南京审计大学硕士研究生毕业论文

蒯强

2021 年 2 月 22 日

目录

第 1 章 绪论 **3**
- 1.1 研究背景及意义 . . . 3
- 1.2 国内外研究现状 . . . 4
  - 1.2.1 宏观经济预测 . . . 4
  - 1.2.2 因子模型 . . . 4
- 1.3 研究内容及技术路线 . . . 4
- 1.4 研究创新 . . . 4
- 1.5 本文结构 . . . 4

第 2 章 L_1 因子分析 **5**
- 2.1 因子模型简介 . . . 5
  - 2.1.1 经典因子模型 . . . 5
  - 2.1.2 近似因子模型 . . . 5
- 2.2 因子分析的求解方法 . . . 5
  - 2.2.1 奇异值分解 . . . 6
  - 2.2.2 主成分分析 . . . 6
- 2.3 L_1 因子分析 . . . 6
- 2.4 L_1 范数及其稳健性 . . . 6
  - 2.4.1 L_1 因子分析算法 . . . 6
  - 2.4.2 缺失值处理 . . . 7
  - 2.4.3 L_1 因子分析算法 . . . 7
- 2.5 稳健性实验 . . . 8
  - 2.5.1 数据准备 . . . 8

第 3 章 L_1 范数最小化算法的性能研究 **10**
- 3.1 线性规划方法 . . . 10
- 3.2 转化为二次优化问题 . . . 10
- 3.3 一种新的方法 . . . 10
  - 3.3.1 理论推导 . . . 10
  - 3.3.2 算法步骤 . . . 12

第 4 章 L_1 因子分析在宏观经济预测中的应用研究 **14**

第 1 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，丛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。

本文的研究意义在于，从主流的基于历史数据进行向前预测的方法中发现了问题，即使用近似因子模型进行宏观经济预测中存在稳健性方面的缺陷，通过引入 L_1 稳健因子模型，给处理带有大量缺失值和离群值的宏观经济数据提供了更好的预测效果，在一定意义上提高了使用因子模型进行经济预测的准确性。

1.2 国内外研究现状

1.2.1 宏观经济预测

宏观经济预测

1.2.2 因子模型

1.3 研究内容及技术路线

1.4 研究创新

1.5 本文结构

第 **2** 章　L_1 因子分析

2.1　因子模型简介

2.1.1　经典因子模型

设 p 维随机向量 $X=(X_1,X_2,...,X_p)^T$ 的数学期望为 $\mu=(\mu_1,\mu_2,...,\mu_p)^T$，协方差矩阵为 Σ，假设 X 线性依赖于少数几个不可观测的随机变量 $f_1,f_2,...,f_m(m<p)$，和 p 个随机误差项 $e_1,e_2,...,e_p$，一般称 $f_1,f_2,...,f_m$ 为公共因子，称 $e_1,e_2,...,e_p$ 为特殊因子或误差，因子模型有如下表达式：

$$X=\mu+Af+e \tag{2-1}$$

其中 $f=(f_1,f_2,...,f_p)^T$ 为因子，A 是因子载荷矩阵，$e=(e_1,e_2,...,e_p)^T$ 是特殊因子。在式 $(2-1)$ 中，随机向量 X 围绕均值的波动由公共因子的线性组合加上一个特殊因子解释。经典因子模型假设相互独立，$e_1,e_2,...,e_p$ 相互独立并且 f 和 e 的样本之间相互独立。

2.1.2　近似因子模型

动态因子模型是经典因子模型在时间序数据上的延伸，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。通过放宽模型假设和形式变换，可以将动态因子模型转变为在经济变量预测中更加实用的静态近似因子模型。令 $X_t=(X_{1t},X_{2t},...,X_{pt})^T$ 为一组宏观经济变量在 t 时刻的水平，并且 X_t 可以表达为如下形式：

$$X_t=\lambda(L)f_t+e_t \tag{2-2}$$

其中 f_t 是 q 维的动态相依因子向量，$\lambda(L)$ 为由 s 阶滞后多项式算子组成的 $p\times q$ 矩阵。近似因子模型的模型假设允许 f_t 的各个分量 $f_1,f_2,...,f_m$ 具有相依性，$e_1,e_2,...,e_p$ 可以不独立，并且允许 f_t，e_t 具有时间序列相依性。称 $(2-2)$ 为动态近似因子模型。

