BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel—2024-13

Last updated: May 10, 2024

Parallel Tempering for DSGE Estimation

by Joshua Brault

Corporate Services Department
Bank of Canada
JBrault@bankofcanada.ca

DOI: https://doi.org/10.34989/swp-2024-13 | **ISSN** 1701-9397

Acknowledgements

Thanks to Dave Campbell, Maryam Haghighi, Hashmat Khan, Alex Murray, and Louis Phaneuf for comments and suggestions on previous versions of this paper. Any remaining errors are my own.

Abstract

In this paper, I develop a population-based Markov chain Monte Carlo (MCMC) algorithm known as parallel tempering to estimate dynamic stochastic general equilibrium (DSGE) models. Parallel tempering approximates the posterior distribution of interest using a family of Markov chains with tempered posteriors. At each iteration, two randomly selected chains in the ensemble are proposed to swap parameter vectors, after which each chain mutates via Metropolis-Hastings. The algorithm results in a fast-mixing MCMC, particularly well suited for problems with irregular posterior distributions. Also, due to its global nature, the algorithm can be initialized directly from the prior distributions. I provide two empirical examples with complex posteriors: a New Keynesian model with equilibrium indeterminacy and the Smets-Wouters model with more diffuse prior distributions. In both examples, parallel tempering overcomes the inherent estimation challenge, providing extremely consistent estimates across different runs of the algorithm with large effective sample sizes. I provide code compatible with Dynare mod files, making this routine straightforward for DSGE practitioners to implement.

Topics: Econometric and statistical methods, Economic models
JEL codes: C11, C15, E10

Résumé

Dans cette étude, j'élabore un algorithme de Monte-Carlo par chaînes de Markov (MCMC) fondé sur une population, soit une atténuation parallèle, pour estimer des modèles d'équilibre général dynamique et stochastique (EGDS). L'atténuation parallèle fait une approximation de la distribution d'intérêt a posteriori à l'aide d'une famille de chaînes de Markov à distributions a posteriori tempérées. À chacune des itérations, deux chaînes sélectionnées de façon aléatoire dans l'ensemble sont proposées pour échanger les vecteurs des paramètres, après quoi chaque chaîne subit une mutation par le biais de l'algorithme de Metropolis-Hastings. L'algorithme crée une méthode MCMC à mixage rapide, qui convient particulièrement bien aux problèmes avec des distributions a posteriori irrégulières. De plus, à cause de sa nature globale, l'algorithme peut être initialisé directement à partir des distributions a priori. Je fournis deux exemples empiriques contenant des distributions a posteriori complexes : un modèle de type nouveau keynésien se caractérisant par des équilibres indéterminés et un modèle de Smets-Wouters où les distributions a priori sont plus diffuses. Dans les deux exemples, l'atténuation parallèle surpasse le défi inhérent à l'estimation, fournissant des estimations extrêmement cohérentes lors de différentes exécutions de l'algorithme où la taille effective de l'échantillon est de grande taille. Je fournis un code compatible avec des fichiers de modélisation Dynare, ce qui facilite l'implantation de cette routine pour les utilisateurs de modèles EGDS.

Sujets : Méthodes économétriques et statistiques, Modèles économiques
Codes JEL : C11, C15, E10

1 Introduction

Modern dynamic stochastic general equilibrium (DSGE) models are predominantly estimated using Bayesian methods. However, the posterior distribution of these models typically cannot be evaluated analytically and instead relies on simulation methods to sample from the posterior. The most common simulation method used by applied macroeconomists is the random walk Metropolis-Hastings (RWMH) algorithm, which belongs to a class of Markov chain Monte Carlo (MCMC) methods.

RWMH remains popular even though it has a number of well-documented shortcomings.[1] These include slow convergence to true moments of the posterior and limited capacity for parallelization, but most concerning is that the algorithm can get stuck at local modes and fail to explore the entire posterior distribution. These issues are especially acute for DSGE models used in policy institutions, which are characterized by a large number of shocks and propagation mechanisms, leading to posterior distributions which are often hard to sample from. Further, as these models continue to expand and improve the set of empirically relevant channels, these issues are likely to become more pronounced.

In this paper I develop a population-based MCMC method known as parallel tempering (PT-MCMC) to estimate DSGE models. Parallel tempering was originally proposed for physical systems (Geyer, 1991) but is now widely used in the sciences and computational statistics.[2] Parallel tempering approximates a target posterior distribution using a family of Markov chains running at different temperatures and exchanging information between them. Chains with hotter temperatures afford relatively more weight in the posterior distribution to the prior. The family of Markov chains are by construction related, but chains with hotter temperatures are easier to sample from and can both provide and store valuable information on the target distribution.

The PT-MCMC algorithm iterates over two types of updates: exchange and mutation. Exchange involves randomly selecting two chains at each iteration and proposing a swap between their parameter vectors, which is accepted according to a Metropolis criterion. The exchange step allows for more global moves, resulting in a faster-mixing MCMC with less serial correlation, and avoids getting stuck in local modes. After the exchange update, each chain performs a mutation update based on a pre-specified number of RWMH steps. The target posterior of interest is simply the coldest chain in the family, and parameter vectors which provide a good fit to the data can effectively bubble sort their way to this chain.

[1] For example, the Bank of Canada uses an adaptive Metropolis-Hastings algorithm to estimate its main DSGE model, ToTEM III, which features 130 estimated parameters. See Corrigan et al. (2021, pg. 55).

[2] Introductions to parallel tempering can be found in Liu (2001, pg. 212) and Gelman et al. (2021, pg. 300).

The algorithm has a number of desirable features for DSGE practitioners, particularly for those working with large-scale DSGE models. First, PT-MCMC is well suited to sample from complex distributions often associated with DSGE models, such as posterior surfaces which are multimodal or contain discontinuities. Second, because of its global nature, PT-MCMC does not require finding the posterior mode or Hessian, which can be tedious and prone to failures. The algorithm can instead be initialized directly from the prior distributions. Third, the algorithm produces large-efficiency gains compared to RWMH because it mixes faster and is able to make use of parallel computing since each chain operates independently during the mutation stage. Fourth, the algorithm is straightforward to tune and operationalize. A moderate number of chains ($\sim$ 8-12) based on functional temperature schemes produce reasonable exchange acceptance rates for the examples in this paper. Fifth, I provide a MATLAB script which takes any Dynare mod file as an input and runs this routine, making it easy to implement for practitioners.

In this paper I make two primary contributions. To my knowledge, I am the first to show how parallel tempering can be adopted to estimate DSGE models. I report estimation results under a variety of tuning settings and provide MATLAB codes which can be used to implement PT-MCMC for DSGE models written in Dynare, offering a population-based MCMC simulation technique which is not currently available in the Dynare software. I provide different versions of the code which allow for parallelization of the mutation step. The parallelization version is particularly useful for large-scale DSGE models, where the solution of the model and evaluation of the likelihood are computationally costly. These should be of interest to practitioners interested in more powerful simulation techniques, while at the same time maintaining the functionality of Dynare.

My second contribution is to illustrate the strength of this method using two empirical applications characterized by complex posterior distributions. The first application estimates a small-scale New Keynesian (NK) DSGE model permitting the possibility of equilibrium indeterminacy. Models with indeterminant equilibria are often challenging to estimate because the likelihood can be discontinuous along much of the boundary between the determinacy and indeterminacy regions. Consequently, algorithms like RWMH can often remain stuck in the region in which they are initialized, failing to explore the entire posterior.

Using PT-MCMC, I estimate the NK model on US data over the years 1960-1979, a period in which a long literature has argued that the US economy was in a state of indeterminacy due to passive monetary policy (Clarida, Galí and Gertler, 2000; Lubik and Schorfheide, 2004), and/or because the level of trend inflation was too high (Coibion and Gorodnichenko, 2011; Hirose, Kurozumi and Van Zandweghe, 2020). I find strong support for the interpreta-

tion that the US economy was in a state of equilibrium indeterminacy and that the volatility of this period was in part due to self-fulfilling expectations-driven phenomena. Estimates of the Federal Reserve's (the Fed's) reaction function suggest they were most likely not satisfying the Taylor principle, with a mean coefficient response to inflation of 0.69. However, even if they had satisfied the Taylor principle, this would not have guaranteed determinacy due to the high level of trend inflation, equal to nearly 4% annualized. I run the PT-MCMC algorithm a number of times and show that regardless of the initial positions of the parameter vectors, PT-MCMC delivers consistent estimates with large effective sample sizes and has no difficulty crossing the boundary between the determinacy and indeterminacy regions.

The second empirical application estimates the Smets and Wouters (2007) DSGE model over the 1966:2004 period, but under more uninformative prior distributions. As documented in Herbst and Schorfheide (2014), the empirical performance of the Smets-Wouters model relies quite heavily on arguably implausibly tight prior distributions. They show that relaxing the tight prior distributions leads to a substantial improvement in model fit but also makes the posterior surface multimodal. They show that RWMH performs poorly, getting stuck around local modes and producing widely different parameter estimates across different runs of the algorithm. Using PT-MCMC, I find very consistent estimates across different runs of the algorithm. With 750,000 draws (excluding warm-up), the algorithm produces effective sample sizes for the most difficult to estimate parameters in the hundreds, with others well into the thousands. The method also captures the multimodality inherent in the wage and price dynamics documented by Herbst and Schorfheide. Further, the application illustrates that this method is reliable and efficient for estimating large-scale DSGE models.

Before proceeding, it is worth distinguishing the conceptual difference between PT-MCMC and simply running a number of parallel chains of RWMH. The primary difference stems from how the target posterior is approximated. It is possible to run a number of parallel chains using RWMH and pool draws from the chains, but for the approximated posterior to be correct, it requires that each chain has converged to the ergodic distribution. If some chains have not, perhaps due to multimodal posteriors or simply a failure to converge, then aggregating samples from each chain will result in an incorrect approximation to the true posterior.[3] Parallel tempering does not suffer from such a dilemma because the target posterior of interest is a single chain in the family with the coldest temperature. The other chains simply facilitate exploration of the parameter space and the option for the coldest chain to visit these areas if they provide a good fit to the data. Additionally, since the auxiliary chains have tempered posteriors which tend to smooth out the posterior surface, they are much less

[3]For an example, see the results from RWMH using multiple chains in An and Schorfheide (2007, Section 4.3).

susceptible to being trapped in local modes.

