., Swensson, B., and Wretman, J. (2003). *Model assisted survey sampling.* Springer.

Scott, S. L., Blocker, A. W., Bonassi, F. V., Chipman, H., George, E., and McCulloch, R. (2013). Bayes and big data: the consensus Monte Carlo algorithm. In *EFaBBayes 250" conference*, volume 16.

Wang, X. and Dunson, D. B. (2013). Parallel MCMC via Weierstrass sampler. *arXiv preprint arXiv:1312.4605.*

Appendix A. Proof of Theorem 1

Let $l_k(\theta_c, \theta_p) = l_k(\theta_c) - l_k(\theta_p)$ be the difference in log-likelihood contribution of observation k at the current and proposed θ. Denote by

$$\hat{l}_m^{(j)}(\theta_c, \theta_p), \quad j = 1, 2,$$

the estimates of $l(\theta_c, \theta_p)$ based on a sample of size m (see Definition 1 below). Denote the ratios corresponding to (2.3) by

$$R_m^{(j)} = \exp\left(\hat{l}_m^{(j)}(\theta_c, \theta_p) - l(\theta_c, \theta_p)\right), \quad j = 1, 2. \tag{A.1}$$

Definition 1. *Consider simple random sampling with replacement and define the following estimators of $l(\theta_c, \theta_p)$ based on a sample of size m:*

i. *The difference estimator*:

$$\hat{l}_m^{(1)}(\theta_c, \theta_p) \quad = \quad w(\theta_c, \theta_p) + \frac{1}{m}\sum_{i=1}^{m}\zeta_i, \quad \textit{with } w(\theta_c, \theta_p) = \sum_{k=1}^{n} w_k(\theta_c, \theta_p),$$

where $w_k(\theta_c, \theta_p) = w_k(\theta_c) - w_k(\theta_p)$ ($w_k(\cdot)$ is an approximation of $l_k(\cdot)$) and the ζ_i's are iid. with

$$\Pr\left(\zeta_i = n\left(l_k(\theta_c, \theta_p) - w_k(\theta_c, \theta_p)\right)\right) = 1/n, \quad \textit{for } i = 1, \ldots m.$$

ii. *The estimator in Payne and Mallick (2014):*

$$\hat{l}_m^{(2)}(\theta_c, \theta_p) = \frac{1}{m} \sum_{i=1}^{m} \eta_i,$$

where the η_i's are iid. with

$$\Pr\left(\eta_i = n l_k(\theta_c, \theta_p)\right) = 1/n, \quad \textit{for } i = 1, \dots m.$$

Lemma 1. *The following results hold for the estimators in Definition 1:*

i.

$$E[\hat{l}_m^{(1)}(\theta_c, \theta_p)] = l(\theta_c, \theta_p) \quad \textit{and} \quad \sigma_1^2 = V[\hat{l}_m^{(1)}(\theta_c, \theta_p)] = \frac{\sigma_\zeta^2}{m},$$

where

$$\sigma_\zeta^2 = n \sum_{k \in F} \left(D_k(\theta_c, \theta_p) - \bar{D}_F(\theta_c, \theta_p)\right)^2 \quad \textit{with} \quad D_k = l_k(\theta_c, \theta_p) - w_k(\theta_c, \theta_p)$$

and $\bar{D}_F$ denotes the mean over the population.

ii.

$$E[\hat{l}_m^{(2)}(\theta_c, \theta_p)] = l(\theta_c, \theta_p) \quad \textit{and} \quad \sigma_2^2 = V[\hat{l}_m^{(2)}(\theta_c, \theta_p)] = \frac{\sigma_\eta^2}{m},$$

where

$$\sigma_\eta^2 = n \sum_{k \in F} \left(l_k(\theta_c, \theta_p) - \bar{l}_F(\theta_c, \theta_p)\right)^2.$$

Proof. The proofs are straightforward and are therefore omitted. □

Lemma 2. *The ratios in* (A.1) *have the following asymptotic (in terms of m) distributions:*

i.

$$R_m^{(1)} \sim \log \mathcal{N}\left(0, \sigma_1^2\right),$$

ii.