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：

$$X_t=AF_t+e_t \tag{2-3}$$

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。

令某一个经济指标 y 在 $t+h$ 时刻的水平为 y_{t+h}，则其预测值由式（$2-4$）给出：

$$X_t=AF_t+e_t$$

$$y_{t+h}=\beta F_t+\beta_y y_t+\varepsilon_{t+h} \tag{2-4}$$

其中 βF_t 包括了近似因子 t 时刻的当期项和滞后项，$\beta_y y_t$ 代表了指标 y 受自身滞后项的影响，ε_{t+h} 为预测误差。

2.2　因子分析的求解方法

述因子模型中，需要估计因子载荷矩阵 A 和因子得分矩阵 F^T。容易发现，这个问题等价于以下矩阵分解问题：

$$X_{p\times n}=\hat{A}_{p\times m}\hat{F^T}_{m\times n} \tag{2-5}$$

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

2.2.1 奇异值分解

2.2.2 主成分分析

2.3 L_1 因子分析

使用最小二乘具有计算简便的特点，尽管在求解某些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。这里我们提出基于 L_1 范数的方法，来增加估计的健壮性。

2.4 L_1 范数及其稳健性

设 $\boldsymbol{X} = (X_1, X_2, ..., X_p)^T$ 为一 p 维随机变量，Y 为响应变量，$\boldsymbol{\beta}$ 为回归系数。假设我们观察到 $i.i.d.$ 样本 $(\boldsymbol{X_1}, \boldsymbol{X_2}, ..., \boldsymbol{X_n})$ 和 $(Y_1, ..., Y_n)$，我们一般使用 $\boldsymbol{\beta}$ 的最小二乘估计量 ¥¥

2.4.1 L_1 因子分析算法

将式（$2-6$）中的目标函数更换为使用 L_1 范数，可得

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

其中 $\|.\|_{L_1}$ 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 A 或者 F^T 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 A 或者 F^T 的值，来求解另一个参数。

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

$$A^{(t)} = \underset{A}{\arg\min}\|X - AF^{T(t)}\|_{L_1} \quad (2-8b)$$

我们改写式（$2-8a$）中的目标函数，

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

其中 x_j 是矩阵 X 的第 j 列，f_j 是 F^T 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:

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

式（$2-10$）的全局最优解可以通过求解下面的线性规划问题得到：

$$\min_{\theta,t} 1^T t$$

$$s.t. -t \leq A^{(t-1)}\theta - x_j \leq t \qquad (2-11)$$

线性规划问题的计算性能取决于未知变量的个数和和约束的个数。这里的 n 个子优化问题共享了 $A^{(t-1)}$，减少了一定的计算量。但是每一个子优化问题都要面临 p 个约束条件和一个 p 维未知变量，当 x_j 是一个维数非常大的向量时 p 很大），每一个子问题的求解仍然需要大量计算。虽然现在已经发展出了有效的处理多变量多约束线性规划问题的方法，但是在因子分析的场合下我们仍给出一个性能更好的方法来代替这里的线性规划。这个方法将在第三章中讨论。

2.4.2　缺失值处理

在使用 PCA 和 SVD 时我们需要对矩阵 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（$2-9$），

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

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。

2.4.3　L_1 因子分析算法

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 A 一个初始值 $A^{(0)}$。对于 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 $A^{(0)}$。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 $E(A,F) = \|X - AF^T\|_{L_1}$ 在每一个交替的优化步骤中都递减，并且 $E(A,F)$ 具有下界（≥ 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 A 或者 F 的第 i 列；α 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 A，Σ 的初始值 $A^{(0)}$，$\Sigma^{(0)} = I$，(其中 Σ 为一对角矩阵，I 为单位矩阵)；

2. 交替凸优化：对于迭代次数 $t = 1, ...,$ 收敛：

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

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

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

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

2.5 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。

2.5.1 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

$$X = U_{(:,1:r)}\Sigma_{(1:r,1:r)}V^T_{(:,1:r)}$$

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 30×30 秩为 3 的矩阵在模拟了缺失值和离群值后的情况。

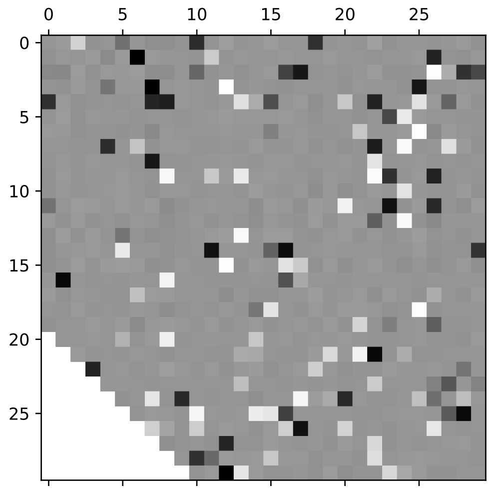

图 2.1: 30×30 的模拟矩阵，左下角白色代表缺失值，图中灰度越深代表元素的绝对值越大，因此深黑色的点代表了离群值

第 3 章 L_1 范数最小化算法的性能研究

在解决式（2 − 10）的优化问题时，我们将问题转化为式（2 − 11）的线性规划问题。前文已经提到，这些子问题都需要通过线性规划解决，需要较大的计算量，这就给 L_1 算法在计算性能上提出了改进的需要。

