Numerical Performance of MCMC Algorithms for Classical Estimation

Anna Kormiltsina[*]and Denis Nekipelov[†]

November 18, 2009

Abstract

In this paper, we study the performance of the Markov Chain Monte Carlo algorithms in estimation of the classical extremum problems. We find that even in well-defined problems, the stability of Markov chains depends on the parameter support and prior distribution. A possible implication is the inferior quality of Markov chains when the guess for the starting point of the MCMC procedure is inadequate. We estimate a simple model using artificial datasets to illustrate that even in a model as simple as ours, the problem with the starting value exists and displays more severely in smaller samples.

[*]Southern Methodist University, e-mail: annak@smu.edu.

[†]UC Berkeley, e-mail: nekipelov@econ.berkeley.edu.

[0]We would like to thank Han Hong and the participants of the All-UC Econometrics Conference in Riverside, CA for valuable comments.

1 Introduction

With the formal introduction of the Laplace type estimators (LTE) by Chernozhukov and Hong (2003), the MCMC approach to classical estimation problems has become increasingly popular. LTEs are attractive for practitioners, because they allow a straightforward frequentist interpretation: point estimates and standard errors of structural parameters can be produced directly as means and standard deviations of Markov chains. Despite these favorable properties, there are pitfalls a researcher should be aware of. As we argue in this paper, even in well-defined structural models the MCMC procedures may fail to produce adequate point parameter estimates. We derive a sufficient condition for the stability of Markov chains generated by the MCMC procedures and show that the parameter support and the choice of prior distributions are critical in producing quality Markov chains.

To illustrate our results, we estimate a simple Dynamic stochastic general equilibrium (DSGE) model using simulated datasets and show that with "flat" priors, parameter estimates depend on a starting value of the Hastings Metropolis algorithm. We also develop a test for a violation of regularity conditions and demonstrate that the set of starting points iducing instability decreases with the number of observations. This indicates that the problem of instability of Markov chain may be more severe in shorter samples. The test suggests that if it is feasible to compute the mean-square Hessian of the quasi-likelihood function in the finite set of points of the parameter space, then a proper proposal density and the starting point of the MCMC procedure can be found such that guarantee the minimum possibility of instability of markov chains.

An attractive feature of the MCMC approach is its reliance on the Bayesian inference for models, for which a quick computation of the likelihood function is not feasible (see (Geweke 1999) and (Chib 2001)). A standard MCMC procedure applied to classical estimation problems requires the formulation of the quasi-likelihood function using the original classical objective function. Hastings-Metropolis algorithms are then used to estimate the structural parameters by sampling from this non-standard distribution. Acceptable performance of the algorithm may only be achieved if a model is identified and the resulting Markov chains are reversible. In this case, the sampling distribution converges to the stationary target distribution. Our theory allows one to study the convergence of sampling distributions in the case of the random walk Hastings-Metropolis algorithm used for estimation of identified models. Such an analysis is complicated for discrete models, because it requires one to solve a non-linear difference equation. A more convenient way to study the stability of the Hastings-Metropolis Monte-Carlo chains is to consider the sampling process in continuous time ((Gelfand and Mitter 1993), (Roberts and Tweedie 1996), (Roberts, Gelman, and Gilks 1997), and (Roberts and Rosenthal 2007), among others). When the number of draws of the Monte-carlo chain approaches to infinity, the process of normalized sample means of the Hastings-Metropolis draws can be described by the Langevin diffusion and thus belong to the class of the so-called Langevin diffusion algorithms (Robert and Casella (2004)). The Langevin diffusion process L_t evolves according to the stochastic differential equation:

$$dL(t) = \frac{1}{2}\nabla \log\, f\left(L(t)\right) dt + dW(t),$$

where $W(t)$ is the Brownian motion, and under appropriate regularity conditions, f is the stationary distribution of the solution $L(t)$. In this paper, we show that the convergence to a stationary distribution of MCMC chains generated by the random walk Hastings-Metropolis algorithm is related to the Lyapunov stability of the corresponding continuous time stochastic dynamic system.

In general the convergence properties of the estimation algorithm will depend on the properties of the objective function which will determine the quality of parameter identification. The problem of weak identification in the parametric setting has been studied in the theoretical econometric literature. Our results are supportive of this literature suggesting that models with weak identification will be hard to estimate with MCMC methods because our sufficient convergence conditions can be violated. However, our results regarding the convergence of the estimation procedure are more general, because we demonstrate that problems may arise even in models without identification problems. We emphasize that the quality of estimates obtained with MCMC procedures may not be necessarily related to identification issues.

From the practical viewpoint, this paper justifies the strategy to search for a "good" starting point of the Hastings-Metropolis algorithm. From the theoretic viewpoint, given appropriate choices of proposal and prior distribution, the choice of the starting values should not have a large impact on the convergence of the MCMC algorithm. However, the estimates produced with the MCMC algorithms often turn out to be sensitive to the choice of a starting point. For instance, An and Schorfheide (2007), suggest the choice of an appropriate starting point as the first step of the Bayesian estimation algorithm. In this paper we provide some theoretical justification for this strategy.

We emphasize that the poor quality of the LTE estimator that may result from the convergence problems that can be intrinsic to the MCMC algorithm rather than a model. However, it is not the goal of this paper to recommend the researchers to refrain from using the MCMC approach. In principal, other algorithms that can be used for finding the extremum estimators can have more complicated problems.

The idea of this paper is closely related to the research on asymptotic convergence of the empirical posterior distribution of MCMC draws (see the summary of this research in Meyn and Tweedie (2009)). We emphasize, however, that this paper has a different focus by treating the MCMC algorithm as a stochastic method to estimate parameters of a structural model rather than the entire posterior parameter distribution. While the Bayesian approach provides conditions that guarantee convergence of the distribution of draws to the target distribution, our theory is less restrictive, as it only looks for the conditions of convergence of estimates to a minimizer of an empirical objective function. Even though in simple cases our theory may suffice for the convergence of the resulting distribution, it is not expected to imply the convergence of distribution in more complicated likelihood structures.

The remainder of this paper is organized as follows. Section 2 introduces the theory of stability of MCMC chains. In Section 3, we apply this theory to the random walk Hastings-Metropolis algorithm. Section 4, presents a quantitative exercise to accompany the theoretical results. Section 5 concludes.

2 Stochastic Stability of a General Diffusion-Driven Stochastic Process

Consider a general diffusion-driven dynamic stochastic process $\theta(t)$, with the dynamics given by

$$d\,\theta(t) = \frac{1}{2}\nabla \log\, f(\theta(t))\, dt + G(t,\, \theta(t))\, dw(t), \tag{1}$$

where $t \geq 0$, $\theta(t)$ is $k \times 1$, and $f(\theta(t))$ and $G(t, \theta(t))$ are a drift and diffusion coefficients respectively. Following Gikhman and Skorokhod (2004), we impose the following set of assumptions on the dynamic behavior of this process to ensure that (1) has a unique solution in the family of non-anticipating stochastic processes on $[0, T]$:

Assumption 1 *Assume that*

- *System (1) has a unique equilibrium* $\theta^* \in int(\Theta) \subset \mathbb{R}^k$*, i.e.*

$$\frac{1}{2}\nabla \log f\ (\theta^*) = 0 \quad \textit{and} \quad G\,(t, \theta^*) = 0;$$

- $\log f(\theta)$ *is bounded and Lipschitz continuous in* $\theta_t \in \Theta \subseteq \mathbb{R}^k$*;*

- $G(t,\, \theta)$ *is Lipschitz continuous in* θ *with a Lipschitz constant* $L \geq 0$*;*

- *For all* $\theta \in \Theta$ *and* $t \in [0,\, T]$ *there exists a constant* $K > 0$ *such that*

$$\|G\,(t,\, \theta)\,\| \leq K\,(1 + \|\theta\|)\,;$$

- $\theta(0)$*, the starting value of the process described by (1) is a second-order random vector independent of the family of* σ*-algebras* $\mathcal{F}_t$ *generated by the Brownian motion* $w(t)$ *for* $t \in [0,\, T]$*.*

Using the analogy with deterministic dynamic systems, we introduce the following definitions of stochastic stability following Hasminskii (1980):

Definition 1 *Assume that* θ^* *in* $\Theta \subset \mathbb{R}^k$ *is a unique equilibrium of (1), and* $\{\theta(t)\}_{t=0}^{\infty}$ *is the stochastic system described by (1). Then* θ^* *is stochastically stable, if for all* $\epsilon > 0$

$$P_{\mathcal{F}_t}\left\{\sup_{t\in\mathbb{R}_+} |\theta(t) - \theta^*| \geq \epsilon \,\middle|\, \theta(0)\right\} \xrightarrow{p} 0,$$

for all $\theta(0) \xrightarrow{p} \theta^*$*. Otherwise, the equilibrium is stochastically unstable.*

Equilibrium point θ^* is locally asymptotically stochastically stable, if it is stochastically stable and

$$P_{\mathcal{F}_t}\left\{\lim_{t\to+\infty}|\theta(t)-\theta^*|=0\,\middle|\,\theta(0)\right\}\xrightarrow{p}1,$$

for all $\theta(0)\xrightarrow{p}\theta^*$. Finally, equilibrium point θ^* is globally stochastically asymptotically stable relative to set $\overline{\Theta}\ni\theta^*$, if it is asymptotically stochastically stable and

$$P_{\mathcal{F}_t}\left\{\lim_{t\to+\infty}|\theta(t)-\theta^*|=0\,\middle|\,\theta(0)\right\}=1,$$

for all $\theta(0)\in\overline{\Theta}$.

The definition of local stability means that for all processes with starting points approaching the equilibrium θ^*, the probability that the stochastic process leaves an arbitrary small neighborhood of equilibrium approaches zero. The definition of asymptotic stability strengthens the notion of stochastic stability by imposing further that a locally stable stochastic process approach the equilibrium with probability tending to one, when a starting point is tending to θ^*. The requirement for a stochastic process (1) to be globally asymptotically stable is that besides being stable, all processes described by the system (1) find themselves in equilibrium with probability 1, when a starting point belongs to some set $\overline{\Theta}$. Global stability is defined relative to a subset of starting values, which means that while a stochastic process may be globally unstable for some starting values, it may still be considered globally stable relative to another set of starting values. Because ultimately we are interested to establish conditions for convergence of LTEs from different starting points, the the concept of global stochastic stability is the most relevant one for this paper.

To approach the problem, we rely on the concept of Lyapunov stability extended to stochastic problems by the following theorem from Hasminskii (1980). With this in mind, we first introduce a relevant definition of a positive-definite in Lyapunov's sense function.

Definition 2 *A function $v(t,\theta)$ is said to be positive definite in Lyapunov sense in a neighborhood of $\theta=\theta^*$, if $v(t,\theta^*)=0$, and in this neighborhood $v(t,\theta)>w(\theta)$, where $w(\theta)>0$, for $\theta\neq\theta^*$.*

Theorem 1 summarizes the result of the global stability of the stochastic system (1):

Theorem 1 *Suppose the process described by system (1) satisfies assumption 2, and there exists a positive definite in Lyapunov's sense function* $v(t,\theta) : \mathbb{R}_+ \times \overline{\Theta} \to \mathbb{R}$, *which is twice continuously differentiable with respect to* θ, *and once continuously differentiable with respect to* t *everywhere except possibly at the equilibrium point* θ^*, *such that*

$$\mathcal{L}v = \frac{\partial v(t,\theta)}{\partial t} + \sum_i \frac{1}{2} \nabla \log f_i(\theta) \frac{\partial v(t,\theta)}{\partial \theta^i} + \frac{1}{2} \sum_{i,j} \{G(t,\theta) G(t,\theta)\}_{i,j} \frac{\partial^2 v(t,\theta)}{\partial \theta^i \partial \theta^j} < 0,$$

for all $(t, \theta) \in \mathbb{R}_+ \times \overline{\Theta}$; *moreover,* $v(t,\theta)$ *has an infinitesimal upper limit, i.e.*

$$\lim_{\theta \to \theta^*} \sup_{t>0} v(t,\theta) = 0,$$

and be radially unbounded, i.e.

$$\lim_{|\theta| \to \infty} \inf_{t>0} v(t,\theta) = \infty.$$

Then the equilibrium point θ^* *is globally asymptotically stochastically stable relative to set* $\overline{\Theta}$.