Related Literature. This paper builds on a long literature using Bayesian methods to estimate linearized DSGE models. A textbook treatment is provided in Herbst and Schorfheide (2016). Early works followed Schorfheide (2000) and used MCMC methods to characterize the posterior distribution of interest.[4] Several attempts to mitigate issues with MH have been proposed in a DSGE context. One approach is to group parameters into blocks in an effort to reduce the inherent persistence in MH Markov chains, where groups are random or chosen by the researchers (Cúrdia and Reis, 2010; Chib and Ramamurthy, 2010). Strid, Giordani and Kohn (2010) propose an adaptive Metropolis-Hastings algorithm where the proposal distribution is a combination of the random walk proposal, an Independence proposal, and a continuously estimated t-copula. However, all of the aforementioned solutions generally place a large burden on researchers in terms of tuning these algorithms correctly, and the correct tuning settings are likely to be context specific. The approach in this paper is easier to tune and only requires changes from my suggested default tuning setting in exceptional cases.

Other works have adopted Sequential Monte Carlo (SMC) sampling techniques. SMC is also a populated-based algorithm, which recursively constructs importance samplers to approximate a posterior distribution. Creal (2007) designs two SMC sampling algorithms and uses them to estimate a small scale New Keynesian model. Herbst and Schorfheide (2014, 2016) provide a number of additional results for theoretical convergence and tuning so that SMC could be applied to the DSGE model of Smets and Wouters (2007). Cai et al. (2020) show how the tempering schedule for SMC can be chosen in an adaptive manner in a DSGE context, making SMC an efficient approach for online estimation and forecasting. While both PT-MCMC and SMC are populated-based algorithms, PT-MCMC may be preferable to SMC because it requires a limited number of tuning choices and can utilize the wide array of convergence tests and diagnostic checks for MCMC methods.[5]

This paper is also related to recent work by Farkas and Tatar (2020) and Childers et al. (2022), who implement variations of Hamiltonian Monte Carlo (HMC) to estimate DSGE models. HMC makes use of information in the gradient of the target distribution to improve sampling performance and generate draws with little serial correlation. While HMC yields

[4] A review of the evolution of Bayesian methods for DSGE models is provided in Fernández-Villaverde et al. (2016). See Chapter 16.

[5] One potential drawback of PT-MCMC is that it does rely on having reasonable prior distributions since the tempering is done towards the prior. If the priors are strongly at odds with the data, PT-MCMC may fail. For example, see Campbell and Steele (2012, Section 4.2.).

improvements in efficiency, the main drawback is that it can perform poorly when the posterior distribution is multimodal. This point is emphasized by Farkas and Tatar (2020), who combine HMC with SMC to deal with this issue. PT-MCMC, on the other hand, is ideally suited for multimodal posteriors.

The remainder of this paper is organized as follows. In Section 2, I describe the parallel tempering algorithm and how it relates to the widely used RMWH algorithm. In Section 3, I estimate two models with complex posterior distributions: a New Keynesian model which permits the possibility of equilibrium indeterminacy, and the Smets-Wouters model under more diffuse priors. Section 4 concludes.

2 Parallel Tempering

The solution to a linearized rational expectations DSGE model has a state-space representation given by

$$s_t = \Phi_1(\theta) s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t, \tag{1}$$
$$y_t = \Psi_0(\theta) + \Psi_1(\theta) s_t, \tag{2}$$

where s_t contains the model's variables, y_t relates the variables to a set of observables (ignoring any measurement error), and ϵ_t is a vector of normally distributed structural shocks. $\Phi_1(\theta)$, $\Phi_\epsilon(\theta)$, $\Psi_0(\theta)$, and $\Psi_1(\theta)$ are coefficient matrices that are functions of the structural parameters, θ, of the DSGE model.

Estimating the model using Bayesian methods involves combining prior distributions, denoted by $p(\theta)$, with the likelihood function $p(Y|\theta)$ to obtain the posterior distribution $P(\theta|Y)$. According to Bayes' Theorem, the posterior distribution is proportional (up to a constant) to the product of the prior density and likelihood function

$$P(\theta|Y) \propto p(Y|\theta)p(\theta). \tag{3}$$

However, in practice, DSGE models do not permit direct analysis of the posterior, and numerical methods are used to characterize it.

Parallel tempering starts by introducing a family $m = 1, \ldots, N_{chain}$ of posterior distributions denoted by $P_m(\theta|Y)$. This ensemble of distributions forms a high-dimensional augmented distribution denoted by

$$\tilde{\pi} = \prod_{m=1}^{N_{chain}} P_m(\theta|Y). \tag{4}$$

Each posterior in the ensemble has an associated inverse temperature given by $0 \leq \xi_1 < \cdots < \xi_{N_{chain}} = 1$, which tempers the likelihood in the posterior. Subsequently, each posterior in the ensemble is approximated by

$$P_m(\theta|Y) \propto (P(Y|\theta))^{\xi_m} P(\theta), \tag{5}$$

where each of the N_{chain} posterior approximations is the target density for the N_{chain} MCMC chains. According to (5), if $\xi_1 = 0$ we have $P_1(\theta|Y) = P(\theta)$, i.e., the chain heated to the highest possible inverse temperature places zero weight on the likelihood, and the posterior distribution is proportional to the prior distribution. Conversely, for $\xi_{N_{chain}} = 1$, we have $P_{N_{chain}}(\theta|Y) = P(Y|\theta)P(\theta)$, which is the target posterior distribution of interest.

The version of parallel tempering implemented in this paper is as follows:

Algorithm 1 (Parallel Tempering)

0. **Initialization.** *Construct a temperature ladder where temperatures are organized according to* $0 \leq \xi_1 < \cdots < \xi_{N_{chain}} = 1$. *Initialize chains with a parameter vector* $\theta_m^{(1)}$ *drawn from the prior distribution* $p(\theta)$ *for* $m = 1, \ldots, N_{chain}$. *Let each* c_m *and* Σ_m *be the associated scale parameter and covariance matrix for the proposal distribution of chain* m.

1. **Iterate.** *While* $i < N_{iter}$:

   (a) **Exchange.** *Randomly draw two integers* j *and* k *from the interval* $[1, \ldots, N_{chain}]$. *Set* $\theta_j^{(i)} = \theta_k^{(i)}$ *and* $\theta_k^{(i)} = \theta_j^{(i)}$ *with probability*

$$\alpha_e = \min\left\{1, \frac{P_j(\theta_k^{(i)}|Y)P_k(\theta_j^{(i)}|Y)}{P_j(\theta_j^{(i)}|Y)P_k(\theta_k^{(i)}|Y)}\right\} \tag{6}$$

and $\theta_j^{(i)} = \theta_j^{(i)}$ *and* $\theta_k^{(i)} = \theta_k^{(i)}$ *otherwise.*

(b) **Mutation.** *For* $m = 1$ *to* N_{chain}*:*

Draw ϑ_m *from the symmetric proposal distribution* $q^m(\vartheta_m|\theta_m^i, c_m, \Sigma_m)$*. Set* $\theta_m^{(i+1)} = \vartheta_m$ *with probability*

$$\alpha_m = \min\left\{1, \frac{P_m(\vartheta_m|Y)}{P_m(\theta_m^{(i)}|Y)}\right\} \tag{7}$$

and $\theta_m^{(i+1)} = \theta_m^{(i)}$ *otherwise.*

$i = i + 1$

2. *Let* $P(\theta|Y)$*, the target posterior distribution of interest, equal* $P_{N_{chain}}(\theta|Y)$*.*

Algorithm 1 outlines that in each iteration, two distinct updates can occur. The first update is the exchange step. Two randomly chosen chains are proposed to swap parameter vectors, which is accepted according to a Metropolis criterion.[6] Since the posterior is evaluated in the previous iteration, this step does not require any additional likelihood evaluations. The second update is the mutation step, where all chains perform a random-walk Metropolis-Hastings step.

Since both the scale parameter and covariance matrix for the proposal density are subscripted by m, they can be specific to each chain. From an efficiency standpoint, this is ideal, as the covariance structure for a chain with a very hot temperature is likely to be quite different from the covariance structure for a chain with a cold temperature. Since the structure cannot be known a priori, I opt for an adaptive approach. The algorithm initially uses a common covariance matrix for the proposal density based on the prior covariance, which is updated to a sample covariance matrix during the warm-up phase. The scale parameters

[6] In some implementations of parallel tempering, only neighbouring chains are proposed to swap. I prefer proposing swaps between all chains because it allows for parameter vectors in highly heated chains to move to the coldest chains quickly.

are also individually updated to ensure reasonable acceptance rates for the mutation steps.

It is straightforward to see that RWMH with parallel chains is a special case of Algorithm 1 if the exchange steps are eliminated, all temperatures are set equal to one, and the posterior is approximated by pooling draws from all the chains. But having parallel chains with tempered likelihoods running in tandem with the target chain has a number of advantages.

First, auxiliary chains of the system contain information on the posterior distribution of interest. Even though these chains have tempered likelihoods, the parameters which yield high likelihood values in a relatively hotter chain should also yield high likelihood values in a relatively colder chain (though this may not always be true for the posterior). This information can then be used in the target chain when a swap occurs between the target chain and one of the hotter chains.

Second, in cases where the posterior may be ill behaved, highly heated chains flatten the posterior surface. Due to this flattening, hotter chains are much less susceptible to becoming trapped in local modes and facilitate an exploration of the entire posterior space.

Third, by definition, highly heated chains are easier to sample from than colder chains. For example, consider a chain heated to the highest possible inverse temperature, such that $\xi_m = 0$. In this case, the posterior distribution is simply the prior distribution, which is straightforward to sample from. Provided that the shape of the temperature schedule is reasonably smooth, one can move from a distribution which is easy to sample from (the prior distribution) to a distribution which may be highly complex (the target posterior).

