

# 基于半参数部分线性混合效应模型的房地产上市公司盈利能力影响因素研究

袁超杰 孙 煜 李志强\*

(北京化工大学 理学院, 北京 100029)

**摘 要:** 基于沪深两市的50家房地产上市公司2006年至2015年期间的纵向数据,从资产结构、负债结构、公司规模、公司风险、发展潜力5个方面分析了影响房地产上市公司盈利能力的因素。通过建立净资产收益率、流动资产率、资产负债率、总资产的对数、主营业务收入的対数5个观测变量的半参数部分线性混合效应模型进行实证分析,并根据贝叶斯推断给出了模型参数的估计值,最后通过将实证模拟结果与房地产行业上市公司的实际情况进行对比分析,证明了所建模型的有效性。

**关键词:** 纵向数据; 半参数混合效应模型; 贝叶斯分析; 房地产上市公司; 盈利分析

**中图分类号:** F064.1 **DOI:** 10.13543/j.bhxbzr.2017.04.018

## 引 言

随着国家不断出台政策调控楼市,房地产行业已然告别高速增长的时代,因此对于房地产上市公司盈利能力的研究尤为重要。通过研究盈利能力,房地产上市公司可以发现自身经营管理中存在的问题,提高自身业绩,而投资者则可以根据盈利能力判断股票价格变化和股息的分配,进行合理的股票投资。

房地产上市公司盈利能力数据是时间序列数据,同时在时间截面上含有多个公司的多个指标,所以应作为纵向数据进行研究。基于纵向数据的特点,目前主要用线性混合效应模型处理纵向数据,但是线性效应模型在很多情况下处理效果不甚理想,本文综合参数混合效应模型与非参数混合效应模型,采用半参数部分线性混合效应模型进行研究。半参数部分线性混合效应模型具有参数模型和非参数模型两者的优点,有效利用数据所含信息的同时考虑了信息不完整的变量,有利于更全面地分析问题。

推断半参数部分线性混合效应模型以往一般采用经典统计方法,主要有截面核法<sup>[1-4]</sup>和基于联合

最大惩罚似然的光滑样条和惩罚样条方法<sup>[5-7]</sup>。相较于这些经典统计方法,贝叶斯统计方法在小样本情况下得到的点估计和区间估计更准确,并且可以通过后验分布将无关的参数通过积分处理掉,有效简化了模型的推断。目前,已经有不少学者利用贝叶斯方法推断统计模型,如王佐仁等<sup>[8]</sup>研究了线性混合效应模型的贝叶斯推断,提出了随机效应协方差矩阵的对数的多元正态逼近方法;Li等<sup>[9]</sup>研究了纵向数据下半参数混合模型的贝叶斯推断,分析比较了随机效应分别服从正态分布和非参数分布时的估计效果。但是基于共轭先验分布并利用贝叶斯方法推断半参数部分线性混合效应模型方面的研究还未见报道。

本文参考文献[8-9]的方法,选择贝叶斯方法进行模型推断。以我国A股市场的50家房地产上市公司为研究对象,首先根据上市公司盈利指标选取的原则<sup>[10]</sup>建立合理的指标体系,将解释变量中与盈利能力的关系比较明确的变量建立参数模型;为了更好地体现时间对于盈利能力的影响,采用非参数函数代替以往文献中常用的线性函数模拟与盈利能力的关系不明确的时间效应,选用半参数部分线性混合效应模型进行分析;选择贝叶斯方法对模型进行推断,利用光滑样条逼近非参数部分,将模型转化为线性混合效应模型,并根据所得后验分布利用Markov Chain Monte Carlo(MCMC)方法<sup>[11]</sup>给出了模型中参数的贝叶斯估计;最后通过实证分析证明了所建模型的

收稿日期: 2017-01-05

第一作者: 男,1991年生,硕士生

\* 通讯联系人

E-mail: li-zhiqiang2000@163.com

准确性。

1 变量选取

本文分别从上海证券交易所和深圳证券交易所各选取 25 家 2005 年底前上市的房地产公司(不含 ST 公司)。以这 50 家具有代表性的来自不同地区的房地产上市公司作为研究对象,截取 2006 年~2015 年共 10 个年度的纵向数据(所有数据取自国泰安 CSMAR 金融经济数据库)进行实证分析。