$$R_m^{(2)} \sim \log \mathcal{N}\left(0, \sigma_2^2\right).$$

Proof. Proof of (i): Define

$$A_m = \hat{l}_m^{(1)}(\theta_c, \theta_p) - l(\theta_c, \theta_p),$$

with (part (i) of Lemma 1)

$$E[A_m] = 0 \quad \text{and} \quad V[A_m] = \frac{\sigma_\zeta^2}{m}.$$

By the central limit theorem (the ζ_i's are iid.)

$$\sqrt{m}A_m \to \mathcal{N}\left(0, \sigma_\eta^2\right).$$

By the continuity of the exponential function it follows that

$$\exp\left(\sqrt{m}A_m\right) \quad \to \quad \log\mathcal{N}\left(0, \sigma_\eta^2\right).$$

Thus, since the power of a lognormal is lognormal, it follows that

$$R_m^{(1)} = \exp(A_m) \sim \log\mathcal{N}\left(0, \sigma_1^2 = \frac{\sigma_\eta^2}{m}\right),$$

which concludes (i). The proof of part (ii) is identical. □

Remark. Before proving the theorem, we note that $\sigma_1^2 < \sigma_2^2$ does not always hold. For example, if $w_k(\cdot)$ is a bad approximation of $l_k(\cdot)$, we can even have the opposite $\sigma_1^2 > \sigma_2^2$. However, it is not difficult to realize that if the approximations are good (i.e. $l_k(\cdot) - w_k(\cdot)$ is small) then $\sigma_\zeta^2 < \sigma_\eta^2$ and consequently $\sigma_1^2 < \sigma_2^2$. Theorem 1 is stated under this assumption.

Proof of Theorem 1. From Lemma 2 it follows that each estimator has a ratio that is asymptotically lognormal and depends only on the variance of the log-ratio (σ_1^2 or σ_2^2 in Lemma 1).

Consider the r.v. $X \sim \log\mathcal{N}(0, \sigma^2)$ with

$$f(x) \quad = \quad \frac{1}{x}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}\log(x)^2\right).$$

The expectation of the acceptance probability $\alpha_2(\theta_c, \theta_p)$ in (2.3) under X is

$$E[\min(1, X)] \quad = \quad \int_0^1 xf(x)dx + \int_1^\infty f(x)dx.$$

Since median$(X) = 1$ we obtain $\int_1^\infty f(x)dx = 0.5$. Now,

$$\begin{aligned}\int_0^1 xf(x)dx &= \int_0^1 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\log(x)^2\right)dx \\ &= \exp\left(\sigma^2/2\right)\int_{-\infty}^0 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\sigma^2)^2\right)dy,\end{aligned}$$

with $y = \log(x)$. The integrand is the pdf of $Y \sim \mathcal{N}(\sigma^2, \sigma^2)$ and thus

$$E[\min(1, X)] = \exp\left(\sigma^2/2\right)(1 - \Phi(\sigma)) + 0.5.$$

We now show that $E[\min(1, X)]$ is decreasing in σ. We have that

$$\frac{d}{d\sigma}E[\min(1, X)] = \exp\left(\sigma^2/2\right)\left(\sigma - \sigma\Phi(\sigma) - \frac{1}{\sqrt{2\pi}}\right),$$

and we can (numerically) compute the maximum of the right-most expression within brackets which is ≈ -0.23. Now, $\exp(\sigma^2/2) > 0$ so it follows that $\frac{d}{d\sigma}E[\min(1, X)] < 0$ and, since $\sigma_1^2 < \sigma_2^2$, we conclude that

$$E[\alpha_2^{(1)}(\theta_c, \theta_p)] > E[\alpha_2^{(2)}(\theta_c, \theta_p)].$$

$\square$

Earlier Working Papers:

For a complete list of Working Papers published by Sveriges Riksbank, see www.riksbank.se