2.0.1 Convergence

It is well known that, despite its added complexity, parallel tempering remains a Markov Chain Monte Carlo method. Since the exchange step is a Metropolis-Hastings move, the algorithm continues to satisfy a detailed balance condition. To see why this is the case, suppose for simplicity that the ensemble is characterized by only two chains, with respective states given by i and j. The detailed balance condition requires that the following holds:

$$\tilde{\pi}(i,j)\Gamma(j,i|i,j) = \tilde{\pi}(j,i)\Gamma(i,j|j,i) \quad \forall \quad i,j. \tag{8}$$

$\tilde{\pi}(.,.)$ is the product of the two posteriors evaluated at their respective states, and $\Gamma(.,.|.,.)$ is the transition probability, which is a function of the distribution proposing swaps between

chains and the acceptance probability. That is, e.g., $\Gamma(j, i|i, j) = Q(2,1) \cdot \min\left(1, \frac{P_1(\theta_j|Y)P_2(\theta_i|Y)}{P_1(\theta_i|Y)P_2(\theta_j|Y)}\right)$.

Now suppose that $\frac{P_1(\theta_j|Y)P_2(\theta_i|Y)}{P_1(\theta_i|Y)P_2(\theta_j|Y)} > 1$. Based on this assumption, equation (8) can be expressed as

$$P_1(\theta_i|Y)P_2(\theta_j|Y)Q(2,1)(1) = P_1(\theta_j|Y)P_2(\theta_i|Y)Q(1,2)\left(\frac{P_1(\theta_i|Y)P_2(\theta_j|Y)}{P_1(\theta_j|Y)P_2(\theta_i|Y)}\right), \tag{9}$$

and since the distribution proposing swaps between chains is symmetric, it is the case that $Q(2,1) = Q(1,2)$, and equation (9) simplifies to

$$P_1(\theta_i|Y)P_2(\theta_j|Y) = P_1(\theta_j|Y)P_2(\theta_i|Y). \tag{10}$$

Thus, the exchange move introduced in parallel tempering satisfies a detailed balance condition with respect to $\tilde{\pi}$, and $\tilde{\pi}$ is a stationary distribution of the chains. A further discussion of convergence properties for population-based MCMC methods can be found in Jasra, Stephens and Holmes (2007a).

2.0.2 Number of distributions and tempering schedule

Parallel tempering requires two additional tuning choices: the number of chains in the ensemble and the shape of the temperature ladder. A primary factor to consider when making these choices is the acceptance rate of exchanges between chains. A general rule of thumb is to target an exchange acceptance rate between neighbouring chains of roughly 50% (Liu, 2001). Both the number of chains and the shape of the temperature ladder will influence this acceptance rate.

At first it may seem natural to use as many chains as possible (given computing capacity), but there is a trade-off that one needs to consider when selecting the number of chains. Increasing the number of chains allows for the algorithm to store more information about the target posterior, which should ultimately be beneficial, and at the same time a large number of chains makes it relatively easy to specify the temperature ladder since the gaps will be small and ensure reasonable exchange acceptance rates. But increasing the number of chains also raises the theoretical convergence time of the Markov chain. Consequently, for

a fixed computation time, there does exist a point where adding more chains can result in inferior performance of the algorithm per unit time (Jasra, Stephens and Holmes, 2007b, see Section 2.4). For reference, in this paper I use 8 chains for the small-scale NK DSGE model and 12 chains for the Smets-Wouters model. However, I report results using an alternative number of chains for the Smets-Wouters DSGE model in Section 3.2.3.

To specify the shape of the temperature ladder, there are a number of functional tempering schemes one could follow. Jasra, Stephens and Holmes (2007b) explore schemes based on uniform, logarithm, and power decay. Each type of scheme may be preferable in specific contexts. If the posterior has many local modes, a temperature ladder with many highly heated chains may be more efficient at exploring the entire posterior space and escaping local modes than a ladder with many cool chains. In preliminary work, I explored a variety of temperature schemes and found that schemes with convex temperature schedules were the most appropriate for producing exchange acceptance rates between neighbouring chains around 50%. The exact schedules in this paper were produced according to

$$\xi_m = \left(\frac{m}{N_{chain} + k} \right)^{\gamma}, \tag{11}$$

for $m = 1, \ldots, N_{chain} - 1$, where k and γ are chosen constants. The chain $m = N_{chain}$ is the target density and has an inverse temperature given by $\xi_{N_{chain}} = 1$.

2.0.3 The transition kernel

To specify the transition kernel, it is common in practice to first run an optimization routine and initialize the MCMC at the posterior mode with a covariance matrix for the proposal density equal to the negative of the Hessian at the mode. But this approach can be prone to failures, especially when the likelihood is ill behaved. In this paper, I initialize the chains with draws from the prior distribution and use the prior covariance matrix for the proposal density. The covariance matrix for each chain is subsequently updated to a sample covariance matrix during the warm-up phase. The scaling parameter is tuned during the warm-up phase such that roughly 23% of RWMH steps are accepted. The scaling parameter is adjusted using the formula described in Algorithm 3 of Herbst and Schorfheide (2014).

3 Empirical Applications

3.1 A New Keynesian model with equilibrium indeterminacy

The New Keynesian framework has become the workhorse model for monetary policy analysis. It is well-known that equilibrium indeterminacy can arise in NK models if the central bank does not respond sufficiently strongly to inflation fluctuations and that the required central bank response to ensure determinacy rises with the level of trend inflation in the economy (Ascari and Ropele, 2009; Khan, Phaneuf and Victor, 2020). When there is equilibrium indeterminacy, business cycle fluctuations can be driven not only by fundamental shocks such as total factor productivity, monetary policy, and government spending, but also by sunspot shocks linked to agents' expectation errors of forward-looking variables.

However, estimating a model jointly over the determinacy and indeterminacy regions of the parameter space is challenging for algorithms like RWMH.[7] The reason for this is twofold. First, within each region of the parameter space there may exist local modes that the algorithm may gravitate around. Second, the boundary between the two regions is often characterized by discontinuities in the likelihood function, which can make it difficult to cross between the two. Algorithms like RWMH can often remain stuck in the region in which they are initialized and fail to explore the entire posterior space. For the interested reader, Appendix B provides an illustration and discussion of these discontinuities in the likelihood function in this model.

To deal with this estimation challenge, the literature has generally taken two different approaches. The first is to estimate the model in each region separately and then make inference by comparing model fit statistics between the two regions (e.g., Lubik and Schorfheide (2004); Bhattarai, Lee and Park (2016); Albonico, Ascari and Haque (2020)). While this approach maintains tractability, inferences could be misleading if the true posterior does not lie firmly in one region or the other. The second approach is to use a more sophisticated sampling method, such as Sequential Monte Carlo sampling. But because this method is not available in packages like Dynare, it is not widely used.

A large literature has argued that the volatility of macroeconomic aggregates, frequent recessions, and rising inflation exhibited by the US economy during the 1960s and 1970s was driven in part by equilibrium indeterminacy, either because the Fed was not respond-

[7] A separate challenge concerns solving the state space of the model under indeterminacy. Much of the work in the New Keynesian literature uses the approach described in Lubik and Schorfheide (2003, 2004), which requires solving the state space representation of the model differently depending on whether the parameters are consistent with determinacy or indeterminacy. In this paper I use a more recent, though equivalent, approach proposed by Bianchi and Nicoló (2021).

ing strongly enough to inflation (Clarida, Galí and Gertler (2000), Lubik and Schorfheide (2004)), the level of trend inflation was too high (Coibion and Gorodnichenko, 2011), or a combination of both of these factors (Hirose, Kurozumi and Van Zandweghe, 2020). I revisit the question of equilibrium indeterminacy in the US during the 1960s and 1970s using PT-MCMC. The model is a small-scale New Keynesian model with positive trend inflation, similar to the model in Ascari and Sbordone (2014).[8] The model features four exogenous disturbances, which include shocks to TFP, preferences, monetary policy, and if in a state of indeterminacy, sunspot shocks. The prior distributions are relatively standard and are reported in Table 7 in Appendix C. The model is estimated using three observables over the period 1960I:1979II, real per capita output growth, inflation, and the Federal Funds rate.[9]

3.1.1 Tuning

Table 1 describes the tuning of the algorithm. I use 8 chains with a convex temperature schedule ranging from 0.04 to 1.0. This schedule is generated using equation (11) with $k = 0.5$ and $\gamma = 1.5$. The algorithm is run for 500,000 iterations and discards the first half as a warm-up. Each chain is initialized with a draw from the prior distribution and the initial covariance matrix, for the proposal distribution is equal to the prior covariance matrix. After 200,000 iterations, I compute the sample covariance matrix for each chain and use these as the covariance matrices for the proposal density in the remaining iterations. During the warm-up phase, the scaling parameters are also tuned so that roughly 23% of mutation steps are accepted. Finally, while the mutation step could be done in parallel because each chain is independent during this step, I found that with my hardware and this model, the overhead costs associated with parallelization outweighed the benefits and resulted in slower performance. Thus the mutation step here is computed serially.

Table 1: PT-MCMC tuning for New Keynesian model with equilibrium indeterminacy

Tuning parameters	
N_{chain}	8
ζ	[0.04, 0.11, 0.21, 0.32, 0.45, 0.59, 0.75, 1]
N_{iter}	500,000
N_{burn}	250,000
Σ	Adaptive covariance

[8] Details on the model and its log linearized equations are reported in the Appendix.

[9] All data was obtained from FRED. Real per capita output is computed by dividing gross domestic product (FRED code: GDP) by the GDP deflator (FRED code: GDPDEF) and a smoothed estimate of the level of population (FRED code: CNP16OV). Inflation is the log difference in the GDP deflator, and the Federal Funds rate is the average quarterly rate (FRED code: FEDFUNDS).

3.1.2 Results

Table 2 reports estimates of the mean parameter values, highest posterior density (HPD) intervals, marginal data density (MDD), and the posterior probability of determinacy, as well as the variability of each of these estimates across 10 different runs of the algorithm. Each run of the algorithm took approximately 65 minutes using an AMD Ryzen 9-5900x 12 core processor. Convergence of the Markov chains was assessed using the scale reduction factor (Gelman et al., 2021, pg. 284).