目前,反映上市公司盈利能力的指标主要包括营业利润率、成本费用利润率、盈余现金保障倍数、总资产报酬率、净资产收益率和资本收益率。实际事务中,上市公司一般采用每股收益、每股股利、市盈率、每股净资产等指标评价其获利能力。本文选择净资产收益率作为反映房地产上市公司盈利能力的指标,并从资产结构、负债结构、公司规模、公司风险、发展潜力 5 个方面选择了流动资产率、资产负债率、总资产、主营业务收入、主营业务收入增长率 5 个影响盈利能力的因素指标。由于总资产和主营业务收入是数量指标,与其他比率指标相差较大,为了方便分析,取二者的对数值进行研究。所选盈利及其影响因素指标变量如表 1 所示。

表 1 盈利及其影响因素指标

Table 1 Variable table

指标变量	定义方法
净资产收益率(ROE)	净利润与股东权益的比率
流动资产率(CTR)	流动资产与总资产的比率
资产负债率(RDA)	总负债与总资产的比率
总资产对数(LTA)	年末总资产的对数
主营业务收入对数(LCR)	当年主营业务收入的対数
主营业务收入增长率(RCR)	当年主营业务收入的增长率

此外,不同年份的房地产价格走势、市场需求特点也有差别,从而不同的房地产上市公司在不同年份的盈利能力可能存在差异。以往的研究一般直接用线性函数模拟时间效应,但是现实中时间对于盈利能力的影响是很复杂的,一般都是非线性关系,因此本文利用非参数函数刻画时间的影响,进而建立半参数部分线性混合效应模型进行分析。

2 模型描述

2.1 半参数部分线性混合效应模型

设  $Y_{ij}$  是第  $i$  个个体在时间点  $t_{ij}$  的观测值,其中

$i = 1, \dots, m, j = 1, \dots, n_i$ , 总样本数为  $n = \sum_{i=1}^m n_i$ 。假设  $Y_{ij}$  服从半参数部分线性混合模型

$$Y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + f(t_{ij}) + \mathbf{Z}_{ij}^T \mathbf{b}_i + \varepsilon_{ij} \tag{1}$$

其中  $\mathbf{X}_{ij} = (X_{ij1}, \dots, X_{ijp})^T$  是与固定效应相关的解释变量构成的  $p \times 1$  维向量;  $\mathbf{Z}_{ij} = (Z_{ij1}, \dots, Z_{ijq})^T$  是与随机效应相关的解释变量构成的  $q \times 1$  维向量;  $\boldsymbol{\beta}$  是  $p \times 1$  维未知参数向量;  $\mathbf{b}_i$  是第  $i$  个个体的  $q \times 1$  维随机效应向量,  $\mathbf{b}_i$  相互独立且服从协方差  $\mathbf{D}$  的正态分布  $N(0, \mathbf{D})$ ;  $f(t)$  是关于观测时间  $t$  的未知二次可微光滑函数,是模型的非参数部分;  $\varepsilon_{ij}$  是随机误差,相互独立,服从正态分布  $N(0, \sigma^2)$  且与  $\mathbf{b}_i$  独立。

2.2 模型参数化

为了对参数  $\boldsymbol{\beta}$ 、 $\mathbf{D}$  和  $\sigma^2$  进行估计,首先采用光滑样条方法对未知光滑函数  $f(\cdot)$  进行参数化。

利用 Li 等<sup>[9]</sup>的思想,假设  $f(t)$  服从 Wiener 积分过程先验<sup>[12]</sup>

$$f(t) = \delta_0 + \delta_1 t + \lambda^{-1/2} \int_0^t (t-u) dW(u), t \in [T_1, T_2] \tag{2}$$