The result in Theorem 1 can be used to test for convergence of an MCMC algorithm with a starting point in the set $\overline{\Theta}$.[1] Because for the limiting diffusion process of the Hastings-Metropolis algorithm, $G(\cdot,\cdot)$ in (1) does not depend on the number of draws, the Lyapunov function $v(t,\theta)$ is independent of time (Gikhman and Skorohod (1977)). Thus, $\frac{\partial v(t,\theta)}{\partial t} = 0$, and under the conditions of Theorem 1 about $v(t,\theta)$, the limiting diffusion process of the Hastings Metropolis MCMC algorithm is stable if the following functional form is negative-definite:

$$\begin{aligned} &\sum_i \nabla \log f_i(\theta) \tfrac{\partial v(\theta)}{\partial \theta^i} + \sum_{i,j} \{G(t,\theta) G(t,\theta)'\}_{i,j} \tfrac{\partial^2 v(\theta)}{\partial \theta^i \partial \theta^j} \\ &= \nabla \log f(\theta)' \nabla v(\theta) + e\left(G(t,\theta) G(t,\theta)' * \tfrac{\partial^2 v(\theta)}{\partial \theta \partial \theta'} \right) e. \end{aligned} \tag{2}$$

In this formula, e is a unit vector and $*$ denotes the Hadamard multiplication.

[1] Provided a Lyapunov function for the stochastic process characterizing the Hastings-Metropolis can be constructed.

We complement the statement of Theorem 1 by the following theorem that provides sufficient conditions for instability:

Theorem 2 *Suppose system (1) satisfies assumption 2, and for some* $\overline{\Theta} \in \Theta$ *there exists a function* $v(t,\theta) : \mathbb{R}_+ \times \overline{\Theta} \to \mathbb{R}$ *which is twice continuously differentiable with respect to* θ*, and once continuously differentiable with respect to* t *everywhere except possibly at the equilibrium point* θ^**, such that for any* $\epsilon > 0$*,*

$$\sup_{t\in\mathbb{R}_+,\,\theta\in\overline{\Theta}\backslash B_\epsilon(\theta^*)} \mathcal{L}v = \frac{\partial v(t,\theta)}{\partial t} + \sum_i \frac{1}{2}\nabla \log f_i(\theta)\frac{\partial v(t,\theta)}{\partial \theta^i} + \frac{1}{2}\sum_{i,j}\{G(t,\theta)\,G(t,\theta)\}_{i,j}\frac{\partial^2 v(t,\theta)}{\partial\theta^i\partial\theta^j} < 0,$$

for $B_\epsilon(\theta^*) = \left\{\theta \in \overline{\Theta} \,\middle|\, \|\theta - \theta^*\| < \epsilon\right\}$*, and*

$$\lim_{\theta\to\theta^*}\inf_{t>0} v(t,\theta) = \infty.$$

Then the equilibrium point θ^* *is asymptotically stochastically unstable, and*

$$P_{\mathcal{F}_t}\left\{\sup_{t\in\mathbb{R}_+}\|\theta_t\| < \rho \,\middle|\, \theta(0)\right\} = 0,$$

for all $\theta(0) \in \overline{\Theta}$ *and* $0 < \rho < diam\left(\overline{\Theta}\right)$*.*

The proof of the theorem can be found in Hasminskii (1980) or Kushner and Dupuis (2001).

To prove the instability of a diffusion process described by Equation (1) with the help of Theorem 2, the positive definite Lyapounov function must approach infinity in the neighborhood of the equilibrium. If such a Lyapunov function can be found, and if there exists at least one $\theta \in \Theta$ such that the following condition holds,

$$\left\{\frac{\partial v(\theta,\,t)}{\partial t} + \sum_i \frac{1}{2}\nabla \log f_i(\theta)\frac{\partial v(t,\theta)}{\partial\theta^i} + \frac{1}{2}\sum_{i,j}\{G(t,\theta)\,G(t,\theta)\}_{i,j}\frac{\partial^2 v(t,\theta)}{\partial\theta^i\partial\theta^j}\right\} < 0,$$

then the diffusion process will not stabilize around the equilibrium point θ^* with probability 1, for any starting point of the process $\theta(0)$.

3 Stability of the Hastings-Metropolis Algorithm with a General Likelihood Structure

We apply the theory developed in Section 2 to the Laplace type estimator (LTE) suggested by Chernozhukov and Hong (2003). This choice is motivated by the fact that the LTE is based on MCMC and Hastings-Metropolis algorithms, and thus the convergence of the estimator is closely related to the stability properties of resulting Markov chains. In addition, the LTE encompasses a broad range of extremum estimation problems. Moreover, the conclusions we make regarding the convergence of LTEs can be useful for Bayesian estimation problems, since the likelihood function in the Bayesian approach can just be considered as a special case of a distance function for the LTE problem.

The LTE defines a (quasi-)likelihood function as the exponential transformation of the moment condition or a distance function $L_N(\theta)$, $e^{L_N(\theta)}$, where N is the number of observations and θ is a $(k \times 1)$ element of the parameter space Θ. The quasi-posterior distribution of the parameter θ is then defined as

$$p_N(\theta) = \frac{e^{L_N(\theta)}\pi(\theta)}{\int_\Theta e^{L_N(\theta)}\pi(\theta)d\theta}, \tag{3}$$

where $\pi(\theta)$ is the weighting function ("prior" distribution).[2] The LTE can then be defined as the mean or a particular quantile of a random variable with the distribution (3). Chernozhukov and Hong show that under some regularity conditions the LTE possesses the same asymptotic properties as the standard classical estimator and is as efficient asymptotically. They also show that under mild regularity conditions, the asymptotically valid parameter estimates and their confidence intervals can be easily obtained as appropriate quantiles of the distribution $p_N(\theta)$.

Because the quasi-posterior distribution in Equation (3) is non-standard, Chernozhukov and Hong (2003) suggest a standard approach to describe it, which consists in generating MCMC chains using a Hastings-Metropolis algorithm. The standard random-walk Hastings-Metropolis

[2]The difference between the LTE and the pure Bayesian approach is that in deriving the posterior distribution, the latter relies on the likelihood function, while the quasi-likelihood function in the LTE approach can be derived from any statistical moment condition or distance function.

algorithm generates a sequence $\{\theta_t\}_{t=1}^T$ according to the process

$$\theta_t = \theta_{t-1} + \mathbf{1}(u_t < \rho_t)\sigma\epsilon_t,$$

where $\mathbf{1}$ is the indicator function, u_t is *i.i.d.* $U[0\,,1]$, ϵ_t is *i.i.d.* $\mathcal{N}(0,1)$, $\sigma > 0$, and $\rho_t = min\{1, \frac{e^{L_N(\theta_{t-1}+\sigma\epsilon_t)}\pi(\theta_{t-1}+\sigma\epsilon_t)}{e^{L_N(\theta_{t-1})}\pi(\theta_{t-1})}\}$. The LTE can then obtained as a simple mean of this Markov chain.

Although Chernozhukov and Hong (2003) acknowledge the fact that a poor choice of a starting value may result in a slower convergence of MCMC chains, they do not elaborate on this problem. In this paper, we demonstrate that a poor choice of the starting value may not only slow down convergence, but may also lead to instability of Markov chains, and as a result produce diverging LTEs. When the number of draws approaches infinity, the process of cumulative means $\{\widehat{\theta}_t\}_{t=1}^T$, where $\widehat{\theta}_t = \frac{1}{\sqrt{t}}\sum_{i=1}^t \theta_i$, can be approximated by the diffusion process as follows:

$$d\,\theta(t) = \frac{1}{2}\nabla\, \log\, p_N\, \left(\theta(t)\right)\, dt + dW(t). \tag{4}$$

Although the stability of this process does not necessarily imply the stability of a Hastings-Metropolis algorithm with a finite number of draws, Markov chains that are long enough will inherit the stability properties of their continuous time counterparts according to the functional central limiting theorem.[3] Thus, the convergence of the LTE identified by the stability of MCMC chains generated with the Hastings-Metropolis algorithm using large number of draws can be established through investigation of (in)stability of the corresponding diffusion process.

Chernozhukov and Hong (2003) establish the asymptotic convergence of the quasi-posterior distribution $p_N(\theta)$, and asymptotic normality and consistency of the LTE under assumptions that the vector of parameters of interest θ is locally identified in population, and the sample

[3]The Markov chain process will inherit the stability property of the corresponding limiting continuous-time diffusion, if the diffusion is a "sufficiently good" approximation of the discrete process behind the LTE, which is determined by the quasi-posterior function and the number of draws used to compute the estimate. Under the smoothness assumption 2, the numerical difference between this diffusion and the process for $\widehat{\theta}_t$ is of stochastic order $o_p\left(\frac{1}{\sqrt{T}}\right)$. Therefore, the stability properties of the continuous-time approximation will translate into the stability properties of the discrete process, when Theorems 1 or 2 hold in the neighborhood of size $\frac{1}{\sqrt{T}}$ around θ^*. The size of this neighborhood shrinks for longer Markov chains.

likelihood function admits a locally linear representation around θ^* in mean square. We continue to rely on the assumptions from Chernozhukov and Hong (2003). In addition, to justify the use of the first-order approximation to the quasi-posterior density $p_N(\theta)$ at any point $\theta \in \Theta$, we strengthen the Expansion assumption (Assumption 4 in Chernozhukov and Hong (2003)) by imposing the uniform convergence of the sample likelihood, and assuming that the sample likelihood has a locally linear $\mathbf{L}^2$-representation in any point of the r-neighborhood of the true parameter estimate θ^*. The modified Expansion assumption has the following formulation:

Assumption 2 *The sample quasi-likelihood $L_N(\cdot)$ admits the local-quadratic approximation in the $\mathbf{L}^2$-norm uniformly in θ, so that for each $\theta \in int(\Theta)$, it is possible find $r > 0$ such that for all θ' such that $\|\theta' - \theta\| < r$,*

$$L_N\left(\theta'\right) - L_N\left(\theta\right) \overset{\mathbf{L}^2}{=} \left(\theta' - \theta\right)' \Delta_{1N}\left(\theta\right) + N\left(\theta' - \theta\right)' \Delta_{2N}\left(\theta\right)\left(\theta' - \theta\right) + R_N\left(r, \theta' - \theta\right),$$

where $\Delta_{1N}(\theta) \xrightarrow{d} N(0, A(\theta))$ and $\Delta_{2N}(\theta)$ is a constant (for a given N) matrix with both elements and determinant uniformly in N bounded from zero. Moreover, for each $\epsilon > 0$, there is a sequence δ_N such that $\delta_N\sqrt{N} \to 0$, and a sufficiently small r, such that for any $\epsilon > 0$,

$$\limsup_{N\to\infty} P^* \left\{ \sup_{\|\theta'-\theta\| \le \delta_N} \frac{\|R_N\left(r, \theta' - \theta\right)\|}{1 + \sqrt{N}\|\theta' - \theta\|} > \varepsilon \right\} = 0,$$

and

$$\limsup_{N\to\infty} P^* \left\{ \sup_{|\theta'-\theta\| \le \delta_N} \frac{\|R_N\left(r, \theta' - \theta\right)\|}{1 + \sqrt{N}\|\theta' - \theta\|} > \varepsilon \right\} < \varepsilon,$$

in any shrinking neighborhood with $\delta_N < r$ and $\delta_N \to 0$.