Table 2: Posterior estimates for NK model

		Parallel tempering		
Parameter	Mean	90% HPD Interval	STD(Mean)	$\hat{n}_{eff}$
σ_z	1.52	[1.13, 1.90]	3.67e-03	3028
σ_b	0.11	[0.02, 0.18]	8.56e-03	1246
σ_v	0.21	[0.17, 0.24]	4.62e-04	3220
σ_s	0.31	[0.26, 0.35]	4.39e-04	4124
$\rho_{z,s}$	-0.15	[-0.36, 0.05]	3.10e-03	3045
$\rho_{b,s}$	-0.11	[-0.96, 0.68]	1.42e-02	2537
$\rho_{v,s}$	-0.09	[-0.43, 0.25]	2.87e-03	2578
h	0.43	[0.29, 0.57]	1.12e-03	3031
ζ_p	0.61	[0.54, 0.68]	9.37e-04	2643
ρ_i	0.69	[0.51, 0.88]	1.38e-03	2712
ϕ_π	0.69	[0.35, 1.02]	3.61e-03	2575
$\phi_{\Delta y}$	0.15	[0.07, 0.22]	6.25e-04	2772
ϕ_y	0.10	[0.02, 0.17]	8.57e-04	2866
ρ_z	0.19	[0.05, 0.33]	1.96e-03	2957
ρ_b	0.51	[0.18, 0.84]	5.74e-03	2690
ρ_v	0.46	[0.29, 0.62]	1.08e-03	3142
$\bar{\pi}$	0.98	[0.71, 1.24]	3.21e-03	3090
$\bar{i}$	1.26	[0.99, 1.53]	3.30e-03	3090
$\bar{g}_A$	0.42	[0.27, 0.57]	2.04e-03	3032
α_{BN}	0.75	[0.52, 0.97]	3.43e-03	2797
Mean(Log MDD)	-145.16			
Std(Log MDD)	0.11			
Mean(Prob. of Determinacy)	7.00e-05			
Std(Prob. of Determinacy)	1.34e-04			
Average Computing Time	65 mins			
Average Exchange Rate	0.51			
Average Mutation Rate	0.23			

Notes: Mean estimates and HPD intervals are computed as the mean across 10 runs of the algorithm. The marginal data density is computed using the harmonic mean estimator (Geweke, 1999). n_{eff} is the effective sample size computed as $\hat{n}_{eff} = \frac{n}{1+2\sum_{t=1}^{T}\hat{\rho}_t}$, where n is the number of draws and T is the first odd positive integer for which the difference in autocorrelations $\hat{\rho}$ at $T+1$ and $T+2$ is negative (Gelman et al., 2021, pg. 287).

Parallel tempering. Parameter estimates from Table 2 show that PT-MCMC provides extremely consistent estimates. Across different runs of the algorithm, the mean estimates are

essentially identical, with standard deviations of the posterior means all falling below 0.01. Additionally, both the MDD and posterior probability of determinacy statistics show little variation across different runs. This finding is important considering that in each run, the chains are initialized with draws from the prior distribution, suggesting that the discontinuities present in the likelihood function in this model do not prevent PT-MCMC from accurately characterizing the posterior distribution. The algorithm also appears to be tuned adequately as the number of chains and temperature schedule produce a reasonable exchange acceptance rate averaging 51% between neighbouring chains.

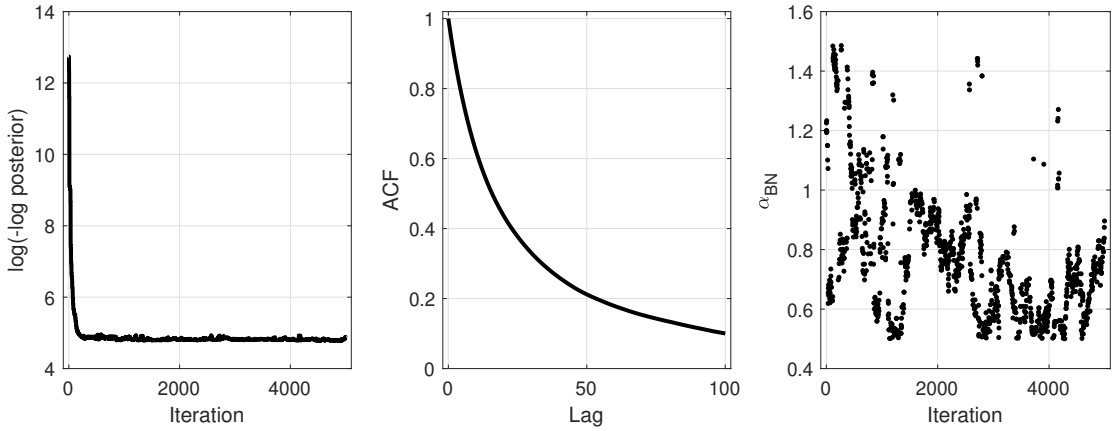
One way to assess the precision of parameter estimates and efficiency of the algorithm is to compute the effective sample size (ESS), which is a measure intended to approximate the number of "independent simulation draws" if one could draw directly from the posterior itself. A general rule of thumb is to aim for a minimum of 10 independent draws (Gelman et al., 2021). ESS is reported under the $\hat{n}_{eff}$ column in Table 2. PT-MCMC exceeds the floor by a wide margin, producing effective sample sizes of well over 1,000 for all parameters, which is consistent with the fact that mean estimates are nearly identical across different runs of the algorithm.

PT-MCMC has a number of other desirable features worth highlighting, particularly as they relate to the shortcomings commonly attributed to RWMH. One criticism of RWMH is that it can be slow to converge to true moments of the posterior distribution because draws are highly correlated. PT-MCMC mixes much faster because the target chain can use the auxiliary chains to find parameter vectors which fit the data well. The left panel in Figure 1 plots the log of the (negative) log posterior over the first 5,000 iterations of the algorithm for the target chain. Within only a few hundred iterations, the algorithm has converged to high-likelihood regions of the parameter space near the global max. The benefit of this fast mixing is twofold. First, PT-MCMC does not require a long warm-up phase, and second, the algorithm can be initialized directly from the prior distributions.

A second criticism of RWMH is that draws are highly correlated, and consequently the algorithm can require a large number of iterations to produce a reasonable ESS. PT-MCMC produces significantly lower autocorrelation in the parameter draws because of the exchange step, meaning that a much lower number of draws is required to produce the same effective sample size. The middle panel in Figure 1 shows the autocorrelation in parameter draws from the target chain averaged across the 20 parameters in the model. The ACF exhibits a sharp drop within the first few lags, falling below 0.4 by lag 25. In comparison, RWMH can often produce ACF values well above 0.9 by lag 25.

Lastly, while the estimates here favour a probability of determinacy of effectively zero,

Figure 1: Properties of PT-MCMC



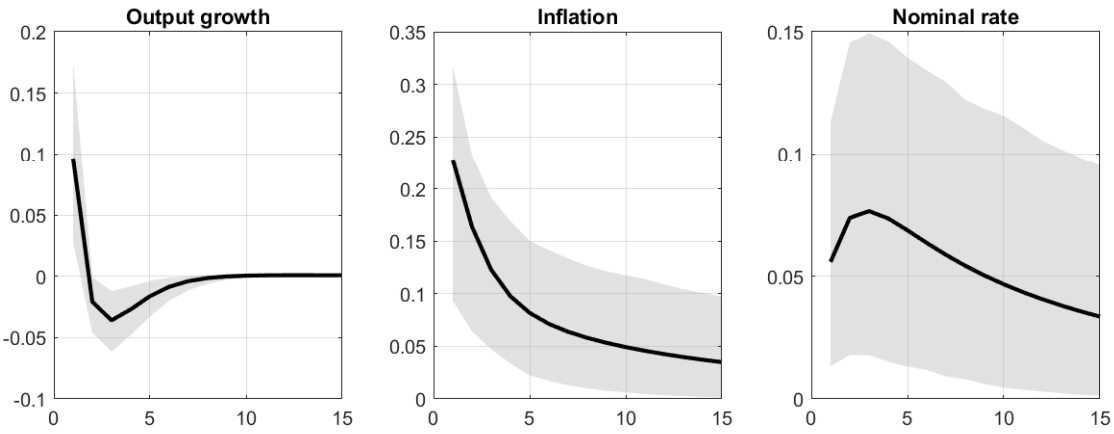
Notes: The reported autocorrelation function is computed by averaging the ACF across all parameters in the model.

there are applications where the probability of determinacy can be closer to 50-50. In these cases, having an algorithm that can efficiently cross between the determinacy and indeterminacy regions of the parameter space becomes important. The rightmost panel of Figure 1 shows the parameter which governs (in)determinacy in the model for the target chain over the first 5,000 iterations. When the parameter is larger than one, the model solution is characterized by determinacy, and when the parameter is less than one, it is characterized by indeterminacy. With PT-MCMC, the target chain is constantly switching between the two regions at the beginning, suggesting that the target chain has no issues traversing the boundary between indeterminacy and determinacy. As a result, PT-MCMC will also be suitable for applications where the target distribution does not lie firmly in the determinacy or indeterminacy region of the parameter space.

Parameter estimates and economic implications. According to the parameter estimates, the data strongly favours the conclusion that the US economy was in a state of equilibrium indeterminacy during this period, with a posterior probability of determinacy of less than 1%. Indeterminacy was primarily driven by the low estimated monetary policy response to inflation, with a mean estimate of the Fed's inflation response parameter of 0.69. The estimates do suggest that there was a roughly 3% posterior probability that the Fed was satisfying the Taylor principle (a response greater than 1), but this would not have ensured determinacy due to the high level of trend inflation, which is estimated to be roughly 4% annualized. The finding that the US economy was in a state of equilibrium indeterminacy means that macroeconomic fluctuations during this period were driven by not only TFP, preference, and monetary policy shocks but also by sunspot shocks.

Figure 2 plots the impulse response functions of output growth, inflation, and the nominal interest rate to a one-standard-deviation sunspot shock. Solid black lines indicate the mean response, and grey bands denote 90% HPD intervals. The sunspot shock generates a comovement between the variables similar to a demand shock, raising output growth, inflation, and the nominal interest rate. The sunspot shock propagates by raising inflation and inflation expectations on impact, and while the central bank raises the policy rate in response, it does so less than one-for-one. As a result there is a fall in the ex-ante real interest rate, generating an increase in output growth.[10] This finding is consistent with Lubik and Schorfheide (2004) and Hirose, Kurozumi and Van Zandweghe (2020), who report qualitatively similar impulse response functions for sunspot shocks using different model ingredients, solution and estimation methods.

