Forecasting with Bayesian Vector Autoregressions

Sune Karlsson

一、背景：传统的VAR模型

- 由Sims（1980）提出
- 简单灵活、拟合和预测
- 参数过多→过拟合→系数不符合实际含义、预测效果不好
- 解决方法
    - 用一组数据作为先验减少参数

二、主要内容

- 1、贝叶斯方法 ▲预测
- 2、Beyes理论计算及预测
- 3 、Minnesota先验分布 etc.
    - 简单模型效果好
- 4 、SVAR——a VAR in structural form
- 5 、VECM——Vector Error Correction Form of VAR（误差修正模型）
    - 4和5 施加约束条件近似正确时——better
    - 实际预测中应用较少
- 6、 条件预测——未来事件
    - 政策分析
- 7 、Bayesian VAR models
    - 参数：常数→时变
    - 随机扰动
    - 二者都能提高预测精度，但参数个数急剧增加→过拟合
- 8 、VAR
    - 选择变量→减少参数：Bayesian shrinkage等方法

- 9 、大数据环境
  - 变量多
  - Bayesian VAR 有竞争力

三、模型：

Y对滞后项及其他解释变量回归：

$$Y_{T*m} = Z_{T*k}\Gamma_{k*m} + U_{T*m} \qquad U \sim N(0,\Psi)$$

$$k = mp + d : Y(m\text{个变量，}p\text{阶滞后项}) + \text{其他解释变量}(d)$$

$$L(Y|\Gamma,\Psi) = (2\pi)^{-mT/2}|\Psi|^{-T/2}exp\{-\frac{1}{2}tr[\Psi^{-1}(Y-Z\Gamma)'(Y-Z\Gamma)]\}$$

并且假设Γ服从均匀分布，根据【Jeffreys' prior】有 $\pi(\Gamma,\Psi) \propto |\Psi|^{-(m+1)/2}$

应用贝叶斯公式：$p(a|b) \propto p(b|a)p(a)$

$$\begin{aligned} p(\Gamma,\Psi|Y_T) &\propto p(Y_T|\Gamma,\Psi)p(\Gamma,\Psi) \\ &\propto |\Psi|^{-T/2}exp\{-\frac{1}{2}tr[\Psi^{-1}(Y-Z\Gamma)'(Y-Z\Gamma)]\}|\Psi|^{-(m+1)/2} \\ &\propto |\Psi|^{-T/2}exp\{-\frac{1}{2}tr[\Psi^{-1}(Y-Z\widehat{\Gamma})'(Y-Z\widehat{\Gamma})]\} \\ &\quad \times exp\{-\frac{1}{2}tr[\Psi^{-1}(\Gamma-\widehat{\Gamma})'Z'Z(\Gamma-\widehat{\Gamma})]\}|\Psi|^{-(m+1)/2} \end{aligned}$$

其中应用到OLS估计: $\widehat{\Gamma} = (Z'Z)^{-1}Z'Y$

$$\rightarrow p(\Psi|Y_T) \ \& \ p(\Gamma|Y_T,\Psi)$$

对Y_{T+H}进行预测时，如果H值较大，写出其分布情况很难，但利用其分布产生随机数就很容易。

Algorithm 1 Simulating the predictive distribution with a normal-Wishart posterior

For $j = 1, \ldots R$

1. Generate $\mathbf{\Psi}^{(j)}$ from the marginal posterior $\mathbf{\Psi}|\mathbf{Y}_T \sim iW_m(\mathbf{S}, T-k)$ distribution using, e.g. the algorithm of Geweke (1988).
2. Generate $\mathbf{\Gamma}^{(j)}$ from the conditional posterior $\mathbf{\Gamma}|\mathbf{Y}_T, \mathbf{\Psi}^{(j)} \sim N\left(\widehat{\mathbf{\Gamma}}, \mathbf{\Psi}^{(j)}, (\mathbf{Z}'\mathbf{Z})^{-1}\right)$
3. Generate $\mathbf{u}_{T+1}^{(j)}, \ldots, \mathbf{u}_{T+H}^{(j)}$ from $\mathbf{u}_t \sim N\left(0, \mathbf{\Psi}^{(j)}\right)$ and calculate recursively

$$\widetilde{\mathbf{y}}_{T+h}^{(j)\prime} = \sum_{i=1}^{h-1} \widetilde{\mathbf{y}}_{T+h-i}^{(j)\prime} \mathbf{A}_i^{(j)} + \sum_{i=h}^{p} \mathbf{y}_{T+h-i}' \mathbf{A}_i^{(j)} + \mathbf{x}_{T+h}' \mathbf{C}^{(j)} + \mathbf{u}_{T+h}^{(j)\prime}. \tag{10}$$

Discarding the parameters yields $\left\{\widetilde{\mathbf{y}}_{T+1}^{(j)}, \ldots \widetilde{\mathbf{y}}_{T+H}^{(j)}\right\}_{j=1}^{R}$ as a sample of independent draws from the joint predictive distribution.

拉普拉斯近似

基本思想是**使用一个高斯分布来近似复杂分布**，求解过程即为求正态分布的期望μ和方差σ^2

首先看一下多维变量$\mathbf{x}$的高斯分布公式的形式:

$$\mathcal{N}(\mathbf{x}|\mu, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\mathbf{\Sigma}|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathrm{T}} \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\} \tag{4}$$

其中D为$\mathbf{x}$的维度。

假设有一个变量$\mathbf{z}$（M维向量，粗体），其分布为$p(\mathbf{z})$，该分布的具体形式未知，定义如下:

$$p(\mathbf{z}) = \frac{1}{\mathbf{Z}} f(\mathbf{z})$$
$$\mathbf{Z} = \int f(\mathbf{z}) dz$$

其中$\mathbf{Z}$为归一化项。

令$\mathbf{z}_0$为$f(\mathbf{z})$的驻点，

$$\ln f(\mathbf{z}) \simeq \ln f(\mathbf{z}_0) - \frac{1}{2}(\mathbf{z}-\mathbf{z}_0)^{\mathrm{T}} \mathbf{A}(\mathbf{z}-\mathbf{z}_0) \tag{5}$$

其中$\mathbf{A}$为$M \times M$的矩阵，是$\ln f(\mathbf{z})$在驻点$\mathbf{z}_0$处的海森矩阵(Hessian matrix)。

$$\mathbf{A} = \nabla\nabla \ln f(\mathbf{z})\Big|_{\mathbf{z}=\mathbf{z}_0}$$

对于公式5两边去掉$\mathbf{ln}$，则：

$$f(\mathbf{z}) \simeq f(\mathbf{z}_0)\exp\{-\frac{1}{2}(\mathbf{z}-\mathbf{z}_0)^{\mathrm{T}}\mathbf{A}(\mathbf{z}-\mathbf{z}_0)\} \tag{6}$$

对比式4中的指数项与式6中的指数项可以发现，如果$A = \Sigma^{-1}$，则可得归一化的高斯函数$q(\mathbf{z})$为下式：

$$q(\mathbf{z}) = \frac{A^{1/2}}{(2\pi)^{M/2}}\exp\{-\frac{1}{2}(\mathbf{z}-\mathbf{z}_0)^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{z}-\mathbf{z}_0)\}$$

其中$|\mathbf{A}|$为矩阵$\mathbf{A}$的行列式。