北京大学

本科生毕业论文

题目：资产定价模型的比较：中国股票市场的实证结果

Comparison of Asset Pricing Models:

Empirical Results in Chinese Stock Market

姓　　名：陈泊帆

学　　号：1900011030

院　　系：经济学院

专　　业：金融学

导师姓名：王熙

论文成绩：

二〇二三年五月

版权声明

摘要

随着资产定价理论的蓬勃发展，近年来越来越多的因子模型被提出，因子模型进入大数据时代。因此，如何比较不同的因子模型并选择最优的（平均）因子模型对学术的研究和行业的应用具有着重要意义。本文利用了贝叶斯统计的方法，通过计算不同因子模型的后验概率，给出了对这一问题的解决方案。(1）在模型参数满足 Jeffery 先验分布的假设下，本文给出了模型后验概率的解析表达式；(2）对于其他先验分布，本文利用马尔可夫链蒙特卡洛方法（MCMC）给出了模型后验概率的数值解法；(3）本文将贝叶斯方法应用到中国股票市场中，基于 2000 至 2020 年时间段内的数据，六因子模型 {MKT SMB UMD CMA VMG PMO} 是中国股票市场的最优因子模型。本文还发现中国股票市场的最优模型具备时变特征，这也符合已有文献对我国股票市场市场风格变迁的发现。最后，本文对此方法进行了稳健性的检验，并对未来进一步的研究方向进行了展望。

关键词：资产定价理论，因子模型，贝叶斯方法

Comparison of Asset Pricing Models: Empirical Results in Chinese Stock Market

Bofan Chen (Finance)
Directed by: Xi Wang

ABSTRACT

In asset pricing theory, a large number of factor models have been proposed in previous literature. It is crucial to compare and select the best model among them using quantitative methods. In this paper, we introduce a Bayesian approach to identify the model with the highest posterior probability. Assuming that the model parameters follow the Jeffery prior, we derive an analytical expression for the posterior probability. For other priors, we also present a numerical method using the Monte Carlo Markov chain technique (MCMC). We apply this method to the Chinese stock market and conclude that the six-factor model {MKT SMB UMD CMA VMG PMO} is the optimal model as of December 2020. Our results also suggest that the optimal model for the Chinese stock market is likely to change over the long run, as market styles evolve. We further test the robustness of our approach and discuss potential directions for future research. By employing a systematic and objective method for selecting the optimal factor model, we contribute to the existing literature on asset pricing theory. This can provide important insights for investors, policymakers, and other stakeholders in financial markets, ultimately leading to a better understanding of asset pricing dynamics.

KEY WORDS: Asset pricing theory, Factor models, Beyesian method

目录

第一章　引言

1.1　研究背景和意义

因子模型作为现代资产定价理论中广泛使用的模型，其理论内涵相当于使用数个因子近似随机定价核。通常，因子是市场上某个特定投资组合不同时间下的收益率，往往反应着市场上各类系统性风险。基于上述理论内涵，因子模型认为通过选取某些特定的因子，用因子的线性组合解释风险资产超额收益率的差异，并据此为风险资产进行定价。前人通过实践和分析，发现有效因子数量达到了上千个甚至上万个，因此也诞生出一个又一个因子模型。因此，已有文献也开始出现利用各类统计或机器学习方法来判定这些模型的好坏的研究。以往的大量研究聚焦于因子模型的解释力上。这类研究往往针对给定的单一模型，研究者利用统计结果来判断模型的定价是否准确，模型因子的构成是否包含了所有系统性风险，在这个模型下又是否有套利的可能。但是这类研究却没有把重点放在多个因子模型的比较上。同时，仅仅用解释力的筛选模型有可能会使模型变得更加复杂。基于简洁性的考量，文献开始向稀疏类因子模型发展。在保证模型解释力的同时，研究者希望排除冗余因子。对于时变的市场，不同阶段适用的模型也会随着时间变化，研究者进而提出了模型的稀疏性的概念。它们认为在不同时刻的适用模型只需要从少数个优秀的因子模型中选择。但这种模型上的稀疏性是否成立，模型究竟应该多么稀疏也值得我们进一步探讨。

本文将利用贝叶斯方法计算出多种因子模型的后验概率，基于此来对潜在的因子模型进行比较，不但考虑到了模型的不确定性和参数的不确定性，还在考虑到这一系列不确定性情况下对模型的稀疏程度进行了进一步解答。本文针对参数不同形式的先验假设，给出了求解因子模型后验概率的解析方法和数值方法。最后，本文将此方法应用到中国股票市场中，并筛选出中国的最佳因子模型（成功概率最高的模型）。

1.2　文献综述

早期的资产定价理论相关文献主要聚焦于对不同资产的风险溢价做出解释。这些文献认为，一种资产所带来的超额收益只与其承担的系统性风险有关。不同资产的风险溢价取决于资产收益率回归到不同风险因子收益率下的系数 beta 或者其他可能的系统性风险度量方式。Sharpe (1964)[17] 和 Lintner (1965)[13]最开始提出经典的资本资产定价模型（CAPM），他们认为市场风险是主要的系统性风险。在 CAPM 模型之后，研究者进一步提出了多因子模型，引入了除市场风险之外的其他系统性风险。如 Fama

和 French (1993) 的三因子模型[8]，Fama 和 French (2015)[7]的五因子模型，Barillas 和 Shanken (2017)[2]的六因子模型，Liu 等研究者 (2019)[14]提出的中国四因子模型等。Fama 和 French (1993) 的三因子模型在原来 CAPM 市场因子的基础上增加了衡量公司成长性风险的价值因子（value factor）和衡量公司规模风险的规模因子（size factor），之后，Fama 和 French (2015) 的五因子模型在原来三因子的基础上又进一步引入了盈利因子和投资风格因子。Barillas 和 Shanken (2017) 的六因子模型又在原来五因子的基础上增加动量因子。而 Liu 等研究者 (2019) 提出的中国四因子模型则针对中国股票市场的特殊性（存在借壳上市的情况），重新构造的价值和规模因子，并提出了换手率因子。就这样，半个世纪以来，为了满足投资分析的需要，成千上万的因子被人们发现，同时也产生了成千上万种潜在的模型，形成“因子动物园”的现象。

与此同时，很多学者也提出了一些方法验证这些资产定价因子模型的正确性。比如，Black, Jensen 和 Scholes (1972)[12] 通过验证市场组合与任意资产组合时间序列下回归后的截距 alpha 是否为零来检验模型。接着，Gibbons, Ross 和 Shanken (1989)[10] 针对这个提出了一种联合 F 检验的方法对资产组合超额回报 alpha 的值进行假设检验，即 GRS 检验。在一个正确的资产定价模型下，超额回报 alpha 应该为零。若假设检验拒绝了 alpha 等于零的原假设，那么这个模型就是不合适的。然而在验证资产定价模型这个问题上，这些假设检验的方法其实只考虑到了一个方面。它们更多地关注 p 值，即有多大可能拒绝原假设 (模型具备解释力)。但如果同时存在多个模型，并且这些模型都可以通过假设检验，单单靠 p 值无法直观地量化一个模型究竟有多么好，即给出不同模型成为最优模型的概率。

在上述资产定价理论研究的历史背景下，人们有着大量的因子可以选择，并且这些因子的组合构成的许多模型都可以保证不被 GRS 检验拒绝。所以在实际的应用中，一个亟需解决的问题就是人们到底该采用哪些因子来构成资产定价模型，哪些因子构成的模型是最佳模型。其实模型比较这个话题，近年来也被许多学者广泛研究。如 Feng 等研究者 (2020)[9]利用 Lasso 回归的思路，同时把多个因子放到一起筛选比较，在模型解释力和因子稀疏性的双重考量下，得到最优的因子模型。但是，这样的模型比较只能筛选出唯一的最优模型，不利于在时变的市场中进行模型稀疏性的考量。

Barillas 和 Shanken (2018)[1]提出了一个贝叶斯分析框架，对解决最优模型问题提供了一种”用模型打败模型“的思路，筛选出了美国股票市场中最具竞争力的模型。Chib 等人 (2020)[4]对 Barillas and Shanken (2018) 的模型中参数先验假设提出了修正，用 CZZ 先验替代 Jeffery 先验，并利用模拟的方法进行验证。这两篇文章研究通过计算不同模型的后验概率来寻找最优模型。这样的贝叶斯分析框架不仅能够选出最终的最优模型，还提供了一个后验概率数值，代表这个最优模型的可信性。在一个时变的市场中，利

