

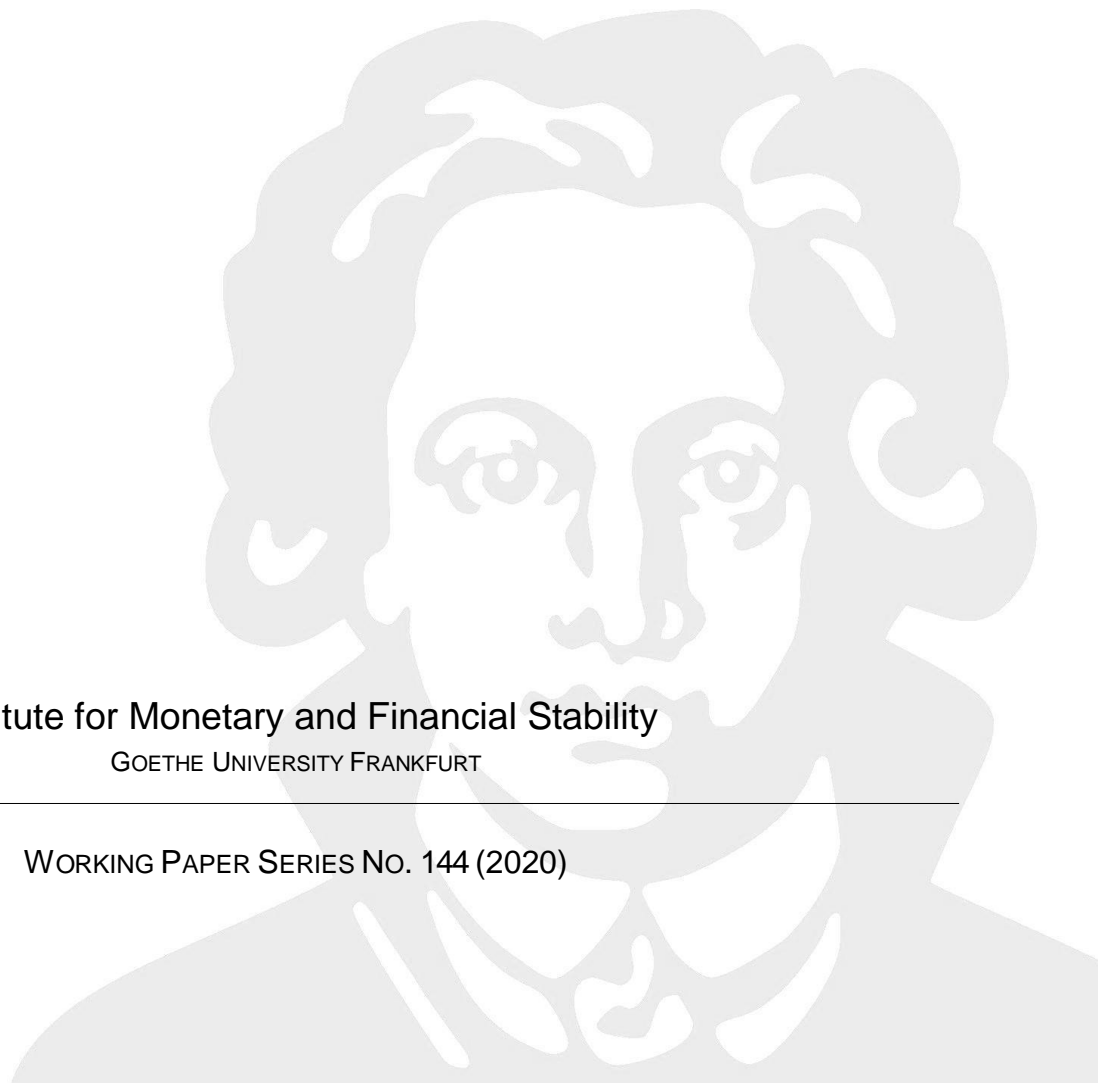
MÁTYÁS FARKAS, BALINT TATAR

Bayesian Estimation of DSGE Models with  
Hamiltonian Monte Carlo

Institute for Monetary and Financial Stability  
GOETHE UNIVERSITY FRANKFURT

---

WORKING PAPER SERIES NO. 144 (2020)



This Working Paper is issued under the auspices of the Institute for Monetary and Financial Stability (IMFS). Any opinions expressed here are those of the author(s) and not those of the IMFS. Research disseminated by the IMFS may include views on policy, but the IMFS itself takes no institutional policy positions.

The IMFS aims at raising public awareness of the importance of monetary and financial stability. Its main objective is the implementation of the “Project Monetary and Financial Stability” that is supported by the Foundation of Monetary and Financial Stability. The foundation was established on January 1, 2002 by federal law. Its endowment funds come from the sale of 1 DM gold coins in 2001 that were issued at the occasion of the euro cash introduction in memory of the D-Mark.

The IMFS Working Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

**Institute for Monetary and Financial Stability**

Goethe University Frankfurt

House of Finance

Theodor-W.-Adorno-Platz 3

D-60629 Frankfurt am Main

[www.imfs-frankfurt.de](http://www.imfs-frankfurt.de) | [info@imfs-frankfurt.de](mailto:info@imfs-frankfurt.de)

# Bayesian Estimation of DSGE Models with Hamiltonian Monte Carlo\*

Mátyás Farkas<sup>†</sup>

Balint Tatar<sup>‡</sup>

European Central Bank

IMFS and Goethe-University Frankfurt

First version: August 31, 2020

This version: October 22, 2020

## Abstract

In this paper we adapt the Hamiltonian Monte Carlo (HMC) estimator to DSGE models, a method presently used in various fields due to its superior sampling and diagnostic properties. We implement it into a state-of-the-art, freely available high-performance software package, STAN. We estimate a small scale textbook New-Keynesian model and the Smets-Wouters model using US data. Our results and sampling diagnostics confirm the parameter estimates available in existing literature. In addition, we find bimodality in the Smets-Wouters model even if we estimate the model using the original tight priors. Finally, we combine the HMC framework with the Sequential Monte Carlo (SMC) algorithm to create a powerful tool which permits the estimation of DSGE models with ill-behaved posterior densities.

*Keywords:* DSGE Estimation, Bayesian Analysis, Hamiltonian Monte Carlo

*JEL-Codes:* C11,C15,E10

---

\*We thank Nils Bertschinger from FIAS, Michael Binder, Alexander Mayer-Gohde, Volker Wieland from IMFS, and Michele Lenza, Luca Dedola, Péter Karádi and participants of RCC5 seminar 11th July 2019 in the ECB for very helpful comments and discussions. This paper builds in parts on the PhD thesis of Balint Tatar. All errors are our own. The views expressed in this paper belong to the authors and are not necessarily shared by the European Central Bank.

<sup>†</sup>matyas.farkas@ecb.europa.eu

<sup>‡</sup>balint.tatar@hof.uni-frankfurt.de

# 1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models have been shaping modern macroeconomic theory since the seminal contribution of Kydland and Prescott (1982) and have since become the workhorse framework for the analysis of economic fluctuations and forecasting. In the past two decades DSGE models were extended along a number of lines to achieve a sufficiently proper fit of the empirical data. As a consequence, likelihood-based estimation increasingly gained attraction.<sup>1</sup> The pioneering work on Bayesian DSGE model estimation in the form as it is conducted today dates back to Schorfheide (2000) and Otrok (2001) and serves as the core of other Markov Chain Monte Carlo (MCMC) based estimation algorithms. Herbst and Schorfheide (2015) provide an outstanding summary of MCMC-based algorithms. To face the challenge of increasing modeling complexity, also further, sequential sampling based estimation methods were developed, e.g. particle filters, see Fernandez-Villaverde and Rubio-Ramirez (2007) and Herbst and Schorfheide (2019) or the Sequential Monte Carlo Method (SMC). The latter algorithm was first applied by Creal (2007), then formalized by Herbst and Schorfheide (2014) and builds also the core of Cai et al. (2020).

The aim of this paper is to contribute to the literature on DSGE estimation by implementing the Hamiltonian Monte Carlo (HMC) algorithm for DSGE models. HMC is widely used in a number of different fields of science due to its superior features compared to other MCMC methods. As an extension, we will combine HMC with SMC, in order to estimate ill-behaved, in our case bimodal, posterior densities in established DSGE models.

Although the baseline MCMC-type estimation framework, also readily available in standard software, e.g. Dynare, is widely used in the DSGE literature, it suffers from several weaknesses which mostly remained poorly or not addressed at all. First, the simulated sample draws often suffer from considerably high autocorrelation, hence the effective sample size will be very small. A common approach to tackle this shortcoming is to run longer and multiple chains and to consider only each

---

<sup>1</sup>Initially, simple DSGE models were calibrated to match only selected moments of the data due to their restrictive nature. A complete review of the methodology and the transition from small scale calibrated models to the state-of-the-art estimation of medium to large scale DSGE models would be beyond the scope of this paper therefore we refer to the excellent summary provided for example in Fernandez-Villaverde et al. (2016).

$n$ -th draw to obtain uncorrelated samples. Thinning the Markov Chain can render an efficient sampling impractical, as it easily becomes time consuming, particularly when the dimension of the parameter space and of the model itself is high.

Second, in higher dimensional spaces the standard random walk MCMC algorithm will explore the typical set only slowly. Large transitions from one point to the other in the typical set will not be possible, as the number of directions to move the chain increases exponentially with the dimension. Therefore, as suggested by Betancourt (2018), "we need a better way of exploring the typical set" and "to better exploit the geometry of the typical set itself".

Third, although in theory the MCMC algorithm converges under certain regularity conditions asymptotically to the target density, in practice the convergence might occur at a very slow pace. Relatedly, a key question whether the Markov chain has already converged to the target distribution, lacks of a clear answer. Unfortunately, there is no single way to address the latter issue, as pointed out by Brooks and Gelman (1998), instead "the idea is to use a wide variety of diagnostics so that if all appear to suggest that convergence has been achieved, then the user can have some confidence in that conclusion". However, even if standard diagnostics suggests that convergence has not been reached yet, it will be challenging to explore the reason for non-convergence, as available diagnostics does not provide any information on irregular regions of the posterior distribution.

Fourth, yet not completely unaddressed, it is common practice to conduct a mode search to specify the starting point for the MCMC-type samplers. This practice is closely followed in the DSGE literature and several algorithms for posterior mode search are readily available in standard software for DSGE estimation. However, mode search algorithms often fail to provide a suitable starting point or their execution might be tedious and time consuming. Furthermore, Betancourt (2018) argues that in general the mode might not be representative for the typical set, as the latter can lie increasingly distant from the mode in higher dimensions.

A straightforward algorithm leveraging the information in the geometry of the typical set and thereby addressing the above issues is the HMC method. Similarly to the Kalman filter and MCMC, it has its roots in physics and dates back to Duane et

al. (1987).<sup>2</sup> It is considered the new standard in high dimensional numerical simulation where the gradient of the target density can be evaluated. Due to the accessibility of an advanced software package for Bayesian estimations, STAN, implementing the HMC algorithm, the methodology is presently applied by many researchers in various fields. HMC has been shown to have significantly better sampling properties than the baseline algorithm, the Random Walk Metropolis Hastings (RWMH), commonly used for the estimation of DSGE models, a fact well documented in the literature, see e.g. Neal (2011). Herbst and Schorfheide (2015) also acknowledges the advantages of HMC and encourages researchers to make progress in this direction.

The current work is the first paper available to apply the HMC algorithm to DSGE models and to present results.<sup>3</sup> As an extension, we also address a main shortcoming of the HMC algorithm combining it with the SMC framework to estimate a DSGE model with an ill-shaped posterior density. We implement HMC in STAN due to its use of C++, a low-level high-performance programming language<sup>4</sup> and automated differentiation enabling the calculation of complicated differentials. It is a suitable tool to deal with complex models and symbolic differentiation, which make the accurate implementation feasible, therefore there is no need to rely on approximations of the gradient of the posterior likelihood function. It also comes along with a set of powerful diagnostics and visualization toolkit readily available and unique to HMC enabling to provide further evidence whether the typical set has been explored appropriately.<sup>5</sup>

In contrast to the RWMH sampler HMC is more efficient, as in the optimum its draws are uncorrelated. This implies, that it is more likely that HMC explores the typical set properly even in higher dimensional problems. A further, practical advantage of HMC is that the time-consuming and potentially erroneous mode-estimation can be abandoned completely as the gradient-based sampling serves as guidance to find and then to explore the typical set appropriately, see Betancourt

---

<sup>2</sup>In the original work of Duane et al. (1987) it was called 'Hybrid Monte Carlo' and designed for the numerical simulation of lattice field theory simulations of quantum chromodynamics

<sup>3</sup>In contemporaneous work Fernandez-Villaverde and Guerrón-Quintana (2020) propose the application of HMC for DSGE estimation, furthermore a working paper by Goodrich and Montes-Galdon (retrieved on 25 August, 2020) exists, although without results.

<sup>4</sup>C++ can be considered as a superset of the C language with features added as e.g. object-oriented programming, exception handling, a rich C++ library and improved memory management.

<sup>5</sup>For a recent review of the work-flow with STAN please see Gabry et al. (2019).

(2018). Lastly, the sampling diagnostics is capable of shedding light on inefficient or improper sampling due to fat tails or high curvature and may also serve as an indication for difficult to sample parameters possibly due to weak identification.<sup>6</sup>

To demonstrate the features of this powerful algorithm we estimate the textbook small scale New-Keynesian model from Herbst and Schorfheide (2015) and the Smets and Wouters (2007) model. We find very similar results to those in existing literature. Yet, for the Smets-Wouters model we find a set of difficult to sample parameters and a second mode, even if tight priors and the entire sample is used for the estimation. In addition, the estimation results showcase two practical advantages of HMC in higher dimensional problems: First, we confirm results of Chib and Ramamurthy (2010) that the steady state of inflation in the original paper of Smets and Wouters (2007) has been over-, while the mean of hours underestimated, possibly due to initializing the Metropolis-Hastings algorithm from a wrong mode. Hence, we highlight that abandoning the mode and using a gradient-based guided sampling is superior to random walk sampling around the mode. Second, based on the HMC diagnostics we highlight that some parameters are difficult to sample from and find that these parameters feature bimodal posterior distributions.

Although the implementation of the HMC algorithm for DSGE models paves the way for a more sophisticated exploration of the typical set and provides access to powerful diagnostics, it has a significant drawback: it fails to deal with multimodal target densities. In case the modes are separated by large energy-barriers from each other, in particular, the posterior likelihood function has no support between modes, the algorithm will get stuck in one mode and the chain will not be able to escape in reasonable time. To address this shortcoming we merge the HMC sampling algorithm into the SMC framework proposed by Herbst and Schorfheide (2014) and explore bimodal densities in the Smets-Wouters model, when less informative priors are used.

The remaining part of this paper is organized as follows. In Section 2 we provide a brief review of the workhorse Bayesian estimation framework. In Section

---

<sup>6</sup>In particular, if the autocorrelations of the sampled draws from the HMC algorithm does not decay to zero after the first lag, it indicates that a specific parameter is hard to sample, this implies that the posterior is either oddly shaped, or the parameters suffers from weak identification. In contrast the high autocorrelation of the Metropolis-Hastings algorithm is an inherent feature, and thus cannot be used to identify problematic parameters.

3 we present the HMC algorithm and summarize the underlying theoretical considerations. Section 4 describes our approach to implement the DSGE estimation framework and discusses some computational issues. Section 5 presents the estimation results of a textbook small scale New-Keynesian DSGE model and the Smets and Wouters (2007) model. Section 6 extends the paper by combining HMC and SMC to estimate ill-behaved posterior densities. Section 7 concludes the paper.

## 2 Bayesian Estimation of DSGE Models:

### A Brief Review

In this chapter we briefly review the main estimation framework used for MCMC-type Bayesian DSGE model estimation. A more extensive treatment can be found in the excellent work of Herbst and Schorfheide (2015).

In order to estimate a Bayesian model, the first step is to specify the joint distribution of the data and the model parameters. The aim is to obtain the posterior density, that is, the distribution of the model parameters given the data, denoted by  $p(\theta|Y)$  which can be expressed by the means of Bayes' rule as follows:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \quad (1)$$

$p(Y|\theta)$  is referred to as the likelihood function and  $p(\theta)$  is the prior distribution. Typically in a Bayesian estimation, the *a priori* beliefs about the parameter vector  $\theta$  are updated using the likelihood function. The posterior distribution then comprises the state of knowledge about  $\theta$  consisting of the *a priori* beliefs and the information available in the data.

To specify a likelihood function conditioned on the parameters and turn a DSGE model into a Bayesian model, a formal representation of the DSGE model is needed. Hence we need to solve for the law of motion of the state variables. There exists a variety of DSGE solution methods, e.g. Blanchard and Kahn (1980), Binder and Pesaran (1997), King and Watson (1998), Uhlig (1999), Klein (2000), Kim (2000), Christiano (2002), Anderson (2010). A popular solution technique for a linearized DSGE model was proposed by Sims (2002) which starts with the following



representation of the DSGE model:

$$\Gamma_0 s_t = \Gamma_1 s_{t-1} + \Psi \epsilon_t + \Pi \eta_t \quad (2)$$

where  $s_t$  is the set of state variables,  $\epsilon_t$  is the vector of structural shocks and  $\eta_t$  the one step ahead rational expectation forecast errors,  $x_t - \mathbb{E}_{t-1} x_t$ .  $\Gamma_0, \Gamma_1, \Psi, \Pi$  are real matrices of appropriate dimensions. The solution is based on the QZ-decomposition, also referred to as the Schur decomposition. If the above system has a unique stable solution then it can be represented in the following VAR-form:

$$s_t = G_0(\theta) s_{t-1} + G_1(\theta) \epsilon_t \quad (3)$$

Applying the solution method proposed by Sims (2002), or any other solution algorithm, a state space representation can be obtained to specify the likelihood function. In this setup the VAR-form from above represents the transition equation which is linked to the data by means of the measurement equation:

$$y_t = H_0(\theta) + H_1(\theta)t + H_2(\theta)s_t + u_t \quad (4)$$

The state space representation allows us to express the joint density function for the the observed data and the DSGE-model variables where the latter are unobserved:

$$p(Y_{1:T}, S_{1:T} | \theta) = \prod_{t=1}^T p(y_t, s_t | Y_{1:t-1}, S_{1:t-1}, \theta) = \prod_{t=1}^T p(y_t | s_t, \theta) p(s_t | s_{t-1}, \theta) \quad (5)$$

where  $p(y_t | s_t, \theta)$  and  $p(s_t | s_{t-1}, \theta)$  are the conditioned probabilities given the observation and the state equation. To obtain the desired likelihood function the unobserved states,  $s_t$ , have to be integrated out. For log-linearized DSGE models with Gaussian disturbance one can use the Kalman filter to obtain the conditional expectations and variances of the observables and finally the log-likelihood function. Once the prior distribution of the parameters is specified one can set up the RWMH algorithm to sample from the posterior density. The algorithm is summarized below:<sup>7</sup>

---

<sup>7</sup>This code summarizes the main steps extensively described in Herbst and Schorfheide (2015).

---

**Algorithm 1** Random Walk Metropolis Hastings

---

1. Maximize  $\ln p(Y|\theta) + \ln p(\theta)$  by a numerical algorithm to obtain the posterior mode, denoted by  $\hat{\theta}$ . (This step involves the application of the solution algorithm of the DSGE model, setting up the state space representation and the calculation of the likelihood by means of the Kalman filter)
2. Compute  $\tilde{\Sigma}$ , the inverse of the Hessian at  $\hat{\theta}$
3. Initialize a starting value or draw  $\theta^{(0)}$  from the proposal density  $q(\theta^{(0)}|\tilde{\theta})$  (in this case  $N(\tilde{\theta}, c_0^2\tilde{\Sigma})$ )
4. For  $n = 1, \dots, N$ 
  - (a) Draw  $\theta'$  from the proposal distribution  $\mathcal{N}(\theta^{(n-1)}, c_0^2\tilde{\Sigma})$ .
  - (b) Solve the DSGE model for  $\theta'$  and build the new state space representation.
  - (c) Calculate  $p(Y|\theta')$  and  $p(\theta')$  (by means of the Kalman filter)
  - (d) Accept  $\theta'$ , that is,  $(\theta^{(n)} = \theta')$ , with probability  $\min\{1, f(\theta^{(n-1)}, \theta'|Y)\}$  and reject  $(\theta^{(n)} = \theta^{(n-1)})$  otherwise where

$$f(\theta^{(n-1)}, \theta'|Y) = \frac{p(Y|\theta')p(\theta')q(\theta'|\theta^{(n-1)})}{p(Y|\theta^{(n-1)})p(\theta^{(n-1)})q(\theta^{(n-1)}|\theta')}$$

5. Estimate the posterior expected value of the function  $h(\theta)$  by  $\frac{1}{N} \sum_{i=1}^N h(\theta^{(i)})$
- 

In the above, the proposal density  $q(\cdot|\cdot)$  is chosen to be the normal distribution with expected value  $\theta^{(n-1)}$  which implies that the proposals follow a random walk. In addition, as the density function of the normal distribution is symmetric, the proposal densities cancel. Also the scaling parameter,  $c_0$ , should be chosen in a way that the acceptance ratio equals to 23.4%, which was proven to be the optimal acceptance ratio, see Roberts et al. (1997). In practice however this parameter is chosen in a way that the acceptance ratio lies between 0.2 and 0.4.