Estimation of an Adaptive Stock Market Model with Heterogeneous Agents *by Henrik Amilon*	2005:177
Some Further Evidence on Interest-Rate Smoothing: The Role of Measurement Errors in the Output Gap *by Mikael Apel and Per Jansson*	2005:178
Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani	2005:179
Are Constant Interest Rate Forecasts Modest Interventions? Evidence from an Estimated Open Economy DSGE Model of the Euro Area *by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani*	2005:180
Inference in Vector Autoregressive Models with an Informative Prior on the Steady State *by Mattias Villani*	2005:181
Bank Mergers, Competition and Liquidity *by Elena Carletti, Philipp Hartmann and Giancarlo Spagnolo*	2005:182
Testing Near-Rationality using Detailed Survey Data *by Michael F. Bryan and Stefan Palmqvist*	2005:183
Exploring Interactions between Real Activity and the Financial Stance *by Tor Jacobson, Jesper Lindé and Kasper Roszbach*	2005:184
Two-Sided Network Effects, Bank Interchange Fees, and the Allocation of Fixed Costs *by Mats A. Bergman*	2005:185
Trade Deficits in the Baltic States: How Long Will the Party Last? *by Rudolfs Bems and Kristian Jönsson*	2005:186
Real Exchange Rate and Consumption Fluctuations follwing Trade Liberalization *by Kristian Jönsson*	2005:187
Modern Forecasting Models in Action: Improving Macroeconomic Analyses at Central Banks *by Malin Adolfson, Michael K. Andersson, Jesper Lindé, Mattias Villani and Anders Vredin*	2005:188
Bayesian Inference of General Linear Restrictions on the Cointegration Space *by Mattias Villani*	2005:189
Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model *by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani*	2005:190
Forecast Combination and Model Averaging using Predictive Measures *by Jana Eklund and Sune Karlsson*	2005:191
Swedish Intervention and the Krona Float, 1993-2002 *by Owen F. Humpage and Javiera Ragnartz*	2006:192
A Simultaneous Model of the Swedish Krona, the US Dollar and the Euro *by Hans Lindblad and Peter Sellin*	2006:193
Testing Theories of Job Creation: Does Supply Create Its Own Demand? *by Mikael Carlsson, Stefan Eriksson and Nils Gottfries*	2006:194
Down or Out: Assessing The Welfare Costs of Household Investment Mistakes *by Laurent E. Calvet, John Y. Campbell and Paolo Sodini*	2006:195
Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models *by Paolo Giordani and Robert Kohn*	2006:196
Derivation and Estimation of a New Keynesian Phillips Curve in a Small Open Economy *by Karolina Holmberg*	2006:197
Technology Shocks and the Labour-Input Response: Evidence from Firm-Level Data *by Mikael Carlsson and Jon Smedsaas*	2006:198
Monetary Policy and Staggered Wage Bargaining when Prices are Sticky *by Mikael Carlsson and Andreas Westermark*	2006:199
The Swedish External Position and the Krona *by Philip R. Lane*	2006:200

Price Setting Transactions and the Role of Denominating Currency in FX Markets *by Richard Friberg and Fredrik Wilander*	2007:201
The geography of asset holdings: Evidence from Sweden *by Nicolas Coeurdacier and Philippe Martin*	2007:202
Evaluating An Estimated New Keynesian Small Open Economy Model *by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani*	2007:203
The Use of Cash and the Size of the Shadow Economy in Sweden *by Gabriela Guibourg and Björn Segendorf*	2007:204
Bank supervision Russian style: Evidence of conflicts between micro- and macro-prudential concerns *by Sophie Claeys and Koen Schoors*	2007:205
Optimal Monetary Policy under Downward Nominal Wage Rigidity *by Mikael Carlsson and Andreas Westermark*	2007:206
Financial Structure, Managerial Compensation and Monitoring *by Vittoria Cerasi and Sonja Daltung*	2007:207
Financial Frictions, Investment and Tobin's q *by Guido Lorenzoni and Karl Walentin*	2007:208
Sticky Information vs Sticky Prices: A Horse Race in a DSGE Framework *by Mathias Trabandt*	2007:209
Acquisition versus greenfield: The impact of the mode of foreign bank entry on information and bank lending rates *by Sophie Claeys and Christa Hainz*	2007:210
Nonparametric Regression Density Estimation Using Smoothly Varying Normal Mixtures *by Mattias Villani, Robert Kohn and Paolo Giordani*	2007:211
The Costs of Paying – Private and Social Costs of Cash and Card *by Mats Bergman, Gabriella Guibourg and Björn Segendorf*	2007:212
Using a New Open Economy Macroeconomics model to make real nominal exchange rate forecasts *by Peter Sellin*	2007:213
Introducing Financial Frictions and Unemployment into a Small Open Economy Model *by Lawrence J. Christiano, Mathias Trabandt and Karl Walentin*	2007:214
Earnings Inequality and the Equity Premium *by Karl Walentin*	2007:215
Bayesian forecast combination for VAR models *by Michael K. Andersson and Sune Karlsson*	2007:216
Do Central Banks React to House Prices? *by Daria Finocchiaro and Virginia Queijo von Heideken*	2007:217
The Riksbank's Forecasting Performance *by Michael K. Andersson, Gustav Karlsson and Josef Svensson*	2007:218
Macroeconomic Impact on Expected Default Freqency *by Per Åsberg and Hovick Shahnazarian*	2008:219
Monetary Policy Regimes and the Volatility of Long-Term Interest Rates *by Virginia Queijo von Heideken*	2008:220
Governing the Governors: A Clinical Study of Central Banks *by Lars Frisell, Kasper Roszbach and Giancarlo Spagnolo*	2008:221
The Monetary Policy Decision-Making Process and the Term Structure of Interest Rates *by Hans Dillén*	2008:222
How Important are Financial Frictions in the U S and the Euro Area *by Virginia Queijo von Heideken*	2008:223
Block Kalman filtering for large-scale DSGE models *by Ingvar Strid and Karl Walentin*	2008:224
Optimal Monetary Policy in an Operational Medium-Sized DSGE Model *by Malin Adolfson, Stefan Laséen, Jesper Lindé and Lars E. O. Svensson*	2008:225
Firm Default and Aggregate Fluctuations *by Tor Jacobson, Rikard Kindell, Jesper Lindé and Kasper Roszbach*	2008:226

