DOI:10.19343/j.cnki.11-1302/c.2016.02.013

第 33 卷第 2 期
2016 年 2 月

统计研究
Statistical Research

Vol. 33，No. 2
Feb. 2016

基于多重尝试 Metropolis 算法的非线性 DSGE 模型贝叶斯推断*

杨 远 林 明

内容提要：本文提出一种改进的多重尝试 Metropolis 算法，用于非线性动态随机一般均衡模型的贝叶斯参数估计和模型选择。多重尝试策略通过每次迭代抽取多个尝试点的方法来提高算法的混合速率，新方法中提出使用近似的方法提高计算速度，并通过接收概率调整偏差。数值实验表明新方法在相同的计算时间内具有更高的估计效率。最后，本文比较了具有不同货币政策设定的模型对中国经济数据的拟合效果，发现中国数据更加支持具有时变通胀目标的模型。

关键词：动态随机一般均衡模型；多重尝试 Metropolis 算法；粒子滤波器；贝叶斯因子

中图分类号：O212　　　**文献标识码**：A　　　**文章编号**：1002－4565（2016）02－0091－08

Bayesian Inference for Nonlinear DSGE Model via Multiple-try Metropolis Algorithm

Yang Yuan & Lin Ming

Abstract: This paper proposes a multiple-try-based Metropolis algorithm for Bayesian parameter estimation and model selection in nonlinear dynamic stochastic general equilibrium（DSGE） models. Multiple-try method generates multiple trial points in each iteration to achieve faster mixing rate. The new approach proposes to use approximation to save computational cost，the approximation bias is adjusted through the acceptance rate. Simulation results demonstrate effectiveness of the proposed algorithm. The algorithm is applied to estimate DSGE models with different monetary policy settings using the macroeconomic data of China. It shows that the data favors the model with time-varying inflation target.

Key words: DSGE Models；Multiple-try Metropolis Algorithm；Particle Filter；Bayes Factor

一、引言

动态随机一般均衡（DSGE）模型是进行宏观经济分析的重要工具。和向量自回归等简约模型相比，DSGE 模型具有微观经济理论基础，能够更好地对经济运行机制、政策传导途径以及外生冲击的动态效应等问题进行研究。此外，包含丰富摩擦及冲击的 DSGE 模型在实际数据拟合和预测方面取得了巨大的成功。基于此，DSGE 模型逐渐被各国政府部门和科研机构所采用，作为探究货币政策与宏观经济波动、国际贸易与金融和劳动市场摩擦等一系列重要问题的有力工具[1][2]。

目前常用的 DSGE 模型为线性化模型，而近年来非线性 DSGE 模型受到了更多的关注。首先，在线性化的政策函数中被忽略的高阶误差会随着模型中观测值数量的增加而不断累积，最终导致线性化模型与真实模型的似然函数之间出现较大的偏差[3]。其次，经验分析表明非线性 DSGE 模型对实际数据具有更好的拟合效果——尤其当数据中存在较强的持续性和对稳态较大的偏离时[4]。最后，非线性政策函数中的高阶项对于最优货币政策、宏观经济模型的资产定价以及外生冲击的时变波动率等问题的研究具有非常关键的作用。

在 DSGE 模型的实际应用中，通常采用贝叶斯方法进行参数估计。一方面，研究人员可以使用所

* 本文获国家自然科学基金项目“状态空间模型参数的高效序贯蒙特卡洛估计及在金融中的应用”（11101341）的资助。

掌握的关于模型参数的先验信息，同时，先验分布的引入也能够防止由不规则似然函数所导致的反常的参数估计结果。另一方面，针对线性 DSGE 模型的贝叶斯计算方法已经很成熟，这些方法能够在各种情形下有效地进行后验抽样。例如，当待估计参数比较多时，Chib 和 Ramamurthy（2010）[5] 提出将参数向量随机分组，然后使用 Metropolis-Hastings 算法更新每一组内的参数；更新过程中所使用的抽样分布，则通过对该组参数的条件后验分布进行近似而得到。Herbst 和 Schorfheide（2014）[6] 使用序贯蒙特卡罗方法进行后验抽样，通过引入抽样分布自适应机制以及对目标分布进行回火（Tempering）来克服多峰后验分布带来的困难。

上述方法主要针对线性化的 DSGE 模型，由于非线性模型的似然函数不是解析可得，如何高效地对其进行后验抽样仍然是一个待解的难题。目前通用的方法是使用粒子滤波器对模型的似然函数实施近似，然后采用随机游走 Metropolis 算法进行抽样。尽管该方法简单易行，但是由于随机游走抽样分布忽略了后验分布的特征，因此在一些情形下相对低效。

针对以上问题，本文提出多重尝试 Metropolis 算法来对非线性 DSGE 模型进行贝叶斯推断。本文第二部分回顾了现有对 DSGE 模型进行贝叶斯估计和后验抽样的方法。第三部分介绍了多重尝试 Metropolis 算法的原理、具体步骤、特点和理论性质。第四部分给出了在后验抽样后计算贝叶斯因子的方法。第五部分开展数值实验对不同算法的效率进行比较，评估不同模型对中国数据的拟合效果，并使用选出的模型进行经济分析。

二、DSGE 模型的状态空间表示与贝叶斯估计

动态均衡模型的均衡条件一般可以表示为[7]：

$$E_t\left[F_\theta(X_t, X_{t+1}, Z_t, Z_{t+1})\right] = 0 \qquad (1)$$

该方程组包括了刻画经济主体最优行为的一阶条件、资源约束、货币政策规则和外生冲击的设定等，其中 F_θ 是从 $R^{2(n_x+n_z)}$ 到 $R\,n_x+n_z$ 的函数、X_t 为状态变量、Z_t 为控制变量、$\theta\in\Theta$ 是反映经济环境与参与者特征的结构参数，E_t 为基于 t 时刻历史信息的条件期望。模型的非随机稳态（$\bar{X},\bar{Z}$）满足条件 $F_\theta(\bar{X},\bar{X},\bar{Z},\bar{Z}) = 0$。

