J_Pfeifer： Full information estimation of linear DSGE models

J. Pfeifer: Full information estimation of linear DSGE models

2021 年 Dynare 线上暑期学校中，Johannes Pfeifer 教授介绍了“线性 DSGE 模型中的完全信息估计 Full information estimation of linear DSGE models”，youtube 链接为：https://youtu.be/GJjSwnikwW4?si=Omt0IKWXNckroJEc

视频中介绍了关于参数估计的原理、诊断、注意事项等，也介绍了先验分布选取的标准，值得学习。

目录

dynare 参数估计的模型文件 Mod file

一个典型的 Dynare 模型估计结构 The structure of a typical Dynare mod-file

在用于生成 IRF 的模型模拟基础上，模型估计添加了以下几个部分用于参数估计。

1. 定义所需估计的参数和其先验分布。

例如下面定义了用于估计政府支出和 TFP 冲击的的自回归系数和标准误

```
estimated_params;
rhog, beta_pdf, 0.7, 0.1;
rhoz, beta_pdf, 0.7, 0.1;
stderr eps_z, inv_gamma_pdf, 0.01, 0.1;
stderr eps_g, inv_gamma_pdf, 0.01, 0.1;
end;
```

注意，估计时可以选择贝叶斯估计 Bayesian estimation，亦可选择最大似然估计 maximum likelihood estimation。若采用贝叶斯估计，此时需要在定义所需的参数后加上先验分布的设定，如 beta_pdf；若采用最大似然估计，那只需要去掉这里先验分布设定即可。**通过加入或去除这里先验分布，可以在贝叶斯估计和最大似然估计之间切换。**

但不能混合：在某些参数上贝叶斯估计，在某些参数上最大似然估计。**要么完全贝叶斯估计，要么完全非贝叶斯估计。**

1. 声明观测变量

参数估计就是基于观测到的数据，寻找合适的参数，因此需要声明观测到的数据对应于哪些变量。

```
varobs g_obs c_obs;
```

这行代码同样适用于如识别 identification 时对观测变量的声明。

1. 估计指令 用于 1）估计模型；2）对数据进行平滑处理，如卡尔曼平滑 Kalman smoother；3）分析贝叶斯脉冲响应、矩、生成预测等

```
estimation(datafile=first_diff, mode_check, mh_replic=200000, mh_nblocks=1, smoother, moments_varendo);
```

指令 datafile = 需要告知观测变量 g_obs c_obs 所存储的文件位置； **mode_check 用于检查计算出的 mode 是否正确，最好总是执行这个检查**；mh_replic 设定 number of replications for Metropolis-Hastings algorithm， **通常这个数值越大结果越可靠，但需要结合 trace plot 进行收敛性诊断**；mh_nblocks 设定估计过程中 Markov chain 的数量，后续介绍中会提及，影响收敛性诊断可以采用的方法（如有些诊断方法只需要一条 chain，有的需要两条或以上）；smoother 指示运行 Kalman smoother；moments_varendo 显示给定数据下，内生变量理论矩的后验分布的计算。

关于 Kalman smoother 的 comments。DSGE 模型的解基于状态空间型 state space form，模型的解实际上是寻找一个状态变量的 policy rule/policy function，因此我们需要知道状态变量 evolve 的过程。估计存在的一个问题是我们有时**无法观测到状态变量**。一个例子是资本存量。为求解模型我们需要这些无法观测的状态变量，此时就需要一些 filter：一个滤波 filter 是一种对数据的观测方式，用处理后的数据可以对状态变量进行反推 inference。在当下的对数线性模型中，冲击都是正态分布的，可以证明**卡尔曼滤波**是最优的 filter。

这个过程类似于 inner regression。首先，观测到数据，对 state 进行猜测。基于猜测的 state，对观测变量进行预测，将预测值和数据（actual realization）进行比较，产生预测误差 forecast error；基于这个预测误差，我们调整对 state 的估计 estimate。

比如说，我认为资本存量比较低（guess the state）。但观察到产出比较高（observation）；基于资本存量比较低的观测，可以预测产出比较低（forecast），因此这一较低的产出预测低估了实际的高产出（forecast error）。这有可能是我们开始对资本存量的估计实际上也低估了，资本存量实际应该更高（adjust the guess of state）。Kalman filter 基于观测数据，提供了一个很好的更新状态的方式。

Kalman filter 基本上是，观测到今天的可观测变量（如产出），推断今天的不可观测的状态变量（如资本）。但有时我们并不关心今天状态变量的猜测，或许我们关心，基于今天看到的产出，我对**过去（如十期前）的** 状态变量的猜测是多少。你可以用今天的观测对过去的变量进行更新；你也就相当于说，可以用**所有的信息**——而不只是**直至今天的信息**——来看状态变量最可能的取值。

filter 和 smoother 的差异 filtering，只采用直至今天的数据和信息。采用所有的信息（包括未来的信息），更新状态变量的猜测，这就是 Kalman smoother。因此，这里的 smoother 可以做的就是，基于所有的数据，告诉我们模型中状态变量（如 TFP 冲击）最有可能的值是多少。

1. 诊断估计是否成功 Diagnostic plots

```
generate_trace_plots(1);%生成第一条chain的trace plot
mh_autocorrelation_function(options_,M_,estim_params,'StructuralShock',1,'eps_z');%如果是模型内参数，type
```

这是对 MCMC 后验抽样方法产生的序列相关问题进行的诊断，后文将提及。这是 MCMC 算法的不足，它不能过于严重，否则将影响估计的准确性。

1. 生成 historical shock decomposition

```
shock_decomposition;
```

Q：如果我先 simulate 了状态变量作为 observations，那么还可以进行估计吗？ A：如果参数识别良好，那么估计结果应该和 simul 时选择的参数一样。但这对你没啥帮助，不会对参数估计有什么新的信息。

参数依赖 Parameter dependence

Always keep track of independent parameters in your model.