用后验概率可以较好地对模型的稀疏性进行考量，也对模型平均（Model Average）有着指导意义。

本文的研究正是基于上述两篇论文提出的贝叶斯框架，推导和分析了不同因子模型的贝叶斯后验概率公式。同时，本文放宽了 Barillas 和 Shanken (2018) 模型对参数先验的条件，首次对任意幂次型参数先验推导出了模型后验概率的解析解，进而对其在幂次型先验分布簇下稳健性进行刻画。在对于给定其他先验无法求解析解的情况下，本文利用马尔可夫链-蒙特卡罗方法（MCMC）给出对应先验分布的数值算法，进而可以计算出任意先验分布下模型的后验概率数值解。

在研究的实证部分中，本文着重分析了中国股票市场下的最优模型。在实证的过程中，本文仅仅考虑可交易的因子作为模型的潜在因子，因为所有可能的因子都可以被可交易的资产组合直接表出或者通过间接模拟获得。根据 Breeden (1979)[3]论文中的结果，尽管像消费增长这类因子并不能直接进行交易，但是他们选择用可以交易的资产组合最大程度地模拟它的变化，这两者在因子模型定价最终效果上的表现相近的。一般来说，本文采用 FF 三因子模型中类似的价差组合（spread-portfolio factors）来替代那些不能交易的因子，本文假设这样的方法附带引入的估计上的偏差是可以忽略的。

国际上对于资产定价的实证研究以美国为主，而中国的资本市场与美国存在一定的区别，因此不能直接把美国的结论直接迁移到中国来。在中国股票市场中，很多传统的因子可能依然有效，但也有可能是无效的。如今，一个被人们广为接受的模型是 Liu 等研究者 (2019)[14]提出的中国四因子模型。他们针对中国股票市场特有的市场结构，构造了适用于中国的四个因子。但是，是否存在着更加优秀的因子模型呢？本文同时考虑了美国和中国的实证研究中一些普遍被认可并经常使用的因子模型中九个代表性因子作为潜在因子，通过理论上的分析求解出了中国市场的最佳因子模型。

1.3　本文内容与结构

本文分为以下五个部分：

第一部分主要介绍研究背景，对金融学中资产定价模型的发展历史和统计上对模型的检验方法进行综述，并点明了本文在资产定价模型比较方法上的边际贡献；

第二部分介绍了本文进行资产定价模型比较的总模型设定，并利用贝叶斯方法求解出在模型参数满足 Jeffery 先验分布下不同模型后验分布的解析解；

第三部分在模型参数满足任意先验分布的情况下使用马尔可夫链-蒙特卡罗方法（MCMC）的数值求解方法求解不同资产定价模型的后验概率；

第四部分介绍了利用本文的分析方法对中国股票市场的资产定价最优模型的实证结果；

第五部分是对全文的总结和未来研究方向的展望。

第二章　模型比较的贝叶斯方法

2.1　因子模型的导出

本节将介绍传统因子模型的假设和推导过程。

假设市场上不同资产组合的超额收益率随机变量 r_i 之间只存在线性相关关系，而不存在非线性相关关系。因此其存在如下数据生成过程 (DGP)

$$r_{i,t} = E[r_i] + \boldsymbol{\beta}_i^\top (\boldsymbol{f}_t - E[\boldsymbol{f}]) + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}\left(0, \sigma_i^2\right)$$

其中向量 $\boldsymbol{f}$ 中的元素被称为因子，向量 $\boldsymbol{\beta}_i$ 代表因子在这个资产组合 i 上的因子暴露。此式中 $\varepsilon_{i,t}$ 与因子线性无关。同时，对于不同的资产组合，当因子的数量足够多时，可以使得不同资产组合的 $\varepsilon_{i,t}$ 没有相关性。

在上述假设的基础上，增加无套利假设，可以得到 APT 模型（详情可见王熙《资产定价导论》第六章[19]），即

$$E[r_i] = \boldsymbol{\beta}_i^\top \boldsymbol{\lambda}$$

其中 $\boldsymbol{\lambda}$ 被称为风险溢价。

设 $\boldsymbol{R}$ 是市场上所有资产的超额收益率构成的随机向量，当所有因子都可以交易时，则存在矩阵 ω，使得

$$\boldsymbol{f} = \omega \boldsymbol{R}.$$

故有

$$E[\boldsymbol{f}] = \omega E[\boldsymbol{R}] = \omega \boldsymbol{B} \boldsymbol{\lambda}$$

其中 $\boldsymbol{B}$ 是因子线性回归到 $\boldsymbol{R}$ 上的回归系数，可以表示成

$$\boldsymbol{B} = \mathrm{Cov}(\boldsymbol{R}, \boldsymbol{f}) \, \mathrm{Var}(\boldsymbol{f})^{-1}.$$

所以

$$E[\boldsymbol{f}] = \omega \boldsymbol{B} \boldsymbol{\lambda} = \mathrm{Cov}(\omega \boldsymbol{R}, \boldsymbol{f}) \, \mathrm{Var}(\boldsymbol{f})^{-1} \lambda. = \lambda$$

代回 DGP，可以得到因子模型

$$r_{i,t} = \boldsymbol{\beta}_i^\top \boldsymbol{f}_t + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}\left(0, \sigma_i^2\right).$$

2.2 贝叶斯框架下模型比较的总模型

本节将介绍本文采用的理论分析模型。为与待比较的潜在因子模型区分，称以下模型为总模型。

假设市场上有 K 个可交易的潜在因子，本文将要比较的每一个潜在模型的因子都从这 K 个因子中进行挑选。对于第 j 个模型，它可以表示成 $\mathscr{M}_j = \{\boldsymbol{f}_j, \boldsymbol{f}_j^*\}$，其中 $\boldsymbol{f}_j$ 是模型中用来定价的风险因子，数量是 L_j，而 $\boldsymbol{f}_j^*$ 是模型中的非风险因子，数量是 $K - L_j$。假设潜在模型的总数是 J。在每个时间点，假设风险因子的收益率满足正态分布。对于非风险因子（或者是市场上的任意资产组合），根据因子模型，其可以由风险因子线性生成。因此，每个潜在模型可以表示如下

$$\begin{aligned} \boldsymbol{f}_{j,t} &= \boldsymbol{\alpha}_j + \boldsymbol{\varepsilon}_{j,t}, \quad \boldsymbol{\varepsilon}_{j,t} \sim \mathscr{N}_{L_j}\left(\mathbf{0}, \Sigma_j\right), \\ \boldsymbol{f}_{j,t}^* &= \boldsymbol{B}_j^* \boldsymbol{f}_{j,t} + \boldsymbol{\varepsilon}_{j,t}^*, \quad \boldsymbol{\varepsilon}_{j,t}^* \sim \mathscr{N}_{K-L_j}\left(\mathbf{0}, \Sigma_j^*\right). \end{aligned}$$

记每个因子 T 期的数据为

$$\boldsymbol{y}_{1:T} = \left(\boldsymbol{f}_1, \boldsymbol{f}_1^*, \cdots, \boldsymbol{f}_T, \boldsymbol{f}_T^*\right).$$

总模型的参数可以表示如下

$$\boldsymbol{\eta}_j = \left(\operatorname{vec}\left(\boldsymbol{B}_j^*\right), \operatorname{vech}\left(\Sigma_j\right), \operatorname{vech}\left(\Sigma_j^*\right)\right)$$

其中 vec(·) 代表对矩阵的向量化，即把每一个列向量堆叠在一起。vech(·) 代表对对称矩阵的半向量化，即仅把对称矩阵的下三角部分堆叠成向量。

在对上述参数选择合适的先验分布后，可以通过积分将参数边缘化，进而计算出给定不同模型下得到数据的条件分布为

$$m\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j\right) \triangleq \int\int p\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j, \boldsymbol{\alpha}_j, \boldsymbol{\eta}_j\right) \pi\left(\boldsymbol{\alpha}_j \mid \mathscr{M}_j, \boldsymbol{\eta}_j\right) \psi\left(\boldsymbol{\eta}_j \mid \mathscr{M}_j\right) \mathrm{d}(\boldsymbol{\alpha}_j, \boldsymbol{\eta}_j)$$