根据 Schmitt-Grohé 和 Uribe（2004）[7] 中的方法求解上述 DSGE 模型，可得到 DSGE 模型的状态空间表示：

$$\begin{aligned} X_t &= \hat{H}_\theta(X_{t-1}, \sigma) + u_t \\ Y_t &= M_0 + M_x X_t + M_z \hat{G}_\theta(X_t, \sigma) + e_t \end{aligned} \qquad (2)$$

其中，X_t 为隐含的状态变量，Y_t 为观测值，$\hat{H}_\theta$ 和 $\hat{G}_\theta$ 为政策函数的二阶泰勒展开，$\sigma\geq 0$ 为尺度参数，$u_t\sim N(0, R_\theta)$ 为经济系统受到的外生冲击，M_x 和 M_z 为选择矩阵，$e_t\sim N(0, Q)$ 为观测误差。观测误差的设定是由于模型中部分变量在现实中缺乏直接对应的观测值，例如货币政策中的通货膨胀目标、潜在产出等；而实际观测到的数据中也常带有观测误差。给定观测值 $Y_{1:T} = \{Y_1, \cdots, Y_T\}$ 之后，模型的似然函数可通过对状态变量 $X_{1:T} = \{X_1, \cdots, X_T\}$ 进行积分而得到：

$$L(\theta) \equiv p(Y_{1:T}\mid\theta) = \int p(Y_{1:T}, X_{1:T}\mid\theta)\, dX_{1:T}$$

在对结构参数 θ 的贝叶斯推断中，我们赋予参数一个先验分布 $p(\theta)$，然后借助马尔可夫链蒙特卡罗（MCMC）算法从后验分布 $p(\theta\mid Y_{1:T}) \propto L(\theta)p(\theta)$ 中抽样，之后用抽样得到的随机样本来计算后验统计量，并进行统计推断。最常用的 MCMC 算法为 Metropolis-Hastings（MH）算法，令 $T(\theta, \theta^*)$ 为给定 θ 后 θ^* 的抽样分布，则 MH 算法步骤如下：

算法一（Metropolis-Hastings 算法）：

（1）当 $n=1$ 时，抽取 $\theta^{(1)}\sim T_1(\cdot)$。

（2）当 $n=2,\cdots,N$ 时，依次实施如下步骤：

①抽取 $\theta^*\sim T(\theta^{(n-1)}, \cdot)$；

②计算 MH 接收概率：

$$r_{\mathrm{MH}} = \min\left\{1, \frac{L(\theta^*)p(\theta^*)T(\theta^*, \theta^{(n-1)})}{L(\theta^{(n-1)})p(\theta^{(n-1)})T(\theta^{(n-1)}, \theta^*)}\right\},$$

以 r_{MH} 的概率接收 $\theta^{(n)} = \theta^*$，否则令 $\theta^{(n)} = \theta^{(n-1)}$。

在线性化的 DSGE 模型中 $L(\theta)$ 可通过 Kalman 滤波器精确地计算出来。而当政策函数含有非线性项时 $L(\theta)$ 不是解析可得，因此无法直接使用上述的 MH 算法。为了解决这个问题，Fernández-Villaverde 和 Rubio-Ramírez（2007）[8] 提出使用粒子滤波器（Particle Filter，PF）来获得对似然函数的近似（记为 $\tilde{L}(\theta)$）并将其代入 MH 算法中，从而能够从一个近似的后验分布中抽样。事实上，尽管基于粒子滤波器的 MH 算法并未使用真实的似然函数 $L(\theta)$，但

是只要$\tilde{L}(\theta)$关于$L(\theta)$无偏，该算法仍然是从真实的后验分布$p(\theta|Y_{1:T})$里抽样。关于粒子滤波近似的无偏性及粒子 MH 方法的进一步阐述可参看 Pitt 等(2012)[9]。

在 MH 算法中，抽样分布$T(\cdot,\cdot)$的选择对于算法的效率至关重要。在实际应用中，随机游走 Metropolis(RWM)是使用最广泛的一个选择。在 RWM 中，$\theta^{*}=\theta^{(n-1)}+C\varepsilon_{n}$，其中$\varepsilon_{n}\sim N(0,\Sigma)$服从多元正态分布。参数$c$调节算法的效率：当$c$很小时，算法的接收率很高，但是生成的样本在参数空间内移动速率缓慢；当c很大时，接收概率低，同样会降低有效样本数。此外，RWM 并未考虑到后验分布的信息，当待估参数个数比较多、方差－协方差矩阵Σ选择不恰当或者后验分布不规则时，算法的混合速率会比较慢。由此生成的随机样本会具有较强的自相关性，从而导致后验估计量具有较大的数值方差和估计结果的不可靠。

三、多重尝试 Metropolis 算法

多重尝试 Metropolis(MTM)算法[10]在每次迭代中抽取k个尝试点$\{\theta_{1}^{*},\cdots,\theta_{k}^{*}\}$，并给每个尝试点赋予权重：$w_{i}=p(\theta_{i}^{*}|Y_{1:T})$，$i=1,\cdots,k$。然后，从这$k$个尝试点中随机选择一个作为提议点，每个点被选中的概率正比于其权重。该提议点随后依据 MH 原则被接收或者拒绝。多重尝试策略相当于对参数空间进行了局部搜索，通过获取后验分布信息来提升提议点的接收概率。MTM 方法在其他提取后验分布信息的技术(Gibbs 抽样、Langevin 算法等)难以实现的情形下十分有用。

在标准的 MTM 算法中同样需要计算似然函数。为了使其能够适用于非线性 DSGE 模型同时也为了节约计算成本，我们使用中心差分 Kalman 滤波器(CDKF)来计算尝试点的权重。CDKF 能够快速、精确地对模型的似然函数进行近似[11]，因此在每次 MTM 迭代中多次运行 CDKF 并不会显著增加计算量。通过 CDKF 得到的似然函数是有偏的，因此在计算 MTM 接收概率的时候需要用 PF 来校正引入的偏差，从而保证算法收敛到正确的后验分布而不是一个有偏差的分布。

