

南京审计大学硕士研究生毕业论文——基于  $L_1$  范数的稳健  
方法在高维宏观经济预测中的应用问题研究

蒯强

2021 年 2 月 26 日

## 目录

第一章 绪论	4
§1.1 研究背景及意义	4
§1.2 国内外研究现状	6
§1.3 研究内容及技术路线	6
§1.4 研究创新	6
§1.5 本文结构	6
第二章 高维最小绝对值回归的性能研究	7
§2.1 简介	7
2.1.1 $L_1$ 范数的稳健性	7
2.1.2 最小绝对值回归的估计方法	8
§2.2 聚类——迭代拆解的算法	8
2.2.1 聚类——迭代拆解算法说明	8
2.2.2 解决最小绝对值回归问题	9
§2.3 基于替代变量最小二乘的方法	10
§2.4 模拟实验	12
§2.5 本章小结	12
第三章 基于 $L_1$ 范数主成分分析的高维因子模型估计	13
§3.1 因子分析简介	13
3.1.1 正交因子模型	13
3.1.2 动态因子模型	13
3.1.3 近似因子模型	14
§3.2 因子模型的估计	14
3.2.1 动态因子模型的估计	14
3.2.2 主成分分析法	15
3.2.3 奇异值分解	15
§3.3 $L_1$ 范数主成分分析	16
3.3.1 使用 $L_1$ 范数	16
3.3.2 缺失值处理	16
3.3.3 算法步骤	17
§3.4 稳健性实验	17
3.4.1 数据准备	17
3.4.2 实验结果	18
3.4.3 结论	19
§3.5 基于国内主要月度宏观数据的实证分析	19
3.5.1 数据说明	19
3.5.2 实证结果	20
§3.6 本章小结	21

第四章 $L_1$ 范数主成分分析算法的性能研究	22
§4.1 基于 $L_1$ 范数最大化算法	22
§4.2 基于 $L_1$ 范数最小化算法	22
§4.3 模拟实验	22

## 第一章 绪论

本章阐述了本篇论文的研究背景和研究意义，综述了国内外学者在相关领域的研究现状、主要的成果、结论和观点。由此展开，叙述了论文的主要研究内容、研究所使用的技术路线、本文的主要创新点和行文的结构。

### §1.1 研究背景及意义

在宏观经济学理论中，政府是指导和调控经济运行的主体。国民经济的发展和政府的指导和调控紧密相关，政府部门需要时刻把握国民经济的方方面面，不仅仅要关心主要产业 GDP 的增长，还要关心经济中的通货膨胀、利率和汇率等等许多复杂的经济变量，并且需要研究这些经济变量发生变化的原因以及各种经济变量之间的作用关系。只有这样，政府才能发现经济中存在的问题，并且给出针对性的指导和调控手段。

在古典经济学理论中，对于国民经济作为一个整体的活动情况一般不予考察，古典经济学往往从围观主体出发来解释观察到的宏观经济问题。然而，在市场经济高度发展的当代，市场规模空前扩大，对人类社会已经产生了深远影响。宏观经济的发展已经密切关系到民生发展，宏观经济的动荡也通过金融危机和失业直接关系到社会稳定。特别是一战后的大萧条使得古典经济学对宏观经济的分析进入了困境，在凯恩斯理论的影响下，现代宏观经济学进入主流。在现代宏观经济学的主张下，政府应该采取随机应变的宏观经济政策，通过财政和货币手段对国民经济发展进行调控和指导。

我国目前仍然是发展中国家，并且经过改革开放的 40 年，经济已经得到了巨大的发展，我国的国民经济仍然处在一个高速发展的上升阶段。在党的指导下，政府制定了保持国民经济又好又快发展的目标，意味着政府部门需要通过政策手段来保持国民经济长期平稳发展。并且，在我国，中央和地方政府享有一定的财政收入，并且能够制定各种不同的收费和税收政策来影响投资、消费和公共事业等。因此我国政府具有很好的条件来采取随机应变的宏观经济政策。我国在改革开放以来，一直保持着惊人的经济增长速度，国家和人民也从经济发展中取得了巨大的红利。然而，发展中也伴随着许多风险和挑战，我国经济受到产能问题、市场化不充分、过度依赖财政政策等问题的困扰，加上金融危机、贸易纠纷等外部冲击的作用，自从 2012 年以来，我国经济已经进入了发展的新常态阶段，我国经济面临很大的下行压力，在以习近平总书记为核心的领导下，我国经济发展需要实现新的目标：减少城乡差距、优化经济产业结构和构建创新驱动型经济，为此，各级政府正在充分发挥宏观手段，试图化解经济中的风险和为经济增长注入新的活力。

政府管理国民经济的关键是制定正确的经济政策，而做决策的前提就是预测。因此制定经济政策不仅需要各级政府对经济发展的当前状态有充分把握，并且能够对经济的未来变化有一定的预测能力。经济预测是根据经济发展的时间和空间纬度的综合信息来试图研究经济变化趋势的技术。政府为了提高经济政策的有效性，就需要通过更加准确的预测来分析和判断信息，因此政府部门需要不断追求更加准确的经济预测技术，才能在复杂的经济形式变化中掌握主动权。

除了经济发展的需要，为了规避国民经济的风险，把控我国宏观经济安全大局，也对政府更加准确地把握和预测国民经济提出了要求。当前我国正处于经济和社会向前发展的关键时期，经济虽然多年持续高速发展，但是，经济结构发展不均衡的现象逐渐显现，社会经济系统的复杂性也日渐增强。这就更需要对经济系统进行实时准确的检测和对未来经济形势准确预测，提前做好准备工作，防范经济增长的重大风险。经济风险不仅仅和宏观经济运行状况紧密关联，同样和国计民生密不可分，不管是通过对经济系统的宏观把控还是对经济运行中某些指标的局部观察，我们都可以了解经济运行的一些趋势，并且

通过对经济中的某些重要运行趋势进行预测,从而规避经济风险,稳定宏观经济安全大局。把控宏观经济安全大局对于国民经济平稳快速发展、改善国计民生、促进经济又好又快发展、促进国家安全稳定具有重要的战略和实际意义。习近平总书记于 2014 年首次提出总体国家安全观,并构建了集政治安全、国土安全、军事安全、经济安全和文化安全等领域为一体的国家安全体系。其中,经济安全是国家安全体系的重要组成部分,是国家安全的基础。宏观经济安全问题需研究 1 个或多个指标(如经济增长量)与其他指标(如资本、劳动力、人口和技术等)相互间的关系,从而揭示经济现象背后的经济规律,发现经济运行风险点,制定经济政策。为了避免宏观经济中的冲击带来的风险,需要对宏观经济总量进行实时预测和短期预测来减少不确定性,为宏观经济调控政策提供决策支持。通过对宏观经济指标进行有效监测和预测,并提前识别和防范相关风险,对保证国民经济稳步增长具有重要意义。

宏观经济预测方法繁多,有的经济学家使用经济学方程组来给宏观经济系统建立模型,从而通过模型来对某些宏观经济指标进行预测并且给出经济学分析。然而国民经济是一个非常复杂的整体,它是一个内部极其复杂的系统,并且经常伴随难以预测的外部冲击,对于复杂的经济变量,想到做到准确预测往往非常艰难。虽然现在传统的经济学理论和现代计算机技术的结合下,已经发展出了许多预测方法,并且学者们仍在不断利用新的理论和技术来完善这些已有的方法。然而,在我国国情下,进行准确的经济预测仍然非常困难。现代经济系统极为复杂,意味着使用这种方法我们就可能需要建立非常复杂的经济分析模型,有可能导致分析困难,并且往往预测结果和真实的经济运行结果有较大出入。因为系统过于复杂,学者们就开始使用基于历史数据进行向前预测的经验方法,这种方法大量应用数据和数据分析、大数据技术和统计学知识,并且融合了成熟的经济学理论和模型,往往能够起到不错的效果。在现代经济系统中,经济社会中每时每刻都在产生大量的数据,这些海量的经济活动数据给经济预测提供了丰富的数据支撑,如何有效地利用大数据时代丰富的数据资源进行宏观经济现时预测已成为一个重要的研究课题。大数据具有高维度、数量多、数据结构复杂和存储、计算困难等特征。但是结合计算机科学和统计学等方面的技术和理论知识,我们能够采取一些方法来操纵数据,从中过挖掘提炼出有效的信息。关于这个方面的研究正在火热进行,人们不断尝试使用更新的技术来处理数据,提取信息,希望利用高维度下丰富多面的经济信息来对未来的经济发展趋势作较为准确的分析。