There are several other modified versions of the MH algorithm. For example, the Block-MH algorithm breaks the parameter vector into blocks and as its name suggests it updates at most only one block of the parameters at once, applied e.g. by Cúrdia and Reis (2010) with fixed blocks. This scheme can be further extended by randomizing the break-up of the parameter vector into blocks in each step, as proposed by Chib and Ramamurthy (2010). A further possibility to improve the algorithm is to apply a more sophisticated proposal density. In particular, the Metropolis-Adjusted Langevin (MALA) algorithm, originally proposed by Besag (1994) and later assessed for its convergence properties by Roberts and Tweedie (1996), suggests to choose again a normal distribution, however the expected value

should be adjusted by one step into the direction of the gradient of the negative log-posterior. Updating the current draw along the gradient, this algorithm accounts for the shape of the posterior density and therefore moves the chain, thus the new proposal for  $\theta$ , into regions with higher probability density. It is common to choose a scaled version of the identity matrix as the variance. Both the step size into the direction of the gradient and the scaling of the variance are subject to fine-tuning. As pointed out in Roberts and Tweedie (1996) using the Langevin-diffusion as an update should be a good choice as it is constructed in a way that under suitable regularity conditions in continuous time it converges to its stationary distribution. Therefore, even before the MH-step the candidate chain itself will approximate the target distribution to be sampled from. Herbst and Schorfheide (2015) uses instead of the normal distribution a  $t$  distribution for the proposal. Furthermore, the algorithm benefits from scaling the step size along the gradient by the Hessian at the posterior mode, also pointed out in Herbst and Schorfheide (2015), based on Roberts and Stramer (2002). The MH-Newton algorithm mainly differs from the latter MALA-algorithm in that instead of the Hessian at the posterior mode the Hessian at  $\theta^{(n-1)}$  is taken and the step size is randomly chosen. For further details we refer to Qi and Minka (2002).<sup>8</sup>

Although first-order linear approximations around the non-stochastic steady state are popular, in a number of cases more elaborate estimation methods are required. For example, when higher order approximations are necessary to capture the impact of shocks on endogenous variables, then the state space will be non linear. To evaluate the likelihood in this more complex case particle filters were proposed in literature, e.g. Fernandez-Villaverde and Rubio-Ramirez (2007). At the same time, particle filters are also applied if the posterior likelihood is ill shaped, e.g. Sequential Monte Carlo Methods (SMC) in Herbst and Schorfheide (2014), which may even occur when standard models are estimated using first order linear approximations. Our extension of the HMC with SMC falls into the latter application.

---

<sup>8</sup>For further discussion and a comparison of the different estimation methods we refer to the work of Herbst and Schorfheide (2015).

### 3 The Hamiltonian Monte Carlo Method

This section of the paper provides an introduction to the HMC framework. It is aimed to offer intuition to the reader and to reveal a few main theoretical aspects of the methodology.<sup>9</sup> Similarly to advanced MCMC methods from above, the HMC algorithm builds on the information provided by the gradient of the log-posterior density function. In particular, it uses the information in the geometry of the target distribution, that is, its shape and the equations characterizing it. Its main advantage is that by means of the Hamiltonian equations, a concept borrowed from physics, the algorithm enables to propose a new parameter draw  $\theta'$  which is distant from the current  $\theta$  while maintaining a sufficiently high acceptance rate.

Before providing a formal descriptive introduction to HMC, let us first illustrate its main concept by applying it to a simple example.<sup>10</sup> In physics researchers usually model the evolution of a mechanical system over time given a particle's position and momentum by functions measuring its potential and kinetic energy. Classical examples of a mechanical system are a bouncing ball, a pendulum or an oscillating spring. Let us assume that one intends to sample from a one dimensional standard normal distribution with density function  $f(q) = 1/(2\pi)^{1/2}\exp\{-q^2/2\}$ . Intuitively, the aim is to generate more samples for  $q$  in those regions of the domain of  $q$ , that is the real line, where the density function  $f(q)$  peaks than in regions of its tails.<sup>11</sup> In particular, one would like to generate samples from each location of the domain of  $q$  in proportion to the value of the density function,  $f(q)$ , at that location. This case is illustrated in Figure 1, where the target density is a normal distribution in black, while the blue circles on the real line represent desired the samples.

---

<sup>9</sup>More extensive treatment is provided e.g. in Neal (2011) or Betancourt (2018).

<sup>10</sup>Our explanation draws on the example in the excellent book of Lambert (2018).

<sup>11</sup>This intuition, that the typical set is close to the mode does not apply in higher dimensional problems, see Betancourt (2018) for additional details.

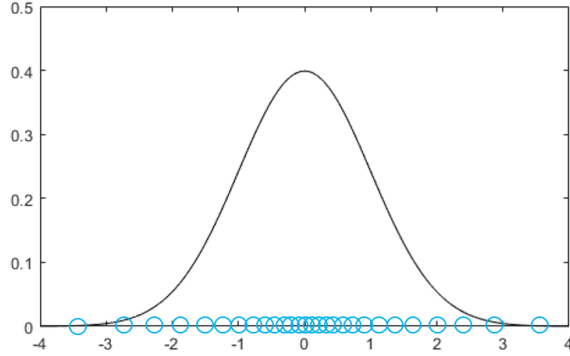


Figure 1: Samples from a normal distribution

To demonstrate the idea of HMC, let us first consider instead of  $f(q)$  the negative logarithm of the density function  $g(q) := -\log f(q)$ . That is, we flip over  $f(q)$  and obtain a valley-shaped function  $g(q)$ , being the new target density. Next, we imagine a person using a sleigh and trying to explore the valley covered in snow, i.e. the new target density defined by  $g(q)$ . We assume that our explorer starts out at some point in the valley, sitting on a sleigh, i.e. starting out somewhere on the surface generated by the function  $g(q)$  and is initially pushed randomly with some impulse, either uphill or downhill, see Figure 2. In contrast, think of the RWMH algorithm as a person exploring the valley on foot and proceeding in equally distant steps in random directions. Assuming that our explorer on the sleigh started its journey downhill, after passing the trough of the valley it will continue sliding uphill. After getting gradually slower it will stop at some point and will start sliding again downhill into the opposite direction. Assuming also that the snow-covered surface of the valley is frictionless, this motion will continue forever. The motion of the sleigh on a frictionless surface can be perfectly described by the Hamiltonian equation in analogue to the mechanical systems from above. The total energy of the sleigh can be described by its potential energy function  $U(q) := g(q)$  and its kinetic energy  $K(p)$ , depending on its velocity  $p$ . By moving up or down the valley the explorer on a sleigh exchanges potential energy  $U(q)$  for kinetic energy  $K(p)$ . As the sleigh slides down (up) the hill, its potential energy will decrease (increase) and its kinetic energy increase (decrease). After some time  $t = T$  we record the position  $q$  of our sleighing explorer. Given the shape of the terrain the person with the sleigh will be more often in lower regions of the valley than in higher regions. In contrast, a RWMH explorer would find it easy to walk downhill to the trough of the valley, but would struggle to make steps upwards. In other words, under RWMH sampling

a proposal is always accepted if the probability of the new proposal  $q'$  is higher compared to the probability of the current draw  $q$ , steps downhill, and only accepted randomly depending on the proportion of the probabilities of  $q'$  and  $q$ , steps uphill. Using RWMH sampling the explorer would walk down to the valley, and explore the region in the bottom, i.e. the target density around the mode, but always struggle to make steps uphill and proceed only very slowly.

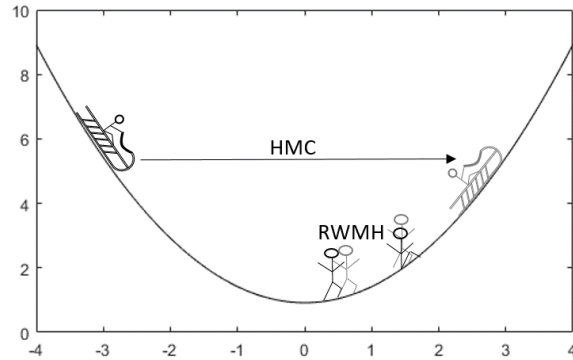


Figure 2: Hamiltonian Monte Carlo vs. Random Walk Metropolis Hastings

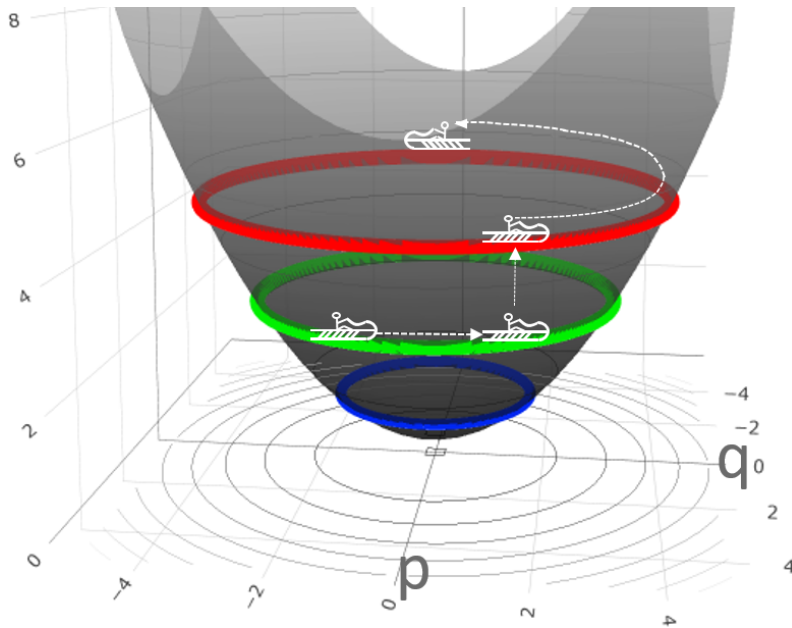


Figure 3: Visualization of Hamiltonian Monte Carlo

A fair question to ask is what is the reason for the explorer on the sleigh sliding and arriving always safely at the other side of the valley without struggling. The core of the concept is that by tracking the velocity  $p$  and the kinetic energy  $K(p)$  the parameter space is extended by the same dimension to measure total energy. This extended space is called the phase space and is fundamental for Hamiltonian dynamics. Defining the kinetic energy by  $K(p) := |p|^2/(2m)$ , where  $m$  corresponds

to the mass of the sleigh with the person, we can describe the system entirely with the Hamiltonian equation  $H(q, p)$ , also referred to as the total energy function. Visually, extending the parameter space and tracking both position and velocity at the same time allows the sleigh moving from one side to the other side of the valley by sliding around on the same (energy) level, hence moving only horizontally in this simple case, see Figure 3. In this extended space, stopping and then pushing the sleigh with a different impulse is equivalent to 'moving' it onto a different height, hence energy level, sliding again along the Hamiltonian, i.e. the contour lines of the extended space. When stopping the sleigh, the explorer 'brakes', throwing away kinetic energy  $K(p)$ , e.g. by wasting it while braking, and records only its position  $q$ .

Considering a higher dimensional parameter space with a more irregular density function to sample from, extending the parameter space and applying the Hamiltonian serves as a sort of secure trail of movement for our explorer. It allows for being able to describe an exact path to move along on a (possibly) complicated extended surface such that one remains always at the same energy level. Hence, the new proposal for the parameter will always be accepted.

Turning now to a more formal description, in classical mechanics the Hamiltonian equation is obtained from Lagrange's equation, a reformulation of the Newtonian mechanics, by a Legendre transformation, where  $H : \mathbb{R}^{2d} \rightarrow \mathbb{R}$  with  $\mathbb{R}^{2d}$  being the phase space and  $d$  the degrees of freedom. This Hamiltonian framework can be easily translated to MCMC applications outside physics, by regarding the position of the sleigh,  $q$ , as the variables of interest of which posterior distribution a sample should be drawn. The main idea is to extend Bayes' Theorem  $p(\theta|Y) \propto p(\theta)p(Y|\theta)$  by an auxiliary vector  $\alpha$  of momentum variables to obtain the joint posterior density  $p(\theta, \alpha|Y) \propto p(\theta, \alpha)p(Y|\theta, \alpha)$  of  $\theta$  and  $\alpha$ . To each parameter  $\theta_i$  one momentum variable  $\alpha_i$  is assigned. The auxiliary variables are a priori independent of  $\theta$  and  $Y$  implying that  $p(\theta, \alpha|Y) \propto p(\theta)p(\alpha)p(Y|\theta)$ .

The change in the current position  $q$  and momentum  $p$ , being both of dimension  $d$  respectively, over time is characterized by the partial derivatives of the Hamiltonian

equation:

$$\frac{dq_i}{dt} = \frac{\partial H(q, p)}{\partial p_i} \quad \forall i = 1, \dots, d \quad (6)$$

$$\frac{dp_i}{dt} = -\frac{\partial H(q, p)}{\partial q_i} \quad \forall i = 1, \dots, d \quad (7)$$

where  $2d$  equals the full dimension of the system. The equations of motion can be presented in a more compact way by defining  $z := (q, p)$  such that

$$\frac{dz}{dt} = J\nabla H(z) \quad (8)$$

with  $\nabla H(z)$  being the gradient of the Hamiltonian system and  $J$  a matrix of dimension  $2d \times 2d$ :

$$J = \begin{bmatrix} 0_{d \times d} & I_{d \times d} \\ -I_{d \times d} & 0_{d \times d} \end{bmatrix}$$

The solution to this system of differential equations can be regarded as a mapping  $F_s : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d$  with  $(q, p)(t) \rightarrow (q, p)(t + s)$  such that the Hamiltonian equations describe the law of motion of the system from  $t$  to  $t + s$ .

The Hamiltonian measures total energy in the system, that is potential energy and kinetic energy, consequently for the HMC algorithm it takes the additive form

$$H(p, q) = U(q) + K(p). \quad (9)$$

The kinetic energy  $K(p)$  is usually defined as

$$K(p) = p^T M^{-1} p / 2 \quad (10)$$

where  $M$  is referred to as the "mass matrix" which is typically diagonal, and is often a scalar multiple of the identity matrix as it often stands for the mass of some bodies or particles.<sup>12</sup>

The Hamiltonian system has four key properties which allow for using it for the construction of an MCMC algorithm. Firstly, the Hamiltonian does not change over

---

<sup>12</sup>In the estimation procedure this assumption can be released.



time, that is,  $dH/dt = 0$ , which is crucial to ensure that the acceptance probability equals always one.

Second, the Hamiltonian system preserves the volume in the phase space. Without venturing too deeply into details of volume measures of a phase space it suffices to state that this property is necessary to avoid accounting for a change in the volume when computing the acceptance probability.

Thirdly, the Hamiltonian system is symplectic. Formally this corresponds to the property that the Jacobian  $B_s := DT_s$  of the mapping  $T_s$ , satisfies the following equation:

$$B_s^T A B_s = A \tag{11}$$

where  $A$  is in general a fixed  $2d \times 2d$ , non-singular, skew symmetric matrix. Usually, the matrix  $J$  from above is chosen for  $A$ . The determinant of the matrix  $J$  is unity and it holds that  $J^{-1} = J^T = -J$ . The symplecticness condition implies that the mapping is volume preserving as from the equation above it immediately follows that  $|\det(B_s)| = 1$ . Yet, the above property is stronger than just volume preservation if  $d > 1$ . This property is important, as in practice Hamiltonian equations can be solved only by numerical integration. Although a large number of numerical integrators exist, most of them are prone to accumulate approximation errors such that the accuracy of the solution will be significantly impaired. However, to solve for the Hamiltonian, symplectic integrators can be applied having the advantage that the approximated trajectory does not drift away from the true one.

Finally, the mapping  $T_s$  defined above, is reversible, that is  $T_s$  has an inverse  $T_{-s}$  which is exactly the negation of the time derivatives in the Hamiltonian equations. Considering again the example with the sleigh one can imagine this as stopping the explorer at  $q(t+s)$  and push it into the opposite direction with the same impulse. In case  $K(p) = p^T M^{-1} p / 2$  one can negate  $K(p)$ , apply  $T_s$  and then negate again  $K(p)$  to obtain the original  $(q, p)(t)$  where the explorer departed from. The reversibility property is crucial when proving the detailed balanced condition in the probabilistic framework which ensures together with ergodicity that the HMC converges to the invariant distribution.

To apply this framework to a probabilistic setting borrowing one further concept

from statistical mechanics is necessary referred to as the "canonical" distribution at a given temperature. This concept describes possible states of a mechanical system which is at thermal equilibrium at temperature  $T$ . For the latter purpose the following distribution is used:

$$P(x) = \frac{1}{Z} e^{-E(x)/T} \quad (12)$$

where we assume that the energy  $E(x)$  and its gradient can be evaluated. Any particular density  $P(x)$  can be adopted to the above scheme by setting  $E(x) = -\log P(x) - \log Z$  and  $T = 1$ . The HMC algorithm translates this framework into an MCMC-sampling algorithm by applying the Hamiltonian equation as the total energy function for the joint state  $(q, p)$  which results in the following canonical distribution:

$$P(q, p) = \frac{1}{Z} e^{-H(q,p)/T} \quad (13)$$

with  $H(q, p) = U(q) + K(p)$  we obtain

$$P(q, p) = \frac{1}{Z} e^{-U(q)/T} e^{-K(p)/T} \quad (14)$$

Setting for  $U(q)$  the negative logarithm of the target density  $-\log(p(Y|\theta)p(\theta))$  and for  $K(p)$  the kinetic energy function allows to define an algorithm which samples from the distribution of interest. The iteration is carried out in three steps:

---

**Algorithm 2** Hamiltonian Monte Carlo

---

1. Draw a momentum vector  $p'$  from its multivariate normal distribution which can be carried out by Gibbs-sampling.
  2. Draw the position vector  $\theta'$  by applying the Hamiltonian equations deterministically.
  3. Metropolis-Hastings step: accept the new proposal and set  $\theta^{(n+1)} = \theta'$  with probability  $\min[1, \exp(-(U(q') - U(q) + K(p') - K(p)))]$
- 

As the total energy in the system remains constant, in theory the proposal obtained by applying the Hamiltonian equations is always accepted. To obtain a sample from the target distribution one simply omits the sampled momenta. It is well known that to show that the resulting Markov chain converges to the target distribution it

has to be ergodic and has to fulfill the detailed balance condition:

$$P(q, p)P_K((q, p) \rightarrow (q', p')) = P(q', p')P_K((q', p') \rightarrow (q, p)) \quad (15)$$

where  $P_K$  is the HMC kernel. The key property that allows to prove that the detailed balance condition holds is reversibility of the Hamiltonian system. In addition, the symplecticness of the numerical integrator to be used ensures that detailed balance holds even if the solution is approximated numerically. A formal proof is available in Duane et al. (1987). As regards ergodicity, the latter paper does not provide any insights, instead it assesses using an example in compact quantum electrodynamics "Whether or not this idea works in practice...". Proving ergodicity for the HMC algorithm involves deep knowledge in probability theory and would go beyond the scope of this paper. Very loosely spoken, ergodicity implies that the Markov chain will not be trapped in a subset of the parameter space, instead it will reach all possible states again and again, hence it will asymptotically converge to the invariant distribution. Neal (2011) points out that in theory it is possible that ergodicity fails once as a fixed number of integration steps is used for the numerical approximation of the solution and illustrates this based on a short example. Mackenzie (1989) proposes that by randomizing the length of the Hamiltonian trajectory this issue can be eliminated while recently more general conditions for ergodicity, and hence for convergence of the HMC algorithm could be proved, see e.g. Livingstone et al. (2018) and Durmus et al. (2019).

## 4 Implementation into STAN

STAN is a state-of-the-art probabilistic programming language for Bayesian inference written in C++ language. It allows users to set up hierarchical Bayesian models in a convenient statistical language and provides thereby an easy to apply interface to the HMC algorithm for complex models. C++ is a machine-oriented programming language and is often applied to perform computationally highly intensive calculations due to its performance, necessary to estimate a DSGE model. Yet, this comes at the cost of complexity in terms of the programming language which the STAN interface remedies and makes this powerful and complex concept available to

researchers, working out-of-the-box.

## 4.1 Features and Calibration

The Hamiltonian equations typically describe the dynamics of a system in continuous time. However, in practice it will be necessary to apply a discrete-time approximation in order to calculate the new position, the momentum and the total energy level, the sum of potential energy and kinetic energy. One of the key challenges lies in the accurate solution of the Hamiltonian equations. As discussed before, the Hamiltonian system is symplectic. Thus a dedicated class of symplectic integrators can be applied enabling the calculation of an accurate discrete time solution for the Hamiltonian trajectory in the phase space. The main advantage of the latter class of integrators is that the approximated trajectory does not drift away from the true one, even if integration is carried out over a long distance, hence a long period of time. STAN uses a simple implementation referred to as “leapfrogging” to solve for the discrete-time approximation of the Hamiltonian equations which is summarized by the following algorithm:

---

**Algorithm 3** Leapfrogging

---

1.  $p_i(t + \epsilon/2) = p_i(t) - (\epsilon/2) \frac{\partial U}{\partial q_i}(q(t))$
  2.  $q_i(t + \epsilon) = q_i(t) + \epsilon \frac{p_i(t + \epsilon/2)}{m_i}$
  3.  $p_i(t + \epsilon) = p_i(t + \epsilon/2) - (\epsilon/2) \frac{\partial U}{\partial q_i}(q(t + \epsilon))$
- 

Although at first glance the above algorithm is easy to implement, it generates a further challenge, especially when applied in the context of DSGE estimation. In general it requires the evaluation of the gradient of the log-posterior which calculation might be extremely difficult and time intensive. Gradients obtained by numerical approximations can be inaccurate or computationally demanding when the parameter space is large. One of the main advantages of STAN is that it applies a reverse-mode automatic differentiation and C++ template metaprogramming. Automatic differentiation requires only a limited number of differentiation rules and

the gradient is constructed via the chain rule by creating an expression tree backwards starting with the last expression in the likelihood function. For example, STAN is capable of differentiating any iterative algorithm which is particularly useful when implementing the estimation of DSGE models. Therefore, there is no need for the user to specify any derivatives manually, yet in theory it is possible to write wrappers if a closed form solution of the partial derivatives is available. Although the derivation of the log-likelihood function which depends on the solution of the DSGE model is computationally involved for a mid-scale DSGE model, the latter is performed by STAN automatically and efficiently due to availability of symbolic differentiation.