其中 $p(\cdot)$ 是观测数据 $\boldsymbol{y}_{1:T}$ 在模型假设下的条件分布。而 $\pi(\cdot)$，$\psi(\cdot)$ 分别是 $\boldsymbol{\alpha}_j$ 和 $\boldsymbol{\eta}_j$ 的先验分布。依据贝叶斯公式，不同模型在观测下的后验分布为

$$\mathbb{P}\left(\mathscr{M}_j \mid \boldsymbol{y}_{1:T}\right) \propto m\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j\right) \cdot \phi(\mathscr{M}_j)$$

其中 $\phi(\mathscr{M}_j)$ 是不同模型的先验分布。

这个后验分布可以理解为在给定参数和模型的先验下，依据观测的数据得出的不同模型的概率，那个后验概率最高的模型就是本文所期望得到的最优模型。

但是，值得注意的是，最优模型的结果是与参数和模型的先验分布息息相关的。这种先验在某种程度上提供了一个看待“最佳模型”的标准，也可以理解成一种人们对在模型选择上除数据之外的偏好。因此，只有选择一个合适的先验，最终的结果才是可

靠与稳健的。

2.3　参数的 CZZ 先验分布

本节将介绍 Siddhartha, Xiaming 和 Lingxiao[4] 针对上述总模型参数 $\boldsymbol{\eta}_j$ 给出的 CZZ 先验。

首先假设潜在模型 $\mathscr{M}_1$ 中，所有 K 个因子都是风险因子，即

$$\boldsymbol{f}_{1,t} = \boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon}_{1,t}, \quad \boldsymbol{\varepsilon}_{1,t} \sim \mathscr{N}_K\left(\mathbf{0}, \Sigma_1\right)$$
$$\boldsymbol{\eta_1} = \operatorname{vech}\left(\Sigma_1\right)$$

并假设 $\boldsymbol{\eta_1}$ 的先验是 $\psi\left(\eta_1 \mid \mathscr{M}_1\right)$。

对于 $\mathscr{M}_j$, 有

$$\boldsymbol{f}^*_{j,t} = \boldsymbol{B}^*_j\boldsymbol{\alpha}_j + \left(\boldsymbol{\varepsilon}^*_{j,t} + \boldsymbol{B}^*_j\boldsymbol{\varepsilon}_{j,t}\right).$$

所以它可以表示成

$$\begin{pmatrix} \boldsymbol{f}_{j,t} \\ \boldsymbol{f}^*_{j,t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_j \\ \boldsymbol{B}^*_j\boldsymbol{\alpha}_j \end{pmatrix} + \boldsymbol{E}_{j,t}, \quad \boldsymbol{E}_{j,t} \sim \mathscr{N}_K\left(\mathbf{0}, \begin{pmatrix} \Sigma_j & \Sigma_j\boldsymbol{B}^{*\mathsf{T}}_j \\ \boldsymbol{B}^*_j\Sigma_j & \Sigma^*_j + \boldsymbol{B}^*_j\Sigma_j\boldsymbol{B}^{*\mathsf{T}}_j \end{pmatrix}\right).$$

不同潜在模型的参数先验之间需要满足 $\boldsymbol{E}_{j,t}$ 的先验是相同的. 因此, 不同潜在模型下参数先验之间存在映射关系（可逆）$g_j : \boldsymbol{\eta}_1 \longmapsto \boldsymbol{\eta}_j$, g_j^{-1} 可以表示为

$$\Sigma_1 = \begin{pmatrix} \Sigma_j & \Sigma_j\boldsymbol{B}^{*\mathsf{T}}_j \\ \boldsymbol{B}^*_j\Sigma_j & \Sigma^*_j + \boldsymbol{B}^*_j\Sigma_j\boldsymbol{B}^{*\mathsf{T}}_{j,f} \end{pmatrix}.$$

进一步地，通过概率密度分布的变量代换公式，可以导出 $\mathscr{M}_j$ 模型下的参数先验分布为

$$\begin{aligned} \psi\left(\boldsymbol{\eta}_j \mid \mathscr{M}_j\right) &= c\psi\left(g_j^{-1}\left(\boldsymbol{\eta}_j\right) \mid \mathscr{M}_1\right)\left|\det\left(\frac{\partial g_j^{-1}\left(\boldsymbol{\eta}_j\right)}{\partial \boldsymbol{\eta}'_j}\right)\right| \\ &= c\psi\left(\begin{pmatrix} \Sigma_j & \Sigma_j\boldsymbol{B}^{*\prime}_{j,f} \\ \boldsymbol{B}^*_{j,f}\Sigma_j & \Sigma^*_j + \boldsymbol{B}^*_{j,f}\Sigma_j\boldsymbol{B}^{*\prime}_{j,f} \end{pmatrix} \mid \mathscr{M}_1\right)\left|\Sigma_j\right|^{K-L_j}. \end{aligned}$$

本节和后续的实证章节四假定 $\mathscr{M}_1$ 选取 Jeffery 先验。它是一种贝叶斯统计中常用的无信息的主观先验分布，可以在一定程度上避免除数据之外的主观性，它的表达式是

$$\psi\left(\eta_1 \mid \mathscr{M}_1\right) = c\left|\Sigma_1\right|^{-\frac{K+1}{2}}$$

其中 c 是一个常数。由 Jeffery 先验，可以推导出

$$\psi\left(\boldsymbol{\eta}_j \mid \mathscr{M}_j\right) = c\left|\Sigma_j\right|^{-\frac{2L_j-K+1}{2}}\left|\Sigma_j^*\right|^{-\frac{K+1}{2}}.$$

由上式可知，在这个先验下，风险因子和非风险因子的残差是相互独立的，它们各自的协方差矩阵的分布是

$$\psi\left(\boldsymbol{\Sigma}_j \mid \mathscr{M}_j\right) = c\left|\Sigma_j\right|^{-\frac{2L_j-K+1}{2}}$$

$$\psi\left(\boldsymbol{\Sigma}_j^* \mid \mathscr{M}_j\right) = c\left|\Sigma_j^*\right|^{-\frac{K+1}{2}}$$

这种现象方便了对边缘分布求解析解，但是却与实际有一定的差别。现实中，两种因子的残差极有可能是相关的，Jeffery 先验有一定的不合理性。

2.4 alpha 的先验分布

每个风险因子的 alpha 代表着承担一单位该风险的平均收益。为了计算 alpha 的先验分布，在实证时需奥取出数据中的一定比例 (记为 tr) 的数据作为测试数据集。本文选择测试集的比例为 $tr = 0.1$，即选择前 1/10 时间段的数据作为测试数据，并利用剩下时间段的数据计算后验概率。

基于前人的文献，本文假设 alpha 满足正态分布如下：

$$\boldsymbol{\alpha}_j \mid \mathscr{M}_j \sim \mathscr{N}_{L_j}\left(\boldsymbol{\alpha}_{j0}, k_j\Sigma_j\right), \quad j = 1, 2, \ldots, J$$

$\boldsymbol{\alpha}_{j0}$ 是风险因子在测试时间段中的均值

$$\boldsymbol{\alpha}_{j0} = n_t^{-1}\sum_{t=1}^{n_t}\boldsymbol{f}_{j,t},$$

其中 n_t 是测试时间段的长度。

对于 $\boldsymbol{\alpha}_j$ 的方差，它代表着不同时间段人们看待同一种风险的态度的变化或者是可能的错误定价。一些之前的资产定价研究认为其应当正比于残差的方差 Σ_j。Dybvig (1983)[6]、Grinblatt 和 Titman (1983)[11] 推导出个体资产从多因子定价模型偏离的界限与其残差方差成正比。从行为经济学的角度来看，Shleifer 和 Vishny (1997)[18] 认为高异质性风险（较大的残差的方差）可能会阻碍套利，从而更有可能存在更多的错误定价（alpha 有较大的方差）。在 Pastor 和 Stambaugh (2000)[15]的论文中，他们认为在先验中使 alpha 的方差与残差的方差之间存在正相关性后会使极大的夏普比率不太可能出现。而市场中出现极大的夏普比率是不合适的，这往往意味着一个可以做到“近似”套利的机会（Shanken (1992)[16]）。Cochrane 和 Saa-Requejo (2000)[5]也对这个先验也提供了进

一步的说明。

而对于 k_j, 本文假设

$$k_j = mult \times L_j^{-1} \operatorname{sum}\left(\operatorname{diag}\left(\hat{V}_{j0}\right) / \operatorname{diag}\left(\hat{\Sigma}_{j0}\right)\right)$$

其中 $\hat{\Sigma}_{j0}$ 是测试集对残差方差的最小二乘估计，$\hat{V}_{j0}$ 是对于 $\boldsymbol{\alpha}_j$ 对于测试集观测数据的负逆 Hessian 矩阵，$mult = \frac{1-tr}{tr}$ 是根据不同测试集比例 $tr = \frac{n_t}{T}$ 用来调节 k_j 大小的乘子。