Wage Adjustment and Productivity Shocks 2011:253
by Mikael Carlsson, Julián Messina and Oskar Nordström Skans

Stylized (Arte) Facts on Sectoral Inflation 2011:254
by Ferre De Graeve and Karl Walentin

Hedging Labor Income Risk 2011:255
by Sebastien Betermier, Thomas Jansson, Christine A. Parlour and Johan Walden

Taking the Twists into Account: Predicting Firm Bankruptcy Risk with Splines of Financial Ratios 2011:256
by Paolo Giordani, Tor Jacobson, Erik von Schedvin and Mattias Villani

Collateralization, Bank Loan Rates and Monitoring: Evidence from a Natural Experiment 2012:257
by Geraldo Cerqueiro, Steven Ongena and Kasper Roszbach

On the Non-Exclusivity of Loan Contracts: An Empirical Investigation 2012:258
by Hans Degryse, Vasso Ioannidou and Erik von Schedvin

Labor-Market Frictions and Optimal Inflation 2012:259
by Mikael Carlsson and Andreas Westermark

Output Gaps and Robust Monetary Policy Rules 2012:260
by Roberto M. Billi

The Information Content of Central Bank Minutes 2012:261
by Mikael Apel and Marianna Blix Grimaldi

The Cost of Consumer Payments in Sweden 2012:262
by Björn Segendorf and Thomas Jansson

Trade Credit and the Propagation of Corporate Failure: An Empirical Analysis 2012:263
by Tor Jacobson and Erik von Schedvin

Structural and Cyclical Forces in the Labor Market During the Great Recession: Cross-Country Evidence 2012:264
by Luca Sala, Ulf Söderström and AntonellaTrigari

Pension Wealth and Household Savings in Europe: Evidence from SHARELIFE 2013:265
by Rob Alessie, Viola Angelini and Peter van Santen

Long-Term Relationship Bargaining 2013:266
by Andreas Westermark

Using Financial Markets To Estimate the Macro Effects of Monetary Policy: An Impact-Identified FAVAR* 2013:267
by Stefan Pitschner

DYNAMIC MIXTURE-OF-EXPERTS MODELS FOR LONGITUDINAL AND DISCRETE-TIME SURVIVAL DATA 2013:268
by Matias Quiroz and Mattias Villani

Conditional euro area sovereign default risk 2013:269
by André Lucas, Bernd Schwaab and Xin Zhang

Nominal GDP Targeting and the Zero Lower Bound: Should We Abandon Inflation Targeting?* 2013:270
by Roberto M. Billi

Un-truncating VARs* 2013:271
by Ferre De Graeve and Andreas Westermark

Housing Choices and Labor Income Risk 2013:272
by Thomas Jansson

Identifying Fiscal Inflation* 2013:273
by Ferre De Graeve and Virginia Queijo von Heideken

On the Redistributive Effects of Inflation: an International Perspective* 2013:274
by Paola Boel