The performance of the algorithm is sensitive to the selection of two parameters: the step size,  $\epsilon$  and the number of steps in time,  $L$ . The selection of the discrete time approximation to calculate the integral,  $\epsilon$ , is of crucial importance. If the approximation is overly fine, then the proposal to update  $\theta$  will be accepted with very high probability, yet the distance  $\|\theta' - \theta\|$  will be small and the chain will explore the parameter space very slowly. If  $\epsilon$  is too high, the approximation of the true solution to the Hamiltonian equation will become imprecise, or may even diverge, and  $\theta'$  will be unlikely to be accepted. Furthermore, it can also happen that the Markov chain will fail completely to explore certain regions of the posterior. Usually, the posterior likelihood function exhibits regions with both lower and larger curvature especially if the model is more complex, therefore one has to strike the right balance when setting  $\epsilon$ . A further strength of STAN lies in the feature that  $\epsilon$  is calibrated automatically during the warm-up period and fixed afterwards, yet the user retains the option to set the parameter manually. STAN aims to calibrate  $\epsilon$  in a way that the acceptance rate lies at 80%, significantly higher than 23.4% in the RWMH algorithm. In case the divergence rate remains still high, the automated calibration mechanism in STAN can be still instructed to target higher acceptance ratios.

It is crucial to select a suitable number of steps,  $L$ , to be conducted by the leapfrogging algorithm in order to explore the state space systematically as pointed out by Neal (2011). An inappropriately low  $L$  will cause  $\theta'$  to be too close to  $\theta$ , hence the algorithm will exhibit random walk behavior and the Markov chain will

explore the parameter space again inefficiently slowly, as also highlighted by Hoffman and Gelman (2014). If  $L$  is too large, computational resources are wasted as the acceptance rate does not depend systematically on the number of steps. A further built-in feature of STAN is that it automatically optimizes the number of steps by means of the No U-Turn Sampling (NUTS) algorithm, see Hoffman and Gelman (2014). The intuition of NUTS is to use the leapfrog integrator to iterate on  $\theta$  both in positive and negative directions, doubling the number of steps each time. That is, first running forwards or backwards 1 step, then forwards or backwards 2 steps, then forwards or backwards 4 steps and so on. The doubling process implicitly builds a balanced binary tree and continues until some proposal moves backwards to its original point of departure, making a U-turn and moving again towards the point of departure,  $\theta$ . STAN applies then a slice sampling algorithm to select randomly a point along the Hamiltonian trajectory which adds complexity, yet it is necessary to preserve the reversibility condition of the generated Markov chain. Finally, it accepts the new proposal with the probability given in the Metropolis-Hastings step.

The mass matrix  $M$ , being typically a diagonal matrix is tuned automatically during the warm-up. Here, the user is allowed to tune  $M$  manually, however the automated tuning process of STAN operates sufficiently well. Furthermore, if desired, a dense matrix with non-zero off-diagonal elements can be also applied for the estimation, which could result in improved efficiency.

A further useful feature which is implemented into STAN is that it is able to remedy the weakness that the HMC algorithm works only if the support of the posterior density spans the entire parameter space. If a proposal is accepted in a region where the mass of the parameter space is zero, the gradient will become zero or undefined and the chain will get stuck. A straightforward approach to avoid this issue is to restrict the parameter space and let the Markov Chain bounce back from the boundary by negating the momentum. However, STAN instead reparametrizes  $\theta$  as a function of unbounded parameters. This occurs typically when standard deviations are estimated. The latter approach obviously involves the calculation of the Jacobian, however this is carried out again automatically by STAN.

As already pointed out, the main advantage of the HMC algorithm is that it uses gradient information to explore suitable paths on which the level of energy remains

constant and finds new proposals  $\theta'$  which are distant from the most recent draw  $\theta$ . However, it comes along with the difficulty that the gradient of the log-likelihood function needs to be evaluated. Recall that the popular solution algorithm to DSGE models proposed by Sims (2002) uses a QZ-decomposition where the entries of the matrices can become complex. A main shortcoming of STAN is that it is not capable of executing calculations with complex numbers, hence a QZ decomposition cannot be implemented. Furthermore, it might be challenging to build the derivatives when complex numbers are involved. To overcome this difficulty we need to rely on a DSGE model solution algorithm which makes it feasible to the automated differentiation implemented in STAN to calculate the gradient. The reverse-mode automatic differentiation relies on the chain rule when building the symbolic derivative, hence it is capable to handle any matrix iteration algorithm where no complex numbers are involved. A straightforward and easy to understand DSGE solution to remedy these shortcomings is the Binder and Pesaran (1997) solution algorithm.

## 4.2 Binder-Pesaran Algorithm

The main idea of the Binder and Pesaran (1997) algorithm is to rewrite  $s_t$  such that the reshuffled form will not contain the  $s_{t-1}$  term and the system can be solved forward in case it has a unique stable solution. A short recap of the main steps of the algorithm looks as follows. Without loss of generality let the system be given in a slightly different form than the Sims (2002) canonical formula:

$$M_{00}s_t = M_{10}s_{t-1} + M_{01}\mathbb{E}_t s_{t+1} + M_s \epsilon_t \quad (16)$$

In the following it is assumed that  $M_{00}$  is invertible which implies that

$$s_t = A s_{t-1} + B \mathbb{E}_t s_{t+1} + W \epsilon_t \quad (17)$$

with  $A = M_{00}^{-1}M_{10}$ ,  $B = M_{00}^{-1}M_{01}$  and  $W = M_{00}^{-1}M_s$ . The assumption that  $M_{00}$  has to be invertible might be slightly restrictive at first sight. However, the matrix can become only non-invertible when a linear combination of future expectations in  $t + 1$  depend only on linear combinations of past values of endogenous variables and shocks which does not seem to be an issue in practice. Anderson (2008) also

compared and benchmarked a handful DSGE solution algorithms on several models and did not report any issues related to the non-invertibility of  $M_{00}$ .<sup>13</sup> Now let  $S_t := s_t - Cs_{t-1}$  with  $S_t$  and  $C$  to be determined.  $s_t$  can be expressed from the definition and substituted above to obtain

$$S_t + Cs_{t-1} = As_{t-1} + B(\mathbb{E}_t S_{t+1} + Cs_t) + W\epsilon_t \quad (18)$$

Collecting and rearranging terms yields

$$(I - BC)S_t = (BC^2 - C + A)s_{t-1} + B(\mathbb{E}_t S_{t+1}) + W\epsilon_t \quad (19)$$

The backward looking component will drop out of the equation if  $BC^2 - C + A = 0$ . The solution of this quadratic matrix equation can be easily obtained by iterating on

$$C_{k+1} = (I - BC_k)^{-1}A \quad (20)$$

Anderson (2008) reported though that solving this quadratic equation system is very costly and in some cases the iteration failed to converge to the correct solution. A potential reason could be for example that inverses of badly conditioned matrices are inaccurate. As a solution one can re-parametrize the model, pre-condition the equation, or use a different iterative method.<sup>14</sup> We choose the latter and present our approach in detail in the next section. Given a solution for  $C$  the system of equations can be written as follows:

$$S_t = \underbrace{(I - BC)^{-1}B(\mathbb{E}_t S_{t+1})}_{=:F} + \underbrace{(I - BC)^{-1}W\epsilon_t}_{=: \zeta_t} \quad (21)$$

If all eigenvalues of the matrix  $F$  are stable the equation can be easily solved forward to obtain

$$S_t = \sum_{i=0}^{\infty} F^i \mathbb{E}_t \zeta_{t+i} \quad (22)$$

To arrive to the unique stable solution of the original model it suffices to plug back

---

<sup>13</sup>Even if this feature of the algorithm presented an issue a number slightly larger than machine precision could be added to the matrix which would not influence the results.

<sup>14</sup>The quadratic matrix equation could be also solved by other techniques from linear algebra, however this would again involve the calculation of generalized eigenvalues.



the definition of  $S_t$ :

$$s_t = Cs_{t-1} + \sum_{i=0}^{\infty} F^i (I - BC)^{-1} W \mathbb{E}_t \epsilon_{t+i} \quad (23)$$

If structural shocks are uncorrelated then the above formula boils down to:

$$s_t = Cs_{t-1} + (I - BC)^{-1} W \epsilon_t \quad (24)$$

For the vast majority of the DSGE models one can thus apply the following short algorithm to obtain the solution:

---

**Algorithm 4** Binder-Pesaran DSGE Solution

---

1. Rewrite the DSGE model into the following form:  
 $M_{00}s_t = M_{10}s_{t-1} + M_{01}\mathbb{E}_t s_{t+1} + M_s \epsilon_t$
  2. Compute the matrices  $A = M_{00}^{-1}M_{10}$ ,  $B = M_{00}^{-1}M_{01}$  and  $W = M_{00}^{-1}M_s$
  3. Iterate the equation  $C_{k+1} = (I - BC_k)^{-1}A$  with an educated guess or simply setting  $C_0 = A$ , until the matrix  $C$  converges.
  4. Calculate  $D := (I - BC)^{-1}W$  to obtain the solution form:  
 $s_t = Cs_{t-1} + D\epsilon_t$
- 

Hence, by applying this algorithm one obtains the solution to a large class of DSGE models by simple matrix iterations and multiplications which can be differentiated such that the solution method can be implemented into the STAN software package.

### 4.3 Further Computational Issues

To find a solution to a DSGE model Binder and Pesaran (1997) propose to iterate the solution to the  $C$  matrix using the following rule:  $C_{k+1} = (I - BC_k)^{-1}A$ . Although STAN is able to cope with the latter iteration types, the calculation of inverses is computationally one of the most expensive operations, therefore it should be generally avoided. Instead one can directly plug in any initial guess into the equation  $C_{k+1} = BC_k^2 + A$  until it converges.<sup>15</sup>

Although the Binder and Pesaran (1997) algorithm is transparent and easy to implement it has a main drawback. While the solution method proposed by Sims

---

<sup>15</sup>Although there is no guarantee for convergence we were not confronted with the latter issue when executed the algorithm.

(2002) provides conditions which are necessary and also sufficient to guarantee that the model has a unique stable solution, for the Binder-Pesaran algorithm only a set of sufficient conditions under which the a unique stable solution exists can be derived. In particular, the matrix iteration will also converge if the model has multiple equilibria however these solutions are commonly excluded when DSGE models are estimated. Therefore, to assess whether the model has a unique stable solution, we rely on the Sims (2002) algorithm. Although STAN is not capable of dealing with complex numbers, external functions can be included into the algorithm and also partial derivatives of external functions could be manually specified. Yet, this is not necessary as the Sims (2002) algorithm is used only to reject the sample draw in case the Hamiltonian sampler enters a point in the parameter space where the model has no unique stable solution. For the latter purpose no calculation of the derivative is needed. To implement several matrix decompositions to execute the Sims (2002) algorithm we rely on the Intel Math Kernel Library (Intel MKL), a collection of BLAS and LAPACK algorithms which also Matlab uses, which we link into our STAN C++ code.

A further computational issue arises when the covariance matrix  $\Sigma$  is initialized for the Kalman filter. Hamilton (1994) proposes to use Kronecker products to solve for  $\Sigma$  which STAN is able to handle, however it is computationally very costly since the dimension of the problem grows quadratically with the number of equations the model consists of. Since the solution of the DSGE model has to be non-explosive we can obtain  $\Sigma$  again by an iterative procedure. However, as  $\Sigma$  has an impact on the log-likelihood the calculation of this part of the gradient is costly once a large number of iterations is necessary to achieve convergence. The initial variance is generally obtained by solving the discrete Lyapunov equation which belongs to the class of Stein matrix equations. Several iterative procedures are proposed in Zhoua et al. (2009) which accelerate the iteration exponentially and enable to calculate parts of the derivative in one step.

Altering the iteration procedure in the Binder-Pesaran algorithm and the adoption of a more efficient calculation to initialize the Kalman filter speeded up the algorithm by a factor of 3-4 for a mid-scale NK-model. In general we can state that calculation of the gradient is costly therefore streamlining the model setup is nec-

essary as far as possible to avoid additional computational burden which increases exponentially with the dimension of the model.

## 5 Estimation Results

In this section we present the results obtained by applying the HMC algorithm to the textbook small scale New-Keynesian model proposed in Herbst and Schorfheide (2015) and subsequently to the Smets and Wouters (2007) model, a medium scale model serving as the core for a wide range of applied policy models.

### 5.1 A Small Scale New Keynesian Model

The most basic DSGE model estimated in Herbst and Schorfheide (2015), similar to the one estimated in An and Schorfheide (2007), is a slightly altered version of the standard three equation textbook New Keynesian model (see e.g. in Clarida et al. (1999) and the references therein). Yet, Herbst and Schorfheide (2015) uses quadratic price adjustment à la Rotemberg (1983) instead of the Calvo (1983) scheme and also adds a government sector to the model. Hence, it consists of the dynamic IS curve, the New Keynesian Phillips curve and a Taylor-type monetary policy rule. Both the technology shock and the government spending shock is  $AR(1)$ . The model can be summarized by the following equations:<sup>16</sup>

$$\hat{y}_t = \mathbb{E}_t[\hat{y}_{t+1}] - \frac{1}{\tau} \left( \hat{R}_t - \mathbb{E}_t[\hat{\pi}_{t+1}] - \mathbb{E}_t[\hat{z}_{t+1}] \right) + \hat{g}_t - \mathbb{E}_t[\hat{g}_{t+1}] \quad (25)$$

$$\hat{\pi}_t = \beta \mathbb{E}_t[\hat{\pi}_{t+1}] + \kappa(\hat{y}_t - \hat{g}_t) \quad (26)$$

$$\hat{R}_t = \rho_R \hat{R}_{t-1} + (1 - \rho_R) \psi_1 \hat{\pi}_t + (1 - \rho_R) \psi_2 (\hat{y}_t - \hat{g}_t) + \epsilon_{R,t} \quad (27)$$

$$\hat{g}_t = \rho_g \hat{g}_{t-1} + \epsilon_{g,t} \quad (28)$$

$$\hat{z}_t = \rho_z \hat{z}_{t-1} + \epsilon_{z,t} \quad (29)$$

To estimate the model, three observables are used: GDP growth, inflation and the

---

<sup>16</sup>For further details we direct the reader to (Herbst and Schorfheide, 2015, pp.15-28.).

nominal interest rate. These are linked to the state equations as follows:

$$YGR_t = \gamma^{(Q)} + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{z}_t) \quad (30)$$

$$INFL_t = \pi^{(A)} + 400\hat{\pi}_t \quad (31)$$

$$INT_t = \pi^{(A)} + 4\gamma^{(Q)} + 400\hat{R}_t \quad (32)$$

In this setup, we do not allow for any measurement error. The small scale model thus has 13 structural parameters to be estimated:

$$\theta = [\tau, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_r, \sigma_g, \sigma_z, r^A, \pi^A, \gamma^Q] \quad (33)$$

The priors we assume are similar to those used in Herbst and Schorfheide (2015) and are summarized in the table below.

Table 1: Prior Distributions

Name	Domain	Distribution	Parameter 1	Parameter 2
$\tau$	$[0, \infty)$	Gamma	2.00	0.50
$\kappa$	$[0, 1)$	Uniform	0.00	1.00
$\psi_1$	$[0, \infty)$	Gamma	1.50	0.25
$\psi_2$	$[0, \infty)$	Gamma	0.50	0.25
$r^{(A)}$	$[0, \infty)$	Gamma	0.50	0.50
$\pi^{(A)}$	$[0, \infty)$	Gamma	7.00	2.00
$\gamma^{(Q)}$	$(-\infty, \infty)$	Normal	0.40	0.20
$\rho_r$	$[0, 1)$	Uniform	0.00	1.00
$\rho_g$	$[0, 1)$	Uniform	0.00	1.00
$\rho_z$	$[0, 1)$	Uniform	0.00	1.00
$100\sigma_r$	$[0, \infty)$	Inv. Gamma	0.40	4.00
$100\sigma_g$	$[0, \infty)$	Inv. Gamma	1.00	4.00
$100\sigma_z$	$[0, \infty)$	Inv. Gamma	0.50	4.00

*Notes:* For the Beta, Gamma and Normal distribution Parameter 1 and Parameter 2 stands for the mean and the standard deviation. For the Uniform distribution the parameters define the bounds of the interval. For the Inverse Gamma distribution they correspond to parameters  $s$  and  $\nu$ , where  $p_{IG}(\sigma) \propto \sigma^{-\nu-1} e^{-\nu s^2/2\sigma^2}$ . See also Herbst and Schorfheide (2015).

For the estimation, we used 10 parallel chains with each 1,000 draws. Due to the efficiency of the HMC a burn-in of 500 draws is sufficient to ensure that the sampler finds regions of high probability which is also confirmed by the diagnostics. To visualize the diagnostics of the HMC method we used ShinyStan Version 3.0 (Gabry and Veen, 2020).

Table 2 shows the statistics describing the sampling efficiency of the HMC for each of the structural parameters and the log-posterior as well. Studying the numerical diagnostics of the sampling efficiency two of the main advantages of the HMC algorithm becomes visible: the high effective sample size, and the high accuracy of the simulation of the target density.

Recall, the first is due to the greatly reduced autocorrelation of the draws, introduced by the random variation in the total energy, i.e. by the random variation of the momentum. The latter is warranted by the smart application of the gradient to set the trajectory in the phase space along the Hamiltonian, i.e. the Hamiltonian equations ensure that all draws, after initial convergence, are from the target distribution.

This improvement in efficiency is why we consider HMC being a significant improvement for DSGE estimation. Herbst and Schorfheide (2015) report the inefficiency factor for the relative risk aversion parameter ( $\tau$ ) for the different RWMH algorithms. The inefficiency factor is the inverse of  $N_{eff}/N$ , and note that the RWMH suffers from high inefficiency due to its high autocorrelation. To grasp the leap in efficiency, we highlight that the naive identity matrix based Metropolis proposal has an inefficiency that translates the "100,000 draws [...] is about as accurate as an approximation obtained from 5.5 *iid* draws" (Herbst and Schorfheide, 2015, p.119.), while the standard, benchmark RWMH algorithm described in Chapter 2 has an inefficiency that increases the effective sample size to 1,137,<sup>17</sup> The 3-Block RWMH algorithm results in an equivalent of 2,440 *iid* draws. In comparison the effective sample size ( $N_{eff}$ ) for the 100,000 draws with HMC is 89,737 for the risk aversion parameter ( $\tau$ ). In other words the HMC estimation represents 78.60 fold efficiency improvement over the standard, 1-Block, RWMH algorithm and a 36.63 fold over the 3-Block RWMH algorithm. However, the efficiency gain comes at a cost in terms of computational time, as the gradient has to be evaluated.

Another advantage of weakly autocorrelated draws is the potential to run fully independent shorter chains in parallel, in other words, STAN based HMC is highly parallelizable. The evaluation of the gradient and its computation for each transi-

---

<sup>17</sup>Herbst and Schorfheide (2015) report the inefficiency factor of 88 for the 1-Block RWMH algorithm for the parameter  $\tau$ . In terms of inefficiency factor the HMC has a 1.12 inefficiency factor.

tion is an increasingly difficult task in the number of structural parameters. The C++ level integration of the automated differentiation and the computational improvements discussed before renders HMC also for larger models feasible.<sup>18</sup>

Lastly, and probably most importantly, we need to highlight the fact that due to the higher convergence of the draws to the typical set, we can abandon the practice of a mode-estimation before sampling. This potentially also improves the reliability of our estimation method in higher dimensional models considerably, as discussed by Betancourt (2018). We are confident that future research will highlight the advantages of HMC in large DSGE models with irregularly shaped posteriors.

Table 2: Sampling Efficiency of the Hamiltonian Monte Carlo

Parameter	$N_{eff}/N$	MCSE/SD	Parameter	$N_{eff}/N$	MCSE/SD
$\tau$	89.37 %	1.05%	$\rho_r$	65.70 %	1.23%
$\kappa$	91.06 %	1.05%	$\rho_g$	94.94 %	1.03%
$\psi_1$	74.18%	1.16%	$\rho_z$	56.72%	1.33 %
$\psi_2$	67.16%	1.22%	$100\sigma_r$	74.72 %	1.16%
$r^{(A)}$	58.19%	1.31%	$100\sigma_g$	94.07 %	1.03%
$\pi^{(A)}$	50.44 %	1.47 %	$100\sigma_z$	89.45 %	1.06 %
$\gamma^{(Q)}$	55.57 %	1.34 %	Log-Posterior	36.24 %	1.66 %

*Notes:* The table summarizes the efficiency of the HMC sampling. The first column ( $N_{eff}/N$ ) displays the effective sample size divided by the total number of draws for the structural parameters of the Small Scale DSGE model and its posterior in percentages (%). A higher number indicates more efficient sampling for the respective parameter. The second column (MCSE/SD) contains the ratio of the Monte Carlo standard error of the mean (MCSE) to the posterior standard deviation (SD), again in percentages (%). Here a lower number indicates a more efficient sampling.

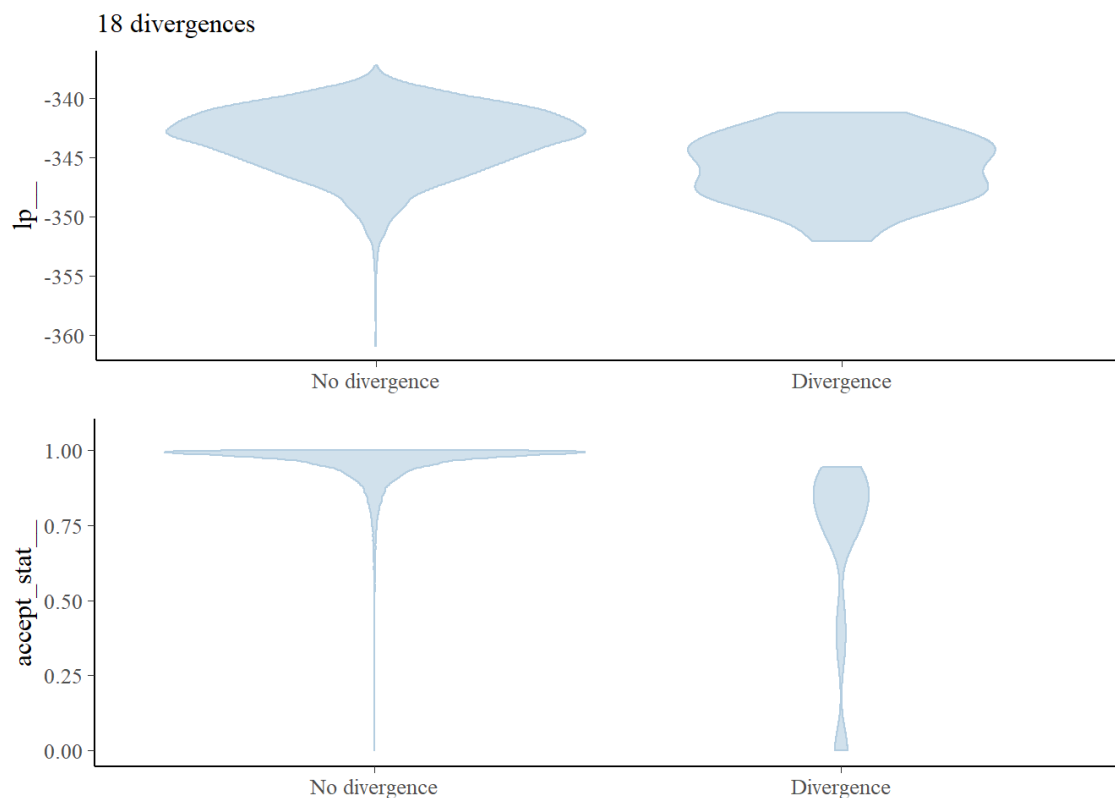
The second column of Table 2 reports the ratio of the Monte Carlo standard error of the mean (MCSE) to the posterior standard deviation (SD). The former is related to the accuracy of the simulation, the smaller the standard error, the closer the estimated parameter value is to the true value. The latter gives the total uncertainty around the structural parameter. The ratio is considered to be small if it is below 5%, thus the values around 1% are indicative of a highly efficient sampling.