Assumption 2 allows to substitute a possibly non-smooth objective function $L_N(\theta)$ with its mean-square representation not only in a neighborhood of equilibrium, but also at any point in the parameter space. It ensures the uniform convergence of the residual, and guarantees existence of a sequence of shrinking radii δ_N and a sufficiently small r, such that for any $\epsilon > 0$ the standard "empirical process condition" holds for the outer probability inside the shrinking neighborhood of θ^*, and outside this shrinking neighborhood, the divergence of the residual is driven by the mean-square expansion terms higher than quadratic.

Under Assumption 2, the first-order approximation to $p_N(\theta)$ from Equation (3) can be written as follows:

$$p_N(\theta) \approx \Delta_{1N}(\theta) + \frac{1}{\pi(\theta)} \frac{\partial \pi(\theta)}{\partial \theta},$$

and the Langevin diffusion process (4) becomes:

$$d\theta(t) = \frac{1}{2} \left\{ \Delta_{1N}(\theta(t)) + \frac{1}{\pi(\theta(t))} \frac{\partial \pi(\theta(t))}{\partial \theta(t)} \right\} dt + dW(t).$$

The sufficient condition for the stability of this diffusion process from Theorem 1 can now be used to formulate the null hypotheses of stability:

$$H_0 \, : \sup_{\theta \in \Theta} \left\{ \left\{ \Delta_{1N}(\theta) + \frac{1}{\pi(\theta)} \frac{\partial \pi(\theta)}{\partial \theta} \right\} \frac{\partial v(\theta)}{\partial \theta'} + e' \frac{\partial^2 v(\theta)}{\partial \theta \partial \theta'} e \right\} < 0.$$

The optimal test of H_0 would include the search over different alternative Lyapunov functions that satisfy the properties listed in Theorem 1, which may be very difficult to implement in practice. Although the stability test we offer is not optimal because it relies on a specific Lyapunov function, it nevertheless is implementable and has a fixed coverage for a particular model and sample size. Following a common strategy, we construct a Lyapunov function candidate to test H_0 for the stability of the continuous-time diffusion process $\theta(t)$ using a quadratic functional form as follows[4]:

$$v(\theta) = \sum_{p=1}^{P} \mathbf{1} \left\{ r_{p-1} \leq (\theta - \theta^*)' \, \Sigma_p \, (\theta - \theta^*) \leq r_p \right\} a_p \left((\theta - \theta^*)' \, \Sigma_p \, (\theta - \theta^*) \right)^{\alpha_p}, \qquad (5)$$

where P is the number of possible discontinuity points for the optimization criterion (the quasi-likelihood function), and for $p = 0, ..., P$, $r_p > 0$, $a_p > 0$, and α_p are real numbers, $r_0 = 0$, $r_P = +\infty$, and matrices Σ_p are positive definite. The lengths of the radii r_p correspond to the discontinuity points of the log-likelihood function. Parameters a_p, α_p and elements of matrices

[4] An attractive feature of such a candidate function is that it implies the Gaussian density as a solution, which simplifies algebraic manipulations for the stability analysis.

Σ_p are calibrated to ensure that $v(\theta)$ is twice continuously differentiable on the ellipsoids of discontinuity $(\theta-\theta^*)' \, \Sigma_p \, (\theta-\theta^*) = r_p$, for $p = 1, \ldots, P$.

In the further analysis, we abstract from problems with discontinuities in the population objective function. In this case, the Lyapunov function candidate,

$$v(\theta) = a_0 \left((\theta-\theta^*)' \, \Sigma_0 \, (\theta-\theta^*) \right)^{\alpha_0}, \tag{6}$$

is a positive-definite quadratic form with the infinitesimal upper limit, radially unbounded and twice continuously differentiable, provided that parameters a_0, and α_0 are greater than 0, and matrix Σ_0 is positive-definite. Assuming without loss of generality that $a_0 = 1$, $\alpha_p = 1$, and Σ is an inverse of the variance of the parameter of interest, $\Sigma_0 = \Sigma_\theta^{-1}$, and taking into account that Σ_θ is symmetric, the first and second derivatives of this Lyapunov function can be written as follows:

$$\frac{\partial v\left(\theta\right)}{\partial \theta'} = 2\Sigma_\theta^{-1}(\theta-\theta^*),$$

and

$$\tfrac{\partial^2 v(\theta)}{\partial\theta\partial\theta'} = 2\Sigma_\theta^{-1}. \tag{7}$$

In this case, the limiting diffusion process of the MCMC algorithm is stable, if for all $\theta \in \Theta$, the following condition holds

$$\sum_{i,j=1}^{k} \left[\Delta_{1N}^{i}\left(\theta\right) + \frac{1}{\pi\left(\theta\right)} \frac{\partial \pi\left(\theta(t)\right)}{\partial \theta^i} \right] \Sigma_{\theta\,\{i,j\}}^{-1} \left(\theta^j(t) - \theta^{j*} \right) + \sum_{i,j} \left(\Sigma_{\theta\,\{i,j\}}^{-1} \right) < 0.$$

Given that the continuous-time stochastic process is a very good approximation for the empirical process of the cumulative average of the posterior Markov chain, we can apply the stability test to the cumulative mean of the posterior draws. Because in empirical applications the values of θ^* and Σ_θ are not known, we suggest a two-step procedure to construct the test statistic to test that the MCMC algorithm is stable and consequently, that the LTE converges. In each step, a Markov chain is generated with the random walk Hastings-Metropolis algorithm. In the first step, the elements of the Markov chain provide the estimates of θ^* and the variance-covariance matrix Σ_θ, $\overline{\theta}^*$ and $\overline{\Sigma}_\theta$ respectively. In the second step, the elements of the Markov chain are used to calculate the following statistic:

$$\mathcal{T}_s = \sup_{t \leq s} \left[\sum_{i,j=1}^{k} \left[\Delta_{1N}^{i} \left(\widehat{\theta}_t \right) + \frac{1}{\pi \left(\widehat{\theta}_t \right)} \frac{\partial \pi \left(\widehat{\theta}_t \right)}{\partial \widehat{\theta^i}_t} \right] \overline{\Sigma}_{\theta \{i,j\}}^{-1} \left(\widehat{\theta^j}_t - \overline{\theta^j}^* \right) + \sum_{i,j}^{k} \left(\overline{\Sigma}_{\theta \{i,j\}}^{-1} \right) \right], \qquad (8)$$

where s is the number of elements of the Markov chain used to calculate the LTE, and $\widehat{\theta}_t$ is the LTE calculated over the first $t < s$ draws of the Markov chain. This statistic is determined by a sequence of stopping times of the considered Markov chain and its asymptotic distribution is defined by a two-sided Brownian bridge. The null hypothesis will fail to reject if the value of $\mathcal{T}_s$ is smaller than the $\alpha-$percent quantile of this distribution.

Alternatively, we can develop the test of non-stability of the MCMC algorithm using the results from Theorem 2. The following Lyapunov candidate function 2:

$$\tilde{v}(\theta) = \log \left((\theta - \theta^*)' \, \Sigma_0 \, (\theta - \theta^*) \right)$$

is explosive when approaching the equilibrium θ^*, as is required by Theorem 2. The null hypothesis of the Markov chain instability can then be derived from condition

$$\left\{ \Delta_{1N}(\theta) + \frac{1}{\pi(\theta)} \frac{\partial \pi(\theta)}{\partial \theta} \right\} \frac{\partial \tilde{v}(\theta)}{\partial \theta'} + e' \, \frac{\partial^2 \tilde{v}(\theta)}{\partial \theta \partial \theta'} \, e < 0.$$

The test statistic for the test of instability can be constructed by analogy to the Markov chain stability test.

3.1 Example

The following example of a simple censoring model demonstrates the divergence of the LTE. In this example, MCMC chains will not stabilize around equilibrium even if Markov chains are infinitely long. It should be noted that the LTE diverges even though the parameter of interest is identified, and the estimator converges at a parametric rate.

Consider a model where the objective function is defined by the moment condition

$$g(x,\theta) = a + (|x-\theta| - a)\, \mathbf{1}\{|x-\theta| \le a\}, \tag{9}$$

where x is the random variable, θ is the parameter of interest, and $a > 0$. In the vicinity of $x = \theta$, the moment condition coincides with the absolute value function, and it turns into a constant when x is distant from θ. Note that the moment condition is symmetric around $x = \theta$. The estimator $\widehat{\theta}$ minimizes the sample objective $Q_N(\theta)$:

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} Q_N(\theta),$$

where $Q_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} g(x_i,\theta)$. If x is uniformly distributed in the interval $[-a, a]$, the population objective $Q(\theta) = E[g(x,\theta)]$, is given by

$$Q(\theta) = \begin{cases} \frac{1}{2a}\left(a^2 + \frac{\theta^2}{2}\right), & if \quad |\theta| < a \\ \frac{1}{2a}\left(2a|\theta| - \frac{\theta^2}{2}\right), & if \quad a \le |\theta| \le 2a \\ a, & if \quad 2a < |\theta| \end{cases}. \tag{10}$$

This function is shown in Figure 1, where the parameter $a = 1$. The figure demonstrates that the population objective has a well-pronounced extremum $\theta^* = 0$. Moreover, because the objective function of the correctly specified model is equivalent to that of the least absolute deviation estimator in the vicinity of θ^*, it inherits the property of parametric convergence from the least absolute deviation estimator. However, in spite of these favorable properties, MCMC chains obtained with a simple Hastings-Metropolis random walk algorithm are unstable.

To see this, consider a standard random walk Hastings-Metropolis algorithm, in which a proposal draw ξ_t is generated from a normal distribution with the mean θ_{t-1}:

$$\xi_t = \theta_{t-1} + \sigma\,\varepsilon_t,$$

where θ_{t-1} is the last element of the MCMC chain, $\sigma > 0$ is the standard deviation of the proposal draws, and $\varepsilon_n \sim N(0,1)$. Under the flat prior distribution for θ, $\theta_t = \xi_t$ with a probability $\rho_t =$

$\min\left\{1, \frac{Q_N(\xi_t)}{Q_N(\theta_{t-1})}\right\}$, and $\theta_t = \theta_{t-1}$ with probability $1-\rho_t$. Notice that whenever $|\theta_{t-1}| > |x_i|^{max}$, θ_{t-1} belongs to the flat part of the objective function. In this case, the next element of the MCMC chain will get outside this area with probability $\Phi(\frac{|x_i|^{max}+|\theta_{t-1}|}{\sigma}) - \Phi(\frac{|x_i|^{max}-|\theta_{t-1}|}{\sigma}) \leq \Phi(\frac{2a}{\sigma}) - \frac{1}{2}$. Thus, the lower bound on the probability for the next element of the MCMC chain to stay in the flat region is $1-(\Phi(\frac{2a}{\sigma})-\frac{1}{2}) = \frac{3}{2}-\Phi(\frac{2a}{\sigma})$. Because $a > 0$, the lower bound of this probability is always between $\frac{1}{2}$ and 1. Thus, an infinitely long MCMC chain will be locked in the flat area with a positive probability, which means the equilibrium $\theta^* = 0$ is not globally asymptotically stable according to Definition 1.