尽管对 simulation 无害，但在估计时要特别重视的，不要对复合参数（依赖于其他独立参数）估计，只去估计那些独立参数。

具体来说：Never define a function of estimated parameters as an independent parameter。也就是说，不要把这些复合参数放在 mod 文件开头的参数声明里。

NK 模型中的经典例子是 Calvo 定价下菲利普斯曲线的斜率$\boldsymbol{\kappa}=\frac{(1-\theta)(1-\theta\beta)}{\theta}$，这里$\boldsymbol{\kappa}$依赖于贴现率$\beta$和 Calvo 参数$\boldsymbol{\theta}$的取值。

不能在 model block 前就声明参数$\boldsymbol{\kappa}$。因为如果想要去估计$\boldsymbol{\theta}$，就要看$\boldsymbol{\theta}$变化后$\boldsymbol{\kappa}$及其他变量的变化。但如果 hard code $\boldsymbol{\kappa}$，那么无论$\boldsymbol{\theta}$取什么值，$\boldsymbol{\kappa}$都将卡在初始设定的值上，不会被更新，这样的估计是错误的，**你甚至不会收到报错信息，但你的估计是错误的。**

两种解决办法

1. 将$\boldsymbol{\kappa}$定义成 model local variable（#-operator），例如

```
#kappa = (1-theta)*(1-theta*beta)/theta;
```

这样，κ得到了正确的更新。

1. 第二种方式是用 steady state file 来更新 dependent parameter。

将观测变量和模型中的状态变量映射起来

模型中的状态变量需要是平稳的，但实际上观测变量可能不平稳，可能有 trending。

解决办法：1）对数据取增长率（first difference filter），和模型中的增长率变量映射起来；2）用 filter 对数据进行 detrend，将其与模型中的偏离平衡增长路径的离差映射起来。

不要使用 two sided filter like the HP filter，这些 filter 利用了未来的信息，但模型是状态空间型，当前的变量只反映当下信息。do not use two-sided filters like the HP filter because it conflicts with recursive nature of model solution

选择对应的 filter 关键取决于研究者对趋势本身的理解。建议阅读：Pfeifer (2013): A Guide to Specifying Observation Equations for the Estimation of DSGE Models（https://drive.google.com/file/d/1r89OU5OE3CBa6tOlj6l3hNVWEaRH5Anv/view?usp=sharing）后续或许会推送。