Figure 2: Impulse response functions to a sunspot shock



Notes: Mean impulse response functions are denoted by the black line, and shaded bands indicate 90% HPD intervals. Mean and 90% HPD intervals were computed using 1,000 draws from the posterior distribution.

Lastly, according to the posterior mean estimates, sunspot shocks played a significant role in inflation and nominal interest rate fluctuations during this period, contributing to nearly 45% of inflation variance and 16% of nominal interest rate variance. However, sunspot shocks played a relatively minor role in output growth fluctuations, contributing to around only 2% of output growth variance.

[10]The solution method does allow for sunspot shocks to be arbitrarily correlated with the other shocks in the model, which could alter the propagation of sunspot shocks. However, posterior estimates do not suggest a strong correlation between sunspot shocks and the other shocks in the model.

3.2 Smets-Wouters with diffuse priors

The Smets-Wouters (SW) model is a large-scale DSGE model featuring a number of real and nominal frictions and seven exogenous shocks. Smets and Wouters (2007) estimated their model using post-war US data and showed that the model fit the data remarkably well, comparable to that of reduced-form econometric models. Their findings also contrasted the narrative that excessive US macroeconomic fluctuations during the 1970s were a result of policy failures and instead pointed towards exogenous shocks.[11] Subsequent debate surrounding the results in the paper partly centred on their choice of prior distributions for the structural parameters (see, e.g., Del Negro and Schorfheide (2013)).

Herbst and Schorfheide (2014) re-estimated this model with more uninformative prior distributions to illustrate their Sequential Monte Carlo sampling algorithm. They showed that with more diffuse prior distributions, the posterior is multimodal. Consequently, the RWMH algorithm could deliver widely different parameter estimates (and policy implications) depending on which mode the algorithm gravitated around. The most glaring differences were obtained for the parameters governing wage and price dynamics in the model. They write (bold characters are my own):

> *For instance, the standard deviation of the estimate of the mean for ρ_w, the autoregressive coefficient for the wage markup, is 0.09. Given the point estimate of* 0.69*, this means that any given run of the simulator* ***(RWMH)*** *could easily return a mean between* 0.5 *and* 0.9*, even after simulating a Markov chain of length 10 million* (Herbst and Schorfheide, 2014, pg. 1088).

To illustrate that PT-MCMC can be used to estimate large scale DSGE models with multimodal posteriors, I also estimate the SW model with the diffuse prior distributions used in Herbst and Schorfheide (2014). These priors are described in Table 8 in Appendix C. The model is estimated on US data over the 1966:I-2004:IV period using the same observables and data used in the original SW paper, and the Dynare mod file for the model was obtained from Johannes Pfeifer's GitHub repo.[12]

[11] Though the solution method and estimation routines in the original SW paper did not permit parameter values which would yield equilibrium indeterminacy.

[12] The data was obtained from the replication files for the paper posted on the American Economic Association's webpage: `https://www.aeaweb.org/articles?id=10.1257/aer.97.3.586`. The mod file can be found at `https://github.com/JohannesPfeifer/DSGE_mod`.

3.2.1 Tuning

Table 3 describes the tuning parameters used for PT-MCMC to estimate the Smets-Wouters model. I use 12 chains with a convex temperature schedule ranging from 0.02 to 1.0. This schedule is generated using equation (11) with $k = 0.25$ and $\gamma = 1.5$. The algorithm is run for 1,000,000 iterations and discards the 250,000 as a warm-up. Each chain is initialized with a draw from the prior distribution, and the initial covariance matrix for the proposal distribution is equal to the prior covariance matrix. After 200,000 iterations, I compute the sample covariance matrix for each chain and use these as the covariance matrices for the proposal density in the remaining iterations. During the warm-up phase, the scaling parameters are also tuned such that roughly 23% of mutation steps are accepted. In this case, because of the size of the model and larger number of data points, I found that it was computationally much more efficient to parallelize the mutation step.

Table 3: PT-MCMC tuning for Smets-Wouters model

Tuning parameters	
N_{chain}	12
ξ	[0.02, 0.07, 0.12, 0.19, 0.26, 0.34, 0.43, 0.43, 0.63, 0.74, 0.85, 1]
N_{iter}	1,000,000
N_{burn}	250,000
Σ	Adaptive covariance

3.2.2 Results

Table 6 reports estimates of the mean parameter values, HPD intervals, marginal data density, and the variability of each of these estimates across 10 different runs of the algorithm. Each run of the algorithm took approximately 8 hours using an AMD Ryzen 9-5900x 12 core processor. Convergence of the Markov chains was assessed using the scale reduction factor (Gelman et al., 2021, pg. 284).

Parallel tempering. PT-MCMC again provides very consistent estimates of the parameters across different runs of the algorithm. For most parameters, the mean estimates are nearly identical across the different runs. Since each run of the algorithm is initialized with draws from the prior distribution, and mean parameter estimates are essentially the same, this suggests that PT-MCMC works well even when the posterior distribution is multimodal.

Effective sample size can again be used to gauge the precision of estimates. But in this case, many of the posterior distributions are far from Gaussian and the method of computing ESS used in the previous section may be misleading. In this case, effective sample size, $\hat{n}_{eff}$, is defined by $\hat{n}_{eff} = \frac{\hat{V}(\theta)}{STD^2}$, where $\hat{V}(\theta)$ is the posterior variance from a given run and

Table 4: Posterior Estimates for Smets-Wouters with Diffuse Priors

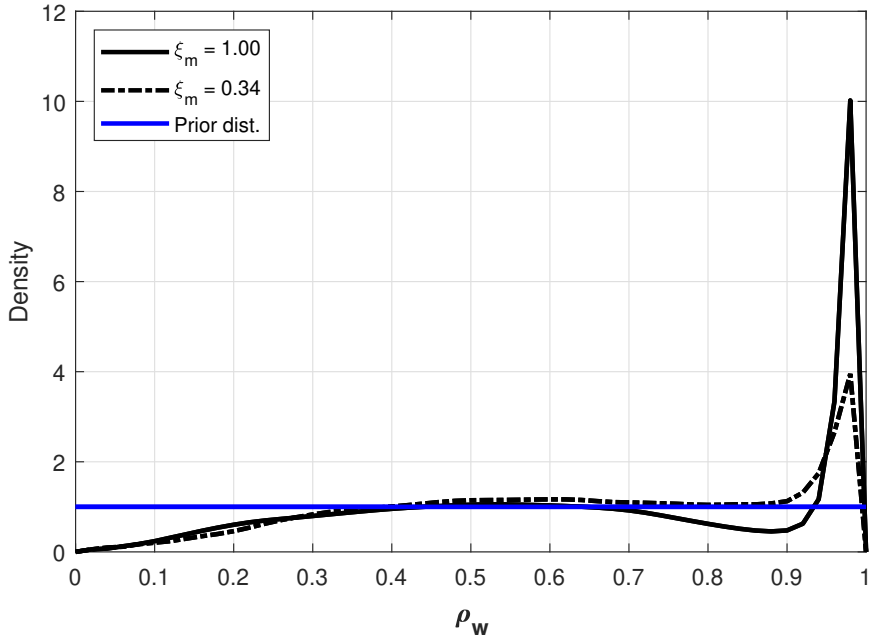
		Parallel tempering		
Parameter	Mean	90% HPD Interval	STD(Mean)	$\hat{n}_{eff}$
φ	8.21	[3.89 , 12.69]	1.51e-01	325
σ_c	1.65	[1.30 , 1.99]	5.42e-03	1551
h	0.69	[0.60 , 0.79]	3.67e-03	307
ξ_w	0.93	[0.84 , 1.00]	4.58e-03	200
σ_l	2.99	[1.12 , 4.76]	4.97e-02	559
ξ_p	0.73	[0.63 , 0.83]	3.64e-03	318
ι_w	0.74	[0.49 , 1.00]	9.31e-03	355
ι_p	0.10	[0.00 , 0.22]	3.92e-03	476
ψ	0.75	[0.56 , 1.00]	5.76e-03	613
Φ	1.70	[1.48 , 1.92]	4.73e-03	806
r_π	2.78	[2.08 , 3.48]	1.10e-02	1529
ρ	0.88	[0.85 , 0.92]	6.65e-04	1170
r_y	0.16	[0.08 , 0.23]	1.41e-03	1180
r_{Δ_y}	0.28	[0.22 , 0.35]	2.43e-03	284
π	0.88	[0.47 , 1.26]	1.32e-02	329
$100(\beta^{-1}-1)$	0.06	[0.00, 0.14]	2.47e-03	590
l	-0.27	[-3.43, 2.98]	9.61e-02	482
γ	0.41	[0.37 , 0.44]	5.45e-04	1441
α	0.17	[0.14 , 0.20]	4.24e-04	2320
ρ_a	0.97	[0.95 , 0.98]	3.99e-04	451
ρ_b	0.21	[0.00 , 0.38]	8.12e-03	336
ρ_g	0.99	[0.98 , 1.00]	1.39e-04	3238
ρ_i	0.72	[0.61 , 0.83]	2.81e-03	559
ρ_r	0.05	[0.00 , 0.12]	1.64e-03	736
ρ_p	0.91	[0.83 , 1.00]	1.96e-03	988
ρ_w	0.68	[0.27 , 1.00]	1.55e-02	329
μ_p	0.75	[0.53 , 1.00]	8.90e-03	368
μ_w	0.62	[0.15 , 0.99]	1.66e-02	357
ρ_{ga}	0.43	[0.26 , 0.61]	4.53e-03	567
σ_a	0.46	[0.41 , 0.51]	8.38e-04	1268
σ_b	0.24	[0.19 , 0.29]	1.75e-03	376
σ_g	0.54	[0.49 , 0.60]	8.06e-04	1631
σ_i	0.46	[0.38 , 0.54]	1.86e-03	764
σ_r	0.24	[0.21 , 0.26]	4.76e-04	1085
σ_p	0.14	[0.08 , 0.20]	3.93e-03	103
σ_w	0.25	[0.21 , 0.29]	9.96e-04	591
Mean(Log MDD)		-879.132		
Std(Log MDD)		0.43		
Computing Time		474 mins		
Average Exchange Rate		0.43		
Average Mutation Rate		0.25		

Notes: Mean estimates and HPD intervals are computed as the mean across 10 runs of the algorithm.