本章我们开拓新的思路来解决（$2-8a$）的优化问题，提出一种近似的方法将问题转化为一个使用牛顿-拉弗森迭代来解决的优化问题。避开了一次迭代需要解决多个线性规划的计算量，并通过模拟实验证明我们的方法能够在较小的误差范围内显著提升计算效率。

3.1 线性规划方法

3.2 转化为二次优化问题

3.3 一种新的方法

3.3.1 理论推导

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

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

考虑如下线性回归问题：

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

其中 $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_{p+1})^T$ 为 $p+1$ 维随机变量，$\boldsymbol{\beta}^* = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_p)^T$ 为回归系数，Y 为响应变量，e 为独立于 $\boldsymbol{X}$ 的噪声。假设我们使用 L_1 损失函数来估计 $\boldsymbol{\beta}$，

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

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

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

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 $\boldsymbol{\beta}^*$，用新的估计量来代替 $\hat{\boldsymbol{\beta}}$。

我们使用牛顿迭代法求解下面的随机优化问题:

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

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

$$\tilde{\boldsymbol{\beta}}_1 = \boldsymbol{\beta}_0 - \boldsymbol{H}(\boldsymbol{\beta}_0)^{-1}\mathbb{E}[g(\boldsymbol{\beta};\boldsymbol{X},Y)] \tag{3-5}$$

其中 $\boldsymbol{\beta}_0$ 是一个初始估计，$g(\boldsymbol{\beta};\boldsymbol{X},Y)$ 为损失函数 $G(\boldsymbol{\beta};\boldsymbol{X},Y)$ 关于 $\boldsymbol{\beta}$ 的梯度。
$\boldsymbol{H}(\boldsymbol{\beta}) := \partial\mathbb{E}[g(\boldsymbol{\beta};\boldsymbol{X},Y)/\partial\boldsymbol{\beta}$ 表示 $\mathbb{E}G(\boldsymbol{\beta};\boldsymbol{X},Y)$ 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，

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

在（$3-6$）的条件下，$g(\boldsymbol{\beta};\boldsymbol{X},Y) = \boldsymbol{X}(\mathbb{I}[Y - \boldsymbol{X}^T\boldsymbol{\beta} < 0] - 0.5)$。
并且，$\boldsymbol{H}(\boldsymbol{\beta}) = \mathbb{E}(\boldsymbol{X}\boldsymbol{X}^T f(\boldsymbol{X}^T(\boldsymbol{\beta}-\boldsymbol{\beta}^*)))$，这里 $f(x)$ 是噪声 e 的密度函数。当初始估计量 $\boldsymbol{\beta}_0$ 和 $\boldsymbol{\beta}^*$ 很接近时，$\boldsymbol{H}(\boldsymbol{\beta}_0)$ 就会很接近 $\boldsymbol{H}(\boldsymbol{\beta}^*) = \boldsymbol{\Sigma}f(0)$，这里 $\boldsymbol{\Sigma} = \mathbb{E}\boldsymbol{X}\boldsymbol{X}^T$ 是 $\boldsymbol{X}$ 的协方差矩阵。使用 $\boldsymbol{H}(\boldsymbol{\beta}^*)$ 替换式（$3-5$）中的 $\boldsymbol{H}(\boldsymbol{\beta}_0)$，可得

$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 - \boldsymbol{H}(\boldsymbol{\beta}^*)^{-1}\mathbb{E}[g(\boldsymbol{\beta};\boldsymbol{X},Y)] = \boldsymbol{\beta}_0 - \boldsymbol{\Sigma}^{-1}f^{-1}(0)\mathbb{E}[g(\boldsymbol{\beta}_0;\boldsymbol{X},Y)] \tag{3-7}$$

我们在 $\boldsymbol{\beta}^*$ 对 $\mathbb{E}[g(\boldsymbol{\beta}_0;\boldsymbol{X},Y)$ 进行泰勒展开，

$$\begin{aligned}\mathbb{E}[g(\boldsymbol{\beta}_0;\boldsymbol{X},Y) &= \boldsymbol{H}(\boldsymbol{\beta}^*)(\boldsymbol{\beta}_0-\boldsymbol{\beta}^*) + O(|\boldsymbol{\beta}_0-\boldsymbol{\beta}^*|_2^2)\\ &= \boldsymbol{\Sigma}f(0)(\boldsymbol{\beta}_0-\boldsymbol{\beta}^*) + O(|\boldsymbol{\beta}_0-\boldsymbol{\beta}^*|_2^2)\end{aligned}$$