Turning to the diagnostics of the sampling, we start with the number and properties of the divergent transitions. Divergent transitions occur when the approximated path calculated by the leapfrog integrator drifts completely away from the original

<sup>18</sup>With the advances of GPU computing in STAN and propagation of higher CPU core counts we anticipate another jump in the computational speed the coming years further advancing the applicability of our solution to estimate DSGE models.

path. Possible reasons for divergent trajectories could be a way too high curvature for the calibrated step size or other irregularities in the posterior likelihood as cliffs or funnels. A prominent example for the latter is presented in Neal (2003), referred to as Neal’s funnel. In general the existence of divergent transitions is a warning sign that results might be invalid, however rejected transitions might be also false positive. In other words, if they do not display a common pattern, and are a low proportion, then they can be neglected. From 10,000 draws we observed approximately 18 divergent iterations, that is 0.2%. The existence of divergent transitions can indicate invalidity of the results, however the fact that their number is very low in relative terms, and they show no systematic pattern, we argue that the results are to be trusted. Figure 4 plots the frequency of divergent transitions against the log-posterior ( $lp_{-}$ ) in the top panel, and the acceptance of statistic ( $accept\_stat_{-}$ ) in the bottom panel.

Figure 4: Small Scale DSGE Diagnostics: Divergence Information



*Notes:* Plots of the divergent transitions (x-axis) against the log-posterior (y-axis top panel) and against the acceptance statistic (y-axis bottom panel) of the Hamiltonian Monte Carlo sampling algorithm.

From the top panel we can see directly the log-posterior distribution. It is worth noting that the divergent transitions are mostly in the medium probability regions, and not in the high, indicating that any divergence could be a false positive, i.e. divergent due to the numerical instability given the complexity of the entire framework. The location of the divergent transitions can provide information which parts of the target distribution is difficult to sample from, albeit comparing the two charts, we can conclude that the sampler did explore the difficult regions of the posterior. Turning to the bottom panel one might be cautious due to the high acceptance rate<sup>19</sup>. In general the intuition applies for the HMC as well that if the acceptance rate is very high it might be indicative of inefficient sampling<sup>20</sup>. To reject this possibility we plot the marginal posterior distributions and the scatter plot of the acceptance rate and the log-posterior in Figure 5.

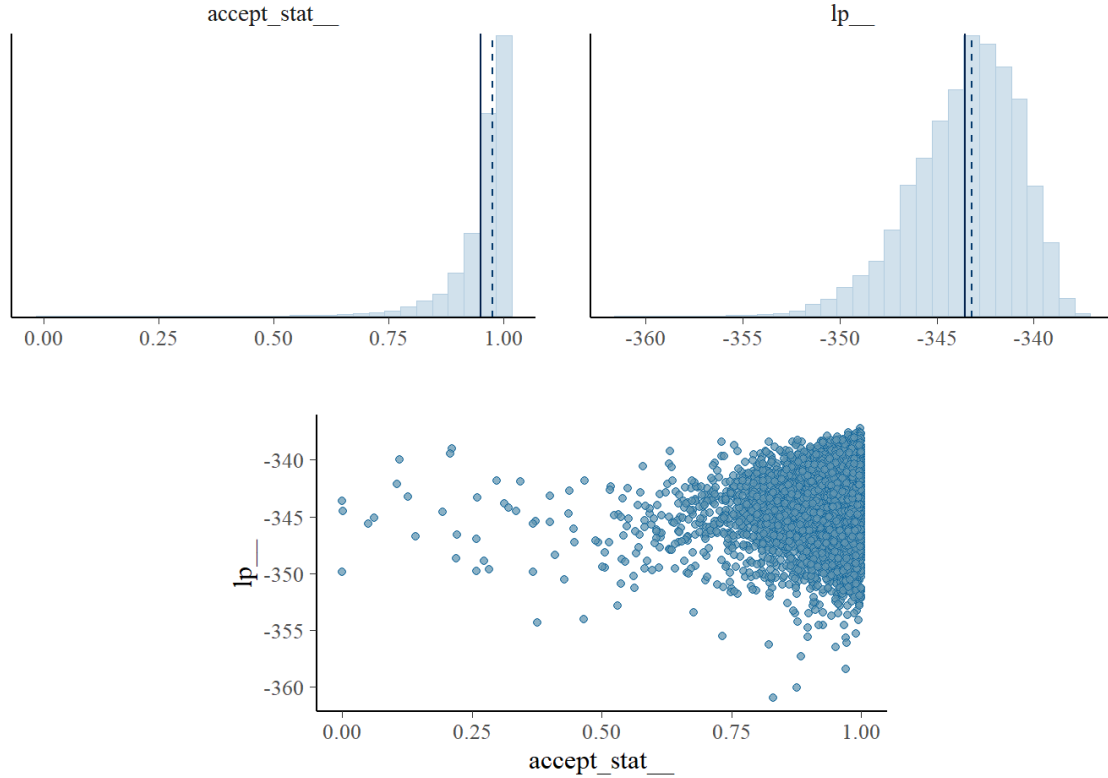
---

<sup>19</sup>The acceptance rate refers to the intermediate Metropolis step in the HMC Algorithm implemented in STAN.

<sup>20</sup>It should be noted that STAN allows to set the target Metropolis acceptance rate with a specific control option that adapts the jump size based on the sampling during the burn in phase.



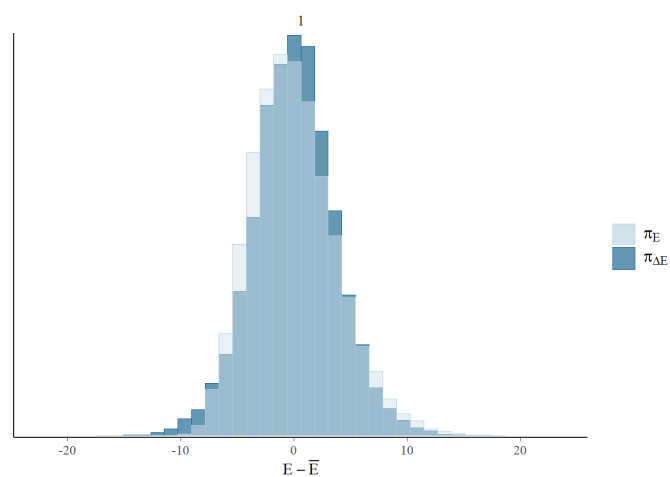
Figure 5: Small Scale DSGE Diagnostics: Acceptance Information



*Notes:* The figure plots the marginal posterior distribution of acceptance statistic (top left panel), marginal posterior distribution of the log-posterior (top right panel), and the scatter plot of acceptance statistic (x-axis bottom panel) against the log-posterior (y-axis bottom panel). The vertical lines indicate the mean (solid line) and median (dashed line). A bad plot would show a relationship between the acceptance statistic and the log-posterior.

The figure shows no relationship of the acceptance rate and the log-posterior, in fact it indicates that the posterior has been adequately explored. This leads us to the discussion of the energy distribution in order to assess robustness of the HMC algorithm, shown in Figure 6. It is desirable that the histograms are "well-matched: [...] The closer  $\pi_{\Delta E}$  is to  $\pi_E$  the faster the random walk explores the energies and the smaller the autocorrelations will be in the chain" (Gabry and Veen, 2020). Figure 6 shows the reason for the low autocorrelation, and thus the high efficiency of the HMC algorithm, the energy levels, and with it the posterior-probability levels of the target distribution, are well explored.

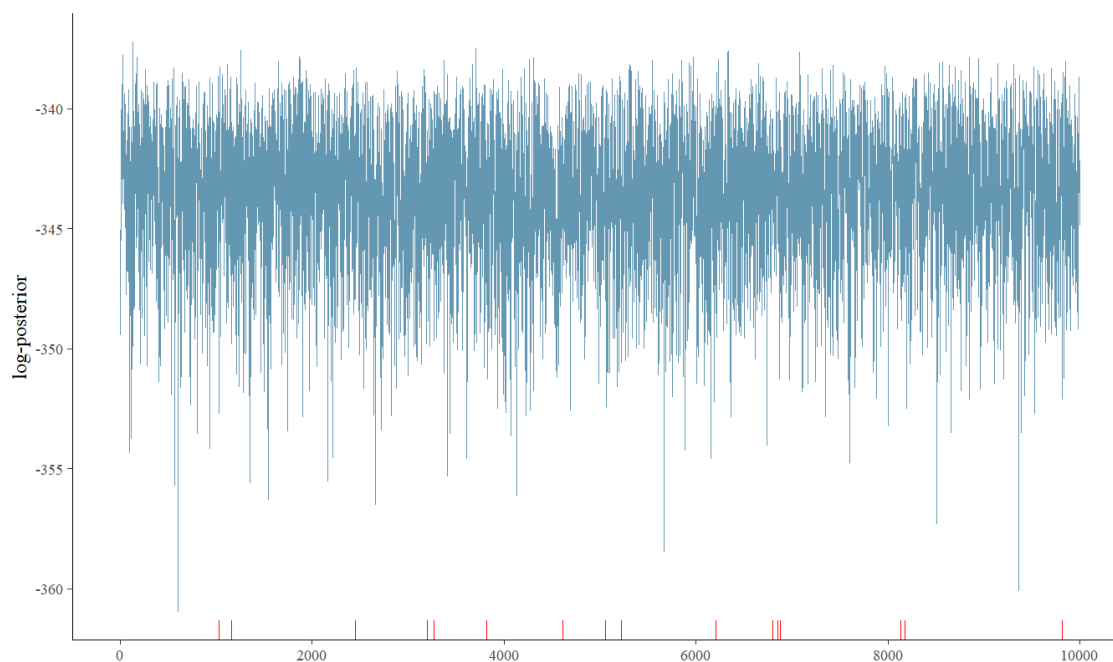
Figure 6: Small Scale DSGE Diagnostics: Energy Distribution



*Notes:* These are plots of the overlaid histograms of the marginal energy distribution ( $\pi_E$ ) and the energy transition distribution ( $\pi_{\Delta E}$ ). See Betancourt (2018) and Carpenter et al. (2017) for more details.

Lastly discussing the trace plot of the log-posterior we can visually inspect the sampling behavior. Figure 7 shows that the chain explored the different parts of the parameter space. This applies to the other chains and structural parameters as well, all indicating a proper sampling, that is the chains are mixing sufficiently well.

Figure 7: Small Scale DSGE Diagnostics: Trace plot



*Notes:* The log-posterior of the draws from the Hamiltonian Monte Carlo are shown in blue. Divergent transitions are marked on the x-axis with red lines.

Turning to the structural parameter estimates, one can verify that the posterior estimates from HMC are the same as the ones obtained with the RWMH algorithm.

This verifies the proper functioning of the algorithm, and tells that in small DSGE models with simple target densities, RWMH sampling works sufficiently well. To explore the properties of HMC in a larger model the next section presents the estimation of the Smets-Wouters model.

Table 3: Posterior Estimates of the Small Scale DSGE Model

Parameter	Hamiltonian Monte Carlo		Random Walk Metropolis Hastings	
	Mean	[0.05, 0.95]	Mean	[0.05, 0.95]
$\tau$	2.43	[1.62, 3.35]	2.37	[1.58, 3.82]
$\kappa$	0.85	[0.62, 0.99]	0.85	[0.62, 0.98]
$\psi_1$	1.95	[1.59, 2.34]	1.92	[1.55, 2.20]
$\psi_2$	0.61	[0.21, 1.15]	0.60	[0.20, 1.21]
$r^{(A)}$	0.42	[0.05, 0.90]	0.44	[0.05, 0.95]
$\pi^{(A)}$	3.41	[2.79, 4.03]	3.38	[2.76, 3.80]
$\gamma^{(Q)}$	0.60	[0.37, 0.83]	0.60	[0.37, 0.74]
$\rho_r$	0.81	[0.76, 0.85]	0.77	[0.71, 0.82]
$\rho_g$	0.98	[0.95, 1.00]	0.98	[0.95, 1.00]
$\rho_z$	0.93	[0.90, 0.97]	0.92	[0.88, 0.92]
$100\sigma_r$	0.19	[0.16, 0.20]	0.22	[0.18, 0.26]
$100\sigma_g$	0.67	[0.59, 0.78]	0.65	[0.57, 0.84]
$100\sigma_z$	0.19	[0.16, 0.23]	0.20	[0.16, 0.36]

*Notes:* The table shows the posterior mean and the 5 and 95 percentile of the posterior from the HMC and the RWMH estimation, respectively. The results for HMC are based collectively on  $N = 10,000$  draws from the posterior, obtained with 10 parallel chains, with a burn in of 500 draws and 1,000 sample draws for each.

The results for the RWMH algorithm are based on the authors' replication of the table reported in Herbst and Schorfheide (2015) using the original code available with 100,000 draws. Please note the slight difference in the posterior estimates and the different notation for the scaling of the shock variances compared to Herbst and Schorfheide (2015). We attribute the former to the inherent random nature of the sampling.

## 5.2 Smets-Wouters Model

The Smets and Wouters (2007) model is a medium-scale closed economy DSGE model. It has become the standard workhorse model for economic policy analysis and served as a basis for newer generations of DSGE models that have followed. It was estimated for the US with the RWMH algorithm for the sample of 1960:1–2004:4 using seven key macroeconomic variables: real GDP, real consumption, real

investment, the GDP deflator, real wages, employment and the nominal short-term interest rate.<sup>21</sup> The model features a deterministic growth rate driven by labour-augmenting technology progress. The model is subject to nominal and real frictions. The former affecting the labour and goods markets as Calvo-type nominal rigidities similar to Christiano et al. (2005). Both wages and intermediate product markets are subject to partial indexation to lagged inflation. The real frictions manifest themselves as investment adjustment and capital utilization costs. Monetary policy follows a Taylor type rule, with interest rate smoothing and the reaction to inflation and output gap, the former defined as the deviation from the estimated steady state inflation, the latter as the distance to the flexible price economy.

Exogenous variation of the model is driven by seven exogenous shock processes: the standard total factor productivity, monetary policy, investment specific technology, exogenous spending, the model features a risk premium shock and wage and price markup shocks with a MA structure. The latter property introduces anticipated, news shocks for both the regular and the wage Phillips curve. All exogenous shocks are iid-normal with zero mean and estimated variance. The model is log-linearized around the steady state and net of deterministic growth rate. Variables are expressed in terms of percentage deviations from steady state. In order to introduce anticipated news shocks we augment the model with auxiliary state variables, similar to Dynare, so the Binder-Pesaran algorithm can be easily applied.

We estimate the Smets-Wouters model with HMC and present the sampling diagnostics in the Appendix. Once again, the efficiency of the HMC algorithm is apparent. Even though we estimate the model with 1000 draws only, it results in an effective sample size of 418.92 for the log-posterior. We failed to find any divergent transitions and a well behaved sampling behaviour that explored the target density well. Comparing the posterior results presented in Table 4 and 5 we can conclude that both estimations deliver similar results. The only exception refers to the steady state of inflation, that is estimated to be slightly lower with HMC. This is a result already documented in the literature, see e.g. Chib and Ramamurthy (2010), and might be also a reason for the poor performance of the Smets-Wouters model for

---

<sup>21</sup>Both real consumption and investments are deflated using the GDP deflator. The hours variable is defined as average weekly hours of all persons in the non-farm business sector times total civilian employment.

inflation forecast, as pointed out in Del Negro et al. (2015).

Table 4: Posterior Estimates of the Smets-Wouters Structural Parameters

Parameter	Hamiltonian Monte Carlo		Random Walk Metropolis Hastings	
	Mean	[0.05, 0.95]	Mean	[0.05, 0.95]
$\varphi$	5.93	[4.38, 7.68]	5.93	[4.26, 7.64]
$\sigma_c$	1.41	[1.20, 1.66]	1.42	[1.19, 1.65]
$h$	0.73	[0.65, 0.80]	0.73	[0.66, 0.80]
$\xi_w$	0.75	[0.67, 0.84]	0.75	[0.66, 0.84]
$\sigma_l$	2.10	[1.21, 3.02]	2.06	[1.11, 2.93]
$\xi_p$	0.65	[0.56, 0.73]	0.64	[0.56, 0.73]
$\iota_w$	0.56	[0.35, 0.77]	0.57	[0.37, 0.78]
$\iota_p$	0.24	[0.11, 0.38]	0.23	[0.09, 0.37]
$\psi$	0.46	[0.30, 0.65]	0.47	[0.30, 0.64]
$\Phi$	1.65	[1.53, 1.79]	1.63	[1.50, 1.76]
$r_\pi$	2.04	[1.77, 2.33]	2.05	[1.78, 2.32]
$\rho$	0.82	[0.77, 0.85]	0.82	[0.78, 0.86]
$r_y$	0.10	[0.07, 0.14]	0.10	[0.06, 0.14]
$r_{dy}$	0.21	[0.17, 0.25]	0.21	[0.17, 0.25]
$\bar{\pi}$	0.67	[0.50, 0.86]	0.77	[0.59, 0.94]
$100(\beta^{-1} - 1)$	0.13	[0.07, 0.22]	0.15	[0.06, 0.23]
$\bar{l}$	0.88	[-0.68, 2.45]	0.74	[-1.00, 2.45]
$\bar{\gamma}$	0.47	[0.43, 0.49]	0.47	[0.44, 0.49]
$\alpha$	0.21	[0.18, 0.23]	0.20	[0.17, 0.23]

*Notes:* The table shows the posterior mean and the 5 and 95 percentile of the posterior from the HMC and the RWMH estimation, respectively. The results for HMC are based on  $N = 1000$  draws from the posterior and a burn in of 500.

The results for the RWMH algorithm are based on the authors' replication of the model using Johannes Pfeiffer's replication files written in Dynare with an acceptance rate of 30.42%, two chains of 500,000 draws and a burn in of 100,000. Thus the resulting number of draws is 800,000.

The results obtained and the HMC diagnostics together seem to indicate that the target density of the Smets-Wouters model is well behaved. Thus the application of the RWMH algorithm should be warranted as long as tight priors are assumed. However, a brief glance at the autocorrelation function of certain parameters suggests that the sample suffers from slight autocorrelation. As HMC is able to travel large distances in the parameter space, in the optimum one should obtain an uncorrelated sample for each parameter. Slight autocorrelation per se does not invalidate sampling, yet it is an indication for inefficient sampling for which feature the reason is not apparent at first sight.

An evident idea to identify the reason for inefficient sampling is to start the au-

Table 5: Posterior Estimates of the Smets-Wouters Model's Shock Processes

Parameter	Hamiltonian Monte Carlo		Random Walk Metropolis Hastings	
	Mean	[0.05, 0.95]	Mean	[0.05, 0.95]
$\sigma_a$	0.48	[0.43, 0.52]	0.47	[0.42, 0.51]
$\sigma_b$	0.24	[0.19, 0.28]	0.23	[0.19, 0.28]
$\sigma_g$	0.52	[0.47, 0.57]	0.51	[0.46, 0.56]
$\sigma_I$	0.46	[0.38, 0.54]	0.45	[0.37, 0.53]
$\sigma_r$	0.23	[0.21, 0.25]	0.23	[0.21, 0.26]
$\sigma_p$	0.13	[0.10, 0.16]	0.13	[0.11, 0.16]
$\sigma_w$	0.25	[0.21, 0.28]	0.24	[0.21, 0.28]
$\rho_a$	0.98	[0.97, 0.99]	0.98	[0.97, 0.99]
$\rho_b$	0.27	[0.12, 0.48]	0.28	[0.10, 0.46]
$\rho_g$	0.97	[0.96, 0.99]	0.97	[0.96, 0.99]
$\rho_I$	0.69	[0.59, 0.78]	0.69	[0.60, 0.79]
$\rho_r$	0.17	[0.07, 0.28]	0.17	[0.06, 0.28]
$\rho_p$	0.96	[0.92, 0.99]	0.96	[0.92, 0.99]
$\rho_w$	0.97	[0.94, 0.99]	0.97	[0.95, 0.99]
$\mu_p$	0.80	[0.66, 0.90]	0.80	[0.69, 0.91]
$\mu_w$	0.89	[0.82, 0.94]	0.89	[0.82, 0.95]
$\rho_{ga}$	0.57	[0.44, 0.70]	0.54	[0.41, 0.68]

*Notes:* The table shows the posterior mean and the 5 and 95 percentile of the posterior from the HMC and the RWMH estimation, respectively. The results for HMC are based on  $N = 1000$  draws from the posterior and a burn in of 500.