其中  $\boldsymbol{\delta} = (\delta_0, \delta_1)^T$  是未知的  $2 \times 1$  维参数向量,  $\lambda$  是调节参数,  $T_1$  和  $T_2$  确定  $t$  的范围,  $\int_0^t (t-u) dW(u)$  是标准 Wiener 过程  $W(t)$  的随机积分。由于  $f(t)$  已经包含截距项和线性项,本文假设方程(1)中  $\mathbf{X}_{ij}$  不存在与 1 或者  $t$  相关的行。

根据 Zhang 等<sup>[5]</sup>的研究,记  $\mathbf{T} = (\mathbf{1}, \mathbf{t}^0)$ , 其中  $\mathbf{t}^0$  是由  $t_{ij} (i = 1, \dots, m, j = 1, \dots, n_i)$  中所有离散观测点排序后构成的  $r \times 1$  维向量( $r$  为观测时点的个数),  $\mathbf{1}$  是由 1 构成的  $r \times 1$  维向量。令

$$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$$

并同样定义  $\mathbf{X}_i$ 、 $\mathbf{Z}_i$  和  $\boldsymbol{\varepsilon}_i$ 。记  $\mathbf{N}_i$  为第  $i$  个个体的关联矩阵,用来映射  $\mathbf{t}_i = (t_{i1}, \dots, t_{in_i})^T$  和  $\mathbf{t}^0$ 。如果  $t_{ij} = t_l^0$ , 则有  $\mathbf{N}_i$  的第  $(j, l)$  个元素为 1, 否则为 0, 从而模型(1)可写成

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{N}_i \mathbf{f} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i \tag{3}$$

其中  $\mathbf{f} = (f(t_1^0), \dots, f(t_r^0))^T$ 。

记  $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_m^T)^T$ , 并同样定义  $\mathbf{X}$ 、 $\mathbf{N}$ 、 $\mathbf{b}$  和  $\boldsymbol{\varepsilon}$ 。同时令  $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_m)$  表示对角元素为  $\mathbf{Z}_1, \dots, \mathbf{Z}_m$  的块对角矩阵,则有

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{N} \mathbf{f} + \mathbf{Z} \mathbf{b} + \boldsymbol{\varepsilon} \tag{4}$$

Li 等<sup>[9]</sup>指出,  $f(t)$  的 Wiener 积分先验假定等价于对有限维光滑样条  $\mathbf{f}$  的先验作以下设定: 令  $\mathbf{K}$  表

示 Wiener 积分过程在  $t^0$  处的协方差, 若  $t_j^0 \geq t_k^0$ , 则矩阵  $\mathbf{K}$  的第  $(j, k)$  个和第  $(k, j)$  个元素为  $(t_j^0)^2 \cdot (3t_j^0 - t_k^0)/6$ 。此外令  $\mathbf{B} = \mathbf{L}(\mathbf{L}^T \mathbf{L})^{-1}$ , 其中  $\mathbf{L}$  为满足  $\mathbf{K} = \mathbf{L}\mathbf{L}^T$  和  $\mathbf{L}^T \mathbf{T} = \mathbf{0}$  的  $r \times (r-2)$  维满秩矩阵。

根据文献[13], 非参数分量  $f$  可通过一一对应的线性变换表示为

$$\mathbf{f} = \mathbf{T}\boldsymbol{\delta} + \mathbf{B}\mathbf{a} \quad (5)$$

其中  $\boldsymbol{\delta}$  是一个  $2 \times 1$  维模型参数向量。假定服从先验分布  $\boldsymbol{\delta} \sim N(0, \phi \mathbf{I})$ ,  $\phi$  为已知的特别大的常数;  $\mathbf{a}$  是一个  $r \times 1$  维随机效应向量, 服从正态分布  $\mathbf{a} \sim N(0, \tau \mathbf{I})$ ,  $\tau$  是未知的光滑参数。将  $\mathbf{f} = \mathbf{T}\boldsymbol{\delta} + \mathbf{B}\mathbf{a}$  代入式(4), 从而可将式(4)表示成如式(6)的线性混合模型

$$\mathbf{Y} = \mathbf{X}_* \boldsymbol{\beta}_* + \mathbf{Z}^{(1)} \mathbf{b} + \mathbf{Z}^{(2)} \mathbf{a} + \boldsymbol{\varepsilon} \quad (6)$$