记$\hat{L}(\theta)$为通过 CDKF 得到的有偏似然函数，$\hat{p}(\theta|Y_{1:T})\propto\hat{L}(\theta)p(\theta)$为相应的有偏后验分布，令$T(\cdot,\cdot)$为对称的抽样分布，定义权重函数$w(\theta)=\hat{p}(\theta|Y_{1:T})$。本文提出的 MTM 算法的具体步骤和理论性质如下：

算法二(多重尝试 Metropolis 算法)：

(1)当$n=1$时，抽取$\theta^{(1)}\sim T_{1}(\cdot)$。

(2)当$n=2,\cdots,N$时，依次实施如下步骤：

①抽取k个 i. i. d 尝试点：$\theta_{1}^{*},\cdots,\theta_{k}^{*}\sim T(\theta^{(n-1)},\cdot)$，运行 CDKF 计算其权重$w(\theta_{1}^{*}),\cdots,w(\theta_{k}^{*})$；

②从$\{\theta_{1}^{*},\cdots,\theta_{k}^{*}\}$中随机选出一个记为$\theta^{*}$，选择概率正比于$\{w(\theta_{1}^{*}),\cdots,w(\theta_{k}^{*})\}$；

③抽取$k-1$个 i. i. d 参考点：$\theta_{1}',\cdots,\theta_{k-1}'\sim T(\theta^{*},\cdot)$，并令$\theta_{k}'=\theta^{(n-1)}$运行 CDKF 计算权重$w(\theta_{1}'),\cdots,w(\theta_{k-1}')$；

④运行 PF 计算似然函数$\tilde{L}(\theta^{*})$和接收概率：

$$r_{MTM}=\min\left\{1,\frac{\tilde{L}(\theta^{*})\hat{L}(\theta^{(n-1)})\sum_{i=1}^{k}w(\theta_{i}^{*})}{\tilde{L}(\theta^{(n-1)})\hat{L}(\theta^{*})\sum_{i=1}^{k}w(\theta_{i}')}\right\},$$

以r_{MTM}的概率接收$\theta^{(n)}=\theta^{*}$，否则令$\theta^{(n)}=\theta^{(n-1)}$。

定理 1 令$\Theta=\{\theta:p(\theta|Y_{1:T})>0\}$，假设$\forall\theta,\theta^{*}\in\Theta$，有：①$T(\theta,\theta^{*})>0$；②$0<\tilde{L}(\theta)<\infty$且$<\hat{L}(\theta)<\infty$；③$E[\tilde{L}(\theta)]=L(\theta)$。则如下结论成立：①MTM算法生成的样本$\{\theta^{(n)}\}_{n\geq 1}$以$p(\theta|Y_{1:T})$为平稳分布；②对于任意可积实值函数$f$，有：

$$\lim_{N\to\infty}N^{-1}\sum_{n=1}^{N}f(\theta^{(n)})=\int f(\theta)p(\theta|Y_{1:T})d\theta$$

证明：使用 Pitt 等(2012)[9]中的辅助变量技巧，记$\tilde{L}(\theta,s)$为 PF 得到的似然函数近似，其中$s\sim p(s)$为粒子滤波过程中生成的所有随机变量。定义如下θ和s的联合分布：

$$\tilde{p}(\theta,s|Y_{1:T})\stackrel{\triangle}{=}\tilde{L}(\theta,s)p(\theta)p(s)/p(Y_{1:T})$$

记$P_{k}(\theta,\theta^{*})$为 MTM 算法中θ^{*}的实际抽样分布，$K_{k}((\theta,s),(\theta^{*},s^{*}))$为 MTM 的转移核，$I\stackrel{\triangle}{=}\{i:\theta_{i}^{*}=\theta^{*},1\leqslant i\leqslant k\}$。因为当$i\neq j$时，$\theta_{i}^{*}$和$\theta_{j}^{*}$是独立和可交换的，我们有：

$$P_{k}(\theta,\theta^{*})=kPr(\{(I=k)\cap(\theta_{k}^{*}=\theta^{*})\}|\theta)$$
$$=k\int Pr(I=k|\theta_{1:k-1}^{*},\theta_{k}^{*}=\theta^{*})Pr(\theta_{1:k-1}^{*},\theta_{k}^{*}=\theta^{*}|\theta)d\theta_{1:k-1}^{*}$$

$$= kT(\theta,\theta^*)\int \frac{w(\theta^*)}{w(\theta^*) + \sum_{j\neq k} w(\theta_j^*)}\prod_{j\neq k} T(\theta,\theta_j^*)\, d\theta^*_{1:k-1}$$

当 $\theta\neq\theta^*$ 时有:

$$K_k((\theta,s),(\theta^*,s^*)) = \int P_k(\theta,\theta^*)\, r_{\mathrm{MTM}} \prod_{j\neq k} T(\theta^*,\theta'_{1:k-1})$$

$$= \int P_k(\theta,\theta^*)\min\left\{1,\frac{\hat{L}(\theta^*,s^*)\hat{L}(\theta)\sum_{i=1}^{k} w(\theta_i^*)}{\hat{L}(\theta,s)\hat{L}(\theta^*)\sum_{i=1}^{k} w(\theta'_i)}\right\}\prod_{j\neq k} T(\theta^*,\theta'_j)\, d\theta'_{1:k-1}$$

结合上边两式,可得:

$$\tilde{p}(\theta,s\mid Y_{1:T})K_k((\theta,s),(\theta^*,s^*))$$

$$= \int \min\left\{\frac{\tilde{p}(\theta,s\mid Y_{1:T})\hat{p}(\theta^*\mid Y_{1:T})}{w(\theta^*)+\sum_{j\neq k} w(\theta_j^*)},\frac{\tilde{p}(\theta^*,s^*\mid Y_{1:T})\hat{p}(\theta\mid Y_{1:T})}{w(\theta)+\sum_{j\neq k} w(\theta'_j)}\right\}$$

$$\times kT(\theta,\theta^*)\prod_{j\neq k} T(\theta,\theta_j^*)T(\theta^*,\theta'_j)\, d\theta^*_{1:k-1} d\theta'_{1:k-1}$$