2.5　潜在模型的后验分布

基于上述讨论，本节将推导不同模型的后验概率公式。对于 Jeffery 先验，风险因子和非风险因子的参数分布是独立的，所以

$$\log m\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j\right) = \log m\left(\boldsymbol{f}_{j,1:T} \mid \mathscr{M}_j\right) + \log m\left(\boldsymbol{f}^*_{j,1:T} \mid \mathscr{M}_j\right)$$

接下来，分别计算上式右端两项的表达式。对于右端第一项:

$$\begin{aligned} m\left(\boldsymbol{f}_{j,1:T} \mid \mathscr{M}_j\right) &= \int \left(\int p\left(\boldsymbol{f}_{j,1:T} \mid \mathscr{M}_j, \boldsymbol{\alpha}_j, \boldsymbol{\Sigma}_j\right) \pi\left(\boldsymbol{\alpha}_j \mid \mathscr{M}_j, \boldsymbol{\Sigma}_j\right) \mathrm{d}\boldsymbol{\alpha}_j \right) \psi\left(\boldsymbol{\Sigma}_j \mid \mathscr{M}_j\right) \mathrm{d}\boldsymbol{\Sigma}_j \\ &= \int p\left(\boldsymbol{f}_{j,1:T} \mid \mathscr{M}_j, \Sigma_j\right) \psi\left(\boldsymbol{\Sigma}_j \mid \mathscr{M}_j\right) \mathrm{d}\boldsymbol{\Sigma}_j \end{aligned}$$

由于 $p\left(\boldsymbol{f}_{j,1:T} \mid \mathscr{M}_j, \Sigma_j\right)$ 是两个正态分布概率密度函数的乘积的积分，可以积分得到解析解为

$$p\left(\boldsymbol{f}_{j,1:T} \mid \mathscr{M}_j, \Sigma_j\right) = \left|Tk_j + 1\right|^{-\frac{L_j}{2}} (2\pi)^{-\frac{TL_j}{2}} \left|\Sigma_j\right|^{-\frac{T}{2}} e^{-\frac{1}{2}\operatorname{tr}\left(\Psi_j \Sigma_j^{-1}\right)}$$

其中

$$\Psi_j = \sum_{t=1}^{T} \left(\boldsymbol{f}_{j,t} - \hat{\boldsymbol{\alpha}}_j\right)\left(\boldsymbol{f}_{j,t} - \hat{\boldsymbol{\alpha}}_j\right)^{\top} + \frac{T}{Tk_j + 1}\left(\hat{\boldsymbol{\alpha}}_j - \boldsymbol{\alpha}_{j0}\right)\left(\hat{\boldsymbol{\alpha}}_j - \boldsymbol{\alpha}_{j0}\right)^{\top},$$

所以

$$\begin{aligned} m\left(\boldsymbol{f}_{j,1:T} \mid \mathscr{M}_j\right) &= \int p\left(\boldsymbol{f}_{j,1:T} \mid \mathscr{M}_j, \Sigma_j\right) \psi\left(\boldsymbol{\Sigma}_j \mid \mathscr{M}_j\right) \mathrm{d}\boldsymbol{\Sigma}_j \\ &= \int \left|Tk_j + 1\right|^{-\frac{L_j}{2}} (2\pi)^{-\frac{TL_j}{2}} \left|\Sigma_j\right|^{-\frac{T}{2}} e^{-\frac{1}{2}\operatorname{tr}\left(\Psi_j \Sigma_j^{-1}\right)} \psi\left(\boldsymbol{\Sigma}_j \mid \mathscr{M}_j\right) \mathrm{d}\boldsymbol{\Sigma}_j \end{aligned}$$

代入 Jeffery 先验 ($\psi\left(\boldsymbol{\Sigma}_j \mid \mathscr{M}_j\right) = c\left|\Sigma_j\right|^{-\frac{2L_j - K + 1}{2}}$), 积分项可以被转化成自由度为 $\nu_j = T + L_j - K$ 的逆 Wishart 分布的概率密度函数乘一个常数。最终可以积出一个解析解

为:

$$\log m\left(\boldsymbol{f}_{j,1:T} \mid \mathscr{M}_j\right) = -\frac{\left(K-L_j\right)L_j}{2}\log 2 - \frac{TL_j}{2}\log\pi - \frac{L_j}{2}\log\left(Tk_j+1\right) - \frac{\left(T+L_j-K\right)}{2}\log\left|\Psi_j\right| + \log\Gamma_{L_j}\left(\frac{T+L_j-K}{2}\right)$$

同理，对于右端第二项

$$m\left(\boldsymbol{f}^*_{j,1:T} \mid \mathscr{M}_j\right) = \int\left(\int p\left(\boldsymbol{f}^*_{j,1:T} \mid \mathscr{M}_j, \boldsymbol{B}^*_j, \boldsymbol{\Sigma}^*_j\right)\mathrm{d}\boldsymbol{B}^*_j\right)\psi\left(\boldsymbol{\Sigma}^*_j \mid \mathscr{M}_j\right)\mathrm{d}\boldsymbol{\Sigma}^*_j = \int p\left(\boldsymbol{f}^*_{j,1:T} \mid \mathscr{M}_j, \Sigma^*_j\right)\psi\left(\boldsymbol{\Sigma}^*_j \mid \mathscr{M}_j\right)\mathrm{d}\boldsymbol{\Sigma}^*_j$$

对 $p\left(\boldsymbol{f}^*_{j,1:T} \mid \mathscr{M}_j, \Sigma^*_j\right)$ 可以积分得到

$$p\left(\boldsymbol{f}^*_{j,1:T} \mid \mathscr{M}_j, \Sigma^*_j\right) = (2\pi)^{-\frac{\left(T-L_j\right)\left(K-L_j\right)}{2}}\left|\Sigma^*_j\right|^{-\frac{T-L_j}{2}}\left|W^*_j\right|^{-\frac{K-L_j}{2}}e^{-\frac{1}{2}\mathrm{tr}\left(\Psi^*_j\Sigma^{*-1}_j\right)}$$

其中

$$W^*_j = \sum_{t=1}^{T}\boldsymbol{f}_{j,t}\boldsymbol{f}^\top_{j,t}, \quad \Psi^*_j = \sum_{t=1}^{T}\left(\boldsymbol{f}^*_{j,t} - \hat{\boldsymbol{B}}^*_{j,f}\boldsymbol{f}_{j,t}\right)\left(\boldsymbol{f}^*_{j,t} - \hat{\boldsymbol{B}}^*_{j,f}\boldsymbol{f}_{j,t}\right)^\top.$$

所以

$$m\left(\boldsymbol{f}^*_{j,1:T} \mid \mathscr{M}_j\right) = \int p\left(\boldsymbol{f}^*_{j,1:T} \mid \mathscr{M}_j, \Sigma^*_j\right)\psi\left(\boldsymbol{\Sigma}^*_j \mid \mathscr{M}_j\right)\mathrm{d}\boldsymbol{\Sigma}^*_j$$
$$= \int (2\pi)^{-\frac{\left(T-L_j\right)\left(K-L_j\right)}{2}}\left|\Sigma^*_j\right|^{-\frac{T-L_j}{2}}\left|W^*_j\right|^{-\frac{K-L_j}{2}}e^{-\frac{1}{2}\mathrm{tr}\left(\Psi^*_j\Sigma^{*-1}_j\right)}\psi\left(\boldsymbol{\Sigma}^*_j \mid \mathscr{M}_j\right)\mathrm{d}\boldsymbol{\Sigma}^*_j$$

同样代入 Jeffery 先验的公式, 积分项也可以被转化成自由度为 $v^*_j = T$ 的逆 Wishart 分布的密度函数乘一个常数。最终可以积出:

$$\log m\left(\boldsymbol{f}^*_{j,1:T} \mid \mathscr{M}_j\right) = \frac{\left(K-L_j\right)L_j}{2}\log 2 - \frac{\left(K-L_j\right)\left(T-L_j\right)}{2}\log\pi - \frac{\left(K-L_j\right)}{2}\log\left|W^*_j\right| - \frac{T}{2}\log\left|\Psi^*_j\right| + \log\Gamma_{K-L_j}\left(\frac{T}{2}\right)$$