其中  $\mathbf{Z}^{(1)} = \mathbf{Z}$ ,  $\mathbf{X}_* = (\mathbf{X}, \mathbf{N}\mathbf{T})$ ,  $\boldsymbol{\beta}_* = (\boldsymbol{\beta}^T, \boldsymbol{\delta}^T)^T$ ,  $\mathbf{Z}^{(2)} = \mathbf{N}\mathbf{B}$ ,  $\mathbf{b}$  和  $\mathbf{a}$  为模型的随机效应。

## 3 贝叶斯分析

### 3.1 固定效应的条件后验分布

假设固定效应  $\boldsymbol{\beta}_*$  的先验为多元正态分布, 即

$$\pi(\boldsymbol{\beta}_*) \sim N_{p+2}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$$

则  $\boldsymbol{\beta}_*$  的后验分布为

$$\pi(\boldsymbol{\beta}_* | \mathbf{b}, \mathbf{a}, \sigma^2, \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2} \left( (\boldsymbol{\beta}_* - \hat{\boldsymbol{\beta}})^T \right. \right.$$

$$\left. \left( \frac{1}{\sigma^2} \mathbf{X}_*^T \mathbf{X}_* \right) (\boldsymbol{\beta}_* - \hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta}_* - \boldsymbol{\beta}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta}_* - \boldsymbol{\beta}_0) \right\}$$

其中  $\hat{\boldsymbol{\beta}} = (\mathbf{X}_*^T \mathbf{X}_*)^{-1} \mathbf{X}_*^T (\mathbf{Y} - \mathbf{Z}^{(1)} \mathbf{b} - \mathbf{Z}^{(2)} \mathbf{a})$ 。

利用多元配方方法<sup>[14]</sup> 配方可得

$$\pi(\boldsymbol{\beta}_* | \mathbf{Y}, \mathbf{b}, \mathbf{a}, \sigma^2) \propto \exp \left\{ -\frac{1}{2} \left( (\boldsymbol{\beta}_* - (\boldsymbol{\beta}_0 - \boldsymbol{\beta}^*))^T \left( \frac{1}{\sigma^2} \mathbf{X}_*^T \mathbf{X}_* + \boldsymbol{\Sigma}_0^{-1} \right) (\boldsymbol{\beta}_* - (\boldsymbol{\beta}_0 - \boldsymbol{\beta}^*)) \right) \right\}$$

其中  $\boldsymbol{\beta}^* = (\mathbf{X}_*^T \mathbf{X}_* + \sigma^2 \boldsymbol{\Sigma}_0^{-1})^{-1} \mathbf{X}_*^T \mathbf{X}_* (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})$ 。通过核函数可得  $\boldsymbol{\beta}_*$  的后验分布为

$$\boldsymbol{\beta}_* | \mathbf{Y}, \mathbf{b}, \mathbf{a}, \sigma^2 \sim N_{p+2}((\boldsymbol{\beta}_0 - \boldsymbol{\beta}^*), (\sigma^{-2} \mathbf{X}_*^T \mathbf{X}_* + \boldsymbol{\Sigma}_0^{-1})^{-1}) \quad (7)$$

### 3.2 随机效应的条件后验分布

#### 3.2.1 随机效应 $\mathbf{b}_i$ 的条件分布

已知  $\mathbf{b}_i$  服从多元正态分布  $N_q(0, \mathbf{D})$ , 不妨令

$$\mathbf{w}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_* - \mathbf{Z}^{(2)} \mathbf{a}$$

则第  $i$  个个体似然方程为

$$l(\boldsymbol{\beta}_*, \mathbf{b}_i, \mathbf{a}, \sigma^2 | \mathbf{Y}_i) \propto \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{b}_i - \hat{\mathbf{b}}_i)^T (\mathbf{Z}_i^{(1)})^T \mathbf{Z}_i^{(1)} (\mathbf{b}_i - \hat{\mathbf{b}}_i) \right\}$$