为什么要采用贝叶斯估计？The Problem

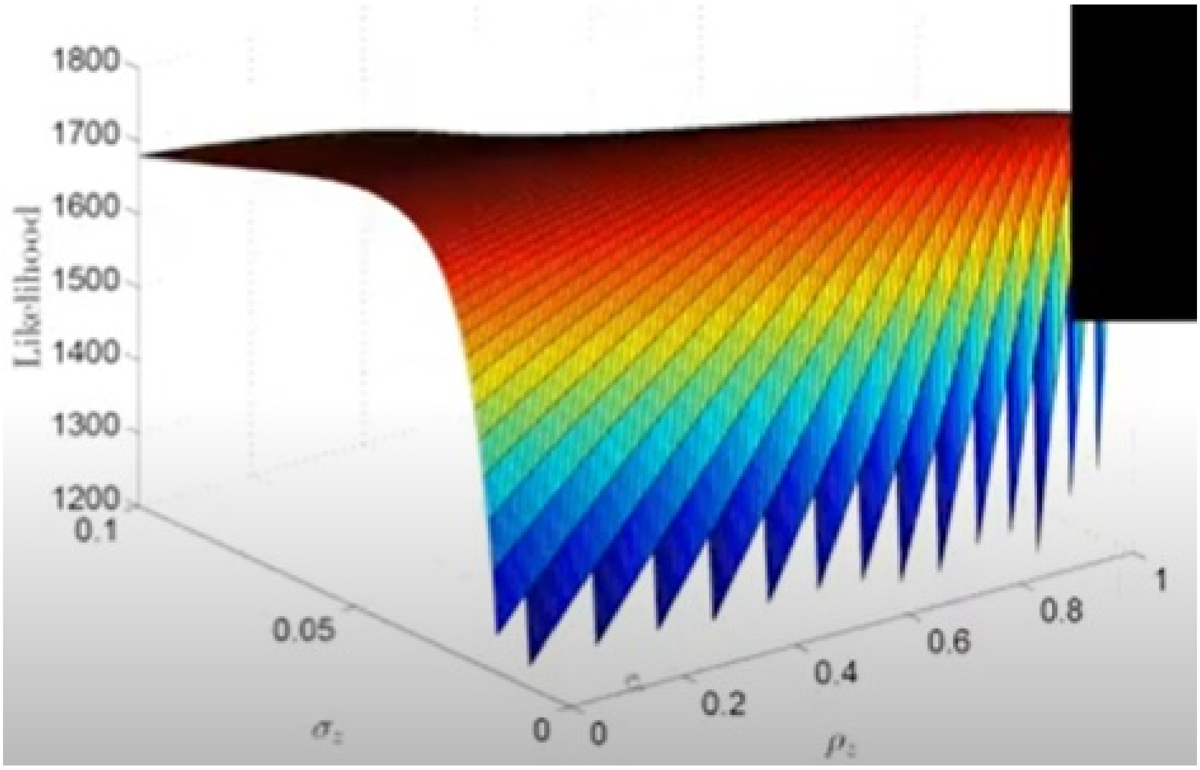
复习 Review

对数线性 DSGE 模型解是 state-space form；但一般并非所有状态变量都可观测，因此采用 Kalman filter 来处理不可观测状态，Kalman filter 将产生一个 likelihood $\mathscr{L}(y^T|\theta)$，其中θ是 deep parameters 向量，y^T是观测的完整数据。

估计 estimating 意味着：找到最大化$\mathscr{L}(y^T|\theta)$的θ。听起来是一个最大似然估计，但有什么困难呢？

Likelihood 似然函数

似然函数是高维的，在参数较多的估计时更是如此。以下图两维为例，在标准差方向，标准差的变动使得似然函数具有较大的变动。这是识别的前提，但反观自相关系数，它的变动是的似然函数变化很小，因此，用数据识别自相关系数非常困难。



高维、缺乏曲度、局部最优，都是最大似然估计面临的麻烦。局部最优就像是，当你爬上上海最高的山，再无法向上攀登时，你可能离最高峰还远得很。

“Dilemma of absurd parameter estimates”(An and Schorfheide 2007)：最大似然估计在给定信息估计时表现奇差often at odds。

Bayesian vs. Frequentist Philosophy

贝叶斯学派和频率学派在计量哲学上有根本性差异。大多数宏观经济学家不是贝叶斯信念者，而是实用主义的信守者，两类方法通常出现在一篇论文中：哪个好用用哪个。

Adopting Bayesian techniques makes our lives easier.

历史上贝叶斯计量因为计算力不足而长期居于次流，但用贝叶斯估计的优势很多，见 Berger and Wolpert 1988，Sims 2007。

两种学派对概率概念的不同见解 concept of probablility

贝叶斯学派的概率通常与信念的强度、知识或认识的程度有关；频率学派的概率通常基于假想的可观测系统的特征，即事件在“较长时间”的频率。

举例来说，考虑两个学派对以下断言的理解：一枚公平的硬币正面朝上的概率是 50%。贝叶斯学派将认为：一枚硬币投掷结果的信念在正面朝上或反面朝上之间均等分配；频率学派则认为，50% 概率是假想无限次抛硬币后正面朝上的频率。

频率学派要求从长期事件发生的相对频率来推断概率，但问题就在于长期“无穷次”是不可行的，不属于研究者的实际兴趣。

因此，贝叶斯学派可以将概率应用于任何与信念和知识相关的东西，比如一个猜想、统计参数、或整个统计模型。

两种学派在条件概率上的不同见解 conditioning set

频率学派对基于一个估计量，$\hat{\beta}$，来估计样本的分布，基于此来做统计推断。

频率学派假设存在一个真实参数β。但其统计推断都基于估计量$\hat{\beta}$，而非真实参数β本身。频率学派只能对估计量量$\hat{\beta}$进行概率性陈述（有 95% 的概率$\hat{\beta}$不为 0），无法直接对真实参数β进行概率性陈述。

如果可以重复无数次，那么频率学派基于估计量而估计的样本分布可以模拟我们所关注的观测到的效应。我们关注，就基于我当前的估计量来说，我应该观测到什么。因此，通过样本分布，数据被认为是承载不确定性的随机变量；参数有一个固定不变的、真实的取值，因此条件概率条件于参数：$P(Y|\theta)$，给定真实参数，观测数据发生的概率。

这个概率不是针对参数的陈述，而是针对数据的陈述——但当我想要估计参数时，这并非我所感兴趣的。

当频率学派进行零假设显著性测试（null hypothesis significance test， NHST），就是对一个我们并不相信的假设进行检验。我们总想去拒绝零假设：我们认为真实参数不为 0，但我们零假设是参数等于 0（我们并不相信如此）。基于这样的我们所不相信的零假设为真，给出一个存在效应的概率 p-value， 但通常零假设却不为真——我们依赖于永不会发生的假想的重复实验。相当吊诡。

贝叶斯学派对效应或参数基于观测到的数据直接进行概率性陈述。认为参数是随机的、并认为数据是固定不变的，因此条件概率条件于数据：$P(\theta|Y)$。

后验概率密度 the posterior density

贝叶斯估计的中心对象是条件于观测数据Y，参数θ的后验概率

$$P(\theta|Y)$$

后验概率是一个 fully-fledged 概率密度函数，和正态分布密度函数一样，积分后等于 1；但通常后验分布的形式是未知的。事实上，要完全确定一个概率密度函数，我们可能需要无穷阶矩。换言之，我们面对的是一个无穷维的函数，我们对其几乎无所知，这就是贝叶斯估计的挑战性。