The results for the RMWH algorithm are based on the authors' replication of the model using Johannes Pfeiffer's replication files written in Dynare with an acceptance rate of 30.42%, two chains of 500,000 draws and a burn in of 100,000. Thus the resulting number of draws is 800,000.

tomatic tuning procedure from a different starting point then before. The gradient based approach is designed to navigate the chain to find the typical set. After starting the chain from a different spot the HMC algorithm indeed found a further mode in spite of setting tight priors and using the data sample 1960Q1-2004Q4. Cai et al. (2020) have already documented that even if setting priors identical to the original paper, yet using a shortened sample, in particular data from 1960Q1 to 1991Q3, will lead to identification problems as the data does not contain sufficient information to pin down properly a handful parameters. While Cai et al. (2020) document that the parameters  $h$  (habit persistence in consumption),  $\iota_p$  (degree of price indexation),  $\rho_p$  (persistence of price markup shock) and  $\rho_{ga}$  (loading of government spending on technology shock innovations) exhibit a multimodal pattern, our results suggest that also several other parameters differ from those documented in the baseline estimation. In particular, we obtain that in addition to further parameters related to price

and wage setting, e.g.  $\xi_p$ ,  $\xi_w$ ,  $\sigma_p$ ,  $\mu_p$ , there seems to be also a significant difference in the parameter estimates for  $\varphi$  (investment adjustment costs),  $\psi$  capacity utilization costs and  $\rho_b$  (persistence of risk premium shock), at least if compared with the differences in the estimates for the rest of the parameters.<sup>22</sup> We also double checked our results obtained using HMC by estimating the model with RWMH in Dynare and starting the estimation from the mean of the parameter estimates from the alternative mode obtained with HMC. The results obtained are fairly similar. The pattern that RWMH estimates steady state inflation by approximately 0.1 higher than HMC remains though. Finally, we also observed that using RWMH and running two parallel chains one of the chains swaps the mode and remains there for the rest of the time. However, estimates obtained by averaging over the entire sample from both chains are likely to be biased and therefore more sophisticated methods are required. In general, it is noteworthy though that improved HMC diagnostics could contribute to uncover irregularities in the posterior likelihood.

---

<sup>22</sup>The results also hold up if we estimate the model on the original sample 1966Q1-2004Q4 used in Smets and Wouters (2007).



Table 6: Posterior Estimates of the Smets-Wouters Structural Parameters - Alternative Mode

Parameter	Hamiltonian Monte Carlo		Random Walk Metropolis Hastings	
	Mean	[0.05, 0.95]	Mean	[0.05, 0.95]
$\varphi$	5.43	[3.76, 7.17]	5.49	[3.67, 7.34]
$\sigma_c$	1.41	[1.18, 1.64]	1.42	[1.18, 1.66]
$h$	0.69	[0.58, 0.77]	0.69	[0.59, 0.79]
$\xi_w$	0.80	[0.73, 0.87]	0.80	[0.72, 0.87]
$\sigma_l$	2.13	[1.19, 3.12]	2.09	[1.14, 3.01]
$\xi_p$	0.80	[0.75, 0.85]	0.78	[0.73, 0.83]
$\iota_w$	0.52	[0.33, 0.72]	0.51	[0.31, 0.71]
$\iota_p$	0.31	[0.17, 0.48]	0.34	[0.17, 0.50]
$\psi$	0.40	[0.24, 0.56]	0.41	[0.24, 0.57]
$\Phi$	1.63	[1.50, 1.77]	1.62	[1.48, 1.54]
$r_\pi$	1.97	[1.68, 2.26]	1.98	[1.70, 2.25]
$\rho$	0.85	[0.82, 0.88]	0.85	[0.81, 0.88]
$r_y$	0.13	[0.08, 0.17]	0.12	[0.08, 0.16]
$r_{dy}$	0.22	[0.18, 0.27]	0.23	[0.18, 0.27]
$\bar{\pi}$	0.67	[0.51, 0.83]	0.77	[0.60, 0.94]
$100(\beta^{-1} - 1)$	0.14	[0.07, 0.24]	0.15	[0.06, 0.24]
$\bar{l}$	0.61	[-0.92, 2.05]	0.46	[-1.11, 2.06]
$\bar{\gamma}$	0.46	[0.43, 0.49]	0.47	[0.44, 0.49]
$\alpha$	0.21	[0.18, 0.24]	0.20	[0.17, 0.23]

*Notes:* The table shows the posterior mean and the 5 and 95 percentile of the posterior from the HMC and the RWMH estimation, respectively. The results for HMC are based on  $N = 1,000$  draws from the posterior and a burn in of 500.

The results for the RWMH algorithm are based on the authors' replication of the model using Johannes Pfeiffer's replication files using Dynare with an acceptance rate of 31.04%, one chain of 1,000,000 draws and a burn in of 200,000. Thus the resulting number of draws is 800,000.

Table 7: Posterior Estimates of the Smets-Wouters Model's Shock Processes - Alternative Mode

Parameter	Hamiltonian Monte Carlo		Random Walk Metropolis Hastings	
	Mean	[0.05, 0.95]	Mean	[0.05, 0.95]
$\sigma_a$	0.48	[0.44, 0.53]	0.48	[0.42, 0.51]
$\sigma_b$	0.21	[0.15, 0.26]	0.21	[0.14, 0.26]
$\sigma_g$	0.52	[0.47, 0.56]	0.51	[0.46, 0.55]
$\sigma_I$	0.45	[0.37, 0.53]	0.44	[0.36, 0.52]
$\sigma_r$	0.23	[0.21, 0.25]	0.23	[0.21, 0.25]
$\sigma_p$	0.21	[0.19, 0.24]	0.21	[0.18, 0.24]
$\sigma_w$	0.23	[0.20, 0.27]	0.23	[0.20, 0.26]
$\rho_a$	0.97	[0.96, 0.98]	0.98	[0.96, 0.99]
$\rho_b$	0.41	[0.19, 0.68]	0.40	[0.16, 0.67]
$\rho_g$	0.97	[0.96, 0.99]	0.98	[0.96, 0.99]
$\rho_I$	0.71	[0.61, 0.80]	0.71	[0.61, 0.81]
$\rho_r$	0.13	[0.05, 0.22]	0.14	[0.04, 0.23]
$\rho_p$	0.93	[0.89, 0.96]	0.92	[0.89, 0.96]
$\rho_w$	0.97	[0.93, 0.99]	0.96	[0.94, 0.99]
$\mu_p$	0.98	[0.97, 0.99]	0.98	[0.97, 1.00]
$\mu_w$	0.91	[0.85, 0.95]	0.90	[0.84, 0.96]
$\rho_{ga}$	0.57	[0.45, 0.70]	0.55	[0.42, 0.68]

*Notes:* The table shows the posterior mean and the 5 and 95 percentile of the posterior from the HMC and the RWMH estimation, respectively. The results for HMC are based on  $N = 1,000$  draws from the posterior and a burn in of 500.

The results for the RWMH algorithm are based on the authors' replication of the model using Johannes Pfeiffer's replication files using Dynare with an acceptance rate of 31.04%, one chain of 1,000,000 draws and a burn in of 200,000. Thus the resulting number of draws is 800,000.

## 6 Extension: Sequential Hamiltonian Monte Carlo

One of the main disadvantages of the HMC algorithm is that it fails to explore multimodal posterior distributions, as documented in existing literature, see e.g. Shiwei et al. (2014). An interesting experiment, which also addresses criticism by researchers with respect to the estimation setup of the original Smets-Wouters model, was carried out in Herbst and Schorfheide (2014). In particular, in the latter work the authors unrestricted the Bayesian model by using uninformative priors for a number of parameters instead of setting tight priors as in Smets and Wouters (2007). Hence, they allow for the information to obtain a larger weight when estimating the model. Herbst and Schorfheide (2014) reports a bimodal shape of the marginal posterior density for a handful parameters once uninformative priors are applied in which case widely used MCMC based samplers as the RWMH algorithm do not mix properly. Instead, commonly used samplers get stuck in one of the modes, depending on the starting point of the chain. To remedy the issue of multimodality several algorithms have already been proposed in the literature, e.g. Neal (2001), Liu and Chen (1998), Gilks and Berzuini (2002) and Del Moral et al. (2006) where the latter works mainly combine three different algorithms: importance sampling and resampling, rejection sampling, and Markov chain iterations. Chopin (2004) derives a central limit theorem for a large class of SMC sampling methods. Herbst and Schorfheide (2014) carried out pioneering work by introducing the SMC algorithm to DSGE models to remedy issues with multimodality. The proposed SMC framework in Herbst and Schorfheide (2014) fits into the scheme described by Chopin (2004) and is in principle a sequential importance sampler. In each step the posterior density  $p(Y|\theta)^{\beta_n}p(\theta)$  at stage  $n$ , where the likelihood is weighted by  $0 \leq \beta_n \leq 1 \forall n$ , serves as a proposal density for the density to be sampled from at the next stage  $p(Y|\theta)^{\beta_{n+1}}p(\theta)$  with  $\beta_{n+1} > \beta_n$ . This framework is commonly referred to as likelihood tempering in existing literature, e.g. Herbst and Schorfheide (2014). Alternatively, one can also carry out data tempering by increasing the number of observations included to calculate the likelihood function at each stage. At each stage the importance weights for all draws  $\{\theta_j^{(n)}\}_{j=1}^J$  at stage  $n$ , that is the fraction of the posterior densities at stage  $n+1$  and  $n$  equaling to  $p(Y|\theta^{(n)})^{\beta_{n+1}-\beta_n}$ , is calculated and serve as the weights for the importance sampling. The swarm of parameter draws and weights  $\{\theta_j^{(n)}, w_j^{(n)}\}_{j=1}^J$

together at each stage are commonly referred to as *particles*. Once the variance of the weights becomes large the draws are resampled using the actual weights and the weights are reset to unity. Finally at each stage the draws are mutated or moved applying a Metropolis-Hastings step, which is alternatively referred to as the 'rejuvenation' step.

A main drawback of using the RWMH sampler to rejuvenate the parameter draws at each stage is again that the MH-proposal  $\theta'$  is either too often rejected or the distance  $\|\theta - \theta'\|$  between the proposal and the current parameter draw is relatively small. In case one targets an acceptance rate of 25 percent each particle will be updated only at each fourth stage on average. The intuition behind likelihood tempering is also that decreasing  $\beta_n$ , the weight of the likelihood function in the posterior density, reduces the energy barrier between distant separated modes which enables to commonly applied MCMC samplers to move between modes. However, this feature can be only exploited when the step size is large enough. For the reason that the HMC algorithm is capable of proposing updates  $\theta'$  to the current draws  $\theta$  which are distant and in theory always accepted, it can exploit this potential when  $\beta_n$  is relatively small. A key question in this context is how large is the probability that the true parameter vector  $\theta$  lies in the region of the posterior density surrounding a particular mode. This probability is measured by the volumes under the posterior density around a particular mode  $\frac{1}{Z} \int_{\theta \in \Theta_i} p(Y|\theta)p(\theta)d\theta$ . A potential issue if using the RWMH algorithm in the rejuvenation step is that particles will tend to get stuck in the same region around the typical set where they started from at stage zero and could potentially bias the estimation. With the number of particles going to infinity this bias will have to disappear even if particles were not rejuvenated at all, when using e.g. annealed importance sampling by Neal (2001), as convergence of these algorithms is warranted. However, with increasing amount of parameters the number of particles necessary will increase exponentially such that a guided approach to rejuvenate the actual parameter draw might be of an advantage. The Sequential Hamiltonian Monte Carlo algorithm has already been applied by Daviet (2018) to logit discrete choice models and reports better convergence properties than the simple SMC method if a leave-one-out approximation of the observed distribution of the particles is used in the correction step. In our work we will use the SMC

framework proposed by Herbst and Schorfheide (2014) with both multinomial and stratified resampling. The next algorithm summarizes the main steps:

---

**Algorithm 5** Sequential Hamiltonian Monte Carlo

---

1. Search for the different modes by starting the HMC algorithm from different parameter settings.
  2. Specify a sequence  $\{\beta_n\}_{n=0}^N$  such that  $1 = \beta_N > \dots > \beta_{n+1} > \beta_n > \dots > \beta_0 \geq 0$
  3. Tune the HMC sampler for each target density  $p(Y|\theta^{(n)})^{\beta_n} p(\theta^{(n)})$  separately, depending on the current position a given particle  $\theta_j^{(n)}$  to be rejuvenated, if necessary.
  4. Run the SMC algorithm by applying the HMC algorithm to execute the rejuvenation step and always use the pretuned sampler at each stage for the target distribution  $p(Y|\theta^{(n)})^{\beta_n} p(\theta^{(n)})$  depending on the current position of the actual draw  $\theta_j^{(n)}$ .
- 

This algorithm fits into the scheme proposed by Chopin (2004), as already pointed out by Daviet (2018). Therefore, under common regularity conditions and assuming that the multinomial resampling is used, almost sure convergence will hold:

$$\frac{1}{J} \sum_{j=1}^J h(\theta_j^{(n)}) \xrightarrow{\text{a.s.}} \mathbb{E}_{\tilde{\pi}_n}(h) \quad (34)$$

$$\frac{\sum_{j=1}^J w_j^{(n)} h(\theta_j^{(n)})}{\sum_{j=1}^J w_j^{(n)}} \xrightarrow{\text{a.s.}} \mathbb{E}_{\pi_t}(h) \quad (35)$$

$$\frac{1}{J} \sum_{j=1}^J h(\hat{\theta}_j^{(n)}) \xrightarrow{\text{a.s.}} \mathbb{E}_{\pi_n}(h) \quad (36)$$

where  $\tilde{\pi}_n(\cdot) := \int \pi_{n-1}(\theta^{(n-1)}) k^{(n)}(\hat{\theta}^{(n-1)}, \cdot) d\theta^{(n-1)}$  with  $k^{(n)}$  being the stochastic kernel density function implied by the HMC algorithm. Furthermore  $\pi_n(\theta^{(n)}) = \frac{1}{Z_n} p(Y|\theta^{(n)})^{\beta_n} p(\theta^{(n)})$ ,  $w_j^{(n)} \propto \nu_j^{(n)} = \pi_n(\theta_j^{(n-1)}) / \tilde{\pi}_n(\theta_j^{(n-1)})$  and  $\hat{\theta}_j^{(n)}$  the particle positions after resampling. As HMC leaves  $\pi_{n-1}$  invariant, it follows that  $\tilde{\pi}_{(n)} = \pi_{n-1}$ , hence  $w_j^{(n)} = p(Y|\theta_j^{(n-1)})^{\beta_n - \beta_{n-1}}$ .

Furthermore, the limit distribution is:

$$J^{1/2} \left\{ \frac{1}{J} \sum_{j=1}^J h(\hat{\theta}_j^{(n)}) - \mathbb{E}_{\pi_n}(h) \right\} \xrightarrow{D} \mathcal{N}(0, \hat{V}_n(h)) \quad \forall n = 1, \dots, N \quad (37)$$

with  $\hat{V}_n(h)$  obtained recursively:

$$\tilde{V}_0(h) = \text{Var}_{\tilde{\pi}_{(0)}}(h) \quad (38)$$

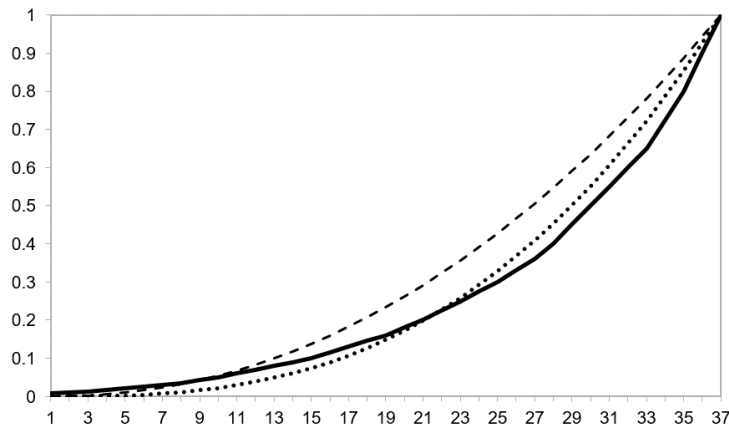
$$\tilde{V}_n(h) = \hat{V}_{n-1}(h) \{\mathbb{E}_{k_n}(h)\} + \mathbb{E}_{\pi_{n-1}}(h) \text{Var}_{k_n}(h) \quad \forall n = 1, \dots, N \quad (39)$$

$$V_n(h) = \tilde{V}_n \{\nu_n \cdot (h - \mathbb{E}_{\pi_n}(h))\} \quad \forall n = 1, \dots, N \quad (40)$$

$$\hat{V}_n(h) = V_n(h) + \text{Var}_{\pi_n}(h) \quad \forall n = 1, \dots, N \quad (41)$$

To apply the algorithm we estimate again the Smets and Wouters (2007) model and loosen the priors in line with Herbst and Schorfheide (2014). We use the same data set as for the estimation of the restricted model. Before executing the estimation code the sampler has to be tuned. In particular, we use  $N = 37$  stages and  $J = 256$  particles in order not to waste computational resources, which amount is rather low if compared SMC frameworks using RWMH for rejuvenation. The tempering schedule  $\{\beta_n\}_{n=1}^N$  was calibrated in a way that  $p(Y|\theta)^{\beta_n}p(\theta)$  serves always sufficiently well as proposal density for  $p(Y|\theta)^{\beta_{n+1}}p(\theta)$ , hence the bridge densities are never too different. Even with such a relatively small amount of stages and particles modes are not absorbed highlighting the power of the SHMC estimator in the sense that the rejuvenation step is guided. The following graph displays the shape of the tempering schedule:

Figure 8: Tempering Schedule



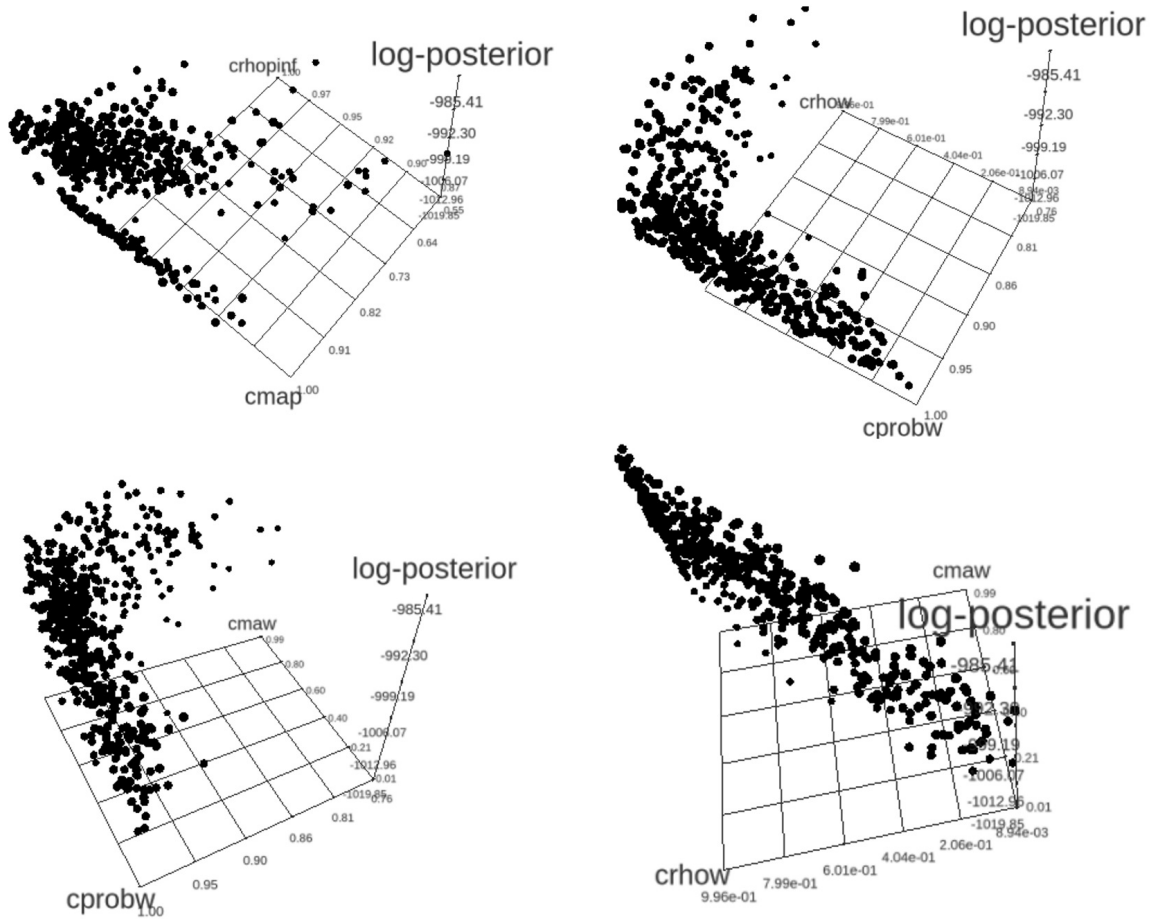
*Notes:* The solid line shows the tempering schedule used for the estimation. The dashed line shows the tempering schedule if  $\beta_n = ((n - 1)/(N - 1))^\lambda$  with  $\lambda = 2.1$  and the dotted line if  $\lambda = 2.75$ .

The solid line shows the tempering schedule, as a dashed line we plotted the original

schedule of Herbst and Schorfheide (2014). At the low end the tempering schedules correspond while after approximately one quarter the schedule used by Herbst and Schorfheide (2014) starts to increase more rapidly. As a comparison we plotted the schedule from Herbst and Schorfheide (2014) with  $\lambda = 2.75$  which provides a better approximation of the schedule used for our estimation framework. Using the HMC sampler there is no need to increase the tempering schedule as rapidly due to the better sampling properties at higher  $\beta_n$  values which allows the particle positions to remain at lower  $\beta_n$ -levels and to mix between the modes for a longer time. However, one should notice that already at relatively low  $\beta_n$  levels mixing is far away from optimal, that is, not jumping often enough between modes. As  $\beta_n$  increases less information can be extracted from the density as regards the ratio of the volumes under the modes by moving the particles in the parameter space, yet one can obtain more information with respect to the exact shape of the modes. Another difference if compared with the tempering schedule used by Herbst and Schorfheide (2014) is that while the latter starts with a draw from the prior distribution our initial sampling stems from a slightly informed distribution with  $\beta_n = 0.005$ , where the HMC sampler is able to mix between the modes. In our initial sample approximately 30 percent of the particles are from the region around the mode which seem be dominated and to encompass less volume. The latter setup is in line with our prior beliefs based on existing literature, see Lanne and Luoto (2018), which augments the SMC algorithm with a non-sequential importance sampling. We also applied the criterion used in Herbst and Schorfheide (2014) to decide whether to resample at a given stage  $n$ , yet we resample when the effective sample size (ESS) drops below 0.7 instead of 0.5. In practice, the algorithm resamples in most of the cases at each second stage. Alternatively, we could have resampled deterministically at each second stage.