由于宏观经济变量往往具有维数很高、分布重尾、数据集不完整等特征,导致在宏观经济的数量研究中遇到稳健性问题。例如线性回归模型的估计、主成分分析、因子模型的估计等场景下,基于最小二乘法的经典方法往往难以得到满意的结果。因此,在宏观经济数据的分析和预测中往往还需要引入更加稳健的方法。使用更加稳健的基于  $L_1$  范数的方法来代替基于  $L_2$  范数的方法例如最小二乘法,是一个很常见的做法。

$L_1$  范数在高维宏观经济数据分析中已经得到了一定的应用,本文的写作意义在于: 1)  $L_1$  范数在宏观经济数据分析中最广泛应用的场景是应用在线性模型,即最小绝对值回归,然而最小绝对值回归在处理高维宏观经济变量往往面临性能不足的问题,本文总结国内外相关的研究成果,比较分析了能够较好解决高维最小绝对值回归性能问题的两种算法,给出了一些具有参考性的结论和建议。2) 相关研究表明,近似因子模型的一种主成分估计量可以用来进行宏观经济指数的构建和有效预测。本文将  $L_1$  范数主成分分析应用在该场景中,通过实证研究论证了  $L_1$  范数可以改进原方法,在含有大量缺失值和离群值的宏观经济数据集上有更好的表现。3) 限制  $L_1$  范数主成分分析广泛应用的主要是其求解性能,近年来,不断有学者提出改进  $L_1$  范数主成分分析的算法,本文总结了一些最具有代表性的研究成果,通过数值实验比较分析了几种改进的  $L_1$  范数主成分分析算法,给出了一些具有参考性的结论和建议。

§1.2 国内外研究现状

§1.3 研究内容及技术路线

§1.4 研究创新

§1.5 本文结构

## 第二章 高维最小绝对值回归的性能研究

基于  $L_1$  范数的方法在宏观经济数据处理中的一个重要应用就是在建立线性模型时使用最小绝对值回归。在宏观经济实证研究中，经济变量往往不能够服从正态分布，而因受到经济冲击呈现出重尾的特征。使用最小二乘法建立的线性回归模型虽然具有较好的解释性，但是对原始数据的分布情况要求高，因此常常在回归中剔除异常点，使得经济变量接近正态分布，然而代价是有可能损失较多有价值的信息。采用最小绝对值回归，是更加稳健的做法，实质上它就是中位数回归，具有很好的解释性，可以很好处理重尾的宏观经济变量。

然而，应对高维宏观经济数据，最小绝对值回归由于需要解决高维线性规划问题，因此在计算的时间和空间复杂度都较高，在一定程度上限制了它的应用。本章介绍了最小绝对值回归的主流估计方法，并介绍了两种性能较高的估计方法：1) 基于聚类——迭代拆解的算法；2) 基于替代变量的牛顿迭代方法。通过模拟数值实验，验证了两种算法的高效性。

### §2.1 简介

#### 2.1.1 $L_1$ 范数的稳健性

$L_2$  范数最优化问题最常见的是最小二乘法。最小二乘法的优点很多，这里不赘述。尽管在求解大部分问题时用最小二乘估计求解可以得到比较令人满意的效果，但最小二乘法也存在一些局限性，比如，当收集的数据较少或者具有较多的缺失数据并且数据中夹杂有异常点时，用最小二乘法所得的结果就令人难以接受，在此情况下应用所得到的回归方程或模型进行预测或者拟合时，则预测或拟合的精度是相当低的，甚至根本不能使用。正因为最小二乘法对数据中的异常值十分敏感，当数据中具有较多离群值时，通过最小二乘、PCA 和 SVD 方法，得到的估计结果也会受到较大影响。

这里可以不使用最小二乘法而选择  $L_1$  范数的方法即最小一乘法，它可以增加估计的健壮性。设  $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$  为一  $p$  维随机变量， $Y$  为响应变量， $\beta$  为回归系数。假设我们观察到 *i.i.d.* 样本  $\mathbf{X}_{n \times p} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$  和  $\mathbf{Y}_{n \times 1} = (y_1, \dots, y_n)^T$ ，我们一般使用  $\beta$  的最小二乘估计量

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \|y_i - \mathbf{x}_i^T \beta\|_{L_2} = \arg \min_{\beta} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \beta)^2 \quad (2.1)$$

在稳健统计中，我们经常使用其他的目标函数，例如使用  $L_1$  范数来代替  $L_2$ ，

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \|y_i - \mathbf{x}_i^T \beta\|_{L_1} = \arg \min_{\beta} \sum_{i=1}^n |y_i - \mathbf{x}_i^T \beta| \quad (2.2)$$

如图所示，通常使用  $L_1$  范数在线性回归中可以有效避免离群值造成的干扰。

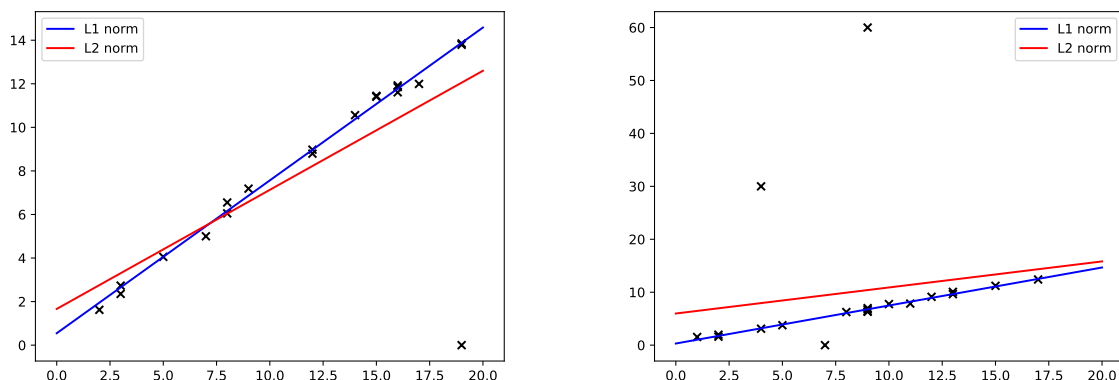


图 2.1: 如图所示, 在简单的线性模型的拟合中, 出现一个离群值就可以导致最小二乘法拟合出现明显的偏差; 而含有较多离群值时最小二乘法拟合变得很不可靠; 而采用  $L_1$  范数则具有相当稳健性。

### 2.1.2 最小绝对值回归的估计方法

对于(2.2), 它是一个凸优化问题, 但是其不具备显式解, 一般求它的数值解。但是其目标函数在  $\mathbf{0}$  点不可导, 因此不能直接使用使用梯度下降法, 一般来说, 该问题的全局最优解可以通过求解下面的线性规划问题得到:

$$\begin{aligned} \min_{\beta, t} \quad & 1^T t \\ \text{s.t.} \quad & -t \leq \mathbf{X}_{n \times p} \beta_{p \times 1} - \mathbf{Y}_{n \times 1} \leq t \end{aligned}$$

目前对于线性规划问题已经有了比较成熟的解决方法, 主要通过单纯形法或者内点法求解, 后者的时间复杂度可以控制在多项式时间, 然而, 一般而言, 当  $n$  和  $p$  均很大时, 上述线性规划问题面临很高的变量和约束维数, 计算速度仍较慢。

由于  $L_1$  范数的目标函数在机器学习领域的大量使用, 已经产生了一些光滑化方法, 做法是用一个接近  $L_1$  的目标函数来替代它, 用来替代的函数往往处处可导, 因而可以使用梯度下降法求解。典型的代表就是使用 Huber's M 统计量近似  $L_1$  范数目标函数,

$$\rho(e) = \begin{cases} \frac{1}{2}e^2, & |e| \leq \gamma \\ \gamma|e| - \frac{1}{2}\gamma^2, & |e| > \gamma \end{cases}$$