STD^2 is the variance of the posterior mean across the 10 runs of the algorithm. PT-MCMC produces a large ESS, with effective sample sizes of well over 500 for many of the parameters, which is consistent with the precision of the parameter estimates across different runs of the algorithm. Even for the most difficult to estimate parameters, the ESS exceeds 100.

To provide some intuition into why parallel tempering is particularly well suited to handle posteriors which are multimodal, Figure 3 plots the kernel density of the posterior distribution for ρ_w, the autoregressive coefficient for the wage markup, obtained for chains with inverse temperatures equal to 1 and 0.34, along with the prior distribution. The chain with $\xi_m = 1$ is the target posterior density and is represented by the solid black line. The density is multimodal, with a large peak near 1 and a smaller peak closer to 0.5. Due to this multimodality, algorithms like RWMH can struggle to accurately characterize the posterior because of the low-density region between these two modes.

Figure 3: The impact of tempering on the posterior distribution



PT-MCMC is able to accurately characterize the target posterior through its use of the auxiliary chains. As the temperature is increased and ξ_m falls, relatively more weight is placed on the uniform prior in the auxiliary distribution's respective posterior. As a result, the peaks in the posterior distribution are increasingly flattened as the temperature rises, and chains with high temperatures have a much easier time moving between these two modes. The parameter vectors in these chains can then be passed to the target chain via the exchange step, allowing the target chain to frequently visit both modes.

Parameter estimates and economic implications. The posterior parameter estimates are essentially identical to the findings in Herbst and Schorfheide (2014). Relaxing the tight prior distributions used in the original SW paper leads to a substantially higher estimate of the marginal data density compared to the original priors, which is evidence in favour

of the diffuse prior model.[13] The improvement is partly a result of substantial differences in the mean estimates of several structural parameters of the model. Notable differences include Calvo wage stickiness (0.93 here vs. 0.73 in SW), investment adjustment costs (8.21 here vs. 5.74 in SW), Frisch elasticity of labour supply (2.99 here vs. 1.83 in SW), and the Fed's inflation response parameter (2.78 here vs. 2.04 in SW).

But as highlighted by Herbst and Schorfheide (2014), some of these mean estimates reflect two quite distinct modes, especially as it relates to wage and price dynamics. One mode is characterized by extremely sticky wages, moderate price stickiness, very persistent price markup shocks (both AR and MA terms), very high wage indexation, with much lower persistence in wage markup shocks. A second mode features slightly lower wage and price stickiness, price markup persistence (both AR and MA terms), and wage indexation, but very persistent wage markup shocks from both the AR and MA terms. These two modes and their associated log posterior values are reported in Table 5.

Table 5: Two modes of Smets-Wouters with diffuse priors

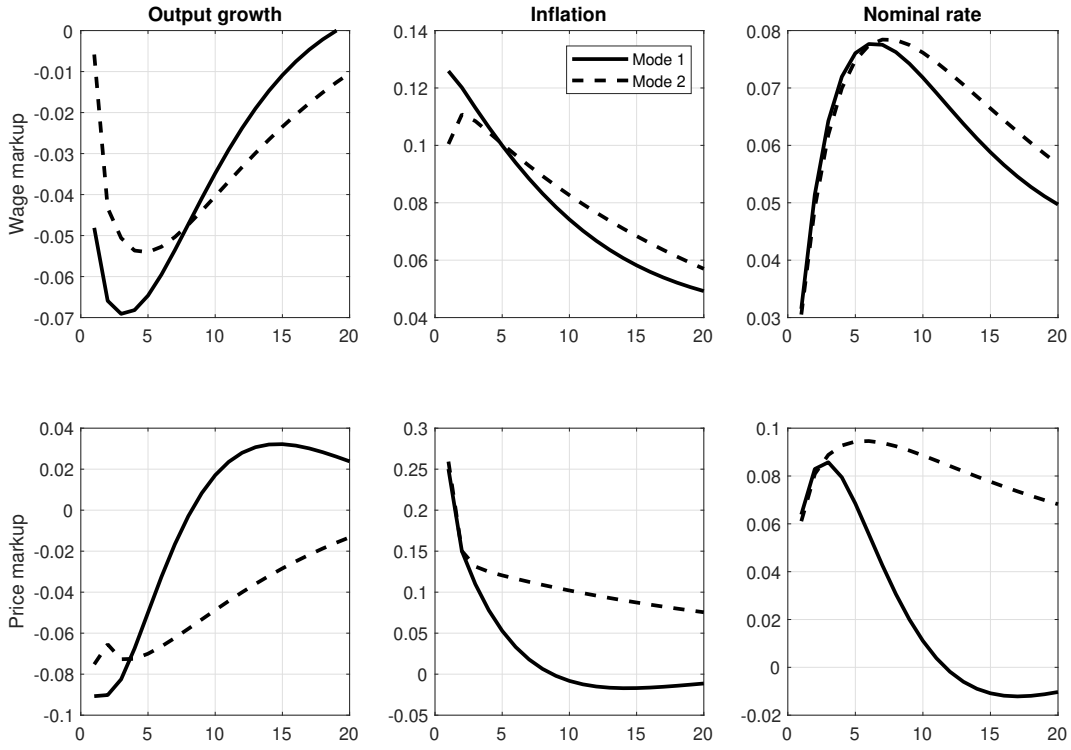
	Mode 1	Mode 2
ξ_p	0.68	0.75
ξ_w	0.87	0.97
μ_p	0.43	0.95
μ_w	0.97	0.19
ρ_p	0.84	0.99
ρ_w	0.99	0.24
ι_p	0.01	0.14
ι_w	0.75	0.99
Log Posterior	-808.93	-808.40

These two modes lead to a very different propagation of markup shocks and explanation of US business cycles. Figure 4 plots the impulse response functions to wage and price markup shocks under the two modes. Under mode 1, wage markup shocks generate more inflation and output movements on impact due to the lower degree of wage stickiness and high persistence. Subsequently, wage markup shocks are the primary contributor to inflation, contributing to around 63% of the unconditional variance. Price markup shocks play a less prominent role, contributing to around 30% of the unconditional variance of inflation.

Under Mode 2, price markup shocks are the dominant driver of inflation fluctuations, contributing to about 58% of the unconditional variance of inflation, while wage markup shocks contribute to around 28%. Inflation is also highly persistent under this mode due to the endogenous dynamics of the model, specifically the high degree of wage indexation to

[13]The exact value of the MDD here should be approached with some caution since it was computed using Geweke's harmonic mean estimator, which is based on the assumption of unimodal posterior.

Figure 4: Impulse response functions to wage and price markup shocks



past inflation. This is evident by the fact that wage markup shocks are more persistent than under mode 1, even though the persistence in the actual shock process is much lower. This also results in all shocks in the model-generating inflation responses which are highly persistent. For reference, the autocorrelation between inflation and inflation 5 quarters previous, Corr(π_t, π_{t-5}), is 0.80 under mode 2, compared to 0.61 under mode 1.

Lastly, these two modes are also likely to reflect very different optimal policy prescriptions. Under mode 2, inflation deviations from target are extremely costly because wages are almost perfectly indexed to the previous period's inflation. Consequently, there is a strong endogenous inertia in inflation from indexation, and returning inflation to target would require a much longer period of below-target economic activity.

3.2.3 Alternative parallel tempering tuning

In the previous section, the number of chains was held fixed at 12. In this subsection, I provide some insight into how changing the number of chains used can impact the performance of the PT-MCMC algorithm. To illustrate this, I use an identical number of iterations but instead use 6 and 18 chains in the ensemble. To generate the temperature ladder, I use the iden-

tical parameters used in the previous section ($k = 0.5$ and $\gamma = 1.5$). This leads to a temperature ladder for the 6 chains given by $\xi = [0.06, 0.18, 0.33, 0.51, 0.72, 1]$ and for 18 chains given by $\xi = [0.01, 0.04, 0.07, 0.10, 0.14, 0.19, 0.24, 0.29, 0.35, 0.41, 0.47, 0.53, 0.60, 0.67, 0.75, 0.82, 0.90, 1]$.

Table 6 reports parameter estimates for a select few parameters which had some variation across different runs in the previous section (i.e., the more challenging ones to estimate). The first three columns report the mean estimates, standard deviation of the mean, and effective sample sizes obtained in the previous section. The last 6 columns report the same outputs for the ensembles with 6 chains and 18 chains. The bottom two rows in the table report the average computing time and accepted exchange rate between neighbouring chains.

Table 6: Posterior Estimates for Smets-Wouters with Diffuse Priors

	12 chains (baseline)			6 chains			18 chains		
Parameter	Mean	STD(Mean)	$\hat{n}_{eff}$	Mean	STD(Mean)	$\hat{n}_{eff}$	Mean	STD(Mean)	$\hat{n}_{eff}$
φ	8.21	1.51e-01	325	8.15	2.11e-01	158	8.10	1.06e-01	631
ρ_p	0.91	1.96e-03	988	0.91	3.83e-03	254	0.91	1.64e-03	1358
ρ_w	0.68	1.55e-02	329	0.68	3.00e-02	86	0.68	7.14e-03	1497
μ_p	0.75	8.90e-03	357	0.75	1.92e-02	80	0.76	5.94e-03	802
μ_w	0.62	1.66e-02	567	0.62	3.30e-02	88	0.62	7.73e-03	1616
Computing Time		474 mins			402 mins			596 mins	
Avg. Exchange		0.43			0.17			0.56	

Notes: Mean estimates for ensembles with 6 and 18 chains are computed as the mean of all mean estimates across 5 runs of the algorithm.

There are a few points from Table 6 worth emphasizing. The first is that even with only 6 chains in the ensemble, the algorithm performs quite well. Mean parameter estimates quite close to those obtained in the baseline case, and the parameters with multimodal posterior distributions, are still captured reasonably well. However, with a smaller number of chains, it becomes more difficult to specify a temperature ladder that yields good exchange acceptance rates. With 6 chains, the average exchange acceptance rates falls to 17%. Consequently, there is a drop-off in the effective sample sizes and more variability in mean estimates across different runs of the algorithm. For example, with 6 chains, the mean estimates for the wage markup parameter μ_w ranged from 0.58 to 0.66.