其中  $\hat{\mathbf{b}}_i = ((\mathbf{Z}_i^{(1)})^T \mathbf{Z}_i^{(1)})^{-1} (\mathbf{Z}_i^{(1)})^T \mathbf{w}_i$ 。

同样利用多元配方法<sup>[14]</sup> 配方可得第  $i$  个个体的随机效应  $\mathbf{b}_i$  的后验分布为

$$\pi(\mathbf{b}_i | \mathbf{Y}_i, \boldsymbol{\beta}_*, \mathbf{D}, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{b}_i - \hat{\mathbf{b}}_i)^T (\mathbf{Z}_i^{(1)})^T \mathbf{Z}_i^{(1)} (\mathbf{b}_i - \hat{\mathbf{b}}_i) \right\} \exp \left\{ -\frac{1}{2} \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i \right\} \propto \exp \left\{ -\frac{1}{2} (\mathbf{b}_i - \mathbf{b}_i^*)^T \left( \frac{1}{\sigma^2} (\mathbf{Z}_i^{(1)})^T \mathbf{Z}_i^{(1)} + \mathbf{D}^{-1} \right) (\mathbf{b}_i - \mathbf{b}_i^*) \right\}$$

其中  $\mathbf{b}_i^* = ((\mathbf{Z}_i^{(1)})^T \mathbf{Z}_i^{(1)} + \sigma^2 \mathbf{D}^{-1})^{-1} (\mathbf{Z}_i^{(1)})^T (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_* - \mathbf{Z}^{(2)} \mathbf{a})$ 。

根据核函数, 可得随机效应  $\mathbf{b}_i$  的后验分布为

$$\mathbf{b}_i | \mathbf{Y}_i, \boldsymbol{\beta}_*, \mathbf{D}, \sigma^2 \sim N_q(\mathbf{b}_i^*, (\sigma^{-2} (\mathbf{Z}_i^{(1)})^T \mathbf{Z}_i^{(1)} + \mathbf{D}^{-1})^{-1}) \quad (8)$$

#### 3.2.2 随机效应 $\mathbf{a}$ 的条件分布

令  $\mathbf{v} = \mathbf{Y} - \mathbf{X}_* \boldsymbol{\beta}_* - \mathbf{Z}^{(1)} \mathbf{b}$ , 则似然方程可写为

$$l(\boldsymbol{\beta}_*, \mathbf{b}, \mathbf{a}, \sigma^2 | \mathbf{Y}) \propto \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{a} - \hat{\mathbf{a}})^T (\mathbf{Z}^{(2)})^T \mathbf{Z}^{(2)} (\mathbf{a} - \hat{\mathbf{a}}) \right\}$$

其中  $\hat{\mathbf{a}} = ((\mathbf{Z}^{(2)})^T \mathbf{Z}^{(2)})^{-1} (\mathbf{Z}^{(2)})^T \mathbf{v}$ 。

同时  $\mathbf{a} \sim N_{r-2}(\mathbf{0}, \tau \mathbf{I})$ , 则  $\mathbf{a}$  的条件后验分布为

$$\pi(\mathbf{a} | \mathbf{Y}, \boldsymbol{\beta}_*, \mathbf{b}, \mathbf{D}, \tau, \sigma^2) \propto \exp \left\{ -\frac{1}{2} (\mathbf{a} - \mathbf{a}^*)^T \left( \frac{1}{\sigma^2} (\mathbf{Z}^{(2)})^T \mathbf{Z}^{(2)} + (\tau \mathbf{I})^{-1} \right) (\mathbf{a} - \mathbf{a}^*) \right\}$$

其中  $\mathbf{a}^* = ((\mathbf{Z}^{(2)})^T \mathbf{Z}^{(2)} + \sigma^2 (\tau \mathbf{I})^{-1})^{-1} (\mathbf{Z}^{(2)})^T (\mathbf{Y} - \mathbf{X}_* \boldsymbol{\beta}_* - \mathbf{Z}^{(1)} \mathbf{b})$ 。