其中  $\gamma$  为某一正数, 该问题可以转化为一个二次规划问题求解。

## §2.2 聚类——迭代拆解算法

近年来针对最小绝对值回归的性能研究发展出了除了光滑化目标函数之外的方法, 可以在不改变目标函数的情况下, 通过寻求新的优化方法进行求解。这样一来, 可不改变最小绝对值回归估计量的统计性质。这里介绍 Park 等人于 2016 年提出的一种基于聚类——迭代拆解算法的最小绝对值回归求解方法。

### 2.2.1 聚类——迭代拆解算法说明

聚类是一种在机器学习中常见的做法, 就是按照某种给定的规则, 将特征接近的样本点归类到一起。聚类——迭代拆解算法的提出受到以下事实的启发: 1) 优化问题规面临的数据集模庞大, 其中许多的样



本点在进行参数估计时的贡献是很接近的；2) 如果对相似的样本点进行聚类，只在每一类别中选则少量样本进入计算，那么就会大大减小问题的规模；3) 假设在聚类后构造的新数据集上不能接近问题的最优解，那么就拆解当前的聚类，更换标准重新进行聚类，构造新的数据集；

采用聚类——迭代拆解算法求解某个优化问题的前提如下：1) 必须能够提出一种规则来对样本点聚类；2) 必须找到合适的聚类和拆解聚类的标准；3) 需要在聚类后构造的新数据集上明确定义新的优化问题；4) 能够判定当前解是否接近最优解。

算法 2.1 给出了任何一个聚类——迭代拆解算法的主要步骤，注意算法 2.1 必然在某处停止，因为每次不断拆解聚类，当聚类个数  $|K^t| = n$  时，相当于计算原问题，此时算法终止。

---

算法 2.1: 聚类——迭代拆解算法 (Aggregate and Iterative Disaggregate, AID)

---

输入：原始数据集  $\mathbf{X}_{n \times p}$ ，样本点的下标集合  $I = \{1, 2, \dots, n\}$ ，数据的特征下标集合  $J = 1, 2, \dots, p$ ，原优化问题  $P$ 。

初始化：对原始数据集  $\mathbf{X}$  聚类，然后按某种规则产生新的优化数据集  $\mathbf{X}^1$ 。

对于  $t = 1, \dots, T$ :

记  $C^t = \{C_1^t, \dots, C_K^t\}$  为聚类的集合， $K^t = 1, \dots, |K^t|$  为当前聚类的下标，

1. 根据当前聚类情况  $C^t$ ，构造新的数据集  $\mathbf{X}^t$ ，求解相应的优化问题  $P^t$ ；
  2. 检查解  $s^t$  是否达到最优条件；
  3. 如果不满足条件，拆解当前聚类。
- 

### 2.2.2 解决最小绝对值回归问题

改写(2.2)的目标函数，

$$E^* = \min_{\beta \in \mathbb{R}^p} \sum_{i \in I} |y_i - \sum_{j \in J} x_{ij} \beta_j| \quad (2.3)$$

首先给出聚类方法。给定  $|K_0|$  为目标聚类个数，初始化  $C_0 = \{C_1^0, C_2^0, \dots, C_K^0\}$ ，我们可以使用任意的聚类方法进行初始化。接下来给出如何根据聚类来产生新的数据集，对于在任一迭代周期内产生的聚类  $C_k^t, k = 1, \dots, K^t$ ，取

$$x_{kj}^t = \frac{\sum_{i \in C_k^t} x_{ij}}{|C_k^t|}, \quad j \in J \text{ 并且 } y_k^t = \frac{\sum_{i \in C_k^t} y_i}{|C_k^t|}$$

对于每一个不同的聚类需要给出一个权重来区分信息量不同的聚类，因此在新的数据集上，我们求解下面的问题

$$F^t = \min_{\beta^t \in \mathbb{R}^p} \sum_{k=1}^{K^t} |C_k^t| |y_k^t - \sum_{j \in J} x_{kj}^t \beta_j^t| \quad (2.4)$$

容易发现，任何(2.4)的可行解都是(2.3)的可行解。记  $\hat{\beta}^t$  为(2.4)的解，在每次迭代，我们计算此时的原目标函数的取值

$$E^t = \sum_{i \in I} |y_i - \sum_{j \in J} x_{ij} \hat{\beta}_j^t| \quad (2.5)$$

接下来给出拆解聚类的准则：设  $t$  步的聚类集合为  $C^t$ ，该步解为  $\hat{\beta}_t$ ，对于  $k = 1, \dots, K^t$ ，计算  $\theta_i = y_i - \sum_{j \in J} x_{ij} \hat{\beta}_j^t$ ，1) 若对于任意  $i \in C_k^t$ ， $\theta_i$  有相同的符号，那么该聚类  $C_k^t$  将保留到下次迭代，即  $C^{t+1} \leftarrow C^{t+1} \cup C_k^t$ ；

### §2.3 基于替代变量最小二乘的方法

2) 若不满足上述条件, 那么根据  $\theta_i$  符号异同, 将  $C_k^t$  分成两个集合,  $C_{k+}^t = \{i \in C_k^t | \theta_i > 0\}$ ,  $C_{k-}^t = \{i \in C_k^t | \theta_i < 0\}$ , 这两个集合在下一步形成新的聚类, 即  $C^{t+1} \leftarrow C^{t+1} \cup \{C_{k+}^t, C_{k-}^t\}$ 。

结合算法 2.1, 到这里已经给出了完成聚类——迭代拆解最小绝对值回归的所有计算步骤, 当聚类无法继续划分时, 迭代终止。

下面证明最后一次迭代的解  $\hat{\beta}_j^T$  就是(2.3)的解  $\beta^*$ ,

$$\begin{aligned} E^* &= \sum_{i \in I} |y_i - \sum_{j \in J} x_{ij} \beta_j^*| = \sum_{k \in K^t} \sum_{i \in C_k^t} |y_i - \sum_{j \in J} x_{ij} \beta_j^*| \\ &\geq \sum_{k \in K^t} \left| \sum_{i \in C_k^t} y_i - \sum_{j \in J} x_{ij} \beta_j^* \right| = \sum_{k \in K^t} |C_k^t| |y_k^t - \sum_{j \in J} x_{kj}^t \beta_j^*| \\ &\geq \sum_{k \in K^t} |C_k^t| |y_k^t - \sum_{j \in J} x_{kj} \hat{\beta}_j^t| = \sum_{k \in K^t} \left| \sum_{i \in C_k^t} y_i - \sum_{i \in C_k^t} \sum_{j \in J} x_{ij} \hat{\beta}_j^t \right| \\ &= \sum_{k \in K^t} \sum_{i \in C_k^t} |y_i - \sum_{j \in J} x_{ij} \hat{\beta}_j^t| = \sum_{i \in I} |y_i - \sum_{j \in J} x_{ij} \hat{\beta}_j^t| = E^t \end{aligned}$$

因为  $\hat{\beta}_t$  是(2.3)的可行解, 又显然  $E^* \leq E^t$ , 这就证明了  $E^* = E^t$ , 注意到  $\sum_{k \in K^t} |C_k^t| |y_k^t - \sum_{j \in J} x_{kj} \hat{\beta}_j^t|$  就是  $F^t$ , 因此  $E^t = F^t$ . 因此最后一次迭代  $F^T$  的最优解  $\hat{\beta}_j^T$  就是原问题的最优解。

### §2.3 基于替代变量最小二乘的方法

首先我们再次观察式 (2-10),

$$f_j = \arg \min_{\theta} |A^{(t-1)} \theta - x_j|$$

其中  $x_j$  为  $p$  维向量, 因而目标函数可以改写成

$$|A^{(t-1)} \theta - x_j| = \sum_{i=1}^p |A_i^{(t-1)} \theta - x_{ji}| = \sum_{i=1}^p \rho(A_i^{(t-1)} \theta - x_{ji})$$

其中  $\rho(x) = x(0.5 - \mathbb{I}[x \leq 0])$ ,  $\mathbb{I}(x)$  为指示函数。  $A_i^{(t-1)}$  为矩阵  $A^{(t-1)}$  的第  $i$  行。即

$$f_j = \arg \min_{\theta} \sum_{i=1}^p \rho(A_i^{(t-1)} \theta - x_{ji}) \quad (3-1)$$