上式对 (θ,s) 和 (θ^*,s^*) 是对称的。因此,细致平衡条件 $\tilde{p}(\theta,s\mid Y_{1:T})K_k((\theta,s),(\theta^*,s^*)) = \tilde{p}(\theta^*,s^*\mid Y_{1:T})K_k((\theta^*,s^*),(\theta,s))$ 成立,故 MTM 算法以 $\tilde{p}(\theta,s\mid Y_{1:T})$ 为平稳分布。由 $\tilde{p}(\theta,s\mid Y_{1:T}) = p(\theta\mid Y_{1:T})\hat{L}(\theta,s)/L(\theta)$ 以及假设③可知 θ 在 $\tilde{p}(\theta,s\mid Y_{1:T})$ 中的边缘分布即为 $p(\theta\mid Y_{1:T})$。

其次,容易验证当假设①和②成立时,对任意 $\theta,\theta^*\in\Theta$,$\theta\neq\theta^*$ 和 $s,s^*\sim p(s)$ 有 $P_k(\theta,\theta^*)>0$ 和 $K_k((\theta,s),(\theta^*,s^*))>0$,即该 MTM 链不可约。根据马尔可夫链的收敛性质和结论①即得到结论②。

可以看出多重尝试策略改进了算法的抽样分布,即从 $k=1$ 时的 $T(\cdot,\cdot)$ 变为 $P_k(\cdot,\cdot)$。当尝试点的数量不断增加时,$P_k(\cdot,\cdot)$ 也越来越接近目标分布。事实上当 $k\to\infty$ 时,有 $P_k(\theta,\theta^*)\to T(\theta,\theta^*)\hat{p}(\theta^*\mid Y_{1:T})/Z(\theta)$,其中 $Z(\theta)$ 为归一化常数。在实际使用过程中,k 值的选取需要对效率与计算成本进行权衡考虑。

四、贝叶斯因子

在 DSGE 模型的应用中,经常使用贝叶斯因子来进行模型比较。假设有两个备选模型 M_1 和 M_2,贝叶斯因子为两个模型的边缘似然之比:$B_{12} = p(Y_{1:T}\mid M_1)/p(Y_{1:T}\mid M_2)$ 在使用 MTM 算法完成后验抽样后,可以根据 Chib 和 Jeliazkov(2001)[12] 的原理来计算边缘似然。

对于每一个模型,根据贝叶斯公式可以得到:

$$p(Y_{1:T}\mid M_i) = p(\bar{\theta}\mid M_i)L(\bar{\theta}\mid M_i)/p(\bar{\theta}\mid Y_{1:T},M_i)$$

其中,$T(\cdot,\cdot)$ 为后验均值。为了计算 θ 的后验概率密度,令 $K(\cdot,\cdot)$ 为以 $p(\cdot\mid Y_{1:T},M_i)$ 为平稳分布的 MH 转移核。

$$K_{\mathrm{MH}}(\theta,\theta^*) = T(\theta,\theta^*)r_{\mathrm{MH}}(\theta,\theta^*) + (1-r(\theta))\delta_\theta(\theta^*)$$

其中,$T(\cdot,\cdot)$ 为 MTM 算法中尝试点的抽样分布,$r(\theta) = \int T(\theta,\theta^*)r_{\mathrm{MH}}(\theta,\theta^*)d\theta^*$。根据平稳分布的定义有:

$$p(\bar{\theta}\mid Y_{1:T},M_i) = \int p(\theta\mid Y_{1:T},M_i)K_{\mathrm{MH}}(\theta,\bar{\theta})d\theta,$$

$$= \int p(\theta\mid Y_{1:T},M_i)T(\theta,\bar{\theta})r_{\mathrm{MH}}(\theta,\bar{\theta})d\theta + (1-r(\bar{\theta}))p(\bar{\theta}\mid Y_{1:T},M_i)$$

整理上式可得:

$$p(\bar{\theta}\mid Y_{1:T},M_i) = \int p(\theta\mid Y_{1:T},M_i)T(\theta,\bar{\theta})r_{\mathrm{MH}}(\theta,\bar{\theta})d\theta\cdot[r(\bar{\theta})]^{-1} \quad (3)$$

用样本的平均来近似式(3)中的积分,即可得到:

$$p(\bar{\theta}\mid Y_{1:T},M_i) \approx \left[\frac{1}{N_1}\sum_{j=1}^{N_1} T(\theta^{(j)},\bar{\theta})r_{\mathrm{MH}}(\theta^{(j)},\bar{\theta})\right]\left[\frac{1}{N_2}\sum_{j=2}^{N_2} r_{\mathrm{MH}}(\bar{\theta},\tilde{\theta}^{(j)})\right]^{-1}$$

其中,$\theta^{(j)}\sim p(\cdot\mid Y_{1:T},M_i)$,$j=1,\cdots,N_1$,通过之前的 MTM 算法进行后验抽样得到;$\tilde{\theta}^{(j)}\sim T(\bar{\theta},\cdot)$,$j=1,\cdots,N_2$ 为新抽出的独立同分布样本。在实际使用中,可以在后验抽样过程中保存通过 PF 计算出来的似然函数,避免在这里重复计算。