最后，依据贝叶斯公式，不同模型在观测下的后验分布为

$$\mathbb{P}\left(\mathscr{M}_j \mid \boldsymbol{y}_{1:T}\right) \propto m\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j\right)\cdot\phi(\mathscr{M}_j)$$

其中 $\phi(\mathscr{M}_j)$ 是不同模型的先验分布。一般来说，当一个模型中有极其相似的因子时，本文令 $\phi(\mathscr{M}_j) = 0$; 对于其他可能的模型，本文令每个模型的先验概率相同。据此，依据不同模型的后验分布本文可以选出不同时期下的最优模型。

第三章　数值方法

对于 Jeffery 先验，模型后验概率的解析解可以很幸运地被求得。然而，很多时候 Jeffery 先验并不是一个恰当的先验，所以本文需要研究任意先验下得到模型后验概率的方法。当考虑到其他先验时，它们很有可能形式复杂。$\mathscr{M}_1$ 诱导出的 $\mathscr{M}_j$ 中风险因子与非风险因子的残差的先验也很有可能是相关的。因此，在使用贝叶斯方法进行模型比较时，更大概率的情况下是推导不出解析解的，人们只能设计数值算法积分求解。同时，当人们考虑更多的潜在模型时，需要积分的方差矩阵的维数也将变得很高，很有可能会遇到积分维数灾难的问题。所以，本文将介绍马尔可夫链蒙特卡罗方法（MCMC）在低计算成本下求解不同模型的后验概率。

由第二节可知，本文的模型 $\mathscr{M}_j$ 可以改写成

$$\boldsymbol{y}_{j,1:T} \sim \mathscr{N}_K\left(C\boldsymbol{\alpha}_j, \Omega_j\right)$$

其中

$$\Omega_j = \begin{pmatrix} \Sigma_j & \Sigma_j \boldsymbol{B}_j^{*\top} \\ \boldsymbol{B}_j^* \Sigma_j & \Sigma_j^* + \boldsymbol{B}_j^* \Sigma_j \boldsymbol{B}_j^{*\top} \end{pmatrix}, \quad C = \begin{pmatrix} \mathbb{1}_{L_j \times L_j} \\ B_j^* \end{pmatrix}.$$

模型参数 $\boldsymbol{\eta}_j$ 的先验概率密度函数满足

$$\psi\left(\boldsymbol{\eta}_j \mid \mathscr{M}_j\right) = \psi\left(\begin{pmatrix} \Sigma_j & \Sigma_j \boldsymbol{B}_j^{*\top} \\ \boldsymbol{B}_j^* \Sigma_j & \Sigma_j^* + \boldsymbol{B}_j^* \Sigma_j \boldsymbol{B}_j^{*\top} \end{pmatrix} \mid \mathscr{M}_1\right) \left|\Sigma_j\right|^{K-L_j}$$

$\boldsymbol{\alpha}_j$ 的先验分布为

$$\boldsymbol{\alpha}_j \mid \mathscr{M}_j \sim \mathscr{N}_{L_j}\left(\boldsymbol{\alpha}_{j0}, k_j \Sigma_j\right).$$

本文需要计算

$$\mathbb{P}(\mathscr{M}_j \mid \boldsymbol{y}_{j,1:T}) \propto \int \left(\int p\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j, \boldsymbol{\alpha}_j, \boldsymbol{\eta}_j\right) \pi\left(\boldsymbol{\alpha}_j \mid \mathscr{M}_j, \boldsymbol{\eta}_j\right) \mathrm{d}\boldsymbol{\alpha}_j \right) \psi\left(\boldsymbol{\eta}_j \mid \mathscr{M}_j\right) \mathrm{d}\boldsymbol{\eta}_j \triangleq \mathscr{I}$$

为了解决这个高维积分 $\mathscr{I}$, 本文利用 MCMC 在 $\psi(\cdot)$ 中对 $\boldsymbol{\eta}_j$ 进行采样，进而根据蒙特卡洛积分计算出 $\mathscr{I}$ 的近似值。

$\boldsymbol{\eta}_j$ 的目标概率分布为

$$P_j(\boldsymbol{\eta}) = \psi\left(\boldsymbol{\eta}_j \mid \mathscr{M}_j\right)$$

采用 Metropolis–Hastings（M-H）算法进行采样，首先设置初始样本为

$$\boldsymbol{\eta}^{(0)} = \left(\hat{\boldsymbol{B}}_j^*, \Psi_j, \Psi_j^*\right)$$

转移概率密度函数 $Q(\boldsymbol{\eta}^{(n+1)} \mid \boldsymbol{\eta}^{(n)})$ 为

$$\Sigma^{(n+1)} \mid \boldsymbol{\eta}^{(n)}, \mathscr{M}_j^{(n)} \sim \mathscr{W}^{-1}(\Sigma^{(n)}, T + L_j - K)$$
$$\Sigma^{*(n+1)} \mid \boldsymbol{\eta}^{(n)}, \mathscr{M}_j^{(n)} \sim \mathscr{W}^{-1}(\Sigma^{*(n)}, T)$$
$$\operatorname{vec}\left(\boldsymbol{B}^{*(n+1)}\right) \mid \boldsymbol{\eta}^{(n)}, \mathscr{M}_j^{(n)} \sim \mathscr{N}_{L_j \times (K-L_j)}\left(\operatorname{vec}\left(\boldsymbol{B}^{*(n)}\right), \hat{\sigma}\left(\operatorname{vec}\left(\boldsymbol{B}^{*}\right)\right)\right)$$

其中 $\mathscr{W}^{-1}$ 代表逆 Wishart 分布，$\hat{\sigma}\left(\operatorname{vec}\left(\boldsymbol{B}^{*}\right)\right)$ 是观测值对在对潜在模型的因子回归时对参数 $\boldsymbol{B}^{*}$ 方差的估计值。

基于上述初始样本和转移概率密度函数，利用 M-H 算法从 $P(\boldsymbol{\eta}_j)$ 分布中采样的算法如下：

Algorithm 1 Metropolis–Hastings 算法

Require: 模型 $\mathscr{M}_j$ 下参数的目标概率分布 $P_j(\boldsymbol{\eta})$，转移概率密度函数 $Q(\boldsymbol{\eta}^{(n+1)} \mid \boldsymbol{\eta}^{(n)})$
Ensure: 对于足够大的 n，$\boldsymbol{\eta}^{(n)}$ 是从目标概率分布 $P_j(\boldsymbol{\eta})$ 中抽取的随机样本

定义样本的初始值 $\boldsymbol{\eta}^{(0)} \Leftarrow \left(\hat{\boldsymbol{B}}_j^{*}, \Psi_j, \Psi_j^{*}\right)$
$n \Leftarrow 0$
迭代次数 $itnum \Leftarrow N$
while $n \leq itnum$ **do**
    利用转移概率密度函数生成候选样本 $\boldsymbol{\eta}' \sim Q(\boldsymbol{\eta}^{(n+1)} \mid \boldsymbol{\eta}^{(n)})$
    计算接受概率 $q = \frac{P_j(\boldsymbol{\eta}')}{P_j(\boldsymbol{\eta}^{(n)})} \frac{Q(\boldsymbol{\eta}^{(n)} \mid \boldsymbol{\eta}')}{Q(\boldsymbol{\eta}' \mid \boldsymbol{\eta}^{(n)})}$
    从 0-1 均匀分布中生成一个随机数 $u \in [0, 1]$
    if $u \leq q$ **then**
        接受候选样本 $\boldsymbol{\eta}^{(n+1)} \Leftarrow \boldsymbol{\eta}'$
    else $u > q$
        拒绝候选样本 $\boldsymbol{\eta}^{(n+1)} \Leftarrow \boldsymbol{\eta}^{(n)}$
    end if
    $n \Leftarrow n + 1$
end while

利用 M-H 算法进行采样后，可以得到从目标概率分布 $P_j(\boldsymbol{\eta})$ 中抽取的随机样本 $\boldsymbol{\eta}^{(n)}$。这种采样方式利用在随机变量采样中可以有效避免高维带来的维数灾难的问题。获得此样本后，可以进一步利用蒙塔卡洛模拟数值计算本文目标后验分布的积分 $\mathscr{I}$

$$\mathscr{I} = \frac{1}{N} \sum_{n=1}^{N} \left(\int p\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j, \boldsymbol{\alpha}_j, \boldsymbol{\eta}_j^{(n)}\right) \pi\left(\boldsymbol{\alpha}_j \mid \mathscr{M}_j, \boldsymbol{\eta}_j^{(n)}\right) \mathrm{d}\boldsymbol{\alpha}_j \right).$$