考虑如下线性回归问题:

$$Y = \mathbf{X}^T \beta^* + e$$

其中  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{p+1})^T$  为  $p+1$  维随机变量,  $\beta^* = (\beta_0, \beta_1, \dots, \beta_p)^T$  为回归系数,  $Y$  为响应变量,  $e$  为独立于  $\mathbf{X}$  的噪声。假设我们使用  $L_1$  损失函数来估计  $\beta$ ,

$$\beta^* = \arg \min_{\beta \in \mathbb{R}^{p+1}} \mathbb{E} |Y - \mathbf{X}^T \beta| = \arg \min_{\beta \in \mathbb{R}^{p+1}} \mathbb{E} \rho(Y - \mathbf{X}^T \beta) \quad (3-2)$$

若已知  $n$  i.i.d. 的样本  $(\mathbf{X}_i, Y_i)$  ( $1 \leq i \leq n$ ), 令  $\hat{\beta}$  为  $\beta^*$  的估计量, 则

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n \rho(Y_i - \mathbf{X}_i^T \beta) \quad (3-3)$$

可以发现式 (3-3) 与 (3-1) 有相同的问题形式，因此我们不妨考虑换一种方法估计  $\beta^*$ ，用新的估计量来代替  $\hat{\beta}$ 。

我们使用牛顿迭代法求解下面的随机优化问题：

$$\beta^* = \arg \min_{\beta \in \mathbb{R}^{p+1}} \mathbb{E}[G(\beta; \mathbf{X}, Y)] \quad (3-4)$$

其中  $G(\beta; \mathbf{X}, Y)$  是损失函数， $\mathbf{X}$  和  $Y$  分别是  $p+1$  维自变量和一元响应变量， $\beta$  为回归系数。使用牛顿-拉弗森迭代来求解，

$$\tilde{\beta}_1 = \beta_0 - \mathbf{H}(\beta_0)^{-1} \mathbb{E}[g(\beta; \mathbf{X}, Y)] \quad (3-5)$$

其中  $\beta_0$  是一个初始估计， $g(\beta; \mathbf{X}, Y)$  为损失函数  $G(\beta; \mathbf{X}, Y)$  关于  $\beta$  的梯度。

$\mathbf{H}(\beta) := \partial \mathbb{E}[g(\beta; \mathbf{X}, Y)] / \partial \beta$  表示  $\mathbb{E}G(\beta; \mathbf{X}, Y)$  的海赛矩阵。特别地，我们这里考虑损失函数为  $L_1$  损失的特殊情形，

$$G(\beta; \mathbf{X}, Y) = \rho(Y - \mathbf{X}^T \beta) \quad (3-6)$$

在 (3-6) 的条件下， $g(\beta; \mathbf{X}, Y) = \mathbf{X}(\mathbb{I}[Y - \mathbf{X}^T \beta < 0] - 0.5)$ 。

并且， $\mathbf{H}(\beta) = \mathbb{E}(\mathbf{X}\mathbf{X}^T f(\mathbf{X}^T(\beta - \beta^*)))$ ，这里  $f(x)$  是噪声  $e$  的密度函数。当初始估计量  $\beta_0$  和  $\beta^*$  很接近时， $\mathbf{H}(\beta_0)$  就会很接近  $\mathbf{H}(\beta^*) = \Sigma f(0)$ ，这里  $\Sigma = \mathbb{E}\mathbf{X}\mathbf{X}^T$  是  $\mathbf{X}$  的协方差矩阵。使用  $\mathbf{H}(\beta^*)$  替换式 (3-5) 中的  $\mathbf{H}(\beta_0)$ ，可得

$$\beta_1 = \beta_0 - \mathbf{H}(\beta^*)^{-1} \mathbb{E}[g(\beta; \mathbf{X}, Y)] = \beta_0 - \Sigma^{-1} f^{-1}(0) \mathbb{E}[g(\beta_0; \mathbf{X}, Y)] \quad (3-7)$$

我们在  $\beta^*$  对  $\mathbb{E}[g(\beta_0; \mathbf{X}, Y)]$  进行泰勒展开，

$$\begin{aligned} \mathbb{E}[g(\beta_0; \mathbf{X}, Y)] &= \mathbf{H}(\beta^*)(\beta_0 - \beta^*) + O(|\beta_0 - \beta^*|_2^2) \\ &= \Sigma f(0)(\beta_0 - \beta^*) + O(|\beta_0 - \beta^*|_2^2) \end{aligned}$$

结合式 (3-7)，可以得到

$$\begin{aligned} |\beta_1 - \beta^*|_2 &= |\beta_0 - \Sigma^{-1} f^{-1}(0)(\Sigma f(0)(\beta_0 - \beta^*) + O(|\beta_0 - \beta^*|_2^2)) - \beta^*|_2 \\ &= O(|\beta_0 - \beta^*|_2^2) \end{aligned}$$

因此，如果我们得到一个  $\beta^*$  的一致估计量  $\beta_0$ ，我们就可以通过式 (3-7) 中的牛顿-拉弗森迭代得到误差更小的估计。

下面我们将 (3-7) 转化成一个最小二乘问题。首先我们重写该式，

$$\begin{aligned} \beta_1 &= \Sigma^{-1}(\Sigma \beta_0 - f^{-1}(0) \mathbb{E}[g(\beta_0; \mathbf{X}, Y)]) \\ &= \Sigma^{-1} \mathbb{E}[\mathbf{X}\{\mathbf{X}^T \beta_0 - f^{-1}(0)(\mathbb{I}[Y \leq \mathbf{X}^T \beta_0] - 0.5)\}] \end{aligned}$$

这里我们定义一个新的响应变量  $\tilde{Y}$ ，

$$\tilde{Y} = \mathbf{X}^T \beta_0 - f^{-1}(0)(\mathbb{I}[Y \leq \mathbf{X}^T \beta_0] - 0.5)$$

那么  $\beta_1 = \Sigma^{-1} \mathbb{E}(\mathbf{X}\tilde{Y})$  就是线性回归问题  $\tilde{Y} = \mathbf{X}^T \beta$  的最优回归系数，即

$$\beta_1 = \arg \min_{\beta \in \mathbb{R}^{p+1}} \mathbb{E}(\tilde{Y} - \mathbf{X}^T \beta)^2$$

假设我们获得了 *i.i.d.* 样本  $(\mathbf{X}_i, Y_i)$ , 则构造

$$\tilde{Y}_i = \mathbf{X}_i^T \hat{\beta}_0 - \hat{f}^{-1}(0)(\mathbb{I}[Y_i \leq \mathbf{X}_i^T \hat{\beta}_0] - 0.5)$$

其中  $\hat{\beta}_0$  为  $\beta^*$  的一个初始估计,  $\hat{f}(0)$  为  $f(0)$  的一个估计, 那么在牛顿迭代的每一步骤,

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - \mathbf{X}_i^T \beta)^2$$

这里我们选择最小二乘估计量作为  $\hat{\beta}_0$ , 并且采用  $f(0)$  的核密度估计作为  $\hat{f}(0)$ ,

$$\hat{\beta}_0 = \arg \min_{\beta \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n (Y_i - \mathbf{X}_i^T \beta)^2$$

$$\hat{f}(0) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{Y_i - \mathbf{X}_i^T \hat{\beta}_0}{h}\right)$$

其中  $K(x)$  为高斯核函数,  $h \rightarrow 0$  是带宽。

---

#### 算法 2 使用牛顿-拉弗森迭代计算 $L_1$ 最小化问题

---

输入:  $Y$  和  $\mathbf{X}$  的样本  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ ,  $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T, \dots, \mathbf{X}_n^T)$ , 迭代次数  $t$ , 高斯核函数  $K$ , 带宽  $h_g (g = 1, \dots, t)$

初始化: 给出

$$\hat{\beta}^{(0)} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n (Y_i - \mathbf{X}_i^T \beta)^2$$

对于  $g = 1, \dots, t$ :

1. 计算  $\hat{f}^{(g)}(0)$ ,

$$\hat{f}^{(g)}(0) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{Y_i - \mathbf{X}_i^T \hat{\beta}^{(g-1)}}{h_g}\right)$$

2. 计算  $\tilde{Y} = (\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_n)$ ,

$$\tilde{Y}_i = \mathbf{X}_i^T \hat{\beta}^{(g-1)} - \hat{f}^{(g)}(0)^{-1}(\mathbb{I}[Y_i \leq \mathbf{X}_i^T \hat{\beta}^{(g-1)}] - 0.5)$$