We performed the estimation both using multinomial and stratified resampling. Our estimation results suggest that the posterior density for a handful parameters is ill-behaved. In particular, we find in line with Herbst and Schorfheide (2014) and Lanne and Luoto (2018) that the joint kernel density estimates of the parameters  $\rho_p$  and  $\mu_p$ , the ARMA(1,1) terms in the exogenous shock process of the Phillips-curve, is bimodal as illustrated below. In our estimations we obtain that approximately between 10 and 30 percent of the particles are concentrated in the area around the

Figure 9: Joint Posterior Density Estimates



*Notes:* The plot show the joint posterior densities of the following parameters:  $[\rho_p, \mu_p]$  (upper left),  $[\xi_w, \rho_w]$  (upper right),  $[\xi_w, \mu_w]$  (lower left) and  $[\rho_w, \mu_w]$  (lower right). Sample size equals 512, where two sample draws of the size  $J = 256$ , respectively, were merged, the first obtained by applying multinomial resampling, the second one by stratified resampling. Divergence rate at the last stage  $\beta_n = 1$  was approximately 2.3 percent and 1.5 percent, respectively, while the overall divergence rate throughout all  $N = 37$  stages amounted to approximately 5.1 percent for both samples.

dominated mode which is higher than the probability of around 5 percent reported in Herbst and Schorfheide (2014). The parameters determining the wage Phillips curve,  $\xi_w$  and  $\rho_w$ , that is the wage rigidity and the AR(1) term of the mark-up shock process exhibit also a bimodal pattern, yet both modes are rather stretched out in length. The joint kernel density of  $\xi_w$  and the MA(1) coefficient of the wage mark-up shock,  $\mu_w$ , is shaped similarly as the joint density of  $[\xi_w, \rho_w]$ . The reason for this feature is that  $\rho_w$  and  $\mu_w$  are highly correlated. The joint kernel density exhibits a long ridge along the 45° line which suggests that the mark-up shock process is overparametrized as also suggested by Lanne and Luoto (2018) and



restricting the model could result in an improved fit. In general we can conclude that by combining the HMC estimator with SMC we obtain a powerful tool which allows for the estimation of complex and ill-behaved posterior densities and delivers results in line with existing literature.

## 7 Conclusion and Outlook

In this paper we review the benchmark DSGE estimation framework, the RWMH algorithm, and present an advanced alternative, the HMC sampler. Subsequently we implement the HMC algorithm for DSGE models in STAN, a state-of-the-art, high-performance software package which has become a workhorse development environment for Bayesian estimation. We estimate a small scale three equation NK textbook model and the Smets-Wouters model using HMC. Our estimation results largely correspond to those from existing literature which underlines the accuracy of the estimation method and the implemented algorithm. In addition, we present in detail the sampling diagnostics which enables to conclude that the target density of the three equation textbook model exhibits a regular shape. We confirm that in this case the RWMH algorithm operates also adequately. However, for the Smets-Wouters model, after the advanced diagnostics unique to HMC exhibited signs of inefficient sampling, we found a second mode, even though the estimation was carried out using the original priors and data. Consequently, we highlight that advanced sampling diagnostics for HMC helps identify irregularities in the posterior and parameters difficult to sample. Furthermore, we would like to stress that HMC does not require any posterior mode search which is a major advantage if compared with the commonly used RMWH algorithm.

Finally, we combine the HMC algorithm with the SMC sampler to address a main shortcoming of HMC that it fails to explore ill-behaved posterior densities properly. We apply this extended framework to estimate the Smets-Wouters model using less informative priors and obtain bimodal posterior densities which results are in line with those in existing literature. Yet, we acknowledge that further effort is needed to increase the speed of the algorithm.

In light of our results, we believe that superior sampling properties and unique

diagnostics to HMC will open new avenues to revisit existing DSGE model estimation exercises. We are also hopeful that additional, already existing, sophisticated gradient-based estimation methods can be implemented for DSGE estimation as computational power available today should render these algorithms feasible. We leave these topics for future research.

## References

- An, A. and F. Schorfheide**, “Bayesian Analysis of DSGE Models,” *Econometric Reviews*, 2007, *26* (2), 113–172.
- Anderson, G.**, “Solving Linear Rational Expectations Models: A Horse Race,” *Computational Economics*, 2008, *31*, 95–113.
- , “A reliable and computationally efficient algorithm for imposing the saddle point property in dynamic models,” *Journal of Economic Dynamics and Control*, 2010, *34* (3), 472–489.
- Besag, J.**, “Discussion of the Paper by Grenander and Miller,” *Journal of the Royal Statistical Society A*, 1994, *56* (4), 591–592.
- Betancourt, M.**, “A Conceptual Introduction to Hamiltonian Monte Carlo,” *arXiv preprint*, 2018, *1701.02434v2*.
- Binder, M. and H. Pesaran**, “Multivariate linear rational expectations models: characterization of the nature of the solutions and their fully recursive computation,” *Econometric Theory*, 1997, *13* (6), 877–888.
- Blanchard, O. J. and C. M. Kahn**, “The solution of linear difference models under rational expectations,” *Econometrica*, 1980, *48* (5), 1305–1312.
- Brooks, S. P. and A. Gelman**, “General Methods for Monitoring Convergence of Iterative Simulations,” *Journal of Computational and Graphical Statistics*, 1998, *7* (4), 434–455.
- Cai, M., M. Del Negro, E. Herbst, E. Matlin, R. Sarfati, and F. Schorfheide**, “Online Estimation of DSGE Models,” *Econometric Journal*, 2020.
- Calvo, G. A.**, “Staggered Prices in a Utility-Maximizing Framework,” *Journal of Monetary Economics*, 1983, *12* (3), 383–398.
- Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell**, “Stan: A probabilistic programming language,” *Journal of Statistical Software*, 2017, *76* (1).
- Chib, S. and S. Ramamurthy**, “Tailored randomized block MCMC methods with application to DSGE models,” *Journal of Econometrics*, 2010, *155* (1), 19–38.
- Chopin, N.**, “Central Limit Theorem for Sequential Monte Carlo Methods and its Application to Bayesian Inference,” *Journal of the American Statistical Association*, 2004, *93* (443), 1032–1044.
- Christiano, L. J.**, “Solving dynamic equilibrium models by a methods of undetermined coefficients,” *Computational Economics*, 2002, *20* (1-2), 21–55.
- , **M. Eichenbaum, and C. L. Evans**, “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy,” *Journal of Political Economy*, 2005, *113* (1), 1–45.

- Clarida, R., J. Galí, and M. Gertler**, “The Science of Monetary Policy: A New Keynesian Perspective,” *Journal of Economic Literature*, 1999, 37, 1661–1707.
- Cúrdia, V. and R. Reis**, “Correlated Disturbances and U.S. Business Cycles,” *NBER Working Papers 15774*, 2010.
- Creal, D.**, “Sequential Monte Carlo Samplers for Bayesian DSGE Models,” *Manuscript, University Chicago Booth*, 2007.
- Daviet, R.**, “Inference with Hamiltonian Sequential Monte Carlo Simulators,” *arXiv*, 2018, 1812.07978v1.
- Del Moral, P., A. Doucet, and A. Jasra**, “Sequential Monte Carlo samplers,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2006, 68 (3), 411–436.
- Del Negro, M., M. P. Giannoni, and F. Schorfheide**, “Inflation in the Great Recession and New Keynesian Models,” *American Economic Journal: Macroeconomics*, 2015, 7 (1), 168–496.
- Duane, S., A.D. Kennedy, B. J. Pendleton, and D. Roweth**, “Hybrid Monte Carlo,” *Physics Letters B*, 1987, 195 (2), 216–222.
- Durmus, A., É. Moulines, and E. Saksman**, “On the convergence of Hamiltonian Monte Carlo,” *arXiv preprint*, 2019, *arXiv:1705.00166v2*.
- Fernandez-Villaverde, J. and J. F. Rubio-Ramirez**, “Estimating Macroeconomic Models: A Likelihood Approach,” *Review of Economic Studies*, 2007, 74, 1059–1087.
- and **P. Guerrón-Quintana**, “Estimating DSGE Models: Recent Advances and Future Challenges,” *NBER Working Paper No. 27715*, 2020.
- , **J. Rubio-Ramirez, and F. Schorfheide**, “Solution and Estimation Methods for DSGE Models,” *In: H. Uhlig and J. Taylor (eds.): Handbook of Macroeconomics, Elsevier, New York*, 2016, 2, 527–724.
- Gabry, J. and D. Veen**, “ShinyStan Version 3.0.0,” *mc-stan.org*, 2020.
- , **D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman**, “Visualization in Bayesian workflow (with discussion),” *Journal of the Royal Statistical Society A*, 2019, 182 (2), 389–402.
- Gilks, W. R. and C. Berzuini**, “Following a moving target - Monte Carlo inference for dynamic Bayesian Models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2002, 63 (7), 127–146.
- Goodrich, B. and C. Montes-Galdon**, “Estimating DSGE Models with Stan,” <https://pdfs.semanticscholar.org/0a9b/35f9f11ce0c489a911622d700d2a4e64f385.pdf>, retrieved on 25 August, 2020.
- Hamilton, J. D.**, *Time Series Analysis*, Princeton, New Jersey: Princeton University Press, 1994.

- Herbst, E. P. and F. Schorfheide**, “Sequential Monte Carlo Sampling for DSGE Models,” *Journal of Applied Econometrics*, 2014, *27* (7), 1073–1098.
- and –, *Bayesian Estimation of DSGE Models*, Princeton, New Jersey: Princeton University Press, 2015.
- Hoffman, M. D. and A. Gelman**, “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo,” *The Journal of Machine Learning Research*, 2014, *15* (1), 1593–1623.
- Kim, J.**, “Constructing and estimating a realistic optimizing model of monetary policy,” *Journal of Monetary Economics*, 2000, *45* (2), 329–359.
- King, R. G. and M. W. Watson**, “The solution of singular linear difference systems under rational expectations,” *International Economic Review*, 1998, *39* (4), 1015–1026.
- Klein, P.**, “Using the generalized Schur form to solve a multivariate linear rational expectations model,” *Journal of Economic Dynamics and Control*, 2000, *24* (10), 1405–1423.
- Kydland, F. E. and E. C. Prescott**, “Time to Build and Aggregate Fluctuations,” *Econometrica*, 1982, *50* (6), 1345–1370.
- Lambert, B.**, *A Student’s Guide to Bayesian Statistics*, SAGE Publications Ltd, 2018.
- Lanne, M. and J. Luoto**, “Data-Driven Identification Constraints for DSGE Models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2018, *80* (2), 236–258.
- Liu, J. S. and R. Chen**, “Sequential Monte Carlo Methods for Dynamic Systems,” *Journal of the American Statistical Association*, 1998, *93* (443), 1032–1044.
- Livingstone, S., M. Betancourt, S. Byrne, and M. Girolami**, “On the geometric ergodicity of Hamiltonian Monte Carlo,” *arXiv preprint*, 2018, *arXiv:1601.08057*.
- Mackenzie, P. B.**, “An improved hybrid Monte Carlo method,” *Physics Letters B*, 1989, *226*, 369–371.
- Neal, R. M.**, “Annealed importance sampling,” *Statistics and Computing*, 2001, *11*, 125–139.
- , “Slice Sampling,” *Annals of Statistics*, 2003, *31* (3), 705–767.
- , “MCMC Using Hamiltonian Dynamics,” in G. L. Jones S. Brooks, A. Gelman and eds. X.-L. Meng, eds., *Handbook of Markov Chain Monte Carlo*, New York: CRC Press, 2011, pp. 30–61.
- Otrok, C.**, “On measuring the welfare cost of business cycles,” *Journal of Monetary Economics*, 2001, *47* (1), 61–92.

- Qi, Y. and T. Minka**, “Hessian-based Markov Chain Monte-Carlo Algorithms,” in “Proceedings of the First Cape Cod Workshop on Monte Carlo Methods” September 2002.
- Roberts, G. O. and O. Stramer**, “Langevin Diffusions and Metropolis-Hastings Algorithms,” *Methodology and Computing in Applied Probability*, 2002, 4, 337–357.
- **and R. L. Tweedie**, “Exponential convergence of Langevin distributions and their discrete approximations,” *Bernoulli*, 1996, 2, 341–363.
- Roberts, G.O., A. Gelman, and W.R. Gilks**, “Weak convergence and optimal scaling of random walk Metropolis algorithms,” *Annals of Applied Probability*, 1997, 7, 110–120.
- Rotemberg, J. J.**, “Aggregate Consequences of Fixed Costs of Price Adjustment,” *American Economic Review*, 1983, 73 (3), 433–436.
- Schorfheide, F.**, “Loss function-based evaluation of DSGE models,” *Journal of Applied Econometrics*, 2000, 15 (6), 645–670.
- Shiwei, L., J. Streets, and B. Shahbaba**, “Wormhole Hamiltonian Monte Carlo,” *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2014, pp. 1953–1959.
- Sims, C. A.**, “Solving linear rational expectations models,” *Computational Economics*, 2002, 20 (1-2), 1–20.
- Smets, F. and R. Wouters**, “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach,” *American Economic Review*, 2007, 97 (3), 586–606.
- Uhlig, H.**, “A toolkit for analyzing non-linear dynamic stochastic models easily,” in R. Marimón and eds. A. Scott, eds., *Computational Methods for the Study of Dynamic Economies.*, Oxford, UK: Oxford University Press, 1999, pp. 30–61.
- Zhoua, B., J. Lamb, and G.-R. Duana**, “On Smith-type iterative algorithms for the Stein matrix equation,” *Applied Mathematics Letters*, 2009, 22, 1038–1044.

## Appendix

# Smets-Wouters Model: Hamiltonian Monte Carlo Estima- tion Diagnostics

## Warnings

[1] "None of the 1000 iterations ended with a divergent transition."

## Numerical diagnostics

	n_eff	Rhat	mean	se_mean	sd
log-posterior	418.92	1.00	-1191.70	0.22	4.59
crpi	1163.52	1.00	2.04	0.00	0.17
crdy	1765.21	1.00	0.21	0.00	0.03
cry	941.05	1.00	0.10	0.00	0.02
crr	995.36	1.01	0.82	0.00	0.02
constelab	1377.36	1.00	0.88	0.03	0.95
constepinf	1071.22	1.00	0.67	0.00	0.10
ctrend	497.64	1.00	0.47	0.00	0.02
constebeta	1084.42	1.00	0.13	0.00	0.05
cgy	1285.63	1.00	0.57	0.00	0.08
cmaw	1015.16	1.00	0.89	0.00	0.04
cmap	532.57	1.01	0.80	0.00	0.08
calfa	1438.32	1.00	0.21	0.00	0.02
czcap	950.76	1.00	0.46	0.00	0.10
csadjcost	892.60	1.00	5.93	0.03	1.01
csigma	730.47	1.00	1.42	0.01	0.14
chabb	1010.00	1.00	0.73	0.00	0.04
cfc	846.70	1.00	1.65	0.00	0.08
cindw	1275.43	1.00	0.57	0.00	0.13

cprobw	632.40	1.00	0.75	0.00	0.05
cindp	867.03	1.00	0.23	0.00	0.09
cprobp	433.91	1.00	0.65	0.00	0.05
csigl	1523.58	1.00	2.10	0.01	0.55
crhoa	754.90	1.00	0.98	0.00	0.01
crhob	498.21	1.00	0.27	0.00	0.11
crhog	1115.33	1.00	0.97	0.00	0.01
crhoqs	775.30	1.00	0.69	0.00	0.06
crhoms	1309.81	1.00	0.17	0.00	0.06
crhopinf	370.24	1.00	0.96	0.00	0.02
crhow	415.38	1.00	0.97	0.00	0.01
sigmaea	952.76	1.00	0.48	0.00	0.03
sigmaeb	665.30	1.00	0.24	0.00	0.03
sigmaeg	1332.55	1.00	0.52	0.00	0.03
sigmaeqs	988.35	1.00	0.46	0.00	0.05
sigmaem	1140.80	1.00	0.23	0.00	0.01
sigmaepinf	865.82	1.00	0.13	0.00	0.02
sigmaew	1255.44	1.00	0.25	0.00	0.02
curvp	NaN	NaN	10.00	NaN	0.00
curvw	NaN	NaN	10.00	NaN	0.00
clandaw	NaN	NaN	1.50	NaN	0.00
LL	727.83	1.00	-961.48	0.20	5.50
cpie	1071.00	1.00	1.01	0.00	0.00
cgamma	496.67	1.00	1.00	0.00	0.00
cbeta	1083.87	1.00	1.00	0.00	0.00
clandap	846.70	1.00	1.65	0.00	0.08
cbetabar	714.02	1.00	0.99	0.00	0.00
cr	956.49	1.00	1.01	0.00	0.00
crk	714.22	1.00	0.03	0.00	0.00
cw	858.44	1.00	0.69	0.00	0.04
cikbar	497.39	1.00	0.03	0.00	0.00
cik	497.64	1.00	0.03	0.00	0.00

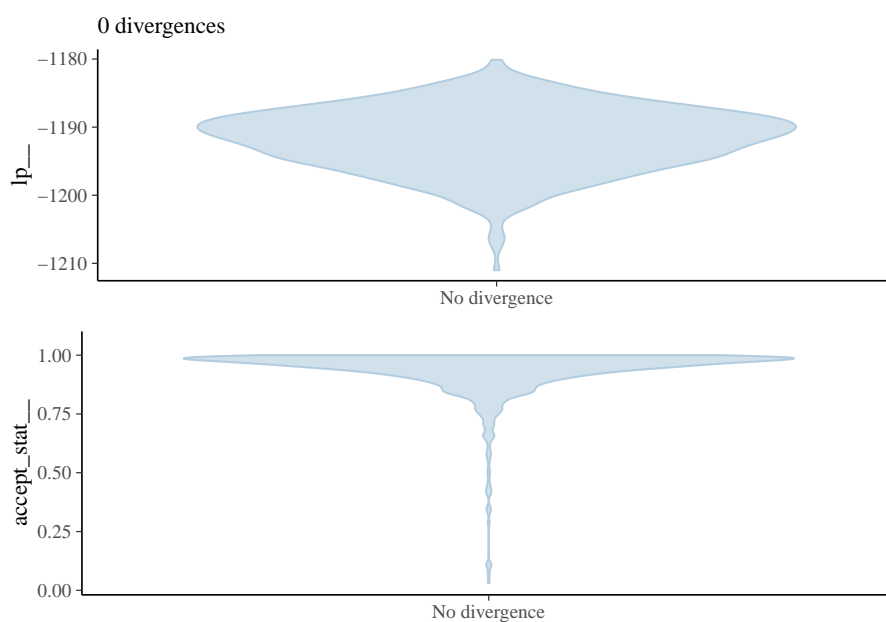


clk	1264.45	1.00	0.19	0.00	0.03
cky	1463.32	1.00	6.26	0.01	0.54
ciy	1404.19	1.00	0.19	0.00	0.02
ccy	1404.19	1.00	0.63	0.00	0.02
crkky	1438.32	1.00	0.21	0.00	0.02
cwhlc	879.78	1.00	0.83	0.00	0.01
cwly	1438.32	1.00	0.79	0.00	0.02
conster	956.99	1.00	1.47	0.00	0.13

## Visual diagnostics

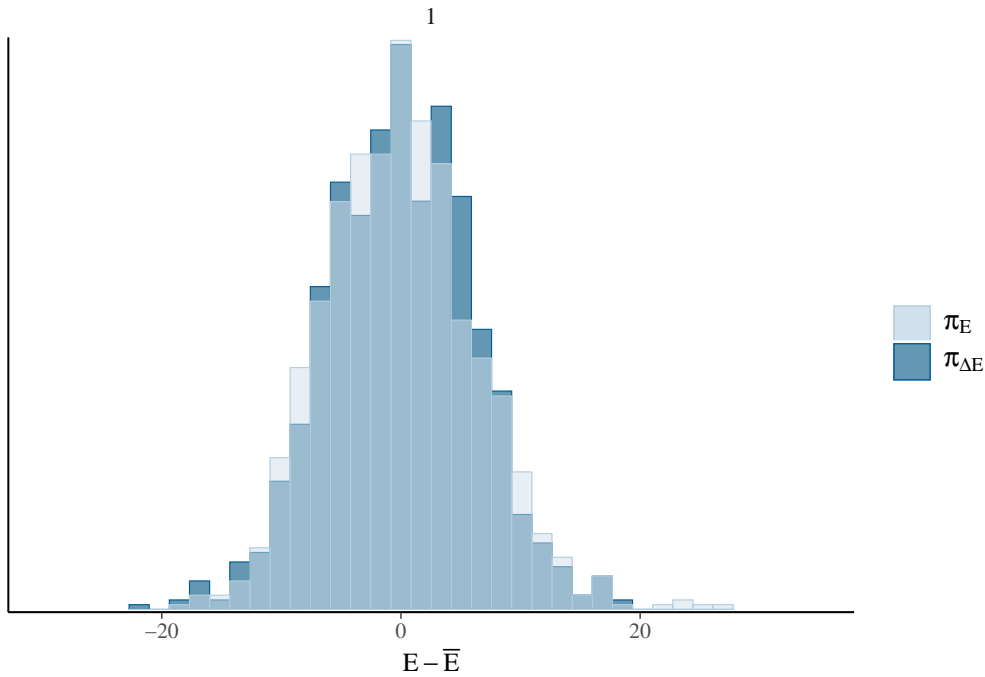
### Divergence Information

These are plots of the *divergent transition status* (x-axis) against the *log-posterior* (y-axis top panel) and against the *acceptance statistic* (y-axis bottom panel) of the sampling algorithm for all chains. Divergent transitions can indicate problems for the validity of the results. A good plot would show no divergent transitions. If the divergent transitions show the same pattern as the non-divergent transitions, this could indicate that the divergent transitions are false positives. A bad plot would show systematic differences between the divergent transitions and non-divergent transitions. For more information see <https://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup>.