但如果我们知道了后验分布，我们就可以对真实参数作如下陈述：有 90% 概率财政乘数在 0.3-1.5 之间——恰是我们所需要的。零假设的显著性测试 NHST 无法做出如此陈述。拒绝零假设并不能让我们对感兴趣的效应有任何描述，我们只知道，在零假设为真的世界，观测数据大概是不会出现的。

如何求后验分布 Posterior

贝叶斯法则 Bayes Rule

条件概率意味着

$$p(A,B) = p(A|B)p(B)$$
$$p(A,B) = p(B|A)p(A)$$

其中，p(A,B) 是联合概率 joint prob，p(AIB) 是条件概率 conditional prob，p(B) 是边际概率 marginal prob。

由此，有贝叶斯法则

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

用观测数据y^T估计参数θ时，我们有

$$p(\theta|A) = \frac{\mathscr{L}(y^T|\theta)p(\theta)}{p(y^T)}$$

其中$p(\theta)$是先验分布 prior，$\mathscr{L}$只是分布的一种写法，和 p 是相同的符号标识。$p(\theta|A)$被称为后验分布 posterior distribution，这是我们对参数θ所知的所有信息，基于我们观测到的数据y^T，以及从先验分布$p(\theta)$中提取的非数据的信息。

$p(y^T)$称为**边际数据密度** marginal data density，是观测到数据的概率，注意，$p(y^T)$与参数θ无关 (对每一个θ，$p(y^T)$都是相同的常数)，也不是我们所关注的。因此，边际密度可以用比例常数处理，我们实际有：

$$p(\theta|A)=\frac{\mathscr{L}(y^T|\theta)p(\theta)}{p(y^T)}\propto \mathscr{L}(y^T|\theta)p(\theta)$$

Posterior is proportional to likelihood times prior!

频率学派关注似然函数$\mathscr{L}(y^T|\theta)$；贝叶斯估计中，先验分布$p(\theta)$是新增的内容，以用于推断后验分布$p(\theta|A)$。

$p(\theta)$被称为**先验分布**，它与数据y^T无关，是我们观测数据前，就对参数拥有的全部的非数据信息。

比如说，贴现率一定在 0-1 之间（关于边界的信息），甚至一般来说，应该在每个季度 1-2% 左右（关于期望和方差的信息）。贝叶斯估计就是用过设定先验 prior 来纳入关于参数的所有数据外信息，你可以从经济直觉中找，也可以从过往文献中归纳信息。

先验分布$p(\theta)$直接进入后验分布，比方说，我的先验告诉我，一个季度名义利率达到 10% 是概率为 0，无论数据如何。那么这个先验乘以 likelihood 得到的后验也一定是 0。因此，先验分布提供了对 likelihood 的修正或约束，或是有助于排除一些不大合理的情况（通过降低其后验概率）。

【注意】边际数据密度$p(y^T)$尽管可以被当成比例常数去掉，但在**模型比较** model comparison 时依然发挥作用。如果我们进行模型比较，就不只是条件于数据，而是条件于整个模型。此时，我们需要一个给定（条件于）模型的数据的分布，如果$p(y^T|model1)>p(y^T|model2)$，我们将认为，model 1 下观测到数据的概率更大，model 1 就在模型比较中胜出。 Koops 2003 的教材第一章介绍了如何比较模型。

通常，我们用卡尔曼滤波得到似然函数，接下来的问题就是，如何计算$p(\theta|A)$，如何设定先验分布。

贝叶斯估计的挑战 The problem with Bayesian estimation

为什么贝叶斯估计有这么多优点，但只是在最近 15 年才开始流行？简单来说，这是因为过去计算力不足。

$p(\theta|A)$是一个 full distribution，通常 *not analytically tractable*。我们或许需要无穷阶矩，但请看

$$E(\theta|y^T)=\int\theta p(\theta|y^T)d\theta$$

and

$$\begin{aligned}var(\theta|y^T)&=E(\theta^2|y^T)-\left[E(\theta|y^T)\right]^2\\&=\int\theta^2p(\theta|y^T)d\theta-\left[E(\theta|y^T)\right]^2\end{aligned}$$

即使一阶矩，我们也需要积分运算——对于未知的分布$p(\theta|A)$，这是无法完成的运算。二阶也同样如此。

估计均值 Getting the mean

介绍一个精巧的想法 *an ingenious idea*。通常我们关注以下形式的积分

$$E(g(\theta)|y^T)=\int g(\theta)p(\theta|y^T)d\theta$$

就算我们不知道概率密度是多少，但可以抽样——蒙特卡洛的思想。我们从后验分布$p(\theta|y^T)$中独立同分布 iid 地抽取θ，应用大数定律，

$$E(g(\theta)|y^T)=\int g(\theta)p(\theta|y^T)d\theta\approx\frac{1}{S}\sum_{s=1}^{S}g(\theta_s)=\hat{g}_S$$