3. 计算  $\hat{\beta}^{(g)}$ ,

$$\hat{\beta}^{(g)} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - \mathbf{X}_i^T \beta)^2$$

输出:  $\beta^{(t)}$

---

## §2.4 模拟实验

## §2.5 本章小结

## 第三章 基于 $L_1$ 范数主成分分析的高维因子模型估计

### §3.1 因子分析简介

因子分析是主成分分析的推广和发展，它也是多元统计中广泛使用的一种降维方法。因子分析研究相关矩阵或者协方差矩阵的内部依赖关系，它将多个变量综合为少数几个因子，以再现原始变量和因子之间的关系。

因子分析最早由英国心理学家 Charles Spearman 提出，最早用于智力测验得分数据的分析中。现在因子分析在心理学、社会学、经济学等学科都取得了广泛应用。

#### 3.1.1 正交因子模型

设  $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$  是可观测的随机向量， $\mathbb{E}\mathbf{X} = \boldsymbol{\mu}$ ， $\mathbb{D}(\mathbf{X}) = \boldsymbol{\Sigma}$ ，设  $\mathbf{f} = (f_1, f_2, \dots, f_m)^T$  ( $m < p$ ) 为不可观测的随机向量，且  $\mathbb{E}\mathbf{f} = \mathbf{0}$ ， $\mathbb{D}(\mathbf{f}) = \mathbf{I}_m$  (即  $\mathbf{f}$  的各分量方差为 1，且互不相关)。最后，设  $\mathbf{e} = (e_1, e_2, \dots, e_p)^T$  和  $\mathbf{f}$  互不相关，并且

$$\mathbb{E}(\mathbf{e}) = \mathbf{0}, \mathbb{D}(\mathbf{e}) = \text{diag}(\sigma_1^2, \dots, \sigma_p^2) \text{ 为对角矩阵。}$$

假设随机向量  $\mathbf{X}$  满足以下模型：

$$\begin{cases} X_1 - \mu_1 = a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m + e_1, \\ X_2 - \mu_2 = a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m + e_2, \\ \dots \\ X_p - \mu_p = a_{p1}f_1 + a_{p2}f_2 + \dots + a_{pm}f_m + e_p, \end{cases}$$

我们称该模型为正交因子模型，矩阵表示为：

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{f} + \mathbf{e} \quad (2-1)$$

其中  $f_1, f_2, \dots, f_m$  称为  $\mathbf{X}$  的公共因子； $e_1, e_2, \dots, e_p$  称为  $\mathbf{X}$  的特殊因子。公共因子一般对  $\mathbf{X}$  的每一个分量都起作用，而  $e_i$  一般仅仅对  $X_i$  起作用。并且各个特殊因子之间以及特殊因子和所有公共因子之间都是不相关的。其中  $\mathbf{A} = (a_{ij})_{p \times m}$  是待估计的系数矩阵，称为因子载荷矩阵。 $a_{ij}$  称为第  $i$  个变量在第  $j$  个因子上的载荷，它反映了第  $i$  个变量在第  $j$  个因子上的相对重要性。

#### 3.1.2 动态因子模型

经典因子模型即正交因子模型提出后，迅速成为在心理学、社会学等多个学科内得到了广泛的使用。然而因子模型在计量经济学研究领域的应用出现较晚，这可能是因为经典因子模型主要用于处理截面数据，而计量经济学，尤其是宏观计量经济学的主要研究对象为时间序列数据。

Geweke 与 Sargent 等人首先在计量经济学领域提出了动态因子模型，动态因子模型是经典因子模型在时间序列数据上的延伸，它提供了从维数众多的经济时间序列数据中提取公共因子来研究和解释经济波动的手段。动态因子模型的基本思想是：宏观经济具有周期性的波动是通过一系列经济变量的活动来传递和扩散的，任何一个经济变量本身的波动都不足以代表宏观经济的整体波动。对于经济波动的研究需要从多个具有相依性的经济变量同时着手。因此，需要从一国许多经济时间序列数据中估计出驱动各变量波动的共同动态因子，并且给出解释。动态因子模型已经成为判别和分析经济周期波动的有效工具。

令  $\mathbf{X}_t = (\mathbf{X}_{1t}, \mathbf{X}_{2t}, \dots, \mathbf{X}_{pt})^T$  为一组宏观经济变量在  $t$  时刻的水平，并且  $\mathbf{X}_t$  可以表达为如下形式：

$$\mathbf{X}_t = \lambda(L)\mathbf{f}_t + \mathbf{e}_t \quad (2-2)$$

其中， $\mathbf{f}_t$  为  $q$  维的动态因子向量， $\lambda(L)$  为由  $s$  阶滞后多项式算子组成的  $p \times q$  矩阵。动态因子模型通常假设动态因子向量服从某一个向量随机过程。即动态因子模型不仅允许观测变量受因子滞后项的影响，而且也允许因子本身具有独立的动态演化过程。

#### 3.1.3 近似因子模型

在近似因子模型出现以前，DFM 主要应用于仅包含几个总体宏观经济变量的小型数据集。这主要是由于经济学家们假定影响所有变量的普遍性结构冲击是经济变量协同变动的原因。然而在包含成百上千个变量的大型经济数据集中，往往存在仅影响某一组变量的部门冲击或局部冲击，若根据经济理论将这些非普遍性冲击归于异质性部分，则会造成异质性部分截面相关，这又违背了经典因子模型中异质性部分相互正交的假定。

因此，Chamberlain 和 Rothschild 放弃异质性部分的协方差矩阵为对角阵的假定，允许异质性部分存在一定程度的截面相关，模型假设允许  $\mathbf{f}_t$  的各个分量  $f_1, f_2, \dots, f_m$  具有相依性， $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$  可以不独立，并且允许  $\mathbf{f}_t, \mathbf{e}_t$  具有时间序列相依性。从而将模型扩展为了为动态近似因子模型。

### §3.2 因子模型的估计

对于正交因子模型，在因子载荷矩阵  $\mathbf{A}$  中，我们计算各列的平方和，记为  $q_j^2$ ，即

$$q_j^2 = \sum_{i=1}^p a_{ij}^2 (j = 1, 2, \dots, m)$$

$q_j^2$  表示第  $j$  个公共因子  $f_j$  对  $\mathbf{X}$  的所有分量的总影响，称为公共因子  $f_j$  对  $\mathbf{X}$  的方差贡献。因此如果我们列出  $\mathbf{A}$  的所有列平方和，按照方差贡献大小就可以选出最有影响的公共因子。因此因子分析的关键步骤就是估计出因子载荷矩阵。

对于正交因子模型可以通过主成分法、主因子法和极大似然法来估计因子载荷矩阵。不可观测的公共因子有时候也需要进行估计，例如用来诊断因子模型或者作为进一步分析的原始数据，这就需要我们给出因子得分，一般在得出了因子载荷矩阵之后，对于因子得分，可以通过加权最小二乘法（巴特莱因子得分）或者通过回归法（汤普森因子得分）进行估计。

但是对于动态因子模型和近似动态因子模型，因为不满足正交因子模型的一些关键模型假设，因此需要不同的估计手段。

#### 3.2.1 动态因子模型的估计

动态因子模型的提出者 Geweke 和 Sargent 等人都是使用频域方法估计模型，这种方法不能够直接估计出动态因子  $\mathbf{f}_t$ ，因此也就不能将因子作为预测或者扩展模型等用途。使得动态因子模型的应用收到限制。由于动态近似因子模型估计困难，因此在预测中，往往使用模型的静态形式：

$$\mathbf{X}_t = \mathbf{A}\mathbf{F}_t + \mathbf{e}_t \quad (2-3)$$

式 (2-3) 中,  $F_t$  是  $m$  维向量, 称为静态因子, 即  $F_t$  仅在当期影响  $X_t$  (因为  $F_t$  包括了动态因子  $f_t$  的当期项和滞后项), 它本身可以不具有经济学含义。 $A$  是因子载荷矩阵。本文采用式 (2-3) 中的模型进行宏观经济指标的预测。

对于静态形式的估计, Stock 和 Watson 于 1989 年最早提出了一种方法: 这种方法采用 Kalman 滤波构造似然函数, 并采用极大似然方法来估计参数。这种方法的优点是: 在误差项服从正态分布的假定下, 能够得到因子的有效估计量。然而, 由于在估计参数过程中会应用到非线性数值优化算法, 而待估参数的个数与变量维数  $p$  成比例, 限于当时的计算能力, 这种方法只能用于处理低维的精确动态因子模型。