结合式（$3-7$），可以得到

$$\begin{aligned}|\boldsymbol{\beta}_1-\boldsymbol{\beta}^*|_2 &= |\boldsymbol{\beta}_0 - \boldsymbol{\Sigma}^{-1}f^{-1}(0)(\boldsymbol{\Sigma}f(0)(\boldsymbol{\beta}_0-\boldsymbol{\beta}^*) + O(|\boldsymbol{\beta}_0-\boldsymbol{\beta}^*|_2^2)) - \boldsymbol{\beta}^*|_2\\ &= O(|\boldsymbol{\beta}_0-\boldsymbol{\beta}^*|_2^2)\end{aligned}$$

因此，如果我们得到一个 $\boldsymbol{\beta}^*$ 的一致估计量 $\boldsymbol{\beta}_0$，我们就可以通过式（$3-7$）中的牛顿-拉弗森迭代得到偏误更小的估计。

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，

$$\begin{aligned}\boldsymbol{\beta}_1 &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}\boldsymbol{\beta}_0 - f^{-1}(0)\mathbb{E}[g(\boldsymbol{\beta}_0;\boldsymbol{X},Y)])\\ &= \boldsymbol{\Sigma}^{-1}\mathbb{E}[\boldsymbol{X}\{\boldsymbol{X}^T\boldsymbol{\beta}_0 - f^{-1}(0)(\mathbb{I}[Y \le \boldsymbol{X}^T\boldsymbol{\beta}_0] - 0.5)\}]\end{aligned}$$

这里我们定义一个新的响应变量 $\tilde{Y}$，

$$\tilde{Y} = \boldsymbol{X}^T\boldsymbol{\beta}_0 - f^{-1}(0)(\mathbb{I}[Y \le \boldsymbol{X}^T\boldsymbol{\beta}_0] - 0.5)$$

那么 $\boldsymbol{\beta}_1 = \boldsymbol{\Sigma}^{-1}\mathbb{E}(\boldsymbol{X}\tilde{Y})$ 就是线性回归问题 $\tilde{Y} = \boldsymbol{X}^T\boldsymbol{\beta}$ 的最优回归系数，即

$$\boldsymbol{\beta}_1 = \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\min}\mathbb{E}(\tilde{Y} - \boldsymbol{X}^T\boldsymbol{\beta})^2$$

3.3.2 算法步骤

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

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

其中 $\hat{\boldsymbol{\beta}}_0$ 为 $\boldsymbol{\beta}^*$ 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，

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

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

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

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

其中 $K(x)$ 为高斯核函数，$h \to 0$ 是带宽。

回到式（$3-1$），我们采用以上介绍的牛顿-拉弗森迭代方法来求解该问题，算法 2 给出了详细的算法步骤，在每一次迭代时，我们需要重新计算 $\hat{f}(0)$，$\cdot$ $(\tilde{Y}_1, \tilde{Y}_2, ..., \tilde{Y}_n)$ 和 $\hat{\boldsymbol{\beta}}$。这几个量都通过显式表达式计算，其中前两者计算的时间复杂度都是 $O(n)$，而计算最小二乘结果涉及到矩阵的逆运算，其运算的时间复杂度和输入变量 $\boldsymbol{X}$ 的维数 p 有关，一般来说时间复杂度为 $O(p^3)$。

考虑我们的算法用来解决（$3-1$）的问题，对于维数特别巨大、样本数量较小的宏观经济数据来说，意味着在算法输入中 $n >> p$。考虑到牛顿-拉弗森迭代算法通常具有迭代迅速的特征。我们认为该算法的时间复杂度小于 $O(n^{3.5})$ 即快于线性规划的方法。

表 1:

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 $\boldsymbol{X}$ 的样本 $\boldsymbol{Y}=(Y_1,Y2,...,Y_n)$，$\boldsymbol{X}=(\boldsymbol{X}_1^T,\boldsymbol{X}_2^T,...,\boldsymbol{X}_n^T)$，迭代次数 t，高斯核函数 K，带宽 $h_g(g=1,...,t)$

初始化：给出

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

对于 $g=1,...,t$:

1. 计算 $\hat{f}^{(g)}(0)$,

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

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

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

3. 计算 $\hat{\boldsymbol{\beta}}^{(g)}$,

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

输出：$\boldsymbol{\beta}^{(t)}$

第 4 章 L_1 因子分析在宏观经济预测中的应用研究