抽取随机θ_s，计算 g(\theta_s)，取平均，这就是蒙特卡洛积分 Monte Carlo integration，用 S 次抽取后计算的平均数代替积分。

如果独立同分布地抽取次数足够多，那么蒙特卡洛积分可以“足够好的”逼近原积分。

标准误的大小 numerical standard error

这类估计精确度如何？从中心极限定理，

$$\sqrt{S}\left(\hat{g}_S - E(g(\theta)|y^T)\right) \to N(0, \sigma_g^2)$$

因此，

$$E(g(\theta)|y^T) \sim N\left(\hat{g}_S, \frac{\sigma_g^2}{S}\right)$$

其中，标准差 $\sigma_g^2 = var(g(\theta)|y^T)$ 可由蒙特卡洛模拟样本的方差来近似估计。

数值标准误 Numerical standard error NSE（样本标准误）

$$\frac{\sigma_g^2}{\sqrt{S}}$$

可以作为一个渐进标准误的度量 measure of approximation error.

收敛速度看，抽取数目变为原来四倍，精确度将变为原来两倍（$\sqrt{S}$）。这也意味着，要达到指定精度，需要的抽取数目可能非常大。但大多数，对于当下的计算机而言还是可行的 feasible。

一些其他术语 Some more jargon

即使你不是一个贝叶斯信念者，下面的术语你大抵也听说过： 定义$\omega = g(\theta)$为定义在某个区域Ω上的某个参数θ的函数向量，那么可信集 credible set 可以定义如下

> 可信集 Credible Set
> The set $C \subseteq Omega$ is a $100(1-\alpha)\%$ credible set with respect to $p(\omega|y)$ if
> $$p(\omega \in C|y) = \int_C p(\omega|y)d\omega = 1-\alpha$$

给定数据，参数在C集合内的概率为$1-\alpha$。

Confidence intervals are derived from frequentist statistics, while credible sets come from the Bayesian framework. 置信区间是频率学派采用的概念，可信集则是来自于贝叶斯学派。

值得注意的是，给定α，$100(1-\alpha)\%$ 可信集不止一个！我们往往最关注的是最小的可信集，这才对应频率学派的置信区间。

> 最大后验密度的可信区间 highest posterior density interval HPDI
> A $100(1-\alpha)\%$ highest posterior density interval for ω is a $100(1-\alpha)\%$ credible interval that has a smaller area than any other $100(1-\alpha)\%$ credible interval for ω.

dynare 结果中的最后两列通常报告的就是 HPDI，这也就是说我们不能错用概念：在贝叶斯估计中，说估计的参数置信区间、可信集、可信区间这样的概念都是误用，我们最关注的是最大后验密度的可信区间。

更多的问题：后验抽样 problem everywhere: posterior sampling

蒙特卡洛积分看起来很棒！可是等一下，怎么从一个未知的、不可求的分布抽样？无法抽样的话后面的所有估计都无从实施。

我们需要的是一类后验分布抽样算法 posterior smapling algorithms，这就是一类从未知分布中抽样的技巧。有三个重要的方法：

1. 重要度抽样 importance sampling

非常高效，但需要很多额外的关于 target distribution 的信息，以至于有时很不可行。

1. Gibbs-Sampling (S. Geman and D. Geman 1984)

在时变向量自回归 time varying vecter autoregression 的估计中常用。Gelman 1992 证明了 Gibbs-Sampling 是 MH 算法的特例。

1. Metropolis-Hastings algorithm (Hastings 1970; Metropolis et al. 1953)

在 DSGE 中常用。

然而，后二者的随机抽样都存在序列相关性问题，属于一个名为蒙特卡洛马尔科夫链 Monte Carlo Markov Chain MCMC 的算法类；而蒙特卡洛积分需要独立同分布 iid 的抽取。

序列相关性有什么问题呢？抽了一个 0.75，序列相关意味着下一次抽取总在 0.75 附近，那么更远处的抽取就没有了，对于建立更远处随机变量的概率分布来说，可用的数据（信息）太少了，大大增加了所需抽样的数量。

先验分布 Prior $p(\theta)$

先验分布和主观判断 prior distributions and subjectivity

Berger 2006 介绍了很多关于先验分布反映研究者主观判断的讨论。许多人认为，先验选取的“主观性”与科学方法的要求不一致，因此该问题本身尚存在争议。但这里我们不考虑其背后的哲学问题。

主观的先验分布选择是必要的，但关键在于，**模型本身比参数的先验分布更重要**！

设定先验 first thing first: specifying the prior

先验分布$p(\theta)$本身是一个完全的多变量分布 multivariate distribution。实际上，人们**通常让各个参数的先验分布彼此独立**，Andrle and Benes 2013 中先验分布是一个例外。功能上，dynare 中的 estimated_params 是可以设定两个变量之间的相关系数的。

先验分布的目的是什么？

1. 引入关于参数的非数据的信息，这通常来自经济直觉、现实约束或过往文献。比如参数范围就是确定先验分布类型的重要来源。

2. 为似然函数提供更大的曲率（平坦的似然函数是难以识别参数的重要原因）。

在选择先验后，总应该做一个先验预测 prior predictive check（根据先验生成数据，以评估先验是否合适），检查这个先验对研究问题的意义、这个参数先验分布看起来科学吗？（见 Leeper, Traum, and T. B. Walker 2017）