五、数值实验及应用

本部分首先构建一个包含价格粘性和时变通货膨胀目标的小型新凯恩斯模型,之后使用中国数据对其进行贝叶斯估计并评估各算法的效率,最后进行模型选择。

(一)模型设定

在每一时期 t,完全竞争厂商使用中间产品 $Y_t(j)$,$j\in[0,1]$ 生产最终消费品 Y_t,其生产技术为:

$$Y_t = \left(\int_0^1 Y_t(j)^{1-\nu}dj\right)^{\frac{1}{1-\nu}}$$

其中,$1/\nu$ 为中间品需求弹性。

中间商品 j 由一个垄断厂商生产,其具有线性生产技术 $Y_t(j) = A_t N_t(j)$,其中 A_t 是外生技术过程,$Nt(j)$ 是厂商 j 的劳动投入。技术过程设定为 $\ln A_t = \ln\gamma + \ln A_{t-1} + \ln Z_t$,其中:

$$\ln Z_t = \rho_z \ln Z_{t-1} + \varepsilon_t^z \tag{4}$$

劳动力通过完全竞争市场以实际工资 W_t 雇佣得到。厂商调整价格过程中存在如下调整成本:

$$AC_t(j) = \frac{\varphi}{2}\left(\frac{P_t(j)}{P_{t-1}(j)} - \pi\right)^2 Y_t(j)$$

其中,$P_t(j)$ 为产品 j 的价格,参数 φ 反映经济中价格粘性的程度,π 是最终产品价格通货膨胀率的稳态。厂商 j 通过选择劳动投入和产品价格来最大化其期望利润:

$$E_t\left[\sum_{s=0}^{\infty}\beta^s Q_{t+s|t}\left(\frac{P_{t+s}(j)}{P_{t+s}}Y_{t+s}(j) - W_{t+s}N_{t+s}(j) - AC_{t+s}(j)\right)\right] \tag{5}$$

上式中,β 为折现因子,P_t 为最终产品价格,$Q_{t+s|t}$ 是一单位 $t+s$ 时刻消费品在 t 期对于家庭户的价值。

代表性家庭户的效用取决于实际货币余量 M_t/P_t 以及消费 C_t,这里假设习惯消费量由技术水平 A_t 给定。家庭户的负效用由工作时间 H_t 决定,其最终的期望效用为:

$$E_t\left[\sum_{s=0}^{\infty}\beta^s\left(\frac{(C_{t+s}/A_{t+s})^{1-\tau} - 1}{1-\tau} + \chi_M \ln\frac{M_{t+s}}{P_{t+s}} - H_{t+s}\right)\right] \tag{6}$$

这里,$1/\tau$ 为跨期替代弹性,χ_M 决定实际货币余额的稳态值。家庭户以给定的实际工资 W_t 向厂商提供完全弹性的劳动、接收来自厂商的剩余利润 D_t 并且缴纳额定税收 T_t、还可以购买利率为 R_t 的政府债券 B_t。家庭户的预算约束可以表示为:

$$P_tC_t + B_t + M_t - M_{t-1} + T_t = P_tW_tH_t + R_{t-1}B_{t-1} + P_tD_t + P_tSG_t \tag{7}$$

其中,SG_t 为通过交易状态相依证券获得的净现金流入。

货币当局依照泰勒规则来实施货币政策:$R_t = R_t^{*\,1-\rho R} R_{t-1}^{\rho R} e^{\varepsilon_t^R}$,其中货币政策冲击 ε_t^R 为白噪声。目标利率 R_t^* 具有如下形式:

$$R_t^* = r\pi^*(\pi_t/\pi_t^*)^{\psi 1}(Y_t/Y_t^*)^{\psi 2} \tag{8}$$

其中,r 为实际利率的稳态,$\pi^* = \pi$ 为通货膨胀目标稳态,$Y^* = A_t y$ 为当名义粘性不存在时的自然产出水平。时变通货膨胀目标 π_t^* 服从自回归过程:

$$\ln\pi_t^* = (1-\rho_\pi)\ln\pi^* + \rho_\pi \ln\pi_{t-1}^* + \varepsilon_t^\pi \tag{9}$$

政府部门预算约束和市场出清条件分别为:

$$R_{t-1}B_{t-1} = T_t + B_t + M_t - M_{t-1} \tag{10}$$

$$H_t = N_t, Y_t = C_t + AC_t \tag{11}$$

(二)观测方程与数据

求解式(5)中的利润最大化问题、式(6)和式(7)中的效用最大化问题,将含有随机增长项的变量去趋势($y_t = Y_t/A_t, c_t = C_t/A_t$),并将所有变量表示为对其稳态的对数差分(对于任一变量 x_t,记 $\hat{x}_t = \log x_t - \log \bar{x}$)。优化问题的一阶条件、外生冲击设定式(4)和式(9)、货币政策设定式(8)、政府预算约束和市场出清条件式(10 - 11)构成式(1)中的均衡条件,其中 $X_t = [\hat{y}_{t-1}, \hat{R}_{t-1}, \varepsilon_t^R, \hat{\pi}_t^*, \hat{Z}_t]$,$Z_t = [\hat{c}_t, \hat{y}_t, \hat{\pi}_t, \hat{R}_t]$。

为了得到式(2)中的状态空间表示,使用 Schmitt-Grohé 和 Uribe(2004)[7] 的二阶近似方法来求解模型,并设立如下观测方程:

$$Y_{1t} = \gamma^Q + 100(\hat{y}_t - \hat{y}_{t-1} + \hat{Z}_t) + e_{1t}$$

$$Y_{2t} = \pi^A + 400\hat{\pi}_t + e_{2t}$$

$$Y_{3t} = \pi^A + r^A + 4\gamma^Q + 400\hat{R}_t + e_{3t}$$

观测误差 $e_t = [e_{1t}, e_{2t}, e_{3t}]$ 的标准差设定为实际数据样本标准差的 25%。

本文使用 1996 年 1 季度到 2014 年 4 季度的中国宏观经济数据来构建观测变量。使用 CPI 指数和"经济活动人口"对名义 GDP 数据进行调整得到实际人均产出,之后通过对数差分得到实际人均产出增长率 Y_{1t}。由于经济活动人口数量只有在年度频率上具有数据,我们通过插值来获取季度数据(假设各季度人口增长率相同)。Y_{2t} 为季度 CPI 增长率,通过对月度数据进行三阶移动平均获得季度数据。名义利率 Y_{3t} 为"全国银行间同业拆借利率"(3 个月加权平均)。所有原始数据均来自中经网数据库和国家统计局数据库,对存在季节效应的数据使用 EViews 中的 X12 方法进行季节调整。