与2.5节类似，由于求和号内的积分是正态分布概率密度函数相乘的形式，依然可以推

导出解析解。积分号内的表达式为

$$p\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j, \boldsymbol{\alpha}_j, \boldsymbol{\eta}_j\right) \pi\left(\boldsymbol{\alpha}_j \mid \mathscr{M}_j, \boldsymbol{\eta}_j\right) = \prod_{i=1}^{T} (2\pi)^{-\frac{K}{2}} \left|\Omega_j\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(y_i - C\alpha_j\right)^{\top} \Omega_j^{-1}\left(y_i - C\alpha_j\right)\right)$$
$$(2\pi)^{-\frac{Lj}{2}} \left|k_j\Sigma_j\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\alpha_j - \alpha_{j0}\right)^{\top} k_j^{-1}\Sigma_j^{-1}\left(\alpha_j - \alpha_{j0}\right)\right).$$

记

$$U = C^{\top}\Omega_j^{-1}C \cdot T + k_j^{-1}\Sigma_j^{-1}$$
$$V = C^{\top}\Omega_j^{-1}\left(\sum_{i=1}^{T} y_i\right) + k_j^{-1}\Sigma_j^{-1}\alpha_{j0}$$
$$W = \sum_{i=1}^{T} y_i^{\top}\Omega_j^{-1}y_i + \alpha_{j0}^{\top}k_i^{-1}\Omega_j^{-1}\alpha_{j0}$$

则有

$$\begin{aligned}
&p\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j, \boldsymbol{\alpha}_j, \boldsymbol{\eta}_j\right) \pi\left(\boldsymbol{\alpha}_j \mid \mathscr{M}_j, \boldsymbol{\eta}_j\right) \\
=&(2\pi)^{-\frac{KT+L_j}{2}} \left|\Omega_j\right|^{-\frac{T}{2}} \left|k_j\Sigma_j\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\alpha_j^{\top}U\alpha_j - 2\alpha_j^{\top}V + W\right)\right) \\
=&(2\pi)^{-\frac{KT}{2}} \left|\Omega_j\right|^{-\frac{T}{2}} \left|k_j\Sigma_j\right|^{-\frac{1}{2}} |U|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(-V^{\top}U^{-1}V + W\right)\right) \\
&\left((2\pi)^{-\frac{L_j}{2}} |U|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\left(\alpha_j - U^{-1}V\right)^{\top} U\left(\alpha_j - U^{-1}V\right)\right)\right)\right)
\end{aligned}$$

其中后面括号中的项刚好是以 $U^{-1}V$ 为均值，U^{-1} 为方差的多元正态分布，积分后为 1。因此求和号内积分的解析表达式为:

$$\begin{aligned}
p\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j, \boldsymbol{\eta}_j\right) &= \int p\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j, \boldsymbol{\alpha}_j, \boldsymbol{\eta}_j\right) \pi\left(\boldsymbol{\alpha}_j \mid \mathscr{M}_j, \boldsymbol{\eta}_j\right) \mathrm{d}\boldsymbol{\alpha}_j \\
&= (2\pi)^{-\frac{KT}{2}} \left|\Omega_j\right|^{-\frac{T}{2}} \left|k_j\Sigma_j\right|^{-\frac{1}{2}} |U|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(-V^{\top}U^{-1}V + W\right)\right).
\end{aligned}$$

综上，每个模型后验概率的数值解为

$$\mathbb{P}(\mathscr{M}_j \mid \boldsymbol{y}_{j,1:T}) \propto \phi(\mathscr{M}_j) \cdot \frac{1}{N}\sum_{n=1}^{N}\left(p\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j, \boldsymbol{\eta}_j^{(n)}\right)\right).$$

第四章　中国股票市场实证结果

4.1　实证因子

本文的实证部分考虑中国股票市场中九个潜在的因子，其大致可以分为七类。

除了已经被广为接受的市场因子 MKT 外，本文选择 SMB 和 SMB_China 作为规模类因子。SMB（Small Minus Big）的构建方法是根据股票在 t 年 6 月的市值进行分类，然后计算从 t 年的 7 月到 t+1 年的 6 月小盘股和大盘股的投资组合按市值加权后的收益率的差异。SMB_China 的构造与 SMB 相似。不同的是，SMB_China 试图避免中国股市中上市公司壳价值的影响，所以它是通过预处理删除底部 30% 小市值的股票后计算的。

本文选择 HML 和 VMG 作为价值类因子。HML（High Minus Low）的构建方法是根据股票在 t-1 年 12 月的账面市值比率进行分类，然后计算从 t 年的 7 月到 t+1 年的 6 月高账面市值股票和低账面市值股票的投资组合按市值加权后的收益率的差异。VMG 的构造与 SMB 相似。不同的是 VMG 在构建过程中使用另一种在中国更加常用的价值指标 EP 来替换了账面市值比率。

本文还选择 UMD、RMW、CMA、PMO 作为其他类型的因子。UMD 是一个动量因子，它的构建基于股票在过去 2 至 12 个月的累积回报。RMW 是一个盈利因子，其构建基于股票在过去一年 12 月的营业收入与资产账面价值的比率。CMA 是一个投资风格因子，其构建基于股票在过去一年年底的总资产增长率。PMO 是一个换手率因子，其构造基于过去一个月的股票换手率除以过去一年的换手率。

上述介绍的因子都分别在前人不同的论文中被提出。Fama 和 French 在他们于 1993 年发表的论文中提出了 SMB 和 HML，在他们于 2015 年发表的论文中提出了 CMA 和 RMW。Carhart 在其 1997 年发表的论文中提出了 UMD 因子。而 Jianan Liu 等在 2019 年发表的论文中提出中国四因子模型，包含了 SMB_China，VMG 和 PMO。由于这些因子在资产定价模型中的优异表现，现在被研究者广泛认可。那么，这些因子在中国股票市场的表现是否经得起本文提出的模型的检验呢？这些因子中是否存在着冗余的因子呢？本文希望通过对这些因子从 2000 年 1 月至 2021 年 6 月的数据进行实证分析，进而给出上述问题的答案。