针对维数很高的宏观经济变量, Stock 和 Watson 于 2002 年给出了一种非参数的方法, 并且得到广泛使用。这种方法使用主成分分析法进行估计, 对式 (2-3), 因子载荷矩阵  $A$  的估计量  $\hat{A}$  即为  $X_t$  的样本协方差矩阵

$$\hat{\Sigma}_X = \frac{1}{n} \sum_{t=1}^n X_t X_t^T \quad (2-4)$$

的前  $r$  个最大特征值所对应的特征向量组成的  $p \times r$  维矩阵, 因子的主成分估计量为:

$$\hat{F}_t = \frac{1}{p} \hat{A} X_t \quad (2-5)$$

主成分估计量具有一系列良好的性质。例如, 在  $p \rightarrow \infty$ ,  $n \rightarrow \infty$  且  $p^2/n \rightarrow \infty$  的条件下, 主成分估计量  $\hat{F}_t$  是因子空间的一致估计量, 且在随后的建模过程中可以当作观测到的数据。

#### 3.2.2 主成分分析法

主成分分析法 (principal component analysis, PCA) 已经是十分常见的数据降维方法, 这里我们不对主成分分析法做详细介绍, 仅针对我们的问题, 下面给出使用主成分分析法计算出估计量的步骤:

记矩阵  $X_{p \times n}$  为  $p$  维宏观经济变量  $X_t$  在  $t_1, \dots, t_n$  的  $n$  次观测,

1. 将  $X$  进行中心化;
2. 计算样本协方差矩阵  $\frac{1}{n} X X^T_{p \times p}$ ;
3. 对样本协方差矩阵特征值分解, 并从小到大排列这些特征值;
4. 因子载荷矩阵估计量  $\hat{A}_{p \times m} (m < p)$  为前  $r$  个特征值对应的特征向量组成的矩阵。

不难看出, 主成分分析法进行估计的关键计算步骤是对样本协方差矩阵的特征值分解。

#### 3.2.3 奇异值分解

这里有必要简要介绍奇异值分解 (singular value decomposition, SVD), 它也是一类常用的降维方法, 在常用统计软件包中 PCA 问题的计算往往转化成奇异值分解问题求解。

奇异值分解解决将矩阵  $A$  分解成正交矩阵  $U$  和对角矩阵  $\Sigma$  和另一正交矩阵  $V^T$  的问题, 即

$$A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T$$

奇异值分解求解的关键就是得到奇异值, 而后者是  $A^T A$  的特征值的平方根, 即奇异值分解的关键计算步骤是对  $A^T A$  进行特征值分解。这里假设我们取  $A = X^T / \sqrt{m}$ , 不难看出和 PCA 问题等价。因此所有的 PCA 问题可以转化为 SVD 进行求解, 这样做的原因是高维矩阵进行特征值分解的效率很低, 而 SVD 问题有不需进行特征值分解的更高效的迭代计算方法。

§3.3  $L_1$  范数主成分分析

Stock 和 Watson 提出的主成分估计量等价于通过下面的  $L_2$  范数优化问题来估计  $\mathbf{A}$  和  $\mathbf{F}$ ，即

$$\hat{\mathbf{A}}_{p \times m}, \hat{\mathbf{F}}_{m \times n} = \arg \min_{\mathbf{A}, \mathbf{F}} \|\mathbf{X} - \mathbf{A}\mathbf{F}\|_{L_2} = \arg \min_{\mathbf{A}, \mathbf{F}} \sum_{i=1}^p \sum_{j=1}^m (x_{ij} - a_i^T f_j)^2 \quad (2-6)$$

上述问题不是一个凸优化问题，我们可以转化为等价的 PCA 问题或者 SVD 问题解决。但是我们采用的等价解决方法具有所有的  $L_2$  范数优化方法普遍的缺陷，那就是对离群值十分敏感。并且 PCA 和 SVD 方法都不能够直接处理具有缺失值的数据，因此对于缺失数据必须进行插补。

然而由于各种原因，高维宏观经济数据中往往具有大量的缺失值和离群值，这就给因子模型的估计带来很多的麻烦，从而进一步使用因子的估计量进行预测可能会变得不够准确。

3.3.1 使用  $L_1$  范数

将式 (2-6) 中的目标函数更换为使用  $L_1$  范数，可得

$$\hat{\mathbf{A}}_{p \times m}, \hat{\mathbf{F}}_{m \times n} = \arg \min_{\mathbf{A}, \mathbf{F}} \|\mathbf{X} - \mathbf{A}\mathbf{F}\|_{L_1} = \arg \min_{\mathbf{A}, \mathbf{F}} \sum_{i=1}^p \sum_{j=1}^m |x_{ij} - a_i^T f_j| \quad (2-7)$$

其中  $\|\cdot\|_{L_1}$  表示矩阵的  $L_1$  范数。

式 (2-7) 中定义的问题不是一个凸优化问题。但是一旦矩阵  $\mathbf{A}$  或者  $\mathbf{F}$  固定为常数，那么该问题就成为了一个凸优化问题，可以找到全局最优解。这启发我们可以使用交替优化的方法求解这个优化问题，即每一步固定  $\mathbf{A}$  或者  $\mathbf{F}$  的值，来求解另一个参数。

$$\mathbf{F}^{(t)} = \arg \min_{\mathbf{F}} \|\mathbf{X} - \mathbf{A}^{(t-1)}\mathbf{F}\|_{L_1} \quad (2-8a)$$

$$\mathbf{A}^{(t)} = \arg \min_{\mathbf{A}} \|\mathbf{X} - \mathbf{A}\mathbf{F}^{(t)}\|_{L_1} \quad (2-8b)$$

我们改写式 (2-8a) 中的目标函数，

$$E(\mathbf{F}) = \|\mathbf{X} - \mathbf{A}^{(t-1)}\mathbf{F}\|_{L_1} = \sum_{j=1}^n |x_j - \mathbf{A}^{(t-1)} f_j| \quad (2-9)$$

其中  $x_j$  是矩阵  $\mathbf{X}$  的第  $j$  列， $f_j$  是  $\mathbf{F}$  的第  $j$  列。于是式 (2-8a) 中的问题可以分解为  $n$  个独立的子优化问题，求解相应的  $f_j$ ：

$$f_j = \arg \min_{\theta} |\mathbf{A}^{(t-1)}\theta - x_j| \quad (2-10)$$

## 3.3.2 缺失值处理

在使用 PCA 和 SVD 时我们需要对矩阵  $\mathbf{X}$  进行缺失值插补，然后才能进行计算。在  $L_1$  算法中我们不需要进行缺失值插补，在式 (2-11) 中，遇到  $x_j$  具有缺失值的场合，我们直接舍弃相应的约束条件即可。我们改写式 (2-9)，

$$E(\mathbf{F}) = \sum_{i=1}^d \sum_{j=1}^n |x_{ij} - a_i^T f_j|$$

如果某个项  $x_{ij}$  缺失，我们直接舍弃目标函数中的对应累加项，在上述算法中对应的做法就是直接删除 (2-11) 中的一个约束条件。



## 3.3.3 算法步骤

我们已经发现可以通过交替优化的方法求解式 (2-7) 中的优化问题, 下面我们更加详细地讨论  $L_1$  因子分析算法一些细节和具体实现步骤。

## (一) 初始化

和其他所有的迭代算法一样, 在算法开始时, 首先我们需要给  $\mathbf{A}$  一个初始值  $\mathbf{A}^{(0)}$ 。对于  $\mathbf{A}$  可以采用简单随机数进行初始化, 这里我们为了加快收敛速度, 可以使用经过缺失值插补 (这里我们使用均值插补) 后通过 PCA 算法进行得到的因子载荷矩阵作为  $\mathbf{A}^{(0)}$ 。在本章后续小结的实验中我们可以发现, 在含有大量缺失值和离群值的条件下, 两种不同的初始化方法最终结果差异并不大。