(三)参数估计

参数的先验分布主要根据 An 和 Schorfheide(2007)[1] 来设定,其中关于稳态的 3 个参数则根据中国数据的具体情况来设定。详细的先验设定见表 1。

本文采用 Yang 和 Wang(2015)[13] 中的辅助变量粒子滤波器来近似似然函数,其中粒子个数为 $M = 100$。该方法通过对观测方程进行线性化来获得条件最优抽样分布并实施辅助变量策略。本文所使用的计算环境为 Intel 5 - 3230M 2.6GHz CUP 和

4GB 内存，使用 MATLAB R2012b 进行编程。在这样的环境中，运行一次 PF 需要 0.45 秒，一次 CDKF 需要 0.1 秒。由于 MTM 中抽取的尝试点相互独立，我们使用 MATLAB 中的 parfor 命令进行并行处理来降低计算时间。使用并行处理后，运行 4 次 CDKF 需要 0.15 秒，9 次需要 0.2 秒。

表 1 先验分布

参数	含义	分布	均值	标准差
τ	1/消费的跨期替代弹性	Gamma	2.00	0.50
ν	1/中间品需求弹性	Beta	0.10	0.05
k	$k=\tau(1-\nu)/(\pi^2\varphi\nu)$	Gamma	0.30	0.20
ψ_1	利率目标对通货膨胀的反应	Gamma	1.50	0.25
ψ_2	利率目标对产出的反应	Gamma	0.50	0.20
ρ_R	货币政策平滑系数	Beta	0.50	0.20
ρ_π	通货膨胀目标自回归系数	Beta	0.80	0.10
ρ_z	技术冲击自回归系数	Beta	0.66	0.15
r^A	实际利率稳态	Gamma	1.00	0.50
π^A	通货膨胀稳态	Gamma	3.00	0.50
γ^Q	技术进步率稳态	Normal	1.00	0.50
$100\sigma_R$	货币政策冲击标准差	Inverse Gamma	0.40	4.00
$100\sigma_\pi$	通货膨胀目标冲击标准差	Inverse Gamma	0.40	4.00
$100\sigma_z$	技术冲击标准差	Inverse Gamma	0.40	4.00

算法中的抽样分布 $T(\cdot,\cdot)$ 选为随机游走抽样分布：$\theta^* \sim N(\theta^{(n-1)}, c^2\Sigma)$。根据 Roberts 和 Rosenthal (2001) [16]，最优的Σ 为后验分布的方差－协方差矩阵。我们通过对后验分布进行高斯近似来找到其方差－协方差矩阵：使用 CDKF 和拟牛顿法对 $\ln p(\theta | Y_{1:T})$ 进行最优化，找到其的最大值（记为 θ），然后令：

$$\Sigma = \left[-\frac{\partial^2 \ln p(\theta | Y_{1:T})}{\partial\theta\partial\theta^T} | \theta = \hat{\theta} \right]^{-1}$$

此外，通过调整 c 的取值使得 RWM 和 MTM 的接收率分别在 0.25 和 0.4 左右。

我们使用 RWM 抽样 32000 个，舍弃前 2000 个。为了保持相近的计算量，MTM 仅抽样 22000 个，舍弃前 2000 个。抽样完成后，计算样本均值$\bar{\theta}$作为对参数的点估计并计算以下统计量来比较算法的效率。首先是平均平方跳跃距离（Averaged Squared Jump Distance，ASJD）：

$$\text{ASJD} = (N-1)^{-1} \sum_{n-2}^{N} \| \theta^{(n)} - \theta^{(n-1)} \|^2$$

一条马尔可夫链具有较大的 ASJD 表明其具有更快的混合速率、在给定的移动次数内对状态空间的探索更充分。对于平稳马尔可夫链，样本均值$\bar{\theta}_i$的渐进方差为 $\text{Var}(\theta_i)(1+2\Sigma_{l}^{\infty}=1\rho_i(l))$。可见生成样本的自相关性直接影响估计量的方差。累积自相关时间（Integrated Autocorrelation Time，IAT）反映了自相关性的强弱：

$$\text{IAT}_i = 1 + 2\sum_{l=1}^{Li} \hat{\rho}_i(l)$$

其中，$\hat{\rho}_i(\cdot)$ 是样本$\{\theta_i^{(1)},\cdots,\theta_i^{(N)}\}$ 的自相关函数，$L_i^* = \min\{l \geqslant 1 : |\hat{\rho}_i(l)| < 2/\sqrt{N}\}$，$L_i = \min\{1000, L_i^*\}$。$\bar{\theta}_i$ 的数值标准误差（Numerical Standard Errors，NSE）为 $\text{NSE}_i = (\hat{S}_i(0)/N)^{1/2}$。其中，$\hat{S}_i(\cdot)$ 为样本$\{\theta_i^{(1)},\cdots,\theta_i^{(N)}\}$ 的谱密度函数。由 $S_i(0) = \text{Var}(\theta_i)(1+2\Sigma_{l=1}^{\infty}\rho_i(l))$ 可见 NSE 是另一个衡量蒙特卡罗方差的统计量。上述统计量中，ASJD 是对整条马尔可夫链收敛性的度量，而 IAT 和 NSE 则反映了每一条一维子链的性质。

表 2 报告了各算法中 c 的取值、接收率以及跳跃距离。可以看出 MTM 算法能够在增大抽样方差的同时提高接收率。平均跳跃距离随着尝试点个数 k 的增加而增大。由于不同算法的迭代次数不同，我们同时也计算了加总平方跳跃距离（SSJD $=(N-1)\times$ ADJS）来考察在相同的计算时间内，马尔可夫链在状态空间内总的游走距离。可以看到在考虑计算成本情形下，MTM 算法依然具有优势——在相同计算时间内，MTM 链对参数空间的探索更加充分。