One can also verify that the sufficient condition of instability is satisfied in this problem when the Lyapunov Candidate is determined as $v(\theta) = log(|\theta|)$. For a one-dimensional problem,

$$\frac{\partial v}{\partial \theta} = \frac{1}{\theta},$$

and

$$\frac{\partial^2 v}{\partial \theta^2} = -\frac{1}{\theta^2}.$$

At the same time, in population $\Delta_1(\theta) = Q'(\theta)$, and

$$Q'(\theta) = \begin{cases} \frac{\theta}{2a} & if \quad |\theta| < a \\ sign(\theta) - \frac{\theta}{2a}, & if \quad a \leq |\theta| \leq 2a \\ 0 & if \quad 2a < |\theta| \end{cases} . \tag{11}$$

Thus, if the prior is uniform, so that $\frac{\partial \pi(\theta)}{\partial \theta} = 0$, then the sufficient condition for instability of the equilibrium $\theta^* = 0$, from Theorem 2 becomes

$$Q'(\theta)\frac{1}{\theta} < \frac{1}{\theta^2},$$

which can be simplified to

$$Q'(\theta) < \frac{1}{\theta},$$

for $\theta > 0$.[5] One can show that if $a < 2$, this inequality holds for all $\theta \in \mathbb{R}$. Thus, the equilibrium $\theta^* = 0$ in this case is stochastically, and thus globally asymptotically unstable.

4 Empirical application

The purpose of the empirical application is to demonstrate that in short samples, the convergence of MCMC algorithm may fail even if models that can be consistently estimated in long data samples. We use a version of the dynamics stochastic general equilibrium (DSGE) model popular in the empirical macroeconomics. A DSGE model is summarized by a dynamic system of aggregate model variables, usually output, consumption, employment, etc. The choice of the model is motivated by the fact that DSGE models, although often enhanced with a number of features to facilitate the fit of the model and the data, are still simpler in terms of dynamics and less computationally intensive than models from the microeconometric literature. The use of a simpler model makes our argument stronger.

We choose to work with artificial datasets to minimize the problem of model misspecification. Specifically, the objective function is the distance function that matches impulse responses from the model and the data. Using the model-generated samples, we estimate the vector autoregression model and obtain impulse responses from this model, following a common approach in the literature. Although this empirical model is misspecified by definition, we show that the resulting bias is not significant. We find that the problem of the LTE convergence is substantial when samples are short. We find that the null hypothesis of Markov chain stability is satisfied more often in larger samples.

In the remaining part of the paper, we first describe the general structure of the model. We proceed by explaining how the model was calibrated, and how data samples were generated. Then, we explain the strategy to estimate model parameters. Finally, we present the results of the empirical exercise.

[5] If this condition holds for $\theta > 0$, the sufficient condition for instability will be automatically satisfied for $\theta < 0$.

4.1 Model

The model is a version of the dynamic stochastic general equilibrium (DSGE) model studied in Christiano, Eichenbaum, and Evans (2005). We introduce real and nominal rigidities, such as variable investment costs and habit formation, and nominal rigidities by assuming Calvo-Yun style price and wage stickiness.[6] For simplicity, we abstract from modeling money, and focus on a cashless economy.

A representative infinitely lived household maximizes the expected lifetime utility

$$\mathcal{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t - bc_{t-1}, 1 - l_t), \tag{12}$$

where $\beta \in [0, 1]$ is the discount factor, and $\mathcal{E}_0$ denotes expectation conditional on information in period $t = 0$. The logarithmic intratemporal utility is defined over consumption and leisure

$$u(c_t, 1 - l_t) = \phi \log(c_t - bc_{t-1}) + (1 - \phi) \log(1 - l_t), \tag{13}$$

where $c_t - bc_{t-1}$ is the adjusted for habit consumption, $b \in [0, 1]$ is the habits parameter, and l_t represents hours of work.

Each household supplies a continuum of differentiated labor types to the labor market in a monopolistically competitive fashion. These labor types are aggregated into homogenous labor by a competitive labor packer firm using a Dixit-Stiglitz aggregating technology where η_p is the intratemporal elasticity of substitution between different labor types. The homogeneous labor is then supplied to intermediate producers to use in the intermediate goods production. In each period, with probability $1 - \alpha_w$ households can freely change the wage rate for differentiated labor supplies. With probability α_w, the wage rate may not be changed freely; however it is allowed to partially adjust to the previous period inflation π_{t-1} according to the formula

$$W_t^j = W_{t-1}^j \pi_{t-1}^{\chi_w}, \tag{14}$$

where $\chi_w \in [0, 1]$ is the parameter of partial wage indexation.

Households invest to accumulate capital, and then they rent it to firms. Besides depreciation

[6] See Calvo (1983) and Yun (1996).

of capital at a rate δ, any adjustment to the level of investment relative to the previous period level is associated with capital loss. Following Christiano, Eichenbaum, and Evans (2005), we assume that the loss per unit of investment is

$$\Phi\left(\frac{i_t}{i_{t-1}}\right) = \frac{\kappa}{2}\left(\frac{i_t}{i_{t-1}} - 1\right)^2, \tag{15}$$

where i_t denotes investment and $\kappa > 0$ is a parameter determining the convexity of the investment cost function.

Besides wage and rental income, households may receive dividends from ownership in firms, and net lump-sum transfers from government.

A continuum of intermediate firms of measure 1 produce differentiated goods according to the Cobb-Douglas technology

$$y_{i,t} \leq z_t k_{i,t}^{1-\theta} {h_{i,t}}^{\theta}, \tag{16}$$

where $k_{i,t}$ denotes capital, $h_{i,t}$ is homogenous labor factor, $\theta \in [0,1]$ is the parameter determining the share of labor in production, and z_t is the exogenous stochastic technology process that evolves according to

$$\log\left(\frac{z_t}{z}\right) = \rho_z \log\left(\frac{z_{t-1}}{z}\right) + \epsilon_{z,t}, \tag{17}$$

where z is a technology level in a non-stochastic steady state, $\rho_z \in [0,1]$ is the autocorrelation parameter, and ϵ_t is an i.i.d.$(0, \sigma_z)$ stochastic process.

Intermediate goods are inputs to the production of the final homogenous good y_t by competitive firms using a Dixit-Stiglitz aggregating technology, with parameter $\eta_p > 1$ determining the intratemporal elasticity of substitution between differentiated inputs. According to the Calvo-Yun price rigidity setting, in every period a firm can reset the price for its product with probability $1 - \alpha_p$. However, with probability α_p the firm cannot choose its price freely, but is allowed to partially adjust it for the previous period inflation π_{t-1} according to the formula

$$P_t^i = P_{t-1}^i \pi_{t-1}^{\chi_p}, \tag{18}$$

where $\chi_p \in [0,1]$ is the parameter of partial indexation for prices.

We assume that monetary policy follows a simple Taylor type interest rate rule

$$log(\frac{R_t}{R}) = \alpha_R log\left(\frac{R_{t-1}}{R}\right) + \alpha_\pi log\left(\frac{\pi_t}{\pi}\right) + \alpha_y log\left(\frac{y_t}{y_{t-1}}\right), \tag{19}$$

where α_R, α_π, and α_y are parameters.

Finally, there is no role for fiscal policy. Government consumption g_t is financed by lump-sum taxes, which implies that the following government budget constraint is satisfied in every period.

As is standard in the literature, we focus on a symmetric equilibrium, where all firms with an opportunity to change prices will set them at the same level. By analogy, all wages that can be changed will be set by households at the same level for each labor type. The model equilibrium is then determined by a nonlinear dynamic system of 14 variables that evolve over time according to the following system of 14 difference equations

$$k_{t+1} = (1-\delta)k_t + i_t\left(1 - \Phi\left(\frac{i_t}{i_{t-1}}\right)\right), \tag{20}$$

$$\frac{y_t}{s_t^p} = c_t + g_t + i_t, \tag{21}$$

$$\lambda_t = u_1(c_t - bc_{t-1}, 1 - l_t) + \beta u_1(c_{t+1} - bc_t, 1 - l_{t+1}), \tag{22}$$

$$\lambda_t q_t = \beta\lambda_{t+1}(mc_{t+1}z_{t+1}f_1(k_{t+1}, h_{t+1}) + q_{t+1}(1-\delta)), \tag{23}$$

$$\lambda_t = q_t\Phi_{i_t}(\frac{i_t}{i_{t-1}}) + \beta\lambda_{t+1}q_{t+1}\Phi_{i_t}(i_{t+1}, i_t), \tag{24}$$

$$\lambda_t = b\beta r_t\frac{\lambda_{t+1}}{\pi_{t+1}}, \tag{25}$$

$$s_{t+1}^p = (1-\alpha_p)\tilde{p}_t^{-\eta_p} + \alpha_p\left(\frac{\pi_t}{\pi_{t-1}^{\chi_p}}\right)^{\eta_p} s_t, \tag{26}$$

$$s_{t+1}^w = (1-\alpha_w)\left(\frac{\tilde{w}_t}{w_t}\right)^{-\eta_w} + \alpha_w \left(\frac{w_{t-1}}{w_t}\right)^{-\eta_w} \left(\frac{\pi_t}{\pi_{t-1}^{\chi_w}}\right)^{\eta_w} s_t^w, \tag{27}$$

$$\mathcal{F}_t = \frac{\eta_w - 1}{\eta_w}\tilde{w}_t\lambda_t \left(\frac{\tilde{w}_t}{w_t}\right)^{-\eta_w} h_t + \alpha_w\beta \left(\frac{\pi_{t+1}}{\pi_t^{\chi_w}}\right)^{\eta_w - 1} \left(\frac{\tilde{w}_{t+1}}{\tilde{w}_t}\right)^{\eta_w - 1} \mathcal{F}_{t+1}, \tag{28}$$

$$\mathcal{F}_t = u_2(c_t, 1-l_t) \left(\frac{\tilde{w}_t}{w_t}\right)^{-\eta_w} h_t + \alpha_w\beta \left(\frac{\pi_{t+1}}{\pi_t^{\chi_w}}\right)^{\eta_w} \left(\frac{\tilde{w}_{t+1}}{\tilde{w}_t}\right)^{\eta_w} \mathcal{F}_{t+1}, \tag{29}$$

$$\mathcal{X}_t = y_t mc_t \tilde{p}^{-\eta_p - 1} + \alpha_p\beta \frac{\lambda_{t+1}}{\lambda_t} \left(\frac{\tilde{p}_{t+1}}{\tilde{p}_t}\right)^{\eta_p + 1} \left(\frac{\pi_{t+1}}{\pi_t^{\chi_p}}\right)^{\eta_p} \mathcal{X}_{t+1}, \tag{30}$$

$$\mathcal{X}_t = \frac{\eta_p - 1}{\eta_p} y_t \tilde{p}_t^{-\eta_p} + \alpha_p\beta \frac{\lambda_{t+1}}{\lambda_t} \left(\frac{\tilde{p}_{t+1}}{\tilde{p}_t}\right)^{\eta_p} \left(\frac{\pi_{t+1}}{\pi_t^{\chi_p}}\right)^{\eta_p - 1} \mathcal{X}_{t+1}, \tag{31}$$

$$log(\frac{R_t}{R}) = \alpha_R log\left(\frac{R_{t-1}}{R}\right) + \alpha_\pi log\left(\frac{\pi_t}{\pi}\right) + \alpha_y log\left(\frac{y_t}{y_{t-1}}\right), \tag{32}$$

and

$$log(\frac{z_{t+1}}{z}) = \rho_z log\left(\frac{z_t}{z}\right) + \epsilon_{z,t+1}, \tag{33}$$

where all the variables in (20) - (33) are defined in Tables 1 and 2, and

$$w_t = mc_t z_t \theta \left(\frac{k}{h}\right)^{1-\theta}, \tag{34}$$

$$y_t = \frac{z_t k_t^{1-\theta} {h_t}^\theta - \psi}{s_{t+1}^p}, \tag{35}$$

$$l_t = s_t^w h_t, \tag{36}$$

$$\tilde{p}_t = \left(\frac{1 - \alpha\left(\frac{\pi_t}{\pi_{t-1}^\chi}\right)^{\eta_p - 1}}{1-\alpha}\right)^{-1/(\eta_p - 1)}, \tag{37}$$

and

$$\tilde{w}_t = \left(\frac{w_t^{1-\eta_w} - \alpha_w w_{t-1}^{1-\eta_w} \left(\frac{\pi_t}{\pi_{t-1}^{\chi_w}} \right)^{1-\eta_w}}{(1-\alpha_w)} \right)^{1/(1-\eta_w)} \tag{38}$$

are the real wage rate, final good output, labor supply, relative optimal price by firms and relative optimal wage by households respectively.