## (二) 收敛性

因为目标函数  $E(\mathbf{A}, \mathbf{F}) = \|\mathbf{X} - \mathbf{A}\mathbf{F}\|_{L_1}$  在每一个交替的优化步骤中都递减, 并且  $E(\mathbf{A}, \mathbf{F})$  具有下界 ( $\geq 0$ )。因此交替优化算法一定收敛。因此我们可以设定一个收敛域值来停止迭代, 这里我们设置终止条件:

$$\theta(a_i^{(t)}, a_i^{(t-1)}) < \alpha$$

这里  $\theta(a, b)$  表示向量  $a$  和  $b$  的夹角; 其中  $a_i$  是  $\mathbf{A}$  或者  $\mathbf{F}$  的第  $i$  列;  $\alpha$  是一个很小的正数。

## (三) 算法步骤

下面我们给出  $L_1$  因子分析算法的计算步骤:

1. 初始化: 给出  $\mathbf{A}$ ,  $\Sigma$  的初始值  $\mathbf{A}^{(0)}$ ,  $\Sigma^{(0)} = \mathbf{I}$ , (其中  $\Sigma$  为一对角矩阵,  $\mathbf{I}$  为单位矩阵);
2. 交替凸优化: 对于迭代次数  $t = 1, \dots$ , 收敛:

$$\mathbf{F}^{(t)} = \arg \min_{\mathbf{F}} \|\mathbf{X} - \mathbf{A}^{(t-1)} \Sigma^{(t-1)} \mathbf{F}^T\|_{L_1}$$

$$\mathbf{A}^{(t)} = \arg \min_{\mathbf{A}} \|\mathbf{X} - \mathbf{A} \Sigma^{(t-1)} \mathbf{F}^{(t)T}\|_{L_1}$$

$$\text{归一化: } \begin{cases} N_a = \text{diag}(\mathbf{A}^{(t)T} \mathbf{A}^{(t)}) \\ N_f = \text{diag}(\mathbf{F}^{(t)T} \mathbf{F}^{(t)}) \\ \mathbf{F}^{(t)} \leftarrow \mathbf{F}^{(t)} N_f^{-1} \\ \mathbf{A}^{(t)} \leftarrow \mathbf{A}^{(t)} N_a^{-1} \\ \Sigma^{(t)} \leftarrow N_a \Sigma^{(t-1)} N_f \end{cases}$$

3. 输出结果:  $\mathbf{F} \leftarrow \mathbf{F} \Sigma^{1/2}$ ,  $\mathbf{A} \leftarrow \mathbf{A} \Sigma^{1/2}$

## §3.4 稳健性实验

为了检验  $L_1$  因子分析算法在处理含有大量离群值和缺失值的数据时的稳健性, 我们进行模拟实验, 来对比采用  $L_1$  因子分析算法和 PCA 因子分析, SVD 因子分析以及采用  $L_2$  目标函数这几种方法的效果。

## 3.4.1 数据准备

为了进行模拟实验, 我们首先需要随机产生一个高维低秩的矩阵来模拟高维宏观经济数据集。我们产生一个  $n$  维方阵  $M$ , 其中每一个随机元素均服从  $[-100, 100]$  的均分分布。然后我们对方阵  $M$  进行奇异值分解,  $M = U \Sigma V^T$ 。假设我们需要产生的低秩矩阵的秩为  $r$ , 则

$$\mathbf{X} = U_{(:, 1:r)} \Sigma_{(1:r, 1:r)} V_{(:, 1:r)}^T$$

即为我们得到的模拟高维低秩矩阵。

之后我们可以设置一定比例的缺失值和离群值，首先我们在矩阵的左下角剔除部分元素。在剩下的元素中，我们随机选取一部分然后重新产生随机元素，每个元素服从  $[-2000, 2000]$  上的均匀分布。图 2.1 展示了一个  $30 \times 30$  秩为 3 的矩阵在模拟了缺失值和离群值后的情况。

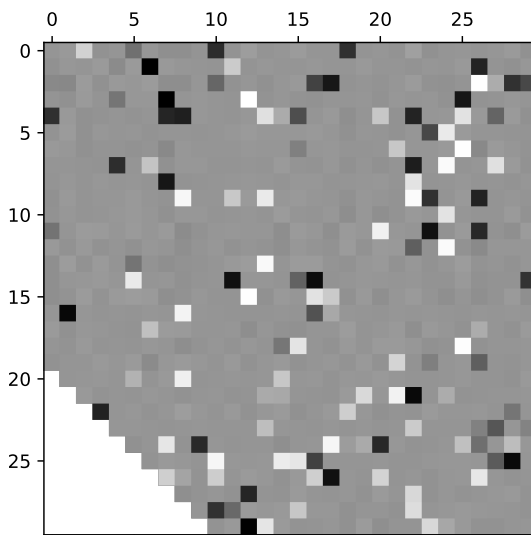


图 3.1:  $30 \times 30$  的模拟矩阵，左下角白色代表缺失值，图中灰度越深代表元素的绝对值越大，因此深黑色的点代表了离群值

### 3.4.2 实验结果

基于获得的高维低秩矩阵，我们主要比较四种获得因子估计量的方法：1) PCA，需要进行缺失值插补，这里我们使用均值插补法；2) SVD，同样使用均值插补；3) 基于  $L_1$  范数的估计方法，可以直接处理原始数据；4) 一种基于加权  $L_2$  范数的估计方法 (Black 和 Rangarajan, 1996)，通过将 (2-6) 损失函数加上随迭代变化的权重，可以直接处理原始数据。

令  $\hat{\mathbf{X}} = \hat{\mathbf{A}}_{N \times M} \hat{\mathbf{F}}_{M \times N}$ ，我们比较因子残差  $\mathbf{E} = \hat{\mathbf{X}} - \mathbf{X}$ 。观察因子残差的平方的分布情况，试验每一种算法作用在无干扰数据上和干扰后数据上的情况。

对于无干扰数据，每一种算法都具有很好的表现；对于进行了缺失值和离群值模拟的数据，我们每组试验（给定  $M, N$ ）下均重复多次取平均值，图 2.3 给出了一个典型结果。

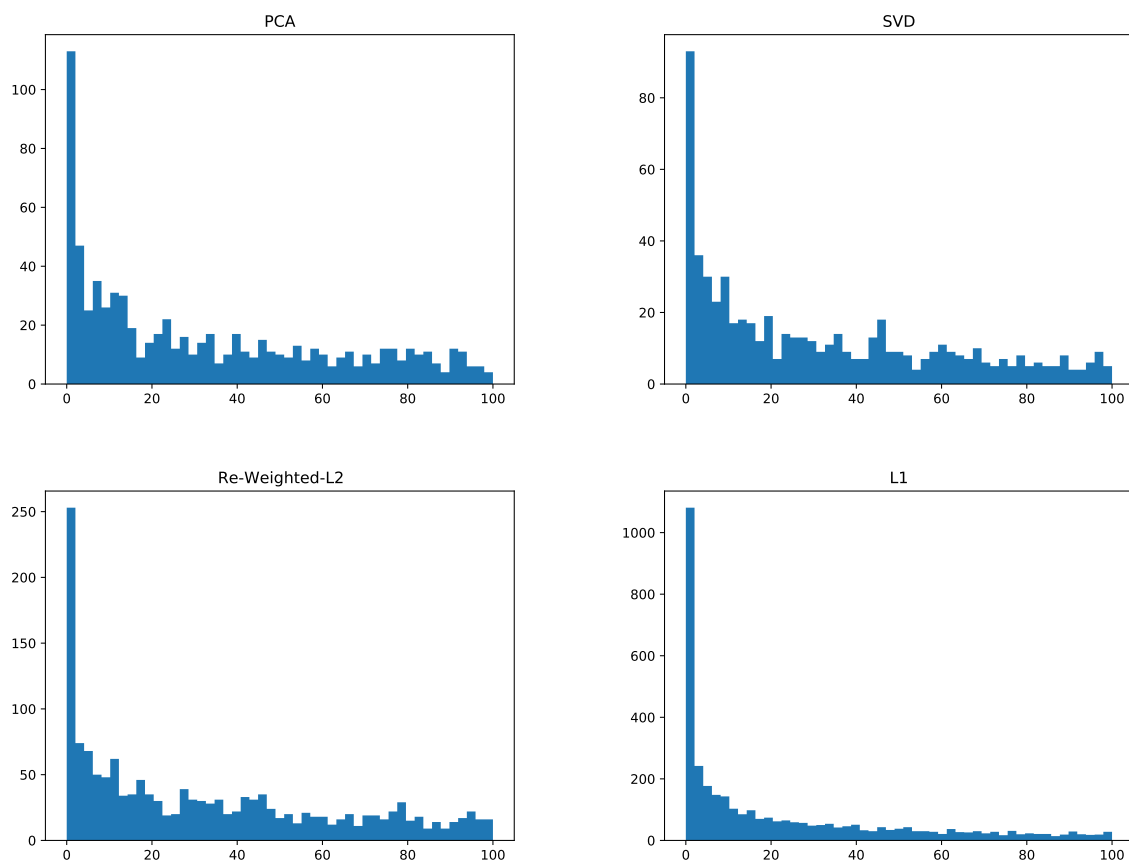


图 3.2: 图示为取  $N = 80, M = 3$ , 缺失值和离群值比例均为 10% 的情况下, 多次实验中平均因子残差的平方的分布情况。可以看出采用基于  $L_1$  范数进行估计, 因子残差表现最好。