表 2 接收率与跳跃距离

算法	N	c	接收率	ASJD	SSJD
RWM	30000	0.6	0.245	0.0713	2137.5
MTM($k=3$)	20000	0.7	0.383	0.1430	2859.9
MTM($k=5$)	20000	0.8	0.394	0.1771	3542.7

表 3 报告了三个算法得到的后验均值以及 IAT 和 NSE 两个统计量。各算法对后验均值的估计很接近，差异大多在 ±0.02 之间。将本文估计结果与使用相近模型的其他文献中的结果进行比较可以发现：消费跨期替代弹性倒数 τ、利率规则平滑系数 ρ_R、利率与通货膨胀稳态 r^A 和 π^A 以及中间参数 k 的估计值与隋建利等(2011) [14] 的结果（分别为 $\tau=3.72$、$\rho_R=0.91$、$r^A=0.93$、$\pi^A=4.34$ 和 $k=0.06$）相近。不过本文中货币政策对通货膨胀和产出缺口的响应系数更大（隋建利等(2011) [14] 中分别为 $\psi_1=1.04$ 和 $\psi_2=0.03$）。中间品需求弹性倒数 ν 与简志宏等(2012) [15] 中校准的 0.17 接近；通货膨胀目标自回归系数 ρ_π 则比简志宏等(2012) [15] 中的 0.89 略大，与 Amisano 和 Tristani(2010) [4] 基于欧洲数据

估计出的 0.99 比较接近。模型估计出的价格调整成本系数很大($\varphi = 445$),反映出中国经济中存在很强的价格粘性。而较强的名义冲击持续性($\rho_R = 0.92$ 和 $\rho_\pi = 0.96$)也和价格粘性具有正向的关联。

在算法的效率方面,MTM 抽样得到各参数的样本自相关性都比 RWM 要弱,并且随着 k 的增加而变得更弱。可以看出各参数的估计“难度”也不尽相同,其中对于泰勒规则中的参数 ψ_1 和 ψ_2,RWM 得到的 IAT 比较大,其与 MTM($k=5$)算法得到的估计结果的差异也较大,分别为 0.03 与 0.02。鉴于 MTM($k=5$)更低的 IAT,其估计结果更加可信。这也显示了对于难以估计的参数,MTM 对于 RWM 的优势更加明显。表 3 中的 NSE 考虑了算法不同的抽样数量,可以看出在相同的计算时间内,MTM 的估计具有更小的蒙特卡罗方差。

表 3　参数估计与收敛诊断统计量

参数	后验均值			IAT			NSE × 100		
τ	2.77	2.76	2.78	54.9	30.5	21.6	1.90	1.80	1.47
ν	0.21	0.21	0.21	48.1	23.5	22.7	0.25	0.21	0.19
k	0.03	0.03	0.03	49.8	43.7	22.7	0.05	0.05	0.04
ψ_1	1.83	1.81	1.80	77.9	35.7	29.4	1.30	0.99	0.92
ψ_2	0.43	0.43	0.45	80.6	47.2	31.3	0.95	0.69	0.77
ρ_R	0.91	0.91	0.92	70.9	30.6	23.3	0.11	0.08	0.07
ρ_π	0.96	0.96	0.96	69.3	41.7	26.1	0.05	0.05	0.04
ρ_z	0.62	0.62	0.61	55.5	38.5	28.0	0.46	0.48	0.44
r^A	0.85	0.86	0.86	52.0	23.0	21.3	0.72	0.60	0.53
π^A	3.82	3.84	3.80	48.5	29.6	20.0	2.68	2.41	2.09
γ^Q	1.09	1.08	1.10	50.2	23.8	19.2	0.57	0.52	0.44
$100\sigma_R$	0.23	0.24	0.24	47.3	32.5	19.8	0.12	0.11	0.09
$100\sigma_\pi$	0.21	0.21	0.21	68.8	33.7	34.3	0.17	0.15	0.13
$100\sigma_2$	1.09	1.10	1.11	46.1	24.5	20.8	0.80	0.72	0.63

注: 每一个统计量栏中的三列从左到右分别对应于 RWM、MTM($k=3$)和 MTM($k=5$)算法。

采用 RWM 和 MTM($k=5$)算法得到的各参数样本的自相关函数显示,对于几乎所有参数,MTM($k=5$)抽取样本的自相关性在 50 阶以后就已经接近于0;对于 r^A、π^A、$100\sigma_R$ 和 $100\sigma_z$ 等参数,在 30 阶以后自相关性就已经很弱。而使用 RWM 抽样时,ψ_1、ψ_2 和 ρ_R 等参数的自相关性在 100 阶以后仍然很明显。

(四)模型选择

在实际应用中,研究人员通常会面临模型不确定的问题。本部分对前文中基准模型的货币政策设定进行修改得到其余两个模型,然后比较这几个模型对中国宏观数据的拟合效果。模型 1 为前文构建的基准模型,由式(4)~(11)组成。在模型 2 中,将式(8)中的货币政策更改为前瞻的泰勒规则:

$$R_t^* = r\pi^*\left(E_t[\pi_{t+1}]/\pi_t^*\right)^{\psi_1}\left(E_t[Y_{t+1}]/Y_t^*\right)^{\psi_2} \quad (12)$$