The 14 variables that constitute the equilibrium dynamics of the model are: mc_t, q_t, iv_t, h_t^d, c_t, λ_t, π_t, f_t, x_t, and R_t, s_t^p, s_t^w, and k_t,and z_t. To solve the model, we use the perturbation method outlined in Schmitt-Grohé and Uribe (2004) to solve for the first order approximation to the equilibrium dynamics of the model.

4.2 Data Generating Process

To generate the data, we calibrate the model as follows. We set the intertemporal discount factor $\beta = 0.9902$, which corresponds to a steady state annualized real interest rate of approximately 4 percent. Depreciation rate δ is set at a conventional value of 2.5 percent. The investment adjustment costs κ is 3, and habits parameter is calibrated at 0.6. The technology parameter θ that reflects the share of labor in output is set at 0.7, implying the capital share of 0.3. Calvo parameters for price (α_p) and wage (α_w) rigidities are 0.6 and 0.8 respectively, which implies that on average prices are set for a period of 2.5 quarter, and wages change every 5 quarters. Both wages and prices are subject to partial indexation, with χ_p and χ_w set at 0.5. The elasticities of substitution between different labor types and differentiated intermediate goods, η_w and η_p, are both set equal to 6 to imply steady state markups of 20 percent. Monetary policy rule coefficients α_R, α_π, and α_y are set at 0.7, $0.3 \cdot 1.5 = 0.45$ and $0.5 \cdot 0.3 = 0.15$ correspondingly, to imply that the rule is inertial, and satisfies the generalized Taylor principle. The steady state inflation target is $\pi = 1.005$, which implies annualized inflation rate of approximately 2 percent. The steady state labor is $h = 0.3$, and the steady state shadow price of capital is $q = 1$. Steady state government consumption is 20 percent of GDP. Finally, the steady state technology is $z = 1$, the autocorrelation of the technology process ρ_z is 0.2, and the standard deviation σ_z is 0.01. Table 2 summarizes the calibration of model parameters.

We generate samples of various sizes for the following 7 variables: GDP, consumption, investment, hours, the real wage rate, inflation, and the interest rate. Each sample is generated by feeding in realizations of the stochastic technology process (17) for N periods. To avoid the problem of stochastic singularity as pointed out by Ireland (2004), we add measurement errors to the data, so that observations z_t^{obs} are related to the underlying model generated data z_t^{model} as follows

$$z_t^{obs} = z_t^{model}(1 + err_t), \tag{39}$$

where err_t is a multivariate (7×1) i.i.d. random variable with mean zero and the standard deviation of 0.01.

We use data sample z_t^{obs} to estimate a vector autoregression model with 2 lags (VAR(2)) in the way an empirical economist would do. We place the productivity variable, $\frac{GDP_t}{H_t}$ first in the VAR. This variable identifies the neutral technology shock. Following the standards of the empirical literature, we order the interest rate last in the VAR. The interest rate represents the reaction of monetary policy to observed information about the current state of the economy. The variables with ordering from 2 to 6 are consumption, investment, labor, wages, and inflation, correspondingly. [7] All variables in the VAR are the logarithms of the data. We use this VAR model to obtain impulse response functions (IRF) of our 7 variables of interest. These sample impulse responses are then used in an IRF matching exercise described in the next subsection.

4.3 Estimation

We use the Laplace type estimator by Chernozhukov and Hong (2003) to estimate the vector of 13 model parameters

$$\theta = \{\alpha_p, \alpha_w, \chi_p, \chi_w, b, \kappa, \theta, \beta, \alpha_R, \alpha_\pi, \alpha_Y, \rho_z, \sigma_z\}. \tag{40}$$

The distance function $L_N(\theta)$ is the weighted average of the difference between theoretical and empirical impulse response functions:

$$L_N(\theta) = (X(\theta) - \hat{X}_N)'\hat{V}_N(X(\theta) - \hat{X}_N), \tag{41}$$

[7] Although the VAR(2) model we estimate with Cholesky identification is misspecifies, we find that at large sample sizes, a VAR(2) model generates impulse responses that accurately match those from the true model.

where $X(\theta)$ denotes impulse responses generated by the model, and $\hat{X}_N$ denotes impulse responses predicted by an empirical model with N data observations. $\hat{V}_N$ is the weighting matrix, which could be a singular matrix.[8] Following Christiano, Eichenbaum, and Evans (2005), we use the diagonal weighting matrix with inverse of variances of impulse responses along the diagonal.

The estimates are obtained by generating a Markov chain of 1 million draws using the random-walk Hastings-Metropolis algorithm. We assume that the prior distribution $\pi(\theta)$ is uniform over the parameter space. The transition kernel is $q(x|y) = f(|x-y|)$, where f is a multivariate 13×1 zero-mean normal distribution with variance $V = \sigma D$, where D is the inverse of the negative numerical hessian of the distance function, $D = -H^{-1}$, $H = \frac{\partial^2 L_N(\theta)}{\partial\theta\partial\theta'}$, and σ is a scaling parameter. The hessian is evaluated in a starting point for the Markov chain, ξ^0, and D is induced to be symmetric by deriving D from the following transformation:

$$D = V \Lambda V, \tag{42}$$

where V is the matrix of eigenvectors of D, and Λ is the diagonal matrix with absolute values of eigenvalues of D along the diagonal. The scaling parameter s is adjusted to achieve the acceptance rate of the algorithm between 30 and 40 percent.

4.4 Results

We test the convergence of the LTE in application to the empirical model by calculating the statistics (8) for the null hypothesis of the stability of the underlying limiting diffusion process. With this purpose, we utilize a two-step procedure outlined in Section 2. In each step, we estimate the model based on 1 million MCMC draws. To reduce correlation between consecutive draws, we record every 1000^{th} element of the Markov chain. In the first step, we set the starting value θ_0 at the true parameter value. We use the draws of the first step to evaluate $\overline{\theta}^*$ and $\overline{\Sigma}_\theta$. The starting value for the MCMC chain in the second step is set at $\overline{\theta}^*$. The elements of the Markov chain obtained in the second step are used to calculate the statistics (8), and its distribution. The statistics is calculated using the last 100 elements of the Markov chain. We obtain the distribution of the statistic by calculating it for all the subsets of the Markov chain

[8] In estimation, we use impulse responses for 20 steps of each variable, which gives rise to 140 points to match. Thus, $X(\theta)$ and $\hat{X}_N$ are vectors 140×1, and $\hat{V}_N$ is 140×140.

with the length of 100 elements. Specifically, we identify the subsets $\{\theta_i\}_{i=t}^{t+99}$, for t from 1 to 991. Once we obtain the distribution of $\mathcal{T}_s$, we can find critical values and test the hypothesis of stability.

Figure 3 shows the distributions of $\mathcal{T}_s$, as well as 5% critical values (red dashed vertical lines), and the value of the statistics for the estimate (green vertical line) for five data samples of lengths 100, 200, 500, $1,000$, and $10,000$ observations. The graphs reveal that the null of stability is accepted for all the samples, which is not surprising, since the starting value of the Hastings-Metropolis algorithm was chosen in the close proximity to the minimizer of the objective function.

The theoretical implication of the stability theorems is that a proper choice for the starting value and the parameter set may considerably improve the performance of the LTE. The choice of the starting point that it is close enough to the true parameter value may lessen the chance of instability, because if the stability set exists, it must surround the true parameter. Chernozhukov and Hong (2003) mention that a proper starting point for the MCMC algorithm may facilitate convergence of the LTE. The strategy to search for a proper starting point has also been widely used in practice (An and Schorfheide (2007)) in Bayesian econometrics, although to the best of our knowledge, there has been no theoretical justification to this strategy.

To demonstrate how the wrong choice of the starting point can negatively affect the outcome of estimation, we run multiple MCMC chains using different starting values. Specifically, we fix all parameters but one at their true values, and vary the starting value for this single parameter. To present the results in a more compact way, we show the resulting estimate for the parameter of interest only, shown with a dot in a corresponding graph in Figures 4 - 6. The figures contain 13 plots, each showing the estimate for the parameter of interest (along the vertical axis) depending on the starting value for this parameter (on the horizontal axis), keeping the starting values of all the other parameters at their true values. If there are no stability issues, one should expect to see the the graphs where all the points lie along the horizontal line. We repeat this exercise for five samples of different sizes: 100, 200, 500, $1,000$, and $10,000$ observations. The graphs reveal that there is a lot of variation in the estimates when samples contain 100 and 200 observations (Figures 4(a) and 4(b)). This is true for almost all the parameters and especially noticeable for parameter α_y. Comparing graphs with different sample sizes, one can notice that convergence of the parameters to true values improves with the number of observations. While a small bias from misspecification of the empirical VAR model remains even at $10,000$ observations, convergence is almost perfect for the longest sample, as can be seen from Figure 6.

To minimize the possibility of instability in MCMC chains, one may also try to shrink the parameter space where the Markov chain will wonder most of the time as much as possible around the true parameter value. While the researcher is usually not aware of the true parameter value, she may still rely on some information from economic theory that can be used to form priors about parameters. Because the prior $\pi(\theta)$ appears in the statistic (8), it can be manipulated with the aim to minimize the possibility of instability in MCMC chains. Since priors are widely used in the Bayesian literature, the results of this paper may help to justify the use of specific prior distributions in Bayesian applications.

A crucial assumption behind our results is that model parameters are identified. Often, the failure of Markov chains to converge is attributed to the absence of identification. In case of identification failure, parameters cannot be estimated consistently without additional information, and MCMC methods will produce diverging Markov chains (Dufour and Hsiao (2008)). Figure 2 shows the profile of the expected distance function across different dimensions of the parameter space. We approximate the expectation with a simple average of empirical distance functions obtained from 100 different data samples. The artificial data samples are created using the strategy outlined in Section 4.2. To reduce the short-sample bias, each sample contains 10,000 observations. The graphs demonstrate that the cross-sections of the expected distance function are hump-shaped for all parameters of interest, which implies that the model is identified across all dimension of the parameter space. A necessary condition for identification of model parameters is non-singularity of the expected hessian of the distance function. We check this condition by evaluating the rank of the expected hessian of the distance function (41), and do not find the evidence of rank deficiency.