#### 3.4.3 结论

实验表明, 基于  $L_1$  范数的因子模型估计相比主成分分析法和其他基  $L_2$  范数优化的方法相比具有更好的稳健性, 特别适合于处理具有大量离群值的数据。

## §3.5 基于国内主要月度宏观数据的实证分析

### 3.5.1 数据说明

我们搜集了来自中国国家统计局、中国人民银行、海关总署、中国人力资源市场监测中心等部门以及国泰安经济研究数据库从 1999 年 9 月至 2019 年 6 月主要公开月度 (部分季度) 宏观经济指标。其中我们人为筛选出涵盖了宏观经济中实际产出 (工业增加值增长率、能源产量增长率等)、价格指数 (如 CPI、PPI、房地产价格指数等)、进出口 (进、出口同比增速等)、财政 (公共财政收入、支出等)、金融 (包括货币、信贷增速以及股票市场成交额和证券发行量等)、各种景气指数 (包括消费者信心指数、消费者预期指数、制造业采购经理指数等)、消费和零售 (各类消费品零售总额增速、居民收入变化率等)、投资 (包括固定资产投资增长、房地产投资增长、外商投资等) 共计 117 个指标的月度时间序列。

对不平稳的增长率数据, 我们对其进行一阶差分 (存在季节效应的进行季节差分), 对于非增长率数

据，我们对其进行对数一阶差分（存在季节效应的数据进行对数季节差分），仍然不能平稳的数据再次进行二阶差分。之后所有数据均进行标准化处理。我们没有剔除任何离群值。

由于某些指标曾多次改变统计口径、统计频率等原因，这些数据中包含许多的缺失数据。另外，在 1999~2019 年这 20 年间，中国经济经历了许多次重大外部冲击，许多经济指标中含有大量离群值（见图 3），从各指标的经验分布的峰度值来看，经济指标的分布存在尖峰厚尾现象。

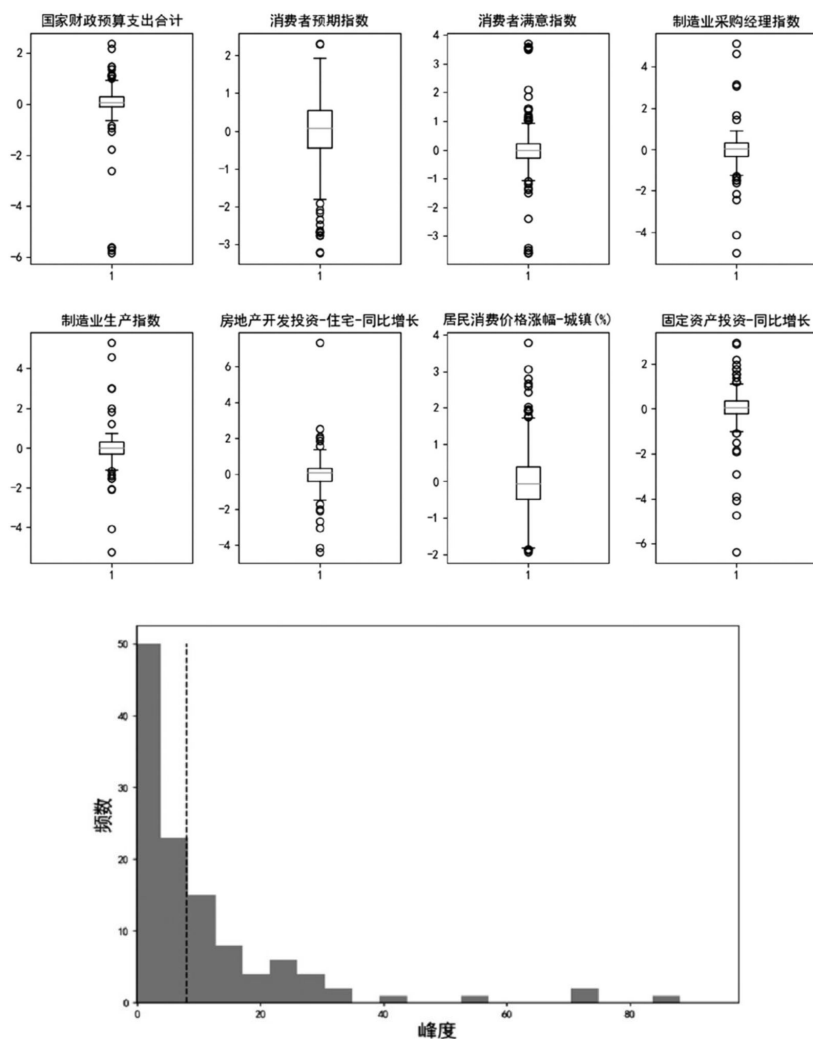


图 3.3: 部分指标的箱形图和各经济指标的峰度分布情况

#### 3.5.2 实证结果

参考由 Bai 与 Ng 提出的用于选择静态因子个数的信息准则，我们将由 117 维月度数据集的共同因子个数定为 8 个。模型中线性预测模型部分滞后阶数由最小 BIC 准则给出。我们分别使用 L1 方法和 PCA 估计因子载荷和因子得分，从可以看出，L1 方法和 PCA 得到的因子载荷是有明显区别的。

接下来比较 PCA 估计得出的因子和 L1 方法得到的因子在指标预测中的表现。我们在处理后的数据集中选取了 10 个经济指标，将固定时间长度（180 个月）的数据集设为训练数据，使用滑动窗口预测，并设置预测步长  $h$  为 1 个月、3 个月和 6 个月。并分别比较了两组因子的预测效果，为比较预测效果选

### §3.6 本章小结

	因子 1	因子 2	因子 3	因子 4	
1	先行指数	工业增加值增长率	海关代征增值税	成交金额——股票	
2	货币供应量同比增长	工业增加值——私营企业	制造业各项税收	成交金额——期货	
3	预警指数	工业增加值——股份合作企业	财政预算收入	成交金额增长率	
4	制造业采购经理指数	工业增加值——国有及国有控股企业	关税	本月流通股换手率	居
5	制造业生产指数	工业增加值——重工业	房屋新开工面积增长率	当月成交金额	居

取了三个主要度量标准，即均方误差（MSE）、平均绝对误差（MAE）和平均绝对百分比误差（MPAE）。每一度量标准均将 PCA 得到的因子相对值设置为 1，见表。

### §3.6 本章小结

本章首先简要介绍了因子模型的基础理论，包括了正交因子模型、动态因子模型和近似因子模型的模型假设和适用场景。近似因子模型的因子估计量可以用于宏观经济预测，我们着重讨论了它的一种非参数估计——主成分估计量。主成分分析、奇异值分解等方法都等价于基于  $L_2$  范数优化问题，都不具有对离群值的稳健性。针对宏观经济数据往往分布重尾、具有大量缺失和离群值的情况，我们提出了引入  $L_1$  范数来改善估计的稳健性，并且提出了基于  $L_1$  范数的因子模型估计方法。最后，通过模拟实验，验证了基于  $L_1$  范数的因子模型估计方法的稳健性。

## 第四章 $L_1$ 范数主成分分析算法的性能研究

### §4.1 基于 $L_1$ 范数最大化算法

### §4.2 基于 $L_1$ 范数最小化算法

### §4.3 模拟实验