内生先验分布的问题 Endogenous Priors

由于我们在做的是全信息贝叶斯估计，而不是简单的对几阶矩进行匹配，因此可能结果中某些矩的匹配并不良好。

有时 referee 会 argue 二阶矩（variance）不够匹配，尤其是估计的模型方差太大（problem of high variance）。

这时，Christiano, Trabandt and Walentin 2011, Del Negro and Schorfheide 2008 提出了一种名为 **内生先验分布 endogenous prior** 的思路。具体来说，在独立先验的估计后，用 pre-sample 中观测到的标准差来更新初始先验：将初始先验和 pre-sample 似然函数的标准差相乘作为新的先验。

毫无疑问，因为引入了二阶矩的信息，它是起作用的：对二阶矩来说结果匹配的更好了。但哲学上（本质上）这个先验是和数据相关的，这不符合贝叶斯估计的精神！

dynare 中的 endogenous_prior 选项可以采用这种方法，但 Pfeifer 不建议这样做。

几种先验分布

主要是根据参数自己的范围、有经济意义的范围来确定分布，boundary 很关键。

	normal	**uniform**	**beta**	**gamma**	**inverse gamma**
support	$(-\infty,\infty)$	$[a,b]$	$[0,1]$	$[0,\infty)$	$(0,\infty)$
mean	0	$\frac{a+b}{2}$	μ	μ	μ
SD	σ	omitted	σ	σ	σ
symmetric	yes	yes	not necessary	no	no
typical use	feedback parameter with unknown sign	knows only the boundary	贴现率、calvo 参数等有界参数	单侧有界，左闭右开，方差，泰勒规则的反应参数	单侧有界，左开右开

值得注意的是，均匀分布 uniform 虽然没有引入任何曲度，但这个分布既指出了参数的边界，也指出在边界内各处概率都均等，因此虽然被称为 uninformtive prior，但本质上是很 informative 的。此外，在模型比较时，不要使用边界为无穷的均匀分布，会有积分问题。

Uniform Prior

- For a variable $Y \sim U(a,b)$, the PDF is given by

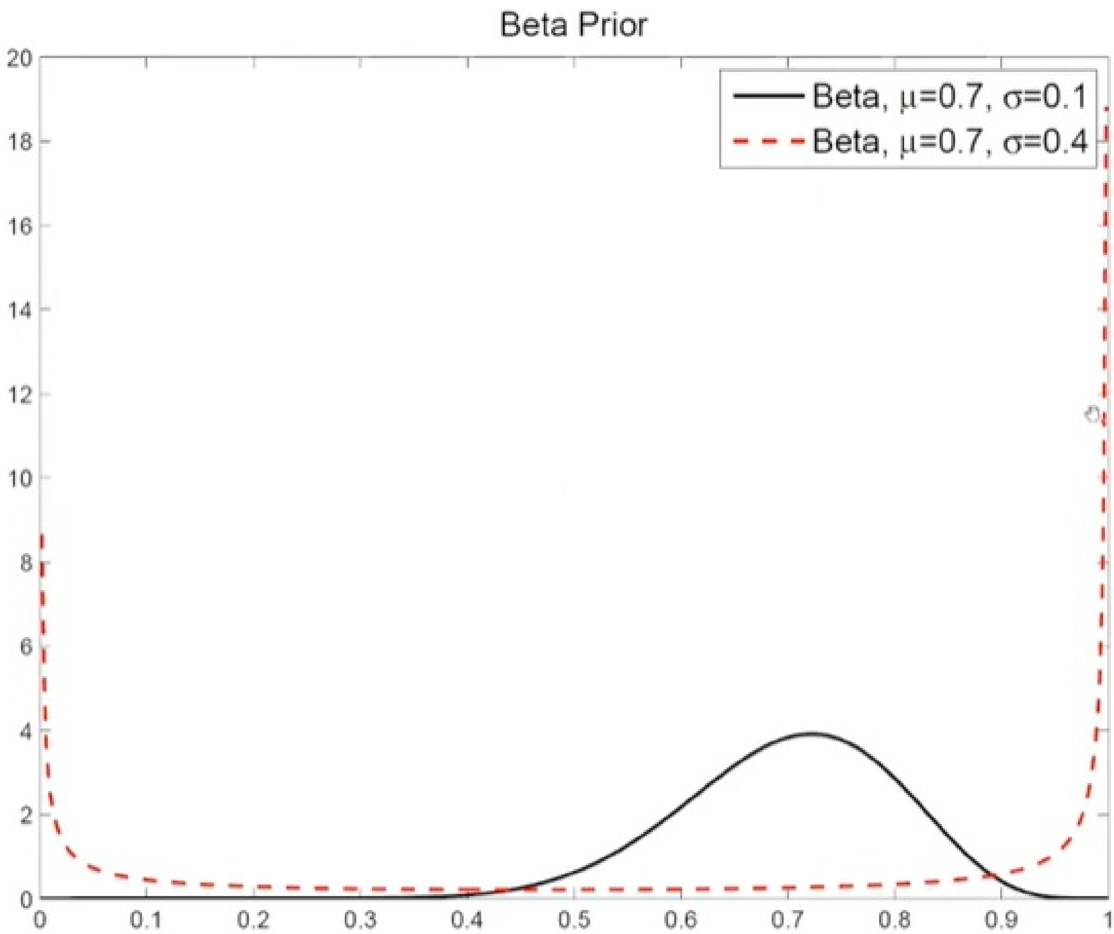
$$f_U(y|a,b) = \begin{cases} \frac{1}{b-a} & \text{if } a \leqslant y \leqslant b \\ 0 & \text{otherwise} \end{cases}$$