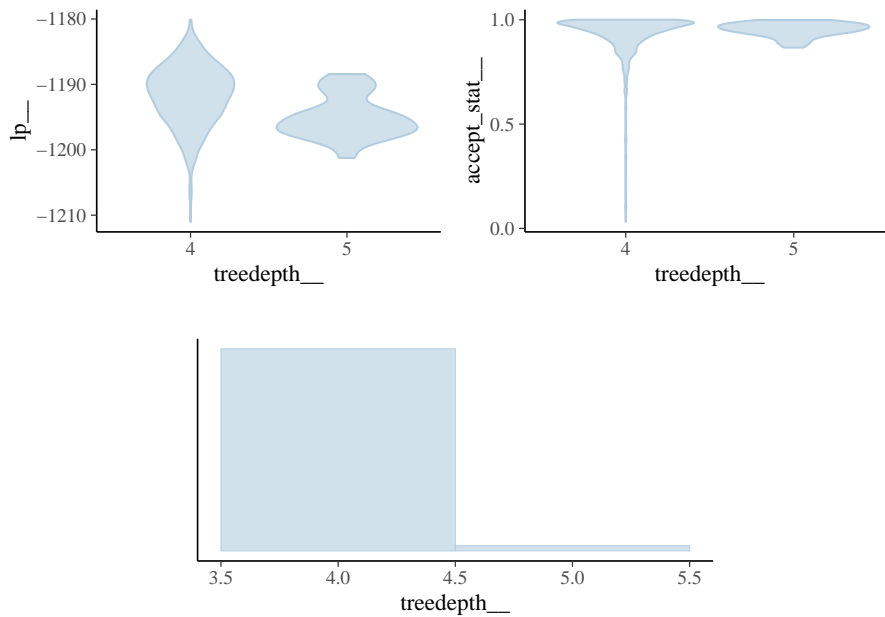
## Energy

These are plots of the overlaid histograms of the marginal energy distribution ( $\pi_E$ ) and the energy transition distribution ( $\pi_{\Delta E}$ ) for all chains. A good plot shows histograms that look well-matched indicating that the Hamiltonian Monte Carlo should perform robustly. The closer  $\pi_{\Delta E}$  is to  $\pi_E$  the faster the random walk explores the energies and the smaller the autocorrelations will be in the chain. If  $\pi_{\Delta E}$  is narrower than  $\pi_E$  the random walk is less effective and autocorrelations will be larger. Additionally the chain may not be able to completely explore the tails of the target distribution. See Betancourt ‘[A conceptual introduction to Hamiltonian Monte Carlo](#)’ and Betancourt ‘[Diagnosing suboptimal cotangent disintegrations in Hamiltonian Monte Carlo](#)’ for the general theory behind the energy plots.



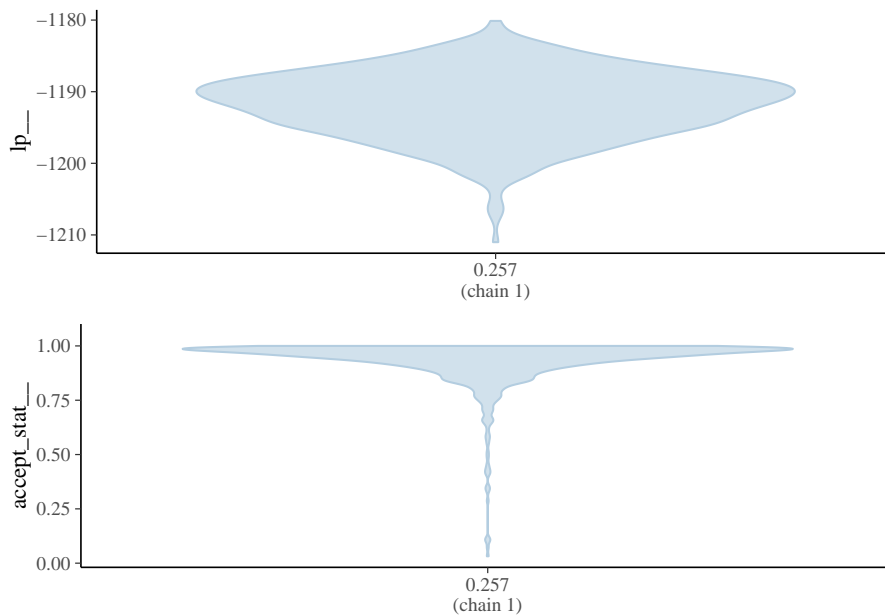
## Tredepth Information

These are plots of the *treedepth* (x-axis) against the *log-posterior* (y-axis top left panel) and against the *acceptance statistic* (y-axis top right panel) of the sampling algorithm for all chains. In these plots information is given concerning the efficiency of the sampling algorithm. Zero treedepth can indicate extreme curvature and poorly-chosen step size. Treedepth equal to the maximum treedepth might be a sign of poor adaptation or of a difficult posterior from which to sample. The former can be resolved by increasing the warmup time, the latter might be mitigated by reparametrization. For more information see <https://mc-stan.org/misc/warnings.html#maximum-treedepth-exceeded> or <https://mc-stan.org/docs/reference-manual/hmc-algorithm-parameters.html>.



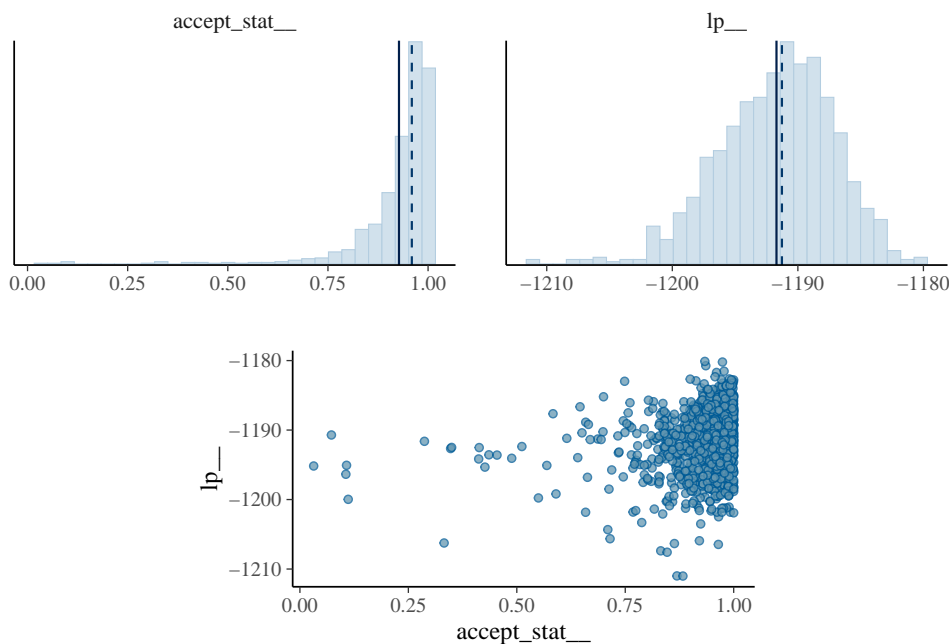
## Step Size Information

These are plots of the *integrator step size per chain* (x-axis) against the *log-posterior* (y-axis top panel) and against the *acceptance statistic* (y-axis bottom panel) of the sampling algorithm. If the step size is too large, the integrator will be inaccurate and too many proposals will be rejected. If the step size is too small, the many small steps lead to long simulation times per interval. Thus the goal is to balance the acceptance rate between these extremes. Good plots will show full exploration of the log-posterior and moderate to high acceptance rates for all chains and step sizes. Bad plots might show incomplete exploration of the log-posterior and lower acceptance rates for larger step sizes.



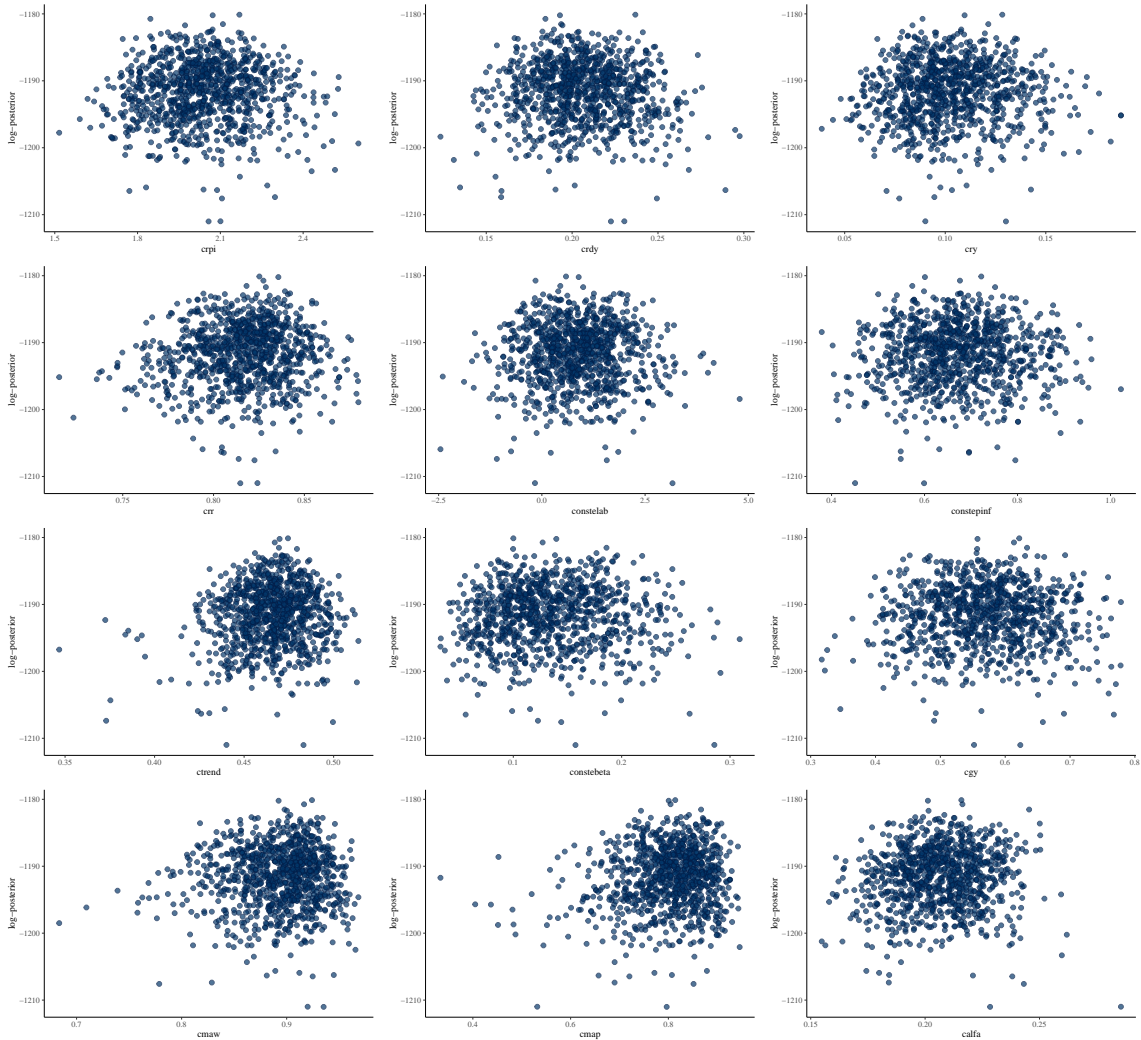
## Acceptance Information

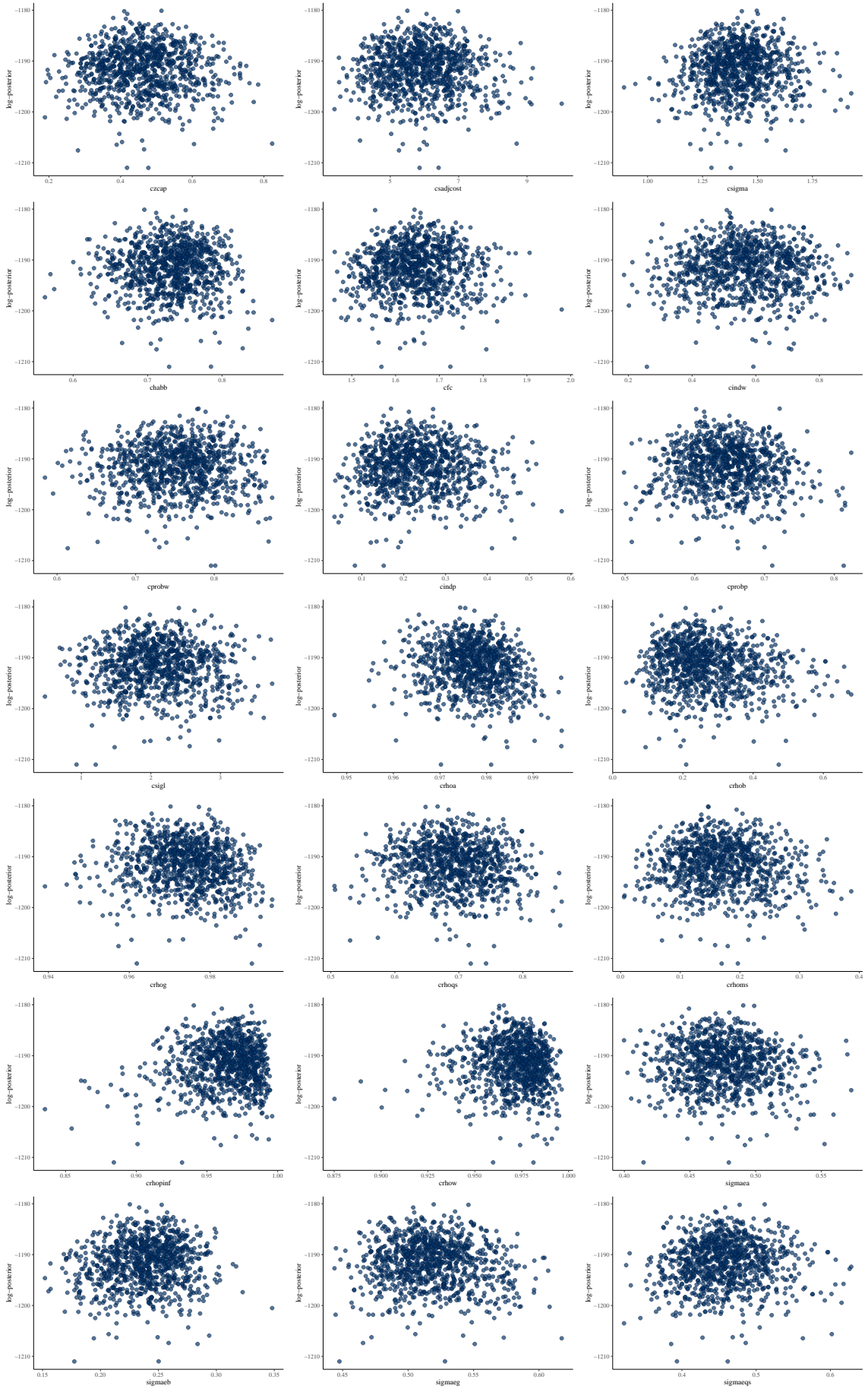
These are plots of the *acceptance statistic* (top left panel), the *log-posterior* (top right panel), and, the *acceptance statistic* (x-axis bottom panel) against the *log-posterior* (y-axis bottom panel) for all chains. The vertical lines indicate the mean (solid line) and median (dashed line). A bad plot would show a relationship between the acceptance statistic and the log-posterior. This might be indicative of poor exploration of parts of the posterior which might be mitigated by reparametrization or adaptation of the step size. If many proposals are rejected the integrator step size might be too large and the posterior might not be fully explored. If the acceptance rate is very high this might be indicative of inefficient sampling. The target Metropolis acceptance rate can be set with the `adapt_delta` control option. For more information see <https://mc-stan.org/docs/reference-manual/hmc-algorithm-parameters.html>.



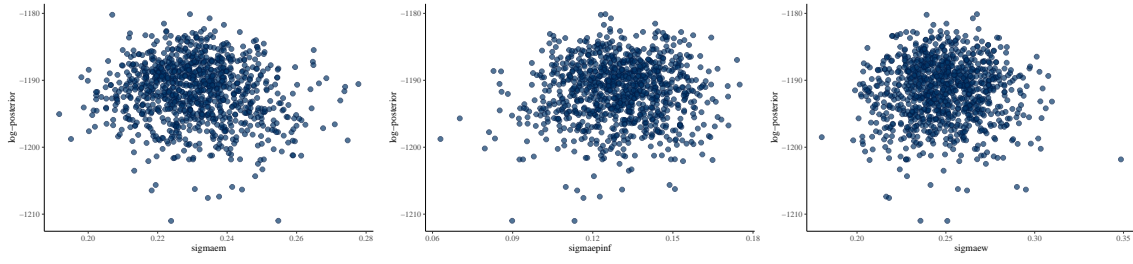
## Scatter plots

These are scatter plots of crpi, crdy, cry, crr, constelab, constepinf, ctrend, conste-beta, cgy, cmaw, cmap, calfa, czcap, csadjcost, csigma, chabb, cfc, cindw, cprobw, cindp, cprobp, csigl, crhoa, crhob, crhog, crhoqs, rhoms, crhopinf, crhow, sigma-maea, sigmaeb, sigmaeg, sigmaeqs, sigmaem, sigmaepinf, sigmaew, plotted against log-posterior. The red dots, if present, indicate divergent transitions. Divergent transitions can indicate problems for the validity of the results. A good plot would show no divergent transitions. A bad plot would show divergent transitions in a systematic pattern. For more information see <https://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup>.