前瞻泰勒规则有助于稳定公众预期、提升社会福利,且被认为与中国货币政策制定与实施的实际情况更加一致。在模型 3 中假设式(8)中的通货膨胀目标是固定的且等于通胀的稳态($\pi_t^* = \pi$)。此外,假设政府部门支出总产出 Y_t 中比例为 $\zeta_t \in [0,1]$ 的部分,其预算约束变更为:

$$P_t\zeta_tY_t + R_{t-1}B_{t-1} = T_t + B_t + M_t - M_{t-1} \quad (13)$$

令 $g_t = 1/(1-\zeta_t)$,$1/g = 0.85$,并设:

$$\ln g_t = (1-\rho g)\ln g + \rho g \ln g_{t-1} + \varepsilon_t^g \quad (14)$$

综上,模型 2 包括式(4)~(7)和式(9)~(12),模型 3 包括式(4)~(8)、式(11)、式(13)或(14)。

我们设模型指标的先验分布为均匀分布: $\Pr(M_i) = 1/3$,$i = 1,2,3$。其后使用 MTM($k=5$)算法对每个模型分别进行后验抽样,并用第四部分中的方法来计算边缘似然和贝叶斯因子,结果列于表 4。可以看出尽管 $B_{12} > 1$,但模型 1 相对于模型 2 的优势非常微弱,很难对这两个模型的优劣程度做出评判;而 $B_{13} = 200.3 > 150$,说明相对于模型 3,数据对模型 1 有强烈的支持。通过比较可以发现带时变通货膨胀目标的两个 DSGE 具有更好的实际数据拟合效果,模型中货币政策的设定更加接近中国的实际情况。

表 4　边缘似然与贝叶斯因子

模型	特征	对数边缘似然	贝叶斯因子(B_{1i})
1	时变通货膨胀目标	−392.9	1
2	前瞻泰勒规则 − 时变通货膨胀目标	−393.4	1.6
3	固定通货膨胀目标 − 政府支出	−398.2	200.3

六、结论

针对现有方法的不足,本文提出了一个高效的 MCMC 方法对非线性 DSGE 模型进行贝叶斯参数估计和模型选择。数值实验显示出该方法在后验抽样中具有更高的计算效率,能够更加准确地进行参数估计和计算贝叶斯因子。通过使用不同模型来拟合中国宏观经济数据,显示出该方法可以很方便的用于对具有不同设定的宏观经济模型进行比较。

参考文献

[1] S An, F Schorfheide. Bayesian analysis of DSGE models [J]. Econometric Reviews, 2007(2-4): 113-172.

[2] 杨农，郭辉铭. 动态随机一般均衡模型理论与实证研究进展[J]. 经济学动态，2013(8): 112-120.

[3] J Fernández-Villaverde, J F Rubio-Ramírez, M S Santos. Convergence properties of the likelihood of computed dynamic models [J]. Econometrica, 2006(1): 93-119.

[4] G Amisano, O Tristani. Euro area inflation persistence in an estimated nonlinear DSGE model [J]. Journal of Economic Dynamics and Control, 2010(10): 1837-1858.

[5] S Chib, S Ramamurthy. Tailored randomized block MCMC methods with application to DSGE models [J]. Journal of Econometrics, 2010(1): 19-38.

[6] E Herbst, F Schorfheide. Sequential Monte Carlo sampling for DSGE models [J]. Journal of Applied Econometrics, 2014(7): 1073-1098.

[7] S Schmitt-Grohé, M Uribe. Solving dynamic general equilibrium models using a second-order approximation to the policy function [J]. Journal of Economic Dynamics and Control, 2004(4): 755-775.

[8] J Fernández-Villaverde, J F Rubio-Ramírez. Estimating macroeconomic models: A likelihood approach [J]. The Review of Economic Studies, 2007(4): 1059-1087.

[9] M K Pitt, R S Silva, P Giordani, et al. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter [J]. Journal of Econometrics, 2012(2): 134-151.

[10] J S Liu, F Liang, W H Wong. The multiple-try method and local optimization in Metropolis sampling [J]. Journal of the American Statistical Association, 2000(1): 121-134.

[11] M M Andreasen. Non-linear DSGE models and the central difference Kalman filter [J]. Journal of Applied Econometrics, 2013(6): 929-955.

[12] S Chib, I Jeliazkov. Marginal likelihood from the Metropolis-Hastings output [J]. Journal of the American Statistical Association, 2001(1): 270-281.

[13] Y Yang, L Wang. An improved auxiliary particle filter for nonlinear dynamic equilibrium models [R], Working paper, 2015.

[14] 隋建利，刘金全，庞春阳. 基于太阳黑子冲击视角的中国货币政策有效性测度[J]. 管理世界，2011(9): 40-52.

[15] 简志宏，朱柏松，李霜. 动态通胀目标，货币供应机制与中国经济波动——基于动态随机一般均衡的分析[J]. 中国管理科学，2012(1): 30-42.

[16] G O Roberts, J S Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms [J]. Statistical Science, 2001(4): 351-367.

作者简介

杨远，男，1987 年生，云南大理人，2010 年毕业于武汉大学，获经济学学士学位，现为厦门大学王亚南经济研究院统计学专业博士研究生。研究方向为状态空间模型的贝叶斯推断。

林明，男，1978 年生，福建福州人，2005 年毕业于北京大学数学科学学院，获应用数学博士学位。现为厦门大学王亚南经济研究院、经济学院副教授，博士生导师，经济学院统计系副主任，福建省统计科学重点实验室副主任。研究方向为蒙特卡罗方法、贝叶斯统计。

（责任编辑：曹 麦）

《统计研究》"统计图版式"要求

基本要求

1. 图的宽度大致以 16 开刊物的分栏宽度为准，长宽比例以美观适度为宜。
2. 图不取外边框。
3. 图内外文字或数字取 6-8 号为宜，宋体。
4. 图底色为白色。
5. 坐标刻度单位如为年份，请用"1997"，勿用"1997 年"或"97"或"97 年"；不得竖排或斜排。如年份较多，可以按间隔一年或数年的方式排出。
6. 坐标刻度单位如为变量值，一般不宜带有小数位，如"10000.0"；最多为 5 位数字，如需用 6 位数字标示，如"100000"，请将其变换为"10"，并改变相应的计量单位。图中标志值一般最多带两位小数，在特殊情况下，如整数位为 0 时，最多可带 4 位小数。
7. 图标题设在图外的图下方，小五号黑体居中，不得设在图内。
8. 图的标注放在图和图标题之间，以图外框左边界为准前空两格，小五号宋体。
9. 图的序号用"图 1"表示，勿用"图 1:"、"图一"、"图 1-1"等，文中只有 1 张图时，可用"图 1"，亦可用"图"。