Second, adding more chains improves the quality of the estimates. Using 18 chains, it becomes significantly easier to specify a temperature ladder that produces reasonable exchange acceptance rates. With 18 chains, the acceptance rate increases to 56% and coincides with a large increase in the effective sample sizes. Additionally, the variability of mean estimates falls for all parameters. In this case, the mean estimates for the wage markup parameter μ_w ranged from 0.61 to 0.63.

Third, when deciding on the number of chains to use, it is important to consider comput-

ing capacity. All computations in this paper were done using an AMD Ryzen 9-5900x 12 core processor. Consequently, moving beyond 12 chains results in a disproportionate slowdown in the average runtime of the algorithm. Relative to the baseline case with 12 chains, reducing the number of chains by 6 reduces the average computing time by roughly 70 minutes, while increasing the number of chains by 6 increases the average computing time by roughly 120 minutes. A general guideline would be to start by using a number of chains equal to the number of available cores and reducing or increasing the number of chains depending on the average exchange acceptance rate, with a target of 50%.

4 Conclusion

In this paper I have proposed a population-based MCMC algorithm known as parallel tempering for estimating DSGE models. PT-MCMC has several appealing properties for macroeconomists. First and foremost, the algorithm is well suited to estimate DSGE models with complex posterior distributions, such as those with discontinuities or multimodal posteriors. Second, the algorithm does not require finding the mode or hessian as a starting point, which can be an error-prone process. Third, the algorithm mixes quickly, meaning that computational time dedicated to a warm-up is reduced and the algorithm produces large effective sample sizes. Fourth, the algorithm is relatively easy to use and operationalize, and it can make efficient use of parallel computing. Estimates herein suggest that a moderate number of chains in the range of 8-12 with a convex temperature ladder worked well for the problems at hand. Further, I provide a MATLAB script which takes a Dynare mod file as an input and implements the proposed algorithm. This makes the routine straightforward to use for DSGE practitioners.

I provided two empirical applications to well-known problems which are typically considered to be challenging to estimate. The first application estimates a New Keynesian model on US data over the 1960-1979 period, allowing for equilibrium indeterminacy. Indeterminacy presents an estimation challenge because the boundary between the determinacy and indeterminacy regions features many discontinuities in the likelihood, making it difficult for algorithms to transition between the two regions. PT-MCMC exploits exchange steps to cross the boundary frequently, and estimates herein suggest that this period of US history was indeed characterized by indeterminacy. This finding is primarily attributed to a low estimated monetary policy response to inflation; however, trend inflation was sufficiently high such that satisfying the Taylor principle would not have ensured equilibrium determinacy.

The second application estimates the model of Smets-Wouters over the period 1966-2004,

but under more diffuse prior distributions. With more uninformative prior distributions, several of the structural parameters feature bimodal posterior distributions, which can present significant challenges for algorithms like RWMH. PT-MCMC provided very consistent estimates across different runs of the algorithm, implying that the algorithm is able to characterize the multimodal posterior extremely well. For a sample of 750,000 draws, the algorithm produced effective sample sizes above 100 for all parameters of the model.

The combination of the algorithm's efficiency, compatibility with the existing Dynare software, and its capability to handle complex posterior distributions with minimal tuning makes PT-MCMC a useful addition to the toolkit of practitioners. These points are particularly salient for those estimating large-scale DSGE models, such as those used in central banks and other policy institutions.

References

Albonico, A., Ascari, G. and Haque, Q.: 2020, The (Ir)Relevance of Rule-of-Thumb Consumers for U.S. Business Cycle Fluctuations, *Economics Discussion / Working Papers 20-26*, The University of Western Australia, Department of Economics.

An, S. and Schorfheide, F.: 2007, Bayesian Analysis of DSGE Models, *Econometric Reviews* **26**(2-4), 113–172.

Ascari, G. and Ropele, T.: 2009, Trend Inflation, Taylor Principle, and Indeterminacy, *Journal of Money, Credit and Banking* **41**(8), 1557–1584.

Ascari, G. and Sbordone, A. M.: 2014, The Macroeconomics of Trend Inflation, *Journal of Economic Literature* **52**(3), 679–739.

Bhattarai, S., Lee, J. W. and Park, W. Y.: 2016, Policy Regimes, Policy Shifts, and U.S. Business Cycles, *The Review of Economics and Statistics* **98**(5), 968–983.

Bianchi, F. and Nicoló, G.: 2021, A generalized approach to indeterminacy in linear rational expectations models, *Quantitative Economics* **12**(3), 843–868.

Cai, M., Del Negro, M., Herbst, E., Matlin, E., Sarfati, R. and Schorfheide, F.: 2020, Online estimation of DSGE models, *The Econometrics Journal* **24**(1), C33–C58.

Campbell, D. A. and Steele, R. J.: 2012, Smooth functional tempering for nonlinear differential equation models, *Statistics and Computing* **22**, 429–443.

Chib, S. and Ramamurthy, S.: 2010, Tailored randomized block MCMC methods with application to DSGE models, *Journal of Econometrics* **155**, 19–38.

Childers, D., Fernández-Villaverde, J., Perla, J., Rackauckas, C. and Wu, P.: 2022, Differentiable State-Space Models and Hamiltonian Monte Carlo Estimation, *Working Paper 30573*, National Bureau of Economic Research.

Clarida, R., Galí, J. and Gertler, M.: 2000, Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory, *Quarterly Journal of Economics* **115**, 147–180.

Coibion, O. and Gorodnichenko, Y.: 2011, Monetary policy, trend inflation and the Great Moderation: An alternative interpretation, *American Economic Review* **101**(1), 341–370.

Corrigan, P., Desgagnés, H., Dorich, J., Lepetyuk, V., Miyamoto, W. and Zhang, Y.: 2021, ToTEM III: The Bank of Canada's Main DSGE Model for Projection and Policy Analysis, *Technical Reports 119*, Bank of Canada.

Creal, D.: 2007, Sequential Monte Carlo samplers for Bayesian DSGE models, *Technical report*, University of Chicago Booth.

Cúrdia, V. and Reis, R.: 2010, Correlated Disturbances and U.S. Business Cycles, *NBER Working Papers 15774*, National Bureau of Economic Research, Inc.

Del Negro, M. and Schorfheide, F.: 2013, DSGE Model-Based Forecasting, *in* G. Elliott, C. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Vol. 2 of *Handbook of Economic Forecasting*, Elsevier, chapter 0, pp. 57–140.

Farkas, M. and Tatar, B.: 2020, Bayesian estimation of DSGE models with Hamiltonian Monte Carlo, *IMFS Working Paper Series 144*, Goethe University Frankfurt, Institute for Monetary and Financial Stability (IMFS).

Fernández-Villaverde, J., Rubio-Ramírez, J. and Schorfheide, F.: 2016, Chapter 9 – Solution and Estimation Methods for DSGE Models, Vol. 2 of *Handbook of Macroeconomics*, Elsevier, pp. 527–724.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D.: 2021, *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.

Geweke, J.: 1999, Using simulation methods for Bayesian econometric models: inference, development,and communication, *Econometric Reviews* **18**(1), 1–73.

Geyer, C. J.: 1991, Markov Chain Monte Carlo Maximum Likelihood.

Herbst, E. P. and Schorfheide, F.: 2016, *Bayesian Estimation of DSGE Models*, number 10612 in *Economics Books*, Princeton University Press.

Herbst, E. and Schorfheide, F.: 2014, Sequential Monte Carlo sampling for DSGE models, *Journal of Applied Econometrics* **29**(7), 1073–1098.

Hirose, Y., Kurozumi, T. and Van Zandweghe, W.: 2020, Monetary policy and macroeconomic stability revisited, *Review of Economic Dynamics* **37**, 255–274.

Jasra, A., Stephens, D. A. and Holmes, C. C.: 2007a, Population-Based Reversible Jump Markov Chain Monte Carlo, *Biometrika* **94**(4), 787–807.

Jasra, A., Stephens, D. and Holmes, C.: 2007b, On population-based simulation for static inference, *Statistics and Computing* **17**, 263–279.

Khan, H., Phaneuf, L. and Victor, J. G.: 2020, Rules-based monetary policy and the threat of indeterminacy when trend inflation is low, *Journal of Monetary Economics* **114**(C), 317–333.

Liu, J.: 2001, *Monte Carlo Strategies in Scientific Computing*, Springer Series in Statistics, Springer.

Lubik, T. A. and Schorfheide, F.: 2003, Computing sunspot equilibria in linear rational expectations models, *Journal of Economic Dynamics and Control* **28**(2), 273–285.

Lubik, T. A. and Schorfheide, F.: 2004, Testing for indeterminacy: An application to U.S. monetary policy, *American Economic Review* **94**(1), 190–217.

Schorfheide, F.: 2000, Loss function-based evaluation of DSGE models, *Journal of Applied Econometrics* **15**(6), 645–670.

Smets, F. and Wouters, R.: 2007, Shocks and frictions in US business cycles: A Bayesian DSGE approach, *American Economic Review* **97**(3), 586–606.

Strid, I., Giordani, P. and Kohn, R.: 2010, Adaptive hybrid Metropolis-Hastings samplers for DSGE models, *SSE/EFI Working Paper Series in Economics and Finance 724*, Stockholm School of Economics.