where $-\infty < a < b < \infty$

- Moreover,

$$E(Y) = \frac{a+b}{2}$$
$$\mathrm{var}(Y) = \frac{(b-a)^2}{12}$$

beta 分布是边界在 0 和 1 内的，因此在均值两边，适当的方差可以让分布在均值附近聚集，但过大的方差将把分布向两端挤过去，这通常不是我们所需要的。如果发现一个 beta 分布的先验有两个 mode，（dynare 会汇报这一点），那么降低标准差是一个需要做的事情。



有界的 beta 分布可以拓展到边界不是 0，1 的 generalized beta dist，对应的只需要在 beta 的均值、标准差后补充上下界即可。例如 要让 home bias 在 1，2 之间分布，可以设置：

```
v, beta_pdf, 1.5, 0,25, 1, 2;
```

Gamma 分布是左闭右开的，适合方差。

Gamma Prior

- For a variable $Y \sim G(\mu, \nu)$, where μ is the mean and ν the degrees of freedom, th is given by

$$f_G(y|\mu,\nu) = \begin{cases} \frac{1}{\left(\frac{2\mu}{\nu}\right)^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} y^{\frac{\nu-2}{2}} e^{-\frac{y\nu}{2\mu}} & \text{if } 0 \leqslant y \leqslant \infty \\ 0 \text{ otherwise} & \end{cases}$$

  where $\Gamma()$ is the Gamma function
- Moreover,

$$E(Y) = \mu$$
$$\text{var}(Y) = \frac{2\mu^2}{\nu}$$

- Beware: Matlab uses $f_G(y|a,b)$ with $a = \frac{\nu}{2}, b = \frac{2\mu}{\nu}$
- Thus, use

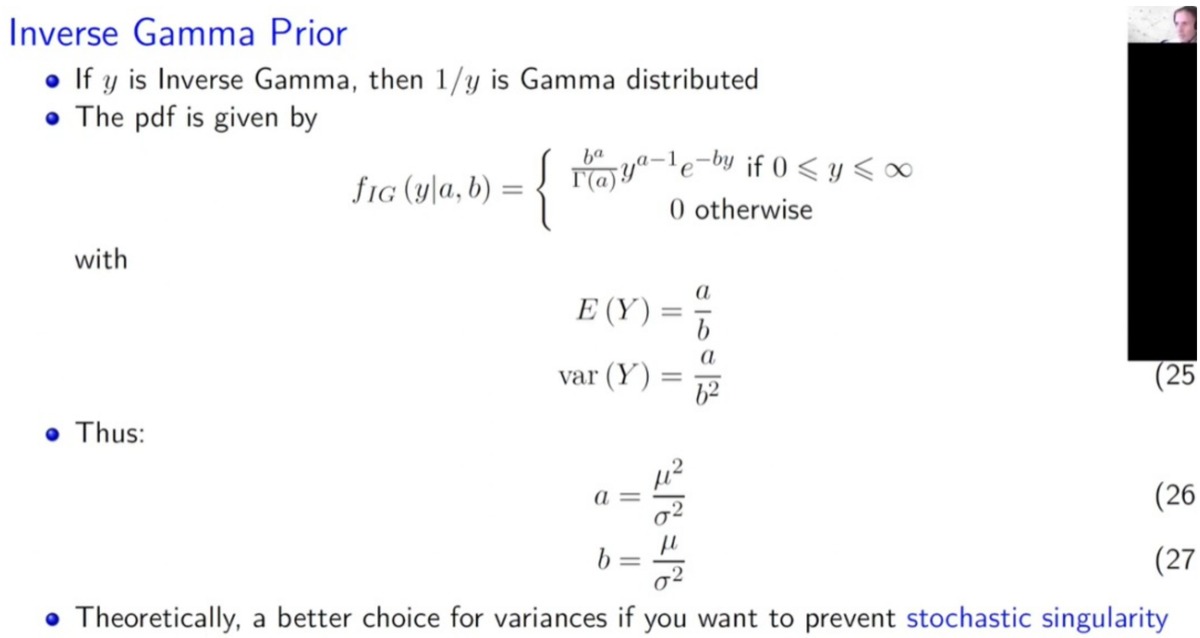
$$a = \frac{\mu^2}{\sigma^2}$$
$$b = \frac{\sigma^2}{\mu}$$