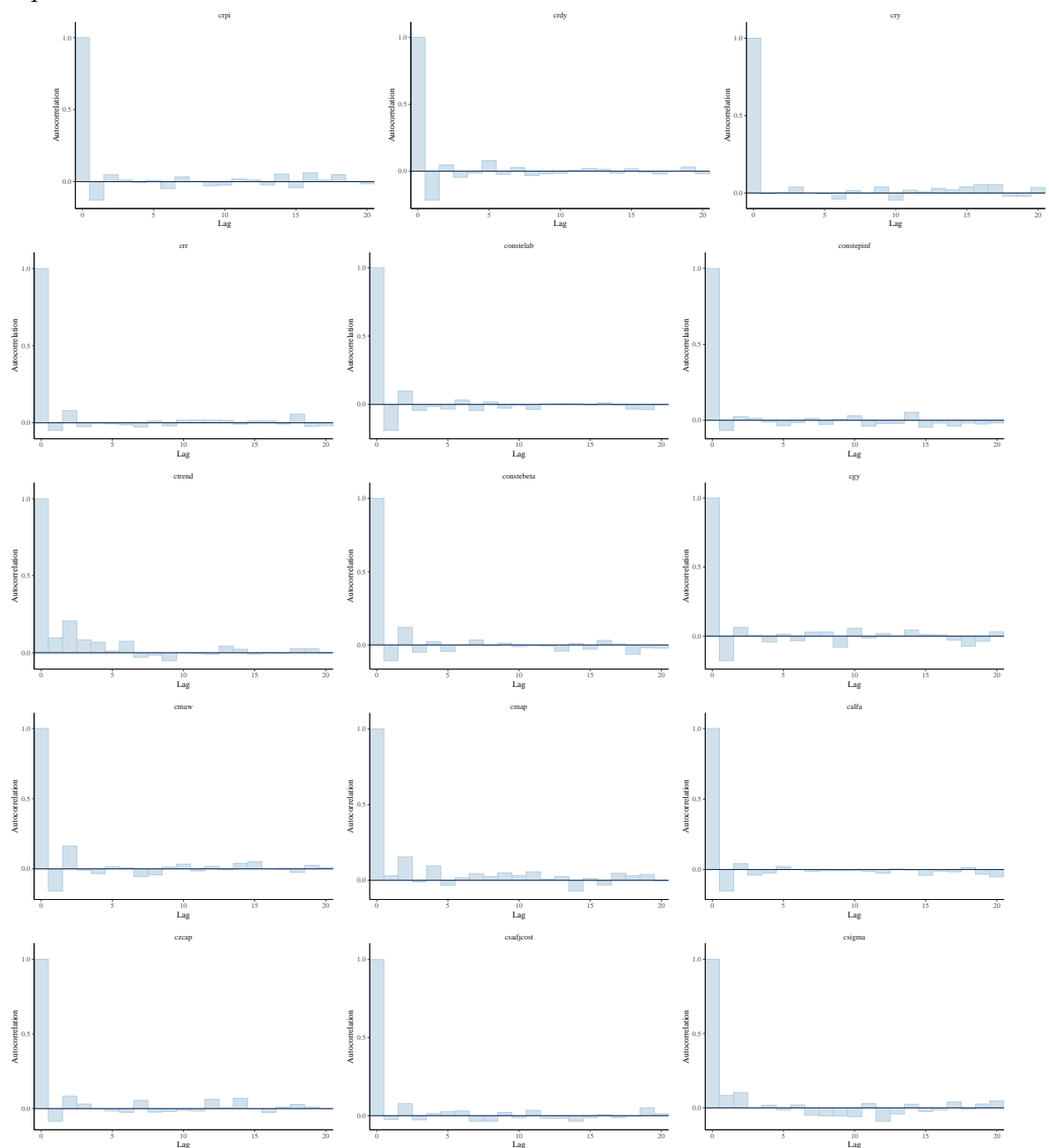


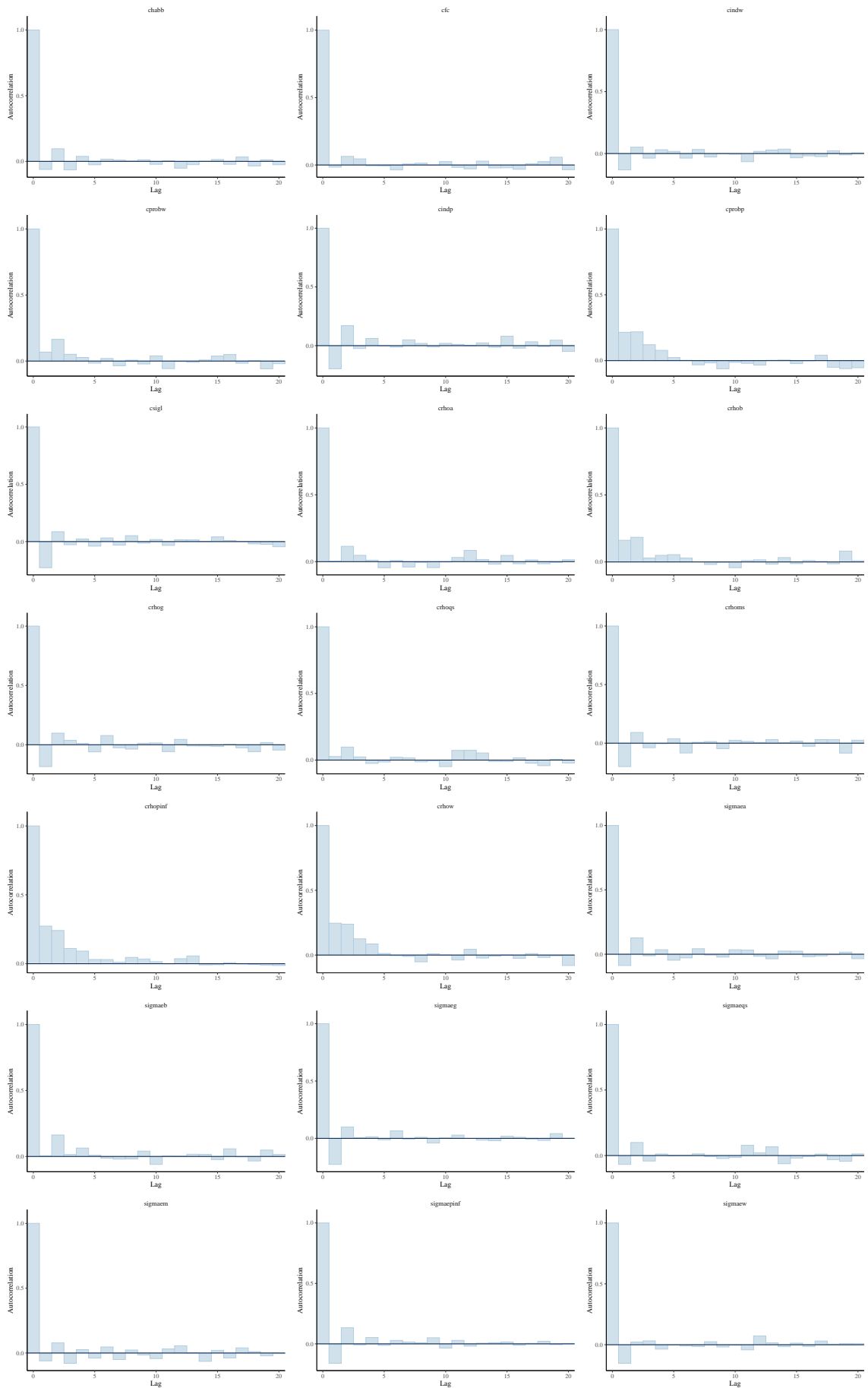




## Autocorrelation

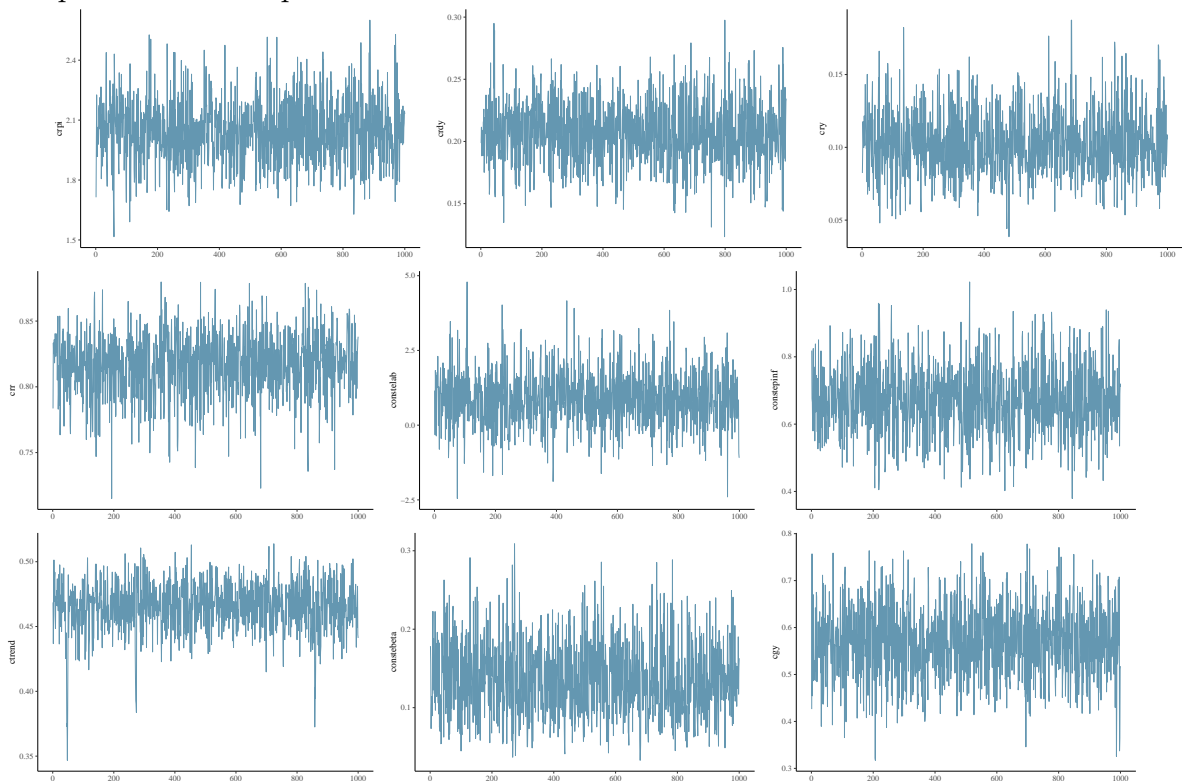
These are autocorrelation plots of `crpi`, `crdy`, `cry`, `crr`, `constelab`, `constepinf`, `ctrend`, `constebeta`, `cgyl`, `cmaw`, `cmap`, `calfa`, `czcap`, `csadjcost`, `csigma`, `chabb`, `cfcc`, `cindw`, `cprobw`, `cindp`, `cprobp`, `csigl`, `crhoa`, `crhob`, `crhog`, `crhoqs`, `crhoms`, `crhopinf`, `crhow`, `sigmaea`, `sigmaeb`, `sigmaeg`, `sigmaeqs`, `sigmaem`, `sigmaepinf`, `sigmaew`. The autocorrelation expresses the dependence between the samples of a Monte Carlo simulation. With higher dependence between the draws, more samples are needed to obtain the same effective sample size. High autocorrelation can sometimes be remedied by reparametrization of the model.

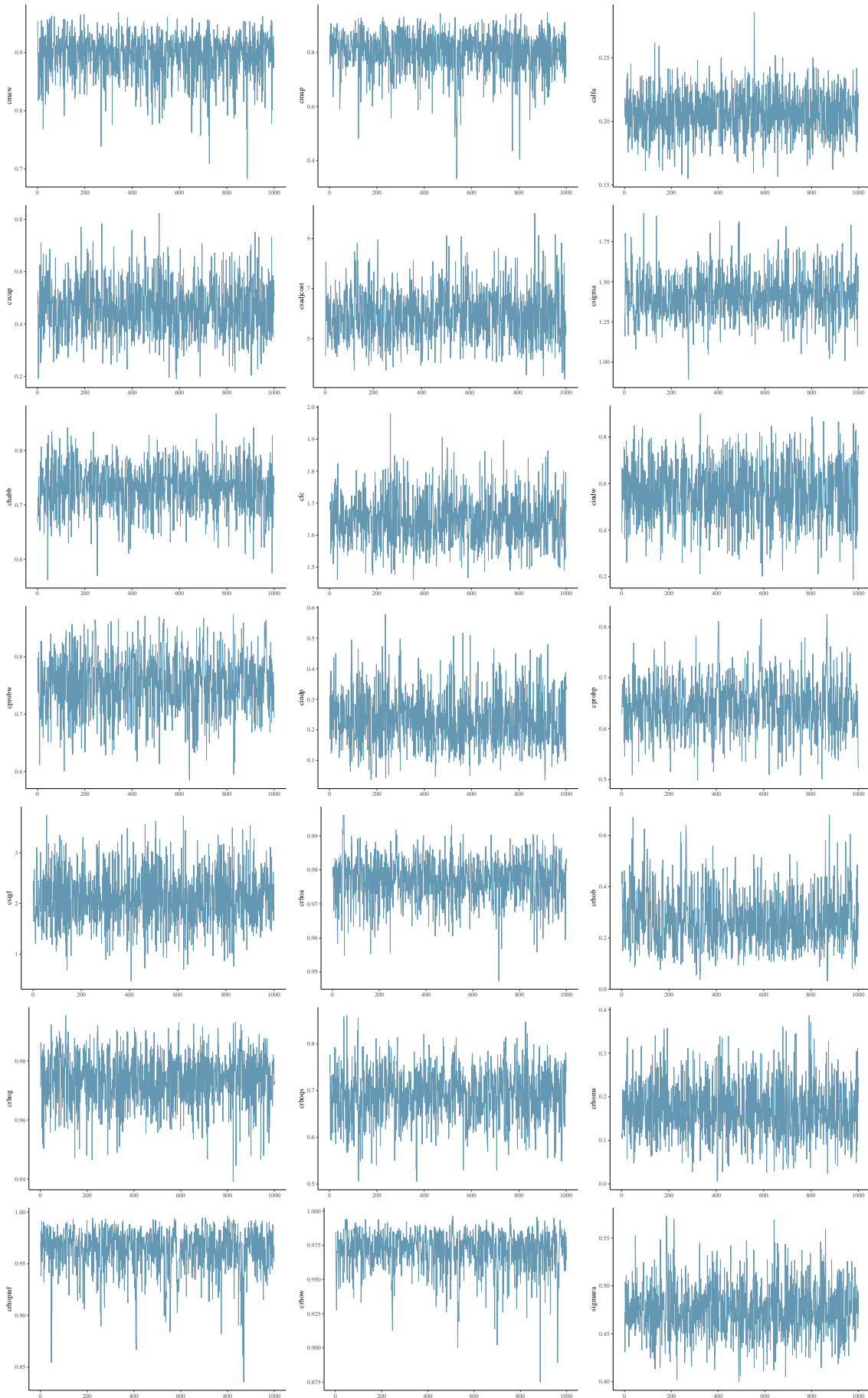


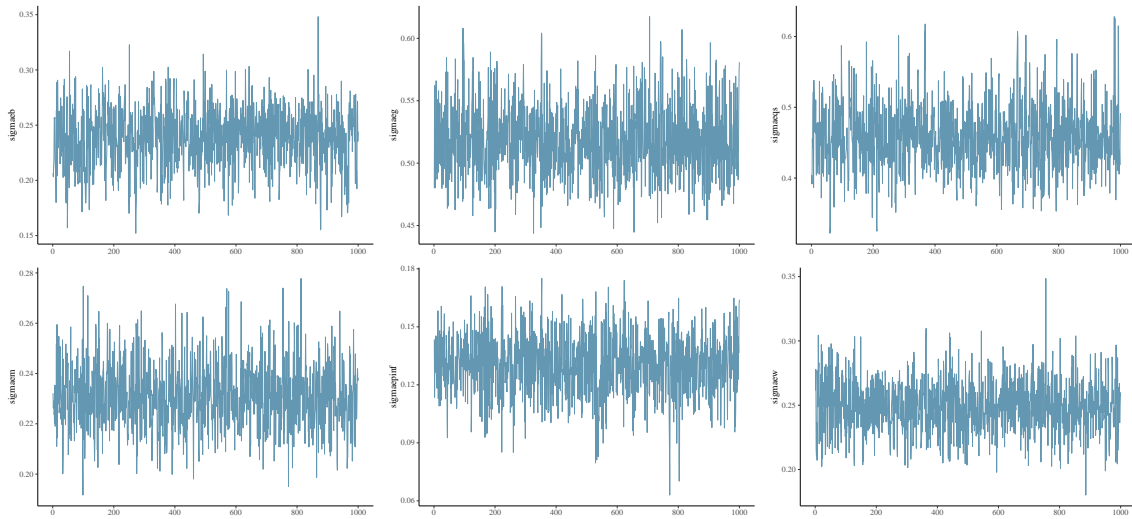


## Trace Plots

These are trace plots of `crpi`, `crdy`, `cry`, `crr`, `constelab`, `constepinf`, `ctrend`, `conste-beta`, `cgy`, `cmaw`, `cmap`, `calfa`, `czcap`, `csadjcost`, `csigma`, `chabb`, `cfc`, `cindw`, `cprobw`, `cindp`, `cprobp`, `csigl`, `crhoa`, `crhob`, `crhog`, `crhoqs`, `crhoms`, `crhopinf`, `crhow`, `sigmaea`, `sigmaeb`, `sigmaeg`, `sigmaeqs`, `sigmaem`, `sigmaepinf`, `sigmaew`. Trace plots provide a visual way to inspect sampling behavior and assess mixing across chains. The iteration number (x-axis) is plotted against the parameter value at that iteration (y-axis). Divergent transitions are marked on the x-axis. A good plot shows chains that move swiftly through the parameter space and all chains that explore the same parameter space without any divergent transitions. A bad plot shows chains exploring different parts of the parameter space, this is a sign of non-convergence. If there are divergent transitions, looking at the parameter value related to these iterations might provide information about the part of the parameter space that is difficult to sample from. Slowly moving chains are indicative of high autocorrelation or small integrator step size, both of which relate to ineffective sampling and lower effective sample sizes for the parameter.







# Smets-Wouters Model: Hamiltonian Monte Carlo Estima- tion Diagnostics for Alternative Mode Estimation

## Warnings

[1] "None of the 1000 iterations ended with a divergent transition."

## Numerical diagnostics

	n_eff	Rhat	mean	se_mean	sd
log-posterior	365.59	1.00	-1199.89	0.23	4.31
crpi	892.49	1.00	1.97	0.01	0.17
crdy	942.12	1.00	0.22	0.00	0.03
cry	1194.44	1.00	0.13	0.00	0.03
crr	868.04	1.00	0.85	0.00	0.02
constelab	846.38	1.00	0.61	0.03	0.91
constepinf	922.57	1.00	0.67	0.00	0.10
ctrend	710.82	1.00	0.46	0.00	0.02
constebeta	962.80	1.00	0.14	0.00	0.05
cgy	1213.01	1.00	0.57	0.00	0.08
cmaw	642.98	1.00	0.91	0.00	0.03
cmap	805.75	1.01	0.98	0.00	0.01
calfa	1054.77	1.00	0.21	0.00	0.02
czcap	898.15	1.00	0.40	0.00	0.10
csadjcost	704.09	1.00	5.43	0.04	1.04
csigma	562.87	1.00	1.41	0.01	0.14
chabb	482.26	1.00	0.69	0.00	0.06
cfc	889.32	1.00	1.63	0.00	0.08
cindw	1312.51	1.00	0.52	0.00	0.12
cprobw	470.28	1.00	0.80	0.00	0.04
cindp	1131.00	1.00	0.31	0.00	0.09

cprobp	788.72	1.01	0.80	0.00	0.03
csigl	1284.08	1.00	2.13	0.02	0.60
crhoa	728.30	1.00	0.97	0.00	0.01
crhob	385.27	1.00	0.41	0.01	0.15
crhog	609.54	1.00	0.97	0.00	0.01
crhoqs	651.13	1.00	0.71	0.00	0.06
crhoms	902.99	1.00	0.13	0.00	0.06
crhopinf	1005.06	1.00	0.93	0.00	0.02
crhow	359.52	1.00	0.97	0.00	0.02
sigmaea	1259.12	1.00	0.48	0.00	0.03
sigmaeb	456.17	1.00	0.21	0.00	0.04
sigmaeg	1121.21	1.00	0.52	0.00	0.03
sigmaeqs	806.33	1.00	0.45	0.00	0.05
sigmaem	1155.58	1.00	0.23	0.00	0.01
sigmaepinf	1240.97	1.00	0.21	0.00	0.02
sigmaew	930.52	1.00	0.23	0.00	0.02
LL	845.43	1.00	-960.54	0.19	5.50
cpie	921.25	1.00	1.01	0.00	0.00
cgamma	712.31	1.00	1.00	0.00	0.00
cbeta	963.16	1.00	1.00	0.00	0.00
clandap	889.32	1.00	1.63	0.00	0.08
cbetabar	845.19	1.00	0.99	0.00	0.00
cr	731.82	1.00	1.01	0.00	0.00
crk	845.41	1.00	0.03	0.00	0.00
cikbar	710.70	1.00	0.03	0.00	0.00
cik	710.81	1.00	0.03	0.00	0.00
clk	1159.86	1.00	0.18	0.00	0.02
cky	1088.85	1.00	6.31	0.02	0.55
ciy	1041.18	1.00	0.19	0.00	0.02
ccy	1041.18	1.00	0.63	0.00	0.02
crkky	1054.77	1.00	0.21	0.00	0.02
cwhlc	860.41	1.00	0.83	0.00	0.01



cwly	1054.77	1.00	0.79	0.00	0.02
curvp	NaN	NaN	10.00	NaN	0.00
curvw	NaN	NaN	10.00	NaN	0.00
clandaw	NaN	NaN	1.50	NaN	0.00
conster	732.03	1.00	1.47	0.00	0.13

## IMFS WORKING PAPER SERIES

### *Recent Issues*

<b>143 / 2020</b>	Gregor Boehl Felix Strobel	U.S. Business Cycle Dynamics at the Zero Lower Bound
<b>142 / 2020</b>	Gregor Boehl Gavin Goy Felix Strobel	A Structural Investigation of Quantitative Easing
<b>141 / 2020</b>	Karl-Heinz Tödter	Ein SIRD-Modell zur Infektionsdynamik mit endogener Behandlungskapazität und Lehren für Corona-Statistiken
<b>140 / 2020</b>	Helmut Siekmann Volker Wieland	The Ruling of the Federal Constitutional Court concerning the Public Sector Purchase Program: A Practical Way Forward
<b>139 / 2020</b>	Volker Wieland	Verfahren zum Anleihekaufprogramm der EZB
<b>138 / 2020</b>	Francisco Gomes Michael Haliassos Tarun Ramadorai	Household Finance
<b>137 / 2019</b>	Martin Kliem Alexander Meyer-Gohde	(Un)expected Monetary Policy Shocks and Term Premia
<b>136 / 2019</b>	Luc Arrondel Hector Calvo-Pardo Chryssi Giannitsarou Michael Haliassos	Informative Social Interactions
<b>135 / 2019</b>	Tiziana Assenza Alberto Cardaci Domenico Delli Gatti	Perceived wealth, cognitive sophistication and behavioral inattention
<b>134 / 2019</b>	Helmut Siekmann	The Asset Purchase Programmes of the ESCB – an interdisciplinary evaluation
<b>133 / 2019</b>	Josefine Quast Maik Wolters	Reliable Real-time Output Gap Estimates Based on a Modified Hamilton Filter
<b>132 / 2019</b>	Galina Potjagailo Maik Wolters	Global Financial Cycles since 1880
<b>131 / 2019</b>	Philipp Lieberknecht Volker Wieland	On the Macroeconomic and Fiscal Effects of the Tax Cuts and Jobs Act
<b>130 / 2019</b>	Eduard Hofert	Regulating Virtual Currencies

<b>129 / 2018</b>	Olga Goldfayn-Frank Johannes Wohlfart	How Do Consumers Adapt to a New Environment in their Economic Forecasting? Evidence from the German Reunification
<b>128 / 2018</b>	Christopher Roth Johannes Wohlfart	How Do Expectations About the Macroeconomy Affect Personal Expectations and Behavior?
<b>127 / 2018</b>	Michael Haliassos Thomas Jansson Yigitcan Karabulut	Financial Literacy Externalities
<b>126 / 2018</b>	Felix Strobel	The Government Spending Multiplier, Fiscal Stress and the Zero Lower Bound
<b>125 / 2018</b>	Alexander Meyer-Gohde Daniel Neuhoff	Generalized Exogenous Processes in DSGE: A Bayesian Approach
<b>124 / 2018</b>	Athanasios Orphanides	The Boundaries of Central Bank Independence: Lessons from Unconventional Times
<b>123 / 2018</b>	Karl-Heinz Tödter Gerhard Ziebarth	Zinsen, Effektivpreise und Lebenskosten – Ein Beitrag zur Konstruktion eines intertemporalen Preisindex
<b>122 / 2018</b>	Helmut Siekmann	Legal Tender in the Euro Area
<b>121 / 2018</b>	Maik H. Wolters	How the Baby Boomers' Retirement Wave Distorts Model-Based Output Gap Estimates
<b>120 / 2017</b>	Helmut Siekmann	Die Einstandspflicht der Bundesrepublik Deutschland für die Deutsche Bundesbank und die Europäische Zentralbank
<b>119 / 2017</b>	Gregor Boehl	Monetary Policy and Speculative Stock Markets
<b>118 / 2017</b>	Gregor Boehl Thomas Fischer	Can Taxation Predict US Top-Wealth Share Dynamics?
<b>117 / 2017</b>	Tobias H. Tröger	Why MREL Won't Help Much
<b>116 / 2017</b>	Tobias H. Tröger	Too Complex to Work – A Critical Assessment of the Bail-in Tool under the European Bank Recovery and Resolution Regime
<b>115 / 2017</b>	Guenter W. Beck Volker Wieland	How to Normalize Monetary Policy in the Euro Area
<b>114 / 2017</b>	Michael Binder Jorge Quintana Philipp Lieberknecht Volker Wieland	Model Uncertainty in Macroeconomics: On the Implications of Financial Frictions