5 Conclusion

In this paper, we study the performance of the MCMC-based estimation algorithm for extremum estimators. This method has become very popular for estimation of complex structural models, since it alleviates many problems associated with alternative gradient-based maximum search procedures. We find that even if an estimation problem is well-defined and the parameter of interest identified, there is no guarantee that Markov chains generated by the MCMC procedure will converge to the true maximizer of the sample objective. In general, if there is a set of parameter values such that the convergence criterion is satisfied for all points in this set, then the MCMC algorithm will produce stable Markov chains, and the LTE will converge. However, the MCMC chains will be unstable and the LTE will not converge to the maximizer of the sample objective, if the sufficient condition of the instability theorem is satisfied. We obtain this result by approximating the Hastings-Metropolis algorithm with the continuous time diffusion process, and using the stability theory for stochastic differential equations. This conclusion can be applied to the LTEs in finite samples as well as to the asymptotic behavior.

We illustrate our results by estimating a structural dynamic stochastic general equilibrium model. We choose a simplified version of the model, where all structural parameters are point-identified. The analysis of the convergence of the MCMC procedure for this model suggests that the LTE converges from almost all points in the parameter support for very large samples. However, for small sample sizes (e.g. smaller than 500 observations) the range of starting values from which the LTE does not converge to the extremum of the sample minimum distance objective is larger. The size of the instability region tends to decrease with the sample size.

References

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

Calvo, G. (1983): "Staggered Prices in a Utility-Maximizing Framework," *Journal of Economic Literature*, 12, 1383–398.

Chernozhukov, V., and H. Hong (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293 – 346.

Chib, S. (2001): "Markov Chain Monte Carlo Methods: Computation and Inference," *Handbook of Econometrics*, 2, 3569–3652.

Christiano, L. J., M. Eichenbaum, and C. A. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1), 1–45.

Dufour, J., and C. Hsiao (2008): "Identification," *The New Palgrave Dictionary of Economics.*

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

Gelfand, S., and S. Mitter (1993): "Metropolis-Type Annealing Algorithms for Global Optimization in R^ d," *SIAM Journal on Control and Optimization*, 31, 111–111.

Geweke, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18(1), 1–73.

Gikhman, I., I., and V. Skorokhod, A. (2004): *The Theory of Stochastic Processes II.* Springer.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2009): "Frequentist inference in weakly identified DSGE models," *Working Papers.*

Hasminskii, R., Z. (1980): *Stochastic Stability of Differential Equations.* Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands.

Ireland, P. (2004): "A Method for Taking Models to the Data," *Journal of Economic Dynamics and Control*, 28(6), 1205–1226.

Kloeden, P., E. Platen, and H. Schurz (1994): *Numerical Solution of SDE Through Computer Experiments.* Springer Verlag.

Kushner, Harold, J., and G. Dupuis, Paul (2001): *Numerical Mathods for Stochastic Control Problems in Continuous Time.* Springer.

Meyn, S., and R. Tweedie (2009): "Markov Chains and Stochastic Stability," *Cambridge University Press New York, NY, USA*, p. 622.

NEWEY, W. (1994): "The asymptotic variance of semiparametric estimators," *Econometrica: Journal of the Econometric Society*, pp. 1349–1382.

ROBERT, C., AND G. CASELLA (2004): *Monte Carlo Statistical Methods*. Springer Verlag.

ROBERTS, G., A. GELMAN, AND W. GILKS (1997): "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms," *The Annals of Applied Probability*, pp. 110–120.

ROBERTS, G., AND J. ROSENTHAL (2007): "Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms," *Journal of Applied Probability*, 44(2), 458.

ROBERTS, G., AND R. TWEEDIE (1996): "Geometric Convergence and Central Limit Theorems for Multidimensional Hastings and Metropolis Algorithms," *Biometrika*, 83(1), 95–110.

SCHMITT-GROHÉ, S., AND M. URIBE (2004): "Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function," *Journal of Economic Dynamics and Control*, 28(4), 755–775.

SMITH, A. (1993): "Estimating nonlinear time-series models using simulated vector autoregressions," *Journal of Applied Econometrics*, 8, 63–84.

STOCK, J., AND J. WRIGHT (2000): "GMM with weak identification," *Econometrica*, 68(5), 1055–1096.

YUN, T. (1996): "Nominal Price Rigidity, Money Supply Endogeneity, and Business Cycles," *Journal of Monetary Economics*, 37(2), 345–370.

6 Appendix

6.1 Tables

Table 1: Model variables

Parameter	Description
k	Capital
y	Production
c	Consumption
g	Government spending
i	Investment
s^p	Price dispersion
s^w	Wage dispersion
λ	Marginal utility of habit-adjusted consumption
l	Labor supply
mc	Marginal cost
w	Wage rate
r	Real interest rate
R	Nominal interest rate
$\tilde{w}$	The wage rate of wage optimizing labor types relative to w
$\tilde{p}$	The relative price of optimizing firms
$\mathcal{F}$	Auxilary variable
$\mathcal{X}$	Auxilary variable

Table 2: Calibration of model parameters

Parameter	Description	
β	Discount factor	$1.04^{-0.25}$
δ	Depreciation rate	0.025
b	Habit parameter	0.6
κ	Investment adjustment cost	3
θ	Labor share in output	0.7
α_p	Price rigidity	0.6
α_w	Wage rigidity	0.8
χ_p	Indexation, price	0.5
χ_w	Indexation, wage	0.5
η_p	Dixit-Stiglitz aggregator, output	6
η_w	Dixit-Stiglitz aggregator, labor	6
α_R	Monetary policy	0.7
α_π	Monetary policy	1.5(1-0.7)
α_Y	Monetary policy	0.5(1-0.7)
π	Inflation target	1.005
h	Labor	0.3
q	Shadow price of capital	1
SG	Government consumption share	0.2
z	Steady state technology	1
ρ_z	Shock, autocorrelation	0.2
σ_z	Shock, standard deviation	1

6.2 Figures

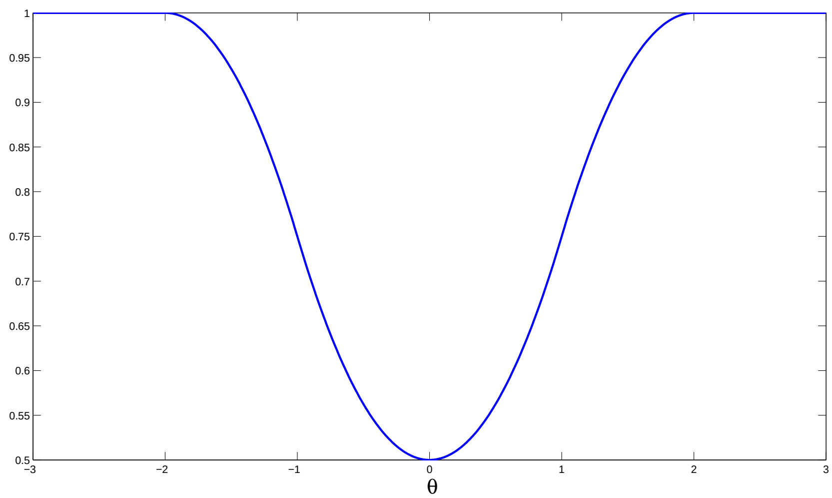


Figure 1: Expected population objective

Notes: The expected population objective is calculated assuming the random variable x in the example from Section 3.1 is uniformly distributed in the interval $[-a\,, a]$, and $a = 1$.

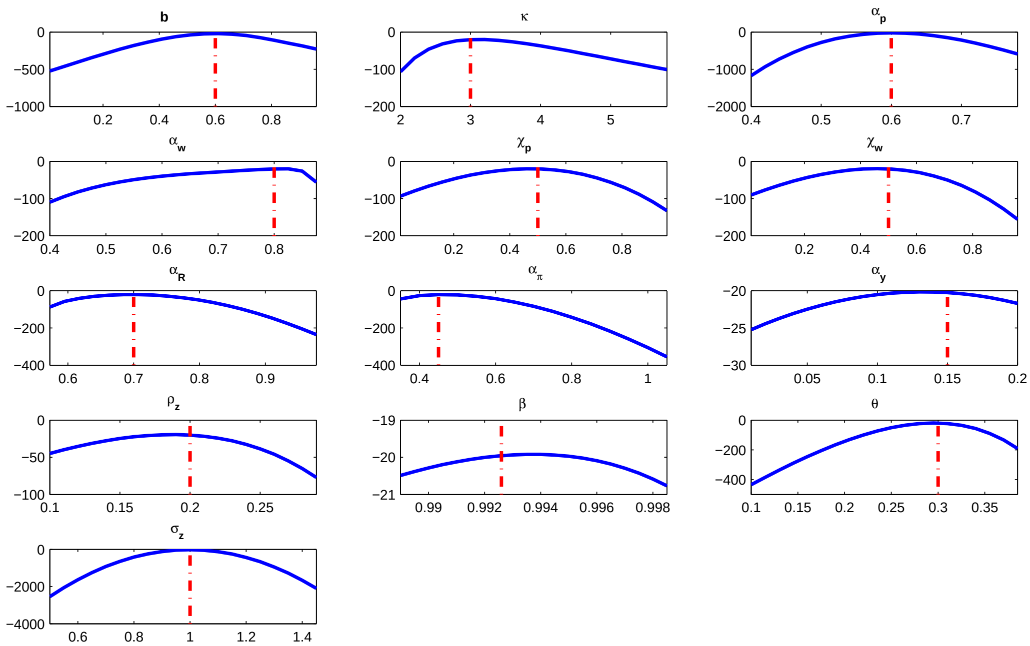


Figure 2: Expected distance function profiles

Notes: The graphs show the cross section of the expected distance function along the dimensions of the parameter space. The expected distance function was approximated by a simple average of distance functions calculated using 100 artificially generated samples with size $N = 10,000$.