Business Cycle Implications of Mortgage Spreads* 2013:275
by Karl Walentin

Approximate dynamic programming with post-decision states as a solution method for dynamic economic models *by Isaiah Hull* 2013:276

A detrimental feedback loop: deleveraging and adverse selection 2013:277
by Christoph Bertsch

Distortionary Fiscal Policy and Monetary Policy Goals 2013:278
by Klaus Adam and Roberto M. Billi

Predicting the Spread of Financial Innovations: An Epidemiological Approach 2013:279
by Isaiah Hull

Firm-Level Evidence of Shifts in the Supply of Credit *by Karolina Holmberg*	2013:280
Lines of Credit and Investment: Firm-Level Evidence of Real Effects of the Financial Crisis *by Karolina Holmberg*	2013:281
A wake-up call: information contagion and strategic uncertainty *by Toni Ahnert and Christoph Bertsch*	2013:282
Debt Dynamics and Monetary Policy: A Note *by Stefan Laséen and Ingvar Strid*	2013:283
Optimal taxation with home production *by Conny Olovsson*	2014:284
Incompatible European Partners? Cultural Predispositions and Household Financial Behavior *by Michael Haliassos, Thomas Jansson and Yigitcan Karabulut*	2014:285
How Subprime Borrowers and Mortgage Brokers Shared the Piecial Behavior *by Antje Berndt, Burton Hollifield and Patrik Sandås*	2014:286
The Macro-Financial Implications of House Price-Indexed Mortgage Contracts *by Isaiah Hull*	2014:287
Does Trading Anonymously Enhance Liquidity? *by Patrick J. Dennis and Patrik Sandås*	2014:288
Systematic bailout guarantees and tacit coordination *by Christoph Bertsch, Claudio Calcagno and Mark Le Quement*	2014:289
Selection Effects in Producer-Price Setting *by Mikael Carlsson*	2014:290
Dynamic Demand Adjustment and Exchange Rate Volatility *by Vesna Corbo*	2014:291
Forward Guidance and Long Term Interest Rates: Inspecting the Mechanism *by Ferre De Graeve, Pelin Ilbas & Raf Wouters*	2014:292
Firm-Level Shocks and Labor Adjustments *by Mikael Carlsson, Julián Messina and Oskar Nordström Skans*	2014:293
A wake-up call theory of contagion *by Toni Ahnert and Christoph Bertsch*	2015:294
Risks in macroeconomic fundamentals and excess bond returns predictability *by Rafael B. De Rezende*	2015:295
The Importance of Reallocation for Productivity Growth: Evidence from European and US Banking *by Jaap W.B. Bos and Peter C. van Santen*	2015:296
SPEEDING UP MCMC BY EFFICIENT DATA SUBSAMPLING *by Matias Quiroz, Mattias Villani and Robert Kohn*	2015:297
Amortization Requirements and Household Indebtedness: An Application to Swedish-Style Mortgages *by Isaiah Hull*	2015:298
Fuel for Economic Growth? *by Johan Gars and Conny Olovsson*	2015:299
Searching for Information *by Jungsuk Han and Francesco Sangiorgi*	2015:300
What Broke First? Characterizing Sources of Structural Change Prior to the Great Recession *by Isaiah Hull*	2015:301
Price Level Targeting and Risk Management *by Roberto Billi*	2015:302
Central bank policy paths and market forward rates: A simple model *by Ferre De Graeve and Jens Iversen*	2015:303
Jump-Starting the Euro Area Recovery: Would a Rise in Core Fiscal Spending Help the Periphery? *by Olivier Blanchard, Christopher J. Erceg and Jesper Lindé*	2015:304
Bringing Financial Stability into Monetary Policy* *by Eric M. Leeper and James M. Nason*	2015:305

SCALABLE MCMC FOR LARGE DATA PROBLEMS USING DATA SUBSAMPLING AND THE DIFFERENCE ESTIMATOR 2015:306
by MATIAS QUIROZ, MATTIAS VILLANI AND ROBERT KOHN

SVERIGES RIKSBANK

Sveriges Riksbank
Visiting address: Brunkebergs torg 11
Mail address: se-103 37 Stockholm

Website: www.riksbank.se
Telephone: +46 8 787 00 00, Fax: +46 8 21 05 31
E-mail: registratorn@riksbank.se