但可能会产生随机奇异 stochastic singularity 问题

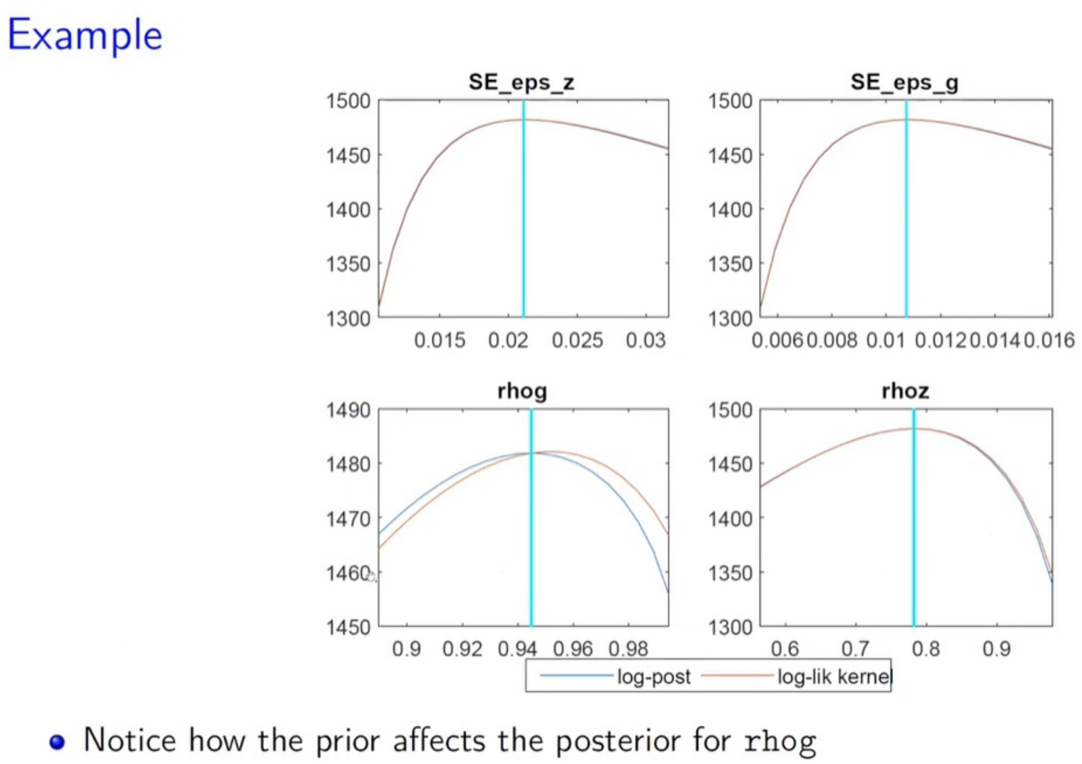
> 随机奇异 stochastic singularity
> 要能够估计参数，你需要至少和观测变量一样多的外生冲击。否则，模型将预测观测变量间是线性相关的，但数据通常并非如此。

Gamma 分布下，若外生冲击的方差取到 0，那就意味着这个冲击是 0，没有变化，就会有线性相关的随机奇异 stochastic singularity 问题（问题不算太大，因为 gamma 分布是连续的，标准差为 0 的概率也是 0）。解决办法是采用左开右开的逆 gamma 分布。

逆 gamma 分布



永远记得检查 prior 是否合理！



这张图告诉我们什么？对于ε_z和ε_g的标准差和ε_z的子回归系数（上、右下），后验分布和似然分布几乎重合，那就意味

着数据信息在后验分布的贡献很大，先验分布几乎没有产生影响。

并不必须要让 posterior 和 likelihood 重合。如果先验没有任何信息量，uniform as uninformative，那么一定重合；如果有无限多的数据，信息很大，那么 prior 的权重也小，也接近重合。

左下角的ε_g自回归系数中，数据（likelihood）分布的均值较大，后验（蓝色，post）较小，说明取了一个均值比较小的先验，并且先验包含的信息显著进入了后验——研究者主观觉得自回归系数应该较小，但数据显示较大，说明政府支出比先验认为的更加持久 persistent。

如果 post 和 likelihood 差的太大，是有什么问题吗？可能并没什么问题，只是 likelihood 对后验贡献的信息太少，数据对估计这个参数没啥用。如果完全没用，那么后验和先验的分布一样。

从特征事实中确定先验，如果不大肯定，就选一个范围宽一点的先验好了。要是真的啥也不知道，uniform。

要是觉得数据不大好，那么映射数据观测变量和模型状态变量时，加一个 measurement error。

再多说一句，存在许多 dogmatic prior 的现象，比如设定对数效用就规避对 risk aversion 的估计（$\sigma = 1$）；或是开放经济中按照文献直接设定$\sigma = 5$而不进行估计。

关键在于，对模型强调的新的机制，大家所不熟悉的，展开估计。

对没把握的，希望数据告诉我的，估计。

其他的，找文献，fix them。

so far，这一篇已经够长了，后面的内容只做 take away summary，原视频仍可在 youtube 学习。

1. 搜寻 mode，algo 6 是 MCMC，时间上效率比较差；algo 8 和 9 提供了全局搜寻 mode 的方法。
2. 对Σ估计中，如果 Hessian 报错非正定，往往暗示模型本身有更深层次问题。
3. 几个参数单独估计不报错；联合估计报错。可能存在识别上的问题：Hessian at the mode may be singular.
4. MH 方法存在自相关问题，过大或过小的 jump step 都会加剧自相关问题。总是检查自相关：mh_autocorrelation_function，trace plot 也是用于检查自相关问题的。

M.H.

omitted

Diagnostics

omitted