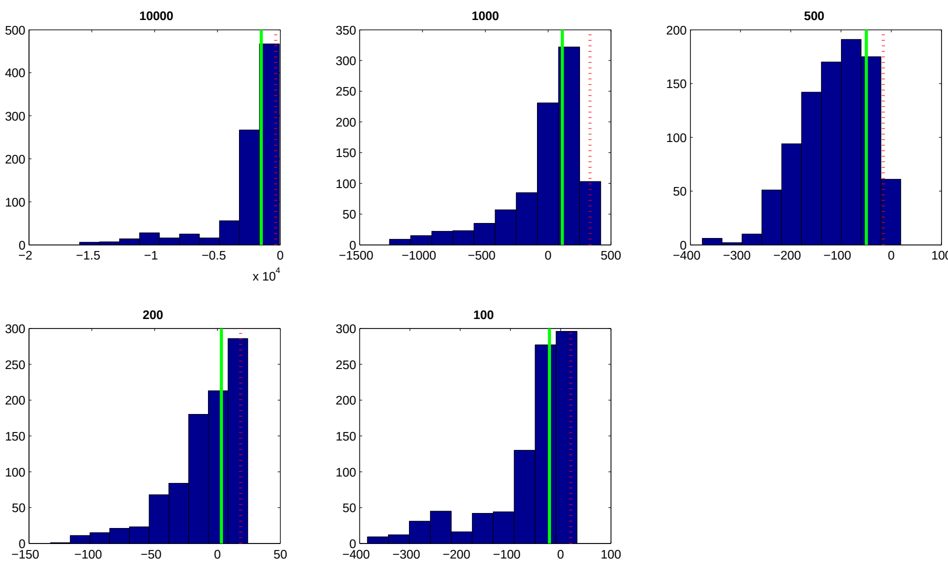
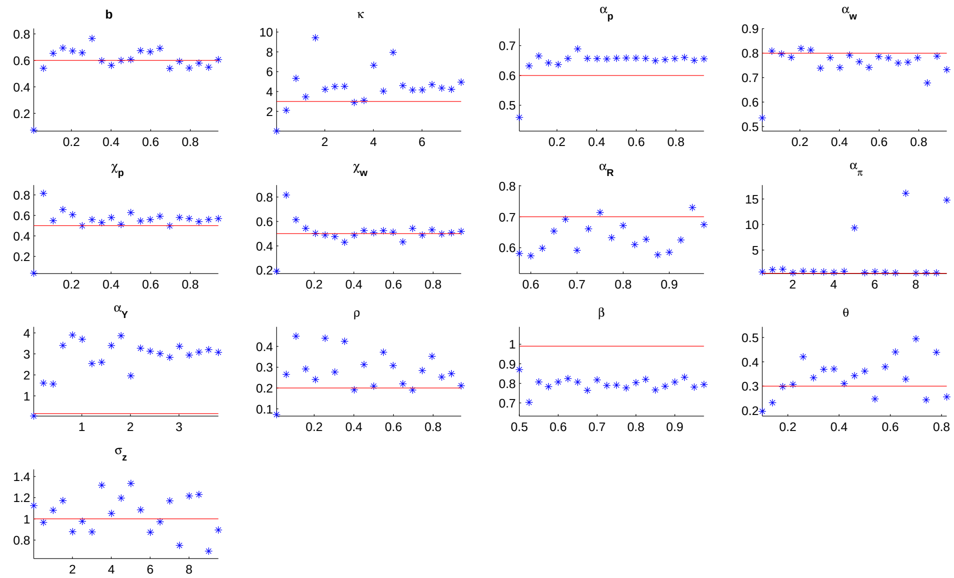
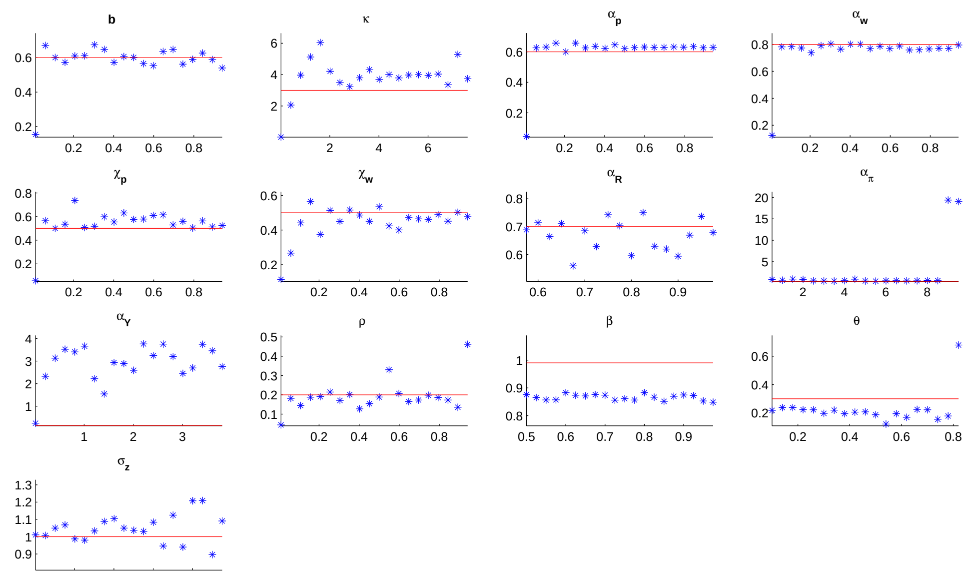


Figure 3: Test statistic for convergence of the estimator

Notes: The graphs show the distribution of the statistics (8) under the null hypothesis of the stability for MCMC chains generated using the data samples of different lengths. The red dashed line determines the 5 percent critical values for rejection of the null hypothesis. The vertical green line shows the statistic for the estimate obtained as the average of the last 100 draws of the Markov chain. A two-step procedure is used to calculate the statistics. In the first step, we run MCMC chain to calculate $\overline{\theta^*}$ as the mean value, and $\overline{\Sigma}_\theta$ as the estimated variance-covariance matrix of the MCMC chain. The MCMC chain obtained in the second step provides θ_t to calculate the statistic. The test reveals that for all sample sizes considered, the null hypothesis of stability cannot be rejected.



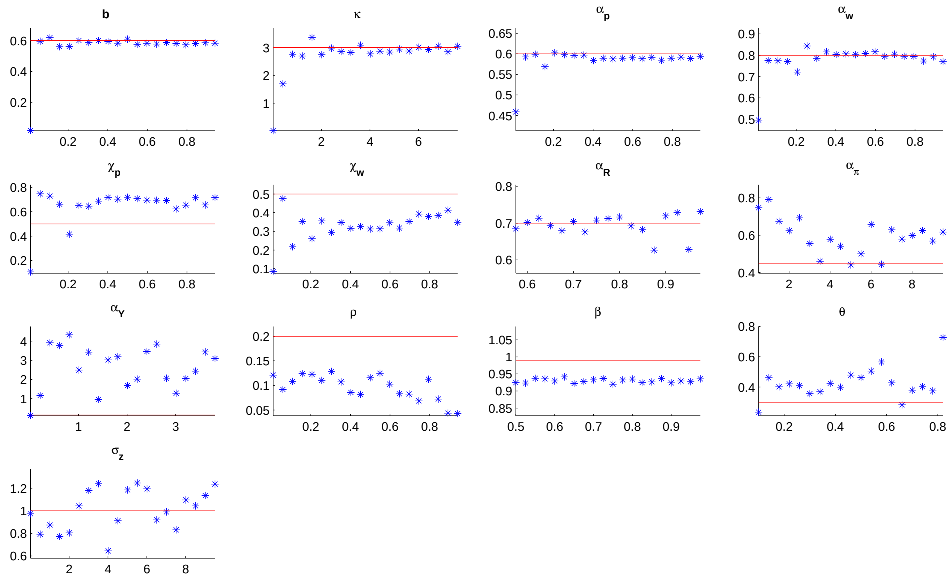
(a) Sample of 100 observations



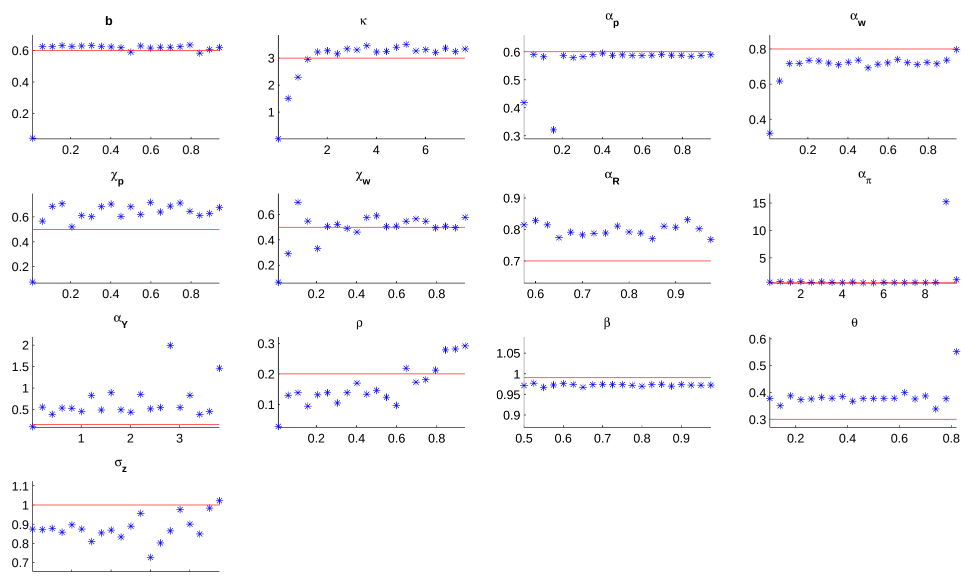
(b) Sample of 200 observations

Figure 4: Dependence of estimates on starting values, part 1

Notes: Each graphs shows the estimate of the parameter of interest depending on the starting value for this parameter, for 13 parameters of the model. The starting value of the parameter of interest is on the horizontal axis, the resulting estimate for this parameter is on the vertical axis, and the star denotes the estimate from one MCMC chain. Every graph presents the results of 20 estimations (260 estimations in total). The starting values for all parameters other than the parameter of interest were kept at true values.The estimate was obtained as a mean of the last 500,000 MCMC draws for this parameter out of the total 1 million draws.



(a) Sample of 500 observations



(b) Sample of 1000 observations

Figure 5: Dependence of estimates on starting values, part 2

Notes: Each graphs shows the estimate of the parameter of interest depending on the starting value for this parameter, for 13 parameters of the model. The starting value of the parameter of interest is on the horizontal axis, the resulting estimate for this parameter is on the vertical axis, and the star denotes the estimate from one MCMC chain. Every graph presents the results of 20 estimations (260 estimations in total). The starting values for all parameters other than the parameter of interest were kept at true values. The estimate was obtained as a mean of the last 500,000 MCMC draws for this parameter out of the total 1 million draws.

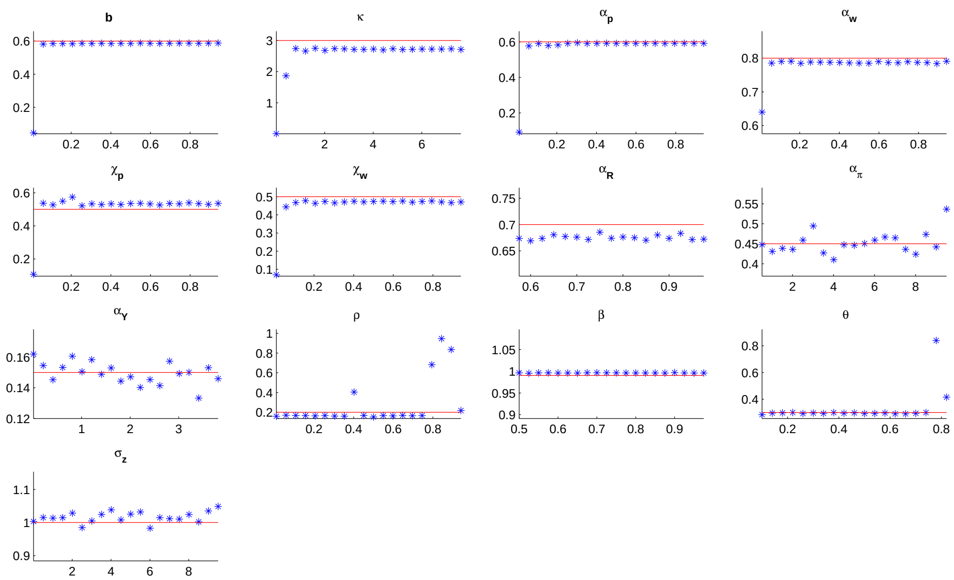


Figure 6: Dependence of estimates on starting values, sample of 10000 observations

Notes: Each graphs shows the estimate of the parameter of interest depending on the starting value for this parameter, for 13 parameters of the model. The starting value of the parameter of interest is on the horizontal axis, the resulting estimate for this parameter is on the vertical axis, and the star denotes the estimate from one MCMC chain. Every graph presents the results of 20 estimations (260 estimations in total). The starting values for all parameters other than the parameter of interest were kept at true values. The estimate was obtained as a mean of the last 500,000 MCMC draws for the parameter of interest out of the total 1 million draws.