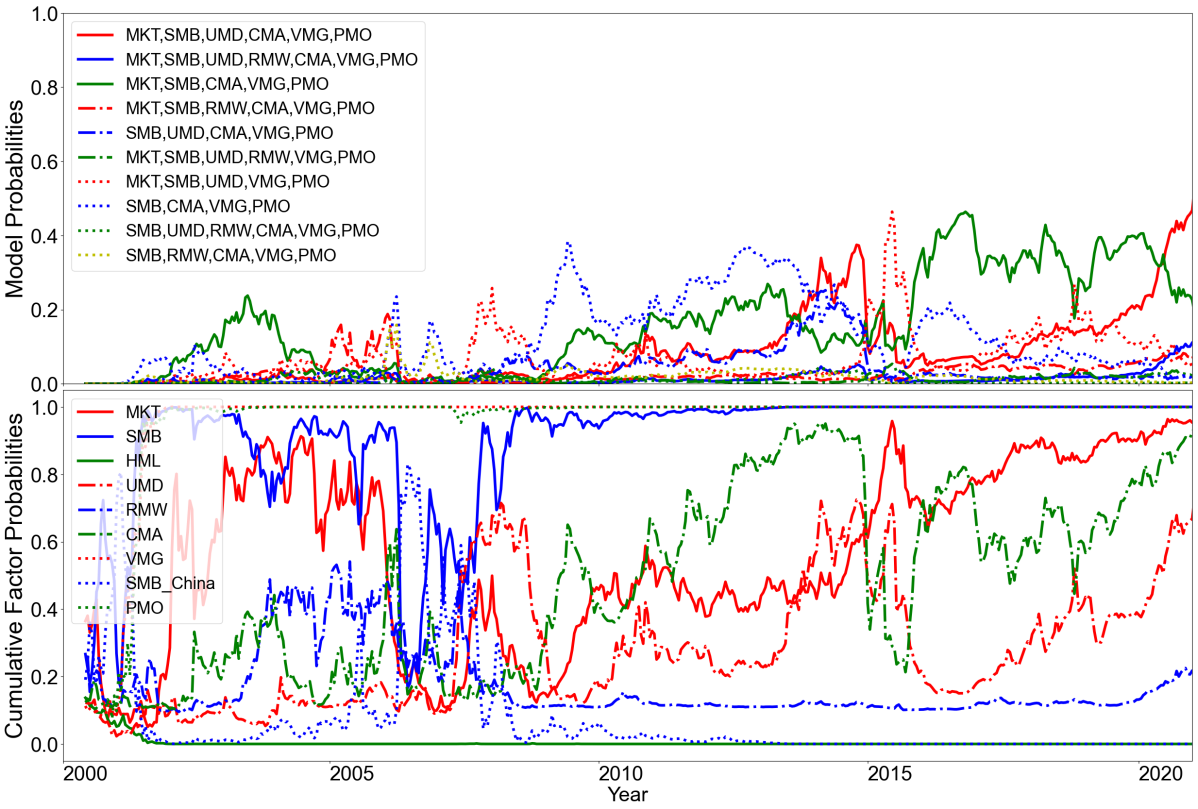

图 4.1　中国股票市场资产定价模型比较的实证结果

4.2　模型比较的实证结果

本文所有的潜在模型基于上面介绍的七类因子，并假设模型中至少有一个风险因子。同时本文要求从每个类别中最多可以选择一个因子加入到风险因子。所以，这九个因子所生成的潜在模型的总数是 $2^5 \times 3^2 - 1 = 287$. 假设这些模型的先验概率相同，即 $\phi(\mathscr{M}_j) = \frac{1}{287}$。在采取 Jeffery 先验的条件下，根据前文第二章的结论，本文从 2000 年 1 月开始，以不同的时刻作为数据集的结尾，从而计算出不同时间长度 $\boldsymbol{T}$ 下每个模型的后验概率如图4.1所示。

图4.1中的第一张图展示了在全时间段后验概率排名前十的模型的后验概率在不同时间长度 $\boldsymbol{T}$ 下的计算结果。如图所示，就算是后验概率排名前十的模型，它们中绝大部分的后验概率依然接近于零。这反映了本文方法在筛选中国股票市场的资产定价模型上可以得到较为明确的结果，也说明这些因子的很多组合构成的模型相比于最佳模型存在明显的缺陷。截至 2020 年末，中国股票市场最佳的（后验概率最高的）模型是六因子模型 {MKT SMB UMD CMA VMG PMO}。这个模型自 2020 年 6 月以来后验概率一直排在所有模型的第一位，其后验概率约为 50%，遥遥领先于剩下的模型。与最佳模型相比，第二优的模型删除了 UMD 因子，它的后验概率在 25% 左右。而第三优的模型相比最佳模型增加了 RMW 因子，它的后验概率只有不到 10%。其余模型的后

验概率则更加不值一提 (小于 10%)，这些模型在 2020 年是不适用的。同时，根据图中的信息，中国股票市场不同模型在不同时间段上交替占据主导地位。一些时间段中表现出色的模型换一个时间段可能并不是最佳的。这个现象很有可能取决于中国股票市场风格的变化。市场风格的轮换让一些原本表现出色的因子丧失解释力，而让一些原本被淘汰的因子重现生机。因此，中国股票市场在长时间的尺度上可能并不存在一个绝对的最佳模型，但是在短时间（几个月）的尺度上却有着最佳模型。换一个角度看模型概率随着时间的变化可以发现，不同模型存在涌现的特征，即很多模型都可以在某个时间段上成为最优模型。这说明因子模型并不是完全稀疏的。

图4.1的第二张图展示了不同时间下的这九个因子的累积因子概率。累积因子概率是包括这个因子的所有模型的后验概率之和，体现了这个因子在最佳模型中被选中的概率大小。一般来说，累积因子概率越高，说明这个因子更有可能出现在最佳模型中，这个因子的价值就越高。截至 2020 年末，MKT、SMB、UMD、CMA、VMG 和 PMO 这六个因子的概率高于 60%，这正好对应着最佳因子模型选出的六个因子，体现了它们在当时表现出色。而 RMW 因子的概率却并不可观（低于 30%），它很有可能是一个冗余的因子。剩下的 HML 和 SMB_China 的概率几乎为零，它们几乎不可能出现在最佳模型中。从长期来看，因子的累积概率并不像最佳模型那样波动频繁。SMB、VMG 和 PMO 这三个因子的累积概率长期都接近 100%，这意味着长期下所有的最佳模型中都会出现这些因子。MKT、CMA 和 UMD 这三个因子的累积概率长期下存在较大的波动，它们是否被选择往往是不同时间下最佳因子模型的主要差别。从某种意义上，这三个因子可以体现出市场风格的变化情况。RMW 在长期下都处于一个较低的概率，说明它很多时候都是冗余的。而 HML 和 SMB_China 在不同时间下概率几乎都为零，这说明它们分别被各自同一类中的 SMB 和 VMG 两个因子长期主导。因此，在中国股票市场上，SMB 和 VMG 分别比 SMB_China 和 HML 更适合作为规模和价值因子。

4.3　样本外模型比较结果

本文发现截至 2020 年末的最优六因子模型中不存在中国四因子模型中的 SMB_China 因子，这与之前的研究似乎有些不符。本节将通过 2021 年 1 月至 2021 年 6 月的数据对本文的最优六因子模型 {MKT SMB UMD CMA VMG PMO} 和中国四因子模型 {MKT SMB_China VMG PMO} 进行样本外的检验。

利用本文做提出的贝叶斯框架，计算得到最优六因子模型 {MKT SMB UMD CMA VMG PMO} 的对数边际似然为

$$\log m\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_{optimal}\right) = 3838.96$$

而中国四因子模型 {MKT SMB_China VMG PMO} 的对数边际似然为

$$\log m\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_{CH4}\right) = 3837.96$$

因此，在假定两个模型先验相同的情况下，本文提出的最优六因子比中国四因子更优。

利用 GRS 检验的框架对两个模型进行检验也可以得到相同的结论。以剩下的两个因子 RMW 和 HML 作为测试资产组合，对于最优六因子模型，计算得到其 p 值为 0.9634；而中国四因子模型 GRS 统计量的 p 值为 0.9609。所以，相比中国四因子模型，GRS 检验更难拒绝本文的最优六因子模型。

4.4 模型的稳健性检验

在采取 Jeffery 先验的条件下，本文介绍的贝叶斯方法为模型比较的问题找到了一个有效的解决办法。但是，这样的结果是否对于不同的先验分布是稳健的呢？因此，本节试图解决不同先验下模型结果稳健性的问题。尝试修改参数的先验分布为

$$\psi\left(\eta_1 \mid \mathscr{M}_1\right) = c\left|\Sigma_1\right|^{-\frac{q}{2}}$$

其中 q 是新增加的超参。特别地，当 $q = K+1$ 时，此先验为 Jeffery 先验；当 $q = 0$ 时，此先验满足均匀分布。进而可以推导出不同模型下参数的先验

$$\psi\left(\boldsymbol{\eta}_j \mid \mathscr{M}_j\right) = c\left|\Sigma_j\right|^{-\frac{2L_j-2K+q}{2}}\left|\Sigma_j^*\right|^{-\frac{q}{2}}.$$

类似第2.5节的做法，可以得到此先验下不同模型后验概率的解析表达式为

$$\log m\left(\boldsymbol{y}_{1:T} \mid \mathscr{M}_j\right) = \log m\left(\boldsymbol{f}_{j,1:T} \mid \mathscr{M}_j\right) + \log m\left(\boldsymbol{f}_{j,1:T}^* \mid \mathscr{M}_j\right)$$

其中

$$\begin{aligned}\log m\left(\boldsymbol{f}_{j,1:T} \mid \mathscr{M}_j\right) = &-\frac{\left(2K-q+1-L_j\right)L_j}{2}\log 2 - \frac{TL_j}{2}\log\pi - \frac{L_j}{2}\log\left(Tk_j+1\right)\\ &-\frac{\left(T+L_j-2K+q-1\right)}{2}\log\left|\Psi_j\right| + \log\Gamma_{L_j}\left(\frac{T+L_j-2K+q-1}{2}\right),\end{aligned}$$

$$\begin{aligned}\log m\left(\boldsymbol{f}_{j,1:T}^* \mid \mathscr{M}_j\right) = &\frac{\left(K-L_j\right)\left(L_j+q-K-1\right)}{2}\log 2 - \frac{\left(K-L_j\right)\left(T-L_j\right)}{2}\log\pi\\ &-\frac{\left(K-L_j\right)}{2}\log\left|W_j^*\right| - \frac{T+q-K-1}{2}\log\left|\Psi_j^*\right| + \log\Gamma_{K-L_j}\left(\frac{T+q-K-1}{2}\right).\end{aligned}$$