Appendix

A New Keynesian model with equilibrium indeterminacy

The log-linearized equations of this model take the following form:

$$\begin{aligned}\tilde{y}_t =& \frac{h}{h+g_A}\left(\tilde{y}_{t-1}-\tilde{g}_{A,t}\right)+\frac{g_A}{h+g_A}\mathbb{E}_t\left(\tilde{y}_{t+1}+\tilde{g}_{A,t+1}\right)\\ &-\frac{g_A-h}{h+g_A}\left(\tilde{i}_t-\mathbb{E}_t\tilde{\pi}_{t+1}-\tilde{b}_t+\mathbb{E}_t\tilde{b}_{t+1}\right)\end{aligned} \tag{12}$$

$$\begin{aligned}\tilde{\pi}_t =&\beta[1+\epsilon(1-\xi_p\pi^{\epsilon-1})(\pi-1)]\mathbb{E}_t\tilde{\pi}_{t+1}+\beta(1-\xi_p\pi^{\epsilon-1})(\pi-1)\mathbb{E}_t\tilde{x}_{1,t+1}\\ &+\left(\frac{(1-\xi_p\pi^{\epsilon-1})(1-\beta\xi_p\pi^{\epsilon})}{\xi_p\pi^{\epsilon-1}}\right)((1+\eta)\tilde{y}_t+\eta\tilde{v}_t^p)+\beta(1-\pi)(1-\xi_p\pi^{\epsilon-1})\tilde{b}_t\\ &+\left(\frac{(1-\beta\xi_p\pi^{\epsilon-1})(1-\xi_p\pi^{\epsilon-1})}{\xi_p\pi^{\epsilon-1}}\right)\left(\frac{h}{g_A-h}\right)\left(\tilde{y}_t-\tilde{y}_{t-1}+\tilde{g}_{A,t}\right)\end{aligned} \tag{13}$$

$$\tilde{x}_{1,t}=(1-\beta\xi_p\pi^{\epsilon})(\tilde{b}_t+(1+\eta)\tilde{y}_t+\eta\tilde{v}_t^p)+\beta\xi_p\pi^{\epsilon}\mathbb{E}_t[\tilde{x}_{1,t+1}+\epsilon\tilde{\pi}_{t+1}] \tag{14}$$

$$\tilde{v}_t^p=\frac{\epsilon\xi_p\pi^{\epsilon-1}(\pi-1)}{1-\xi_p\pi^{\epsilon-1}}\tilde{\pi}_t+\pi^{\epsilon}\xi_p\tilde{v}_{t-1}^p \tag{15}$$

$$\tilde{y}_t^n=\frac{h}{(1+\eta)g_A-h\eta}\left(\tilde{y}_{t-1}^n-\tilde{g}_{A,t}\right) \tag{16}$$

$$\tilde{i}_t =\rho_i\tilde{i}_{t-1}\quad+(1-\rho_i)\left(\phi_\pi\tilde{\pi}_t+\phi_{\Delta y}(\tilde{y}_t-\tilde{y}_{t-1}+\tilde{g}_{A,t})+\phi_y\tilde{x}_t\right)+\tilde{v}_t^r, \tag{17}$$

where $\tilde{x}_t=\tilde{y}_t-\tilde{y}_t^n$. The exogenous shocks evolve according to

$$\tilde{g}_{A,t}=\rho_z\tilde{g}_{A,t-1}+\epsilon_t^z \tag{18}$$

$$\tilde{b}_t=\rho_b\tilde{b}_{t-1}+\epsilon_t^b \tag{19}$$

$$\tilde{v}_t^r = \rho_v \tilde{v}_{t-1}^r + \epsilon_t^v. \tag{20}$$

The equation pinning down determinacy/indeterminacy and the sunspot shock is given by

$$\tilde{\omega}_t = \frac{1}{\alpha_{BN}} \tilde{\omega}_{t-1} + \epsilon_t^s - (\tilde{\pi}_t - \mathbb{E}_{t-1}\tilde{\pi}_t). \tag{21}$$

B Discontinuities in the likelihood function in a New Keynesian model with indeterminacy

Estimating a New Keynesian model with indeterminacy using the Bianchi and Nicoló (2021) approach can be challenging because for many parameter configurations around the boundary between determinacy and indeterminacy, the likelihood function is discontinuous. This can be of the result of (1) A draw where the structural parameters of the model are consistent with determinacy but α_{BN} is consistent with indeterminacy; or (2) A draw where the structural parameters are consistent with indeterminacy but α_{BN} is consistent with determinacy. As a result, it can be difficult for algorithms like RWMH to make the leap from one region to another.

Figure 5: Log likelihood in the determinacy and indeterminacy regions

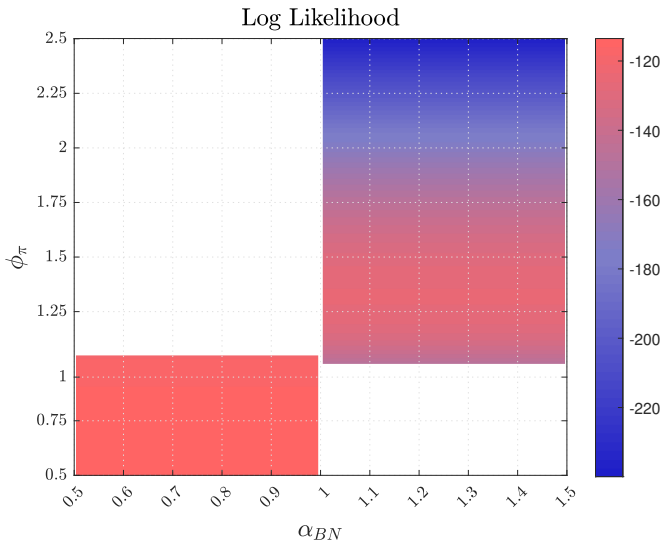


Figure 5 illustrates this challenge. The figure plots the log likelihood for different parameter values for the the central bank's response to inflation, ϕ_π, and for the parameter which governs whether the model solution is characterized by indeterminacy, α_{BN}. The remaining parameters are held fixed at the posterior mean conditional on determinacy and indeterminacy from a run of the algorithm. If $\alpha_{BN} < 1$ and the central bank's inflation response parameter is sufficiently low, the model is characterized by indeterminacy. The resulting likelihood values are in the bottom-left quadrant, and this region contains the global maximum. If $\alpha_{BN} > 1$ and the central bank's inflation response parameter is sufficiently high, the model is characterized by determinacy. The likelihood values for this region appear in the upper-right quadrant of the figure. The white space in the figure is draws of parameters for which there is no unique model solution and hence no value for the likelihood. Since there exists only a narrow pathway from the determinacy region to the indeterminacy region (and

vice versa), algorithms which only make small moves in the parameter space can often have a difficult time crossing from one region to another. To make matters more complicated, the exact boundary between the two regions moves around depending on other structural parameters when the model has positive trend inflation.

C Additional results

Table 7: Prior distributions for small-scale New Keynesian model

		Prior distributions		
Parameter	Domain	Density	Para(1)	Para(2)
σ_b	$\mathbb{R}^+$	InvGamma	0.1	2
σ_z	$\mathbb{R}^+$	InvGamma	0.1	2
σ_v	$\mathbb{R}^+$	InvGamma	0.1	2
σ_s	$\mathbb{R}^+$	InvGamma	0.1	2
$\rho_{z,s}$	[-1,1]	Uniform	-1	1
$\rho_{b,s}$	[-1,1]	Uniform	-1	1
$\rho_{v,s}$	[-1,1]	Uniform	-1	1
h	[0,1)	Beta	0.7	0.15
ζ_p	[0,1)	Beta	0.66	0.05
ρ_i	[0,1)	Beta	0.5	0.2
ϕ_π	$\mathbb{R}^+$	Normal	1.5	0.5
$\phi_{\Delta y}$	$\mathbb{R}^+$	Normal	0.125	0.05
ϕ_y	$\mathbb{R}^+$	Normal	0.125	0.05
ρ_z	[0,1)	Beta	0.5	0.2
ρ_b	[0,1)	Beta	0.5	0.2
ρ_v	[0,1)	Beta	0.5	0.2
$\bar{\pi}$	$\mathbb{R}^+$	Normal	0.75	0.25
$\bar{i}$	$\mathbb{R}^+$	Normal	1.50	0.25
$\bar{g}_A$	$\mathbb{R}^+$	Normal	0.40	0.10
α_{BN}	[0.5,1.5]	Uniform	0.5	1.5

Notes: Para(1) and Para(2) refer to the prior means and standard deviations. For the Uniform densities, Para(1) and Para(2) refer to the lower and upper bounds.

Table 8: Diffuse prior distributions for Smets-Wouters (2007)

		Prior distributions		
Parameter	Domain	Density	Para(1)	Para(2)
φ	$\mathbb{R}^+$	Normal	4.00	4.50
σ_c	$\mathbb{R}^+$	Beta	1.50	1.11
h	[0,1)	Beta	0.00	1.00
ξ_w	[0,1)	Normal	0.00	1.00
σ_l	$\mathbb{R}^+$	Normal	2.00	2.25
ξ_p	[0,1)	Normal	0.00	1.00
ι_w	[0,1)	Normal	0.00	1.00
ι_p	[0,1)	Normal	0.00	1.00
ψ	[0,1)	Uniform	0.00	1.00
Φ	[0,1)	Normal	1.25	0.36
r_π	[1,20]	Normal	1.50	0.75
ρ	[0,1)	Uniform	0.00	1.00
r_y	$\mathbb{R}^+$	Normal	0.12	0.15
r_{Δ_y}	$\mathbb{R}^+$	Normal	0.12	0.15
π	$\mathbb{R}^+$	Gamma	0.625	0.30
$100(\beta^{-1}-1)$	[0.5,1.5]	Gamma	0.25	0.30
l	$\mathbb{R}^+$	Normal	0.00	6.00
γ	[-1,1]	Normal	0.40	0.30
α	[-1,1]	Normal	0.30	0.15
ρ_a	[0,1)	Uniform	0.00	1.00
ρ_b	[0,1)	Uniform	0.00	1.00
ρ_g	[0,1)	Uniform	0.00	1.00
ρ_i	[0,1)	Uniform	0.00	1.00
ρ_r	[0,1)	Uniform	0.00	1.00
ρ_p	[0,1)	Uniform	0.00	1.00
ρ_w	[0,1)	Uniform	0.00	1.00
μ_p	[0,1)	Uniform	0.00	1.00
μ_w	[0,1)	Uniform	0.00	1.00
ρ_{ga}	[-1,1]	Uniform	0.00	1.00
σ_a	$\mathbb{R}^+$	InvGamma	0.10	2.00
σ_b	$\mathbb{R}^+$	InvGamma	0.10	2.00
σ_g	$\mathbb{R}^+$	InvGamma	0.10	2.00
σ_i	$\mathbb{R}^+$	InvGamma	0.10	2.00
σ_r	$\mathbb{R}^+$	InvGamma	0.10	2.00
σ_p	$\mathbb{R}^+$	InvGamma	0.10	2.00
σ_w	$\mathbb{R}^+$	InvGamma	0.10	2.00

Notes: Para(1) and Para(2) refer to the prior means and standard deviations. For the Uniform densities, Para(1) and Para(2) refer to the lower and upper bounds.