Identifiability and the structure of the estimated model

In this paper we aim at studying the performance of the Metropolis-Hastings algorithm that is applied to a quasi-likelihood of a structural model in the context of Markov Chain Monte-Carlo inference. We only focus on the models that are identified. We consider identification on the population level and in this section we introduce the notions of identifiability that we use. We narrow our analysis to the models where we substitute the inference using the likelihood by the inference using a vector of moments constructed from a structural model. We interpret the term "structural" in a broad sense where the econometric model or its part depends on the implicitly defined function that has to be evaluated numerically. We define the structural econometric model as a p-dimensional vector of parameters θ, the structural function g, which is derived from a model that presumably generates the data and the distribution of vector random covariates $F_z(\cdot)$.

In the empirical part of the paper we consider the case where g is associated with an equilibrium of a DSGE model. In DSGE models, maximum likelihood approach to parameter estimation is time-consuming. A convenient way of dealing with this problem is related to the method of indirect inference developed in (Smith 1993) and (Gallant and Tauchen 1996). According to the indirect inference method, rather than searching for the maximum of the entire likelihood profile of the model, one can effectively obtain parameters by maximizing a quasi-likelihood function, determined by only some characteristics of the model. Depending on the goals of the estimation, this set of characteristics could include the second moments of variables of interest, or impulse response functions. Such reduced quasi-likelihood problems belong to the class of standard GMM problems with the moment vector $m(z, \theta, f)$, which depends on the random variable z, finite-dimensional parameter of interest θ, and the numerical procedure f. The numerical procedure may be a likelihood for maximum-likelihood problems, or a set of moment conditions for GMM-type estimations.

In the subsequent discussion we will call a pair (θ, f) - the parameters of the model, and in this pair θ is a Euclidean parameter of interest while f is a nuisance parameter denoting the distribution of the covariate z. Thus, we consider the problem of semiparametric M-estimation where the true parameter $(\theta(0), f_0)$ satisfies a semi-parametric moment condition:

$$E\left[m\left(z, \theta(0), f_0\right)\right] = 0.$$

The inference from the reduced model will only be possible if the numerical procedure is sensitive to the parameter of interest. Moreover, for asymptotic validity of parameter estimates a unique parameter must deliver the maximum to the quasi-likelihood in population. In this case, we say that the parameter is identified by a model and data. We propose the following formal definition of parameter identification.

Definition 3 *The Euclidean parameter $\theta \in \Theta$ (where Θ is a convex compact subset of a Euclidean space $\mathbb{R}^k$) is identified by the observable distribution of the data $F_z(\cdot)$ if the functional correspondence $(\theta,\, f) \mapsto F_z$ is injective.*

For the purposes of this paper we impose a general identification assumption. We do not exclude the possibility of irregular (or weak) identification as in (Stock and Wright 2000). Irregular identification generally leads to the singularity of the information matrix of the model.[9] However, it is still possible to estimate the parameter of interest using estimation procedures converging at non-parametric rates. There is a line of research that studies asymptotic properties of estimators under weak identification (see (Guerron-Quintana, Inoue, and Kilian 2009) among others) In this paper, the focus is on finite-sample properties of identified (possibly weakly) models. The aim of this research is to demonstrate that in finite samples, Bayesian estimation of model parameters may fail even for strongly identified models. Thus, we restrict the analysis to (locally) identified models.Parallel to (Stock and Wright 2000), we impose the following set of assumptions on functions and parameters of the model to ensure the model of interest is identified. We impose the following assumption that allows us to consider estimation separately from the numerical procedure.

Assumption 3 **A.1** *1. $\theta \in \Theta \subset \mathbb{R}^k$, where Θ is compact with respect to the Euclidean norm in $\mathbb{R}^k$*

[9] In the setup of (Stock and Wright 2000) the empirical moment function contains a part that, first, contains a (weakly identified) parameter of interest. second, it vanishes as the sample size increases. This leads to a zero limit for the corresponding components of the information matrix.

2. *We consider a class of numerical procedures* $\mathcal{H}$ *endowed with a norm (e.g. numerical tolerance level) such that for each* f *and* f' *with* $\sup z, \theta \| m(z, \theta, f) - m(z, \theta, f') \| = o(\|f\|)$

3. $z \in \mathbb{R}^p$ *is a random variable with absolutely continuous density.*

A.2 *For each* $\epsilon > 0$ $m(z, \theta, f)$ *is locally Lipschitz continuous in* θ *in the* ϵ*-neighborhood of* $\theta(0)$ *for all* $\|f\| < \epsilon$*. Moreover, the Lipschitz constant does not depend on* f*.*

A.3 *There exists a function* $\Gamma(z)$ *such that* $E\left[\|\Gamma(z)\|^2\right] < \infty$ *and a number* $\|J_0\| < \infty$ *such that*

$$\frac{\partial}{\partial \theta} E\left[\Gamma(z) m(z, \theta, f)\right]\Big|_{\theta=\theta(0)} = J_0.$$

A.4 *Equation* $E[m(z, \theta, f)] = 0$ *has a unique solution in* Θ *at point* $\theta(0)$ *for a fixed* f*.*

Consider a family of local perturbations of parameters $\theta(0)$ in the model with numerical algorithm f. Index these perturbations bt t and form θ_t similar to Ibragimov and Has'minskii(1981) and in the family of numerical procedures f_t we reqire $\|f_t\| \leq \|f_{t'}\|$ for $t < t'$, $f_0 = f$ and:

$$\theta_t = \theta(0) + \Gamma(z) t \Delta_\theta.$$

Parameters θ_t are such that for each $t \in [0, 1]$ and each numerical procedure f_t :

$$E_t\left[m(z, \theta_t, f_t)\right] = 0.$$

The paths along the local perturbations pass through $(\theta(0), f_0)$. We can compute the derivative of the moment condition along the path indicated by local perturbations. In this case, $\lambda(z)$ is related to the score of the model $S_t(z)$ along the parametrization path t. Then we can express Δ_θ as:

$$J_0 \, \Delta_\theta = -E\left[\left(m(z, \theta(0), f_0) + \delta(z)\right) S_t(z)\right]. \tag{43}$$

In this expression, the quantity Δ_θ can be characterized as a normalized derivative of the Euclidean parameter along the path. For the expectation on the right-hand side of (43), the standard projection results in Newey(1994) will hold, from which it follows that the estimator for θ is locally asymptotically linear:

$$\overline{\psi(z)} = -\left\{m\left(z, \beta_0, f_0\right) + \left(\delta(z) - E\left[\delta(z)\right]\right)\right\}. \tag{44}$$

Note that we assumed that the Lipschitz constant does not depend on the choice of the numerical procedure, given that the tolerance $\|f_t\| = O\left(\|f\|\right)$. This means that $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\delta(z_i) - E\left[\delta(z_i)\right]\right) = o_p(1)$. We make an additional assumption which can be verified for a particular model by checking that the model is sufficiently smooth and satisfies the Lindeberg condition.

Assumption 4 *Suppose that a moment vector $m(\cdot)$ forms a system of moments that exactly identifies the parameter θ. Empirical moment function evaluated at the true Euclidean parameter value*

$$\sqrt{n}\hat{m}^{(n)}\left(\theta(0)\right) \xrightarrow{d} N\left(0, \sigma_m^2\right).$$

The estimator $\hat{\theta}$ solves the system of equations:

$$\hat{m}^{(n)}\left(\theta\right) = 0. \tag{45}$$

Theorem 1 summarizes the properties of the estimator of interest:

Theorem 3 *Suppose that Assumptions 3.1-3.5 and 4 are satisfied. Then $\sqrt{n}\left(\hat{\theta} - \overline{\theta}\right)$ is asymptotically linear. Moreover, the asymptotic variance of $\sqrt{n}\left(\hat{\theta} - \overline{\theta}\right)$ does not depend on the computational procedure.*

Proof:

Assumption 4and the definition of the estimator, we obtain that the estimator is asymptotically normal. Then using the proof of Theorem 2.1 in (Newey 1994) we verify that the influence function associated with the estimator will be defined by the locally linear representation above. □

Theorem 3 proves that asymptotic properties of the semiparametric estimate of the finite-dimensional parameters, normalized by the Jacobi matrix do not depend on a particular choice of the numerical procedure. The numerical procedure, however, will affect the variance of unnormalized parameters. We consider local representations for coefficients with

$$\hat{\theta}_n - \overline{\theta}_n = \left(\hat{\theta}_n - \theta(0)\right) - \left(\overline{\theta}_n - \theta(0)\right).$$

We consider a normalization sequence γ_n such that

$$\gamma_n^{-1} J_n = J_0^{\eta} + o_p(1),$$
$$\gamma_n \left(\overline{\theta}_n - \theta(0)\right) = h^{\eta} + o_p(1).$$

We label both the bias and the normalized Jacobi matrix by η to indicate that, generally speaking, these limits depend on the choice of structure used to estimate infinite-dimensional nuisance parameter of the model. Then the asymptotic distribution of parameter of interest can be expressed as:

$$\gamma^n \left(\hat{\theta}_n - \theta(0)\right) \xrightarrow{d} N\left(h^{\eta},\, (J_0^{\eta})^{-1}\, \Omega\, (J_0^{\eta})^{-1}\right)$$

6.1 Technical implementation for continuous metropolis chains: Integration of stochastic differential equations

From the theoretical point of view, Langevin algorithms are more attractive than standard Metropolis-in-Gibbs procedures. However, numerical integration of diffusion-driven stochastic differential equations is not equivalent to simple simulation from the appropriately adjusted normal distribution. To see this, consider a general Itô equation of the form:

$$dX(t) = a\left(X(t)\right) dt + b\left(X(t)\right) dW(t).$$

The "naive" Euler-type approximation to this equation, frequently used in the literature[10] is given by:

$$X_{k+1} - X_k = a\left(X_k\right) \tau + b\left(X_k\right) \omega_k,$$

[10] A detailed treatment of approximations of stochastic differential equations is given in (Kloeden, Platen, and Schurz 1994).

with $\omega_k \sim i.i.d.\ N(0, \tau)$. (Kloeden, Platen, and Schurz 1994) shows that the "naive" discretization with the time step τ evaluates the solution with the absolute error of the order $O(\sqrt{\tau})$, that is:

$$E\{|X_k - X(t_k)| \mid X_0 = x_0\} \leq C\sqrt{\tau}$$

Thus, the error accumulation for integration over long intervals of time can be quite significant. A useful way to deal with this problem of order $O(\tau^2)$ is based on the second-order Taylor-Itô expansion of the solution to the stochastic differential equation, represented as the integral:

$$X(t) = X(0) + \int_0^t [a(X(t))\,dt + b(X(t))\,dW(t)]\,.$$

The second order expansion leads to the discrete difference equation in the form:

$$X_{k+1} - X_k = a(X_k)\,\tau + b(X_k)\,\omega_k + \frac{1}{2} b(X_k)\,\frac{\partial b(X_k)}{\partial x}\left(\omega_k^2 - \tau\right).$$

The precision of this approximation is quadruple the precision of the naive integration algorithm.