采用 2000 年 1 月至 2021 年 6 月全时间段的数据（包含先前讨论的样本外数据），依照解析表达式可以筛选出不同先验分布下的最佳模型，如表4.1：

表 4.1　不同先验下的最佳模型及其后验概率

先验分布参数 (q)	最优模型	最优模型的后验概率（%）
0 （均匀分布先验）	{MKT SMB UMD CMA VMG PMO}	65.15
10 （Jeffery 先验）	{MKT SMB UMD CMA VMG PMO}	65.22
20	{MKT SMB UMD CMA VMG PMO}	65.30
30	{MKT SMB UMD CMA VMG PMO}	65.36
50	{MKT SMB UMD CMA VMG PMO}	65.50
100	{MKT SMB UMD CMA VMG PMO}	65.82
200	{MKT SMB UMD CMA VMG PMO}	66.40
500	{MKT SMB UMD CMA VMG PMO}	67.62
1000	{MKT SMB UMD CMA VMG PMO}	68.32

由表4.1可以发现，本文的方法在这一组先验下表现是非常稳健的。无论如何改变先验分布的参数 q，依照本文的方法最终都会得到 {MKT SMB UMD CMA VMG PMO} 构成的六因子模型是最优模型，其后验概率的数值也较为稳定。

第五章　总结和未来展望

本文通过贝叶斯统计的方法计算出了不同潜在的因子模型的后验概率，以此作为模型比较的依据。对于模型中参数的 Jeffery 先验分布 (以及幂次型先验分布)，本文推导出了模型后验分布的解析解；而对于模型中参数其他可能的先验分布，本文采用了马尔可夫蒙特卡洛（MCMC）方法，通过积分将参数边缘化，得出模型后验概率的数值解最后，本文将此方法应用到中国股票市场的因子数据上，发现在 2000 年至 2020 年期间，最适合中国股票市场的因子模型为 {MKT SMB UMD CMA VMG PMO}。同时本文发现中国股票市场在长时间的尺度上并不存在一个绝对的最佳模型，但是在短时间（几个月）的尺度上却存在着相对最优模型。同时，很多模型在某些时间段上都会存在一个较高的后验概率，在同一时间段也有可能存在多个后验概率较高的模型。换而言之，因子模型并不是完全稀疏的。

长期看来，最佳模型几乎都包含着 SMB、VMG 和 PMO 这三个因子；RMW，HML 和 SMB_China 这三个因子很有可能是中国市场的冗余因子, 这一发现与以往的研究不同；MKT、CMA 和 UMD 这三个因子体现了长期市场的风格轮动，也是决定不同时期最佳模型的关键因子。

当然，本文提出的方法和实证结果依然存在着许多不足之处。比如，本文实证中采用了 Jeffery 先验来得出最优模型，但是这其实十分依赖参数先验分布的选取。尽管本文对一组特定的先验分布族进行了稳健性的检验，但是这并不能保证结果在任意先验下的稳健性。在未来的研究中，我们也将继续探索对相对因子模型稳健的并且有实际意义的先验分布并进行模型平均（Model Average）。

参考文献

[1] Barillas F, Shanken J. Comparing Asset Pricing Models[J]. The Journal of Finance, 2018, 73(2): 715-754.

[2] Barillas F, Shanken J. Which alpha?[J]. The Review of Financial Studies, 2017, 30(4): 1316-1338.

[3] Breeden D T. An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities[J]. Journal of financial Economics, 1979, 7(3): 265-296.

[4] Chib S, Zeng X, Zhao L. On Comparing Asset Pricing Models[J]. The Journal of Finance, 2020, 75(1): 551-577.

[5] Cochrane J H, Saa-Requejo J. Beyond Arbitrage: Good-deal Asset Price Bounds in Incomplete Markets[J]. Journal of Political Economy, 2000, 108(1): 79-119.

[6] Dybvig P H. An Explicit Bound on Individual Assets' Deviations from APT Pricing in a Finite Economy[J]. Journal of Financial Economics, 1983, 12(4): 483-496.

[7] Fama E F, French K R. A Five-factor Asset Pricing Model[J]. Journal of Financial Economics, 2015, 116(1): 1-22.

[8] Fama E F, French K R. Common Risk Factors in the Returns on Stocks and Bonds[J]. Journal of Financial Economics, 1993, 33(1): 3-56.

[9] Feng G, Giglio S, Xiu D. Taming the Factor Zoo: A Test of New Factors[J]. The Journal of Finance, 2020, 75(3): 1327-1370.

[10] Gibbons M R, Ross S A, Shanken J. A Test of the Efficiency of a Given Portfolio[J]. Econometrica: Journal of the Econometric Society, 1989: 1121-1152.

[11] Grinblatt M, Titman S. Factor Pricing in a Finite Economy[J]. Journal of Financial Economics, 1983, 12(4): 497-507.

[12] Jensen M C, Black F, Scholes M S. The Capital Asset Pricing Model: Some Empirical Tests[J]. Social Science Electronic Publishing, 1972.

[13] Lintner J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets[J]. The Review of Economics and Statistics, 1965, 47(1): 13-37.

[14] Liu J, Stambaugh R F, Yuan Y. Size and Value in China[J]. Journal of Financial Economics, 2019, 134(1): 48-69.

[15] Pástor Ľ, Stambaugh R F. Comparing Asset Pricing Models: an Investment Perspective[J]. Journal of Financial Economics, 2000, 56(3): 335-381.

[16] Shanken J. The Current State of the Arbitrage Pricing Theory[J]. The Journal of Finance, 1992, 47(4): 1569-1574.

[17] Sharpe W F. Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of risk[J]. The Journal of Finance, 1964, 19(3): 425-442.

[18] Shleifer A, Vishny R W. The Limits of Arbitrage[J]. The Journal of Finance, 1997, 52(1): 35-55.

[19] 王熙. 资产定价导论[M]. 2023.

致谢

当我写下“致谢”这两个字时，离别的日子就不远了。北大四年的时光留给了我太多的回忆。在即将和我的本科作别之际，我要向所有帮助过我和关心过我的老师、家人、同学、朋友们致以最真挚的谢意。

首先，我想要感谢我本科的老师们，是他们教授我知识，引领我走进学术科研的大门。我要特别感谢我的毕业论文导师王熙老师。从本文的选题和方法到本文的修改和答辩，我都是在王熙老师的悉心指导下完成的。他扎实的学术水平帮助我在学习和研究的过程中厘清了许多疑惑，为我打下了良好的金融学基础。我还要特别感谢我的科研导师李阿明老师。李老师教导我完成了我本科阶段的第一篇完整的科研论文，他几乎是手把手地指导我研究，每周都会抽出时间和我讨论研究进展。在我申请研究生的过程中，李老师也为我提供了很多帮助和非常中肯的建议。另外，我要感谢我在本科金融学和机器人工程双学位学习道路上每一位专业课老师，如经院的施建淮老师、石凡奇老师、王法老师，光华的李辰旭老师，工院的王雪峰老师、唐少强老师、陈光老师，数院的蒋美跃老师，信科的毛新宇老师。他们不厌其烦地在课堂内外为我答疑解惑，让我了解了一个又一个有趣的研究领域。

然后，我想要感谢我的父母和朋友，是他们在我的求学路上给我鼓励和陪伴。感谢我的爸爸妈妈这 23 年来在生活和学习上对我无微不至的照顾和无怨无悔的付出。一路走来，你们总是默默地关心着我，没有你们的鼓励和支持，我无法克服困难和取得今天的成果。感谢我的朋友们，感谢杉杉、小赫、济泽、翰阳、睿哲、博医、聪哥、星宇、杰哥、豪泽、昊天以及每一个和我一起约过饭赶过 ddl 的伙伴们，有一群有趣可爱的你们在身边真的很好。感谢你们在学习和生活中与我相伴，为我提供了无尽的鼓励、帮助和友谊。

最后，我想要感谢北大，是北大兼容并包的思想和轻松自由的氛围造就了今天的我。在北大的求学岁月里，我收获了知识、友谊和成长。我将永远珍视在这个园子里度过的宝贵时光，并将所学所得用于将来的学习和发展中！

北京大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名：

日期：　　年　　月　　日

学位论文使用授权说明

本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：

- 按照学校要求提交学位论文的印刷本和电子版本；
- 学校有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务，在校园网上提供服务；
- 学校可以采用影印、缩印、数字化或其它复制手段保存论文；

论文作者签名：

日期：　　年　　月　　日