根据核函数, 可得  $\mathbf{a}$  的后验分布为

$$\mathbf{a} | \mathbf{Y}, \boldsymbol{\beta}_*, \mathbf{b}, \tau, \sigma^2 \sim N_{r-2}(\mathbf{a}^*, (\sigma^{-2} (\mathbf{Z}^{(2)})^T \mathbf{Z}^{(2)} + \tau^{-1} \mathbf{I})^{-1}) \quad (9)$$

### 3.3 随机效应方差项的条件后验分布

#### 3.3.1 $\mathbf{D}$ 的条件分布

本文考虑逆 Wishart 分布(公式中用  $IW$  表示)作为  $\mathbf{D}$  的先验分布, 即  $\mathbf{D} \sim IW_q(\rho_0, \mathbf{R}_0)$  ( $\rho_0$  和  $\mathbf{R}_0$  为超参数), 则

$$p(\mathbf{D}) \propto |\mathbf{D}|^{-\frac{\rho_0 + q + 1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{D}^{-1} \mathbf{R}_0) \right\}$$

又由  $\mathbf{b}_i|\mathbf{D} \sim N_q(0, \mathbf{D})$  可得

$$p(\mathbf{b}_i|\mathbf{D}) \propto |\mathbf{D}|^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^m \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i \right\}$$

则  $\mathbf{D}$  的后验分布为

$$\pi(\mathbf{D}|\mathbf{b}) \propto |\mathbf{D}|^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^m \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i \right\}$$

$$|\mathbf{D}|^{-\frac{\rho_0+q+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{D}^{-1} \mathbf{R}_0) \right\} \propto |\mathbf{D}|^{-\frac{m+\rho_0+q+1}{2}} \exp$$

$$\left\{ -\frac{1}{2} \text{tr} \left( \mathbf{D}^{-1} \left( \mathbf{R}_0 + \sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i^T \right) \right) \right\}$$

所以  $\mathbf{D}$  的条件后验分布为

$$\mathbf{D}|\mathbf{b}_i \sim IW_q \left( m + \rho_0, \mathbf{R}_0 + \sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i^T \right) \quad (10)$$

3.3.2  $\tau$  的条件分布

对于  $\tau$ , 同样考虑服从先验分布  $IW(\rho_1, \mathbf{R}_1)$  ( $\rho_1$  和  $\mathbf{R}_1$  为已知的超参数), 则

$$p(\tau) \propto |\tau|^{-\frac{\rho_1+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\tau^{-1} \mathbf{R}_1) \right\}$$

又根据  $\mathbf{a}|\tau \sim N_{r-2}(0, \tau \mathbf{I})$ , 则

$$p(\mathbf{a}|\tau) \propto |\tau|^{-\frac{r-2}{2}} \exp \left\{ -\frac{1}{2} \mathbf{a}^T (\tau \mathbf{I})^{-1} \mathbf{a} \right\}$$

通过推导可得  $\tau$  的后验分布为

$$\pi(\tau|\mathbf{a}) \propto |\tau|^{-\frac{r-2+\rho_1+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\tau^{-1} (\mathbf{R}_1 + \mathbf{a} \mathbf{a}^T)) \right\}$$

所以  $\tau$  的条件后验分布为

$$\tau|\mathbf{a} \sim IW(r-2+\rho_1, \mathbf{R}_1 + \mathbf{a} \mathbf{a}^T) \quad (11)$$

3.4 随机误差方差项的条件后验分布

对于随机误差的方差  $\sigma^2$ , 采用扩散先验分布, 即  $\pi(\sigma^2) \sim 1/\sigma^2$ , 则  $\sigma^2$  的条件后验分布为

$$\pi(\sigma^2|\boldsymbol{\beta}_*, \mathbf{b}, \mathbf{a}, \mathbf{Y}) \propto (\sigma^2)^{-\left(\frac{n}{2}+1\right)} \exp \left\{ -\frac{1}{\sigma^2}$$

$$\left( \frac{1}{2} \mathbf{u}^T \mathbf{u} \right) \right\}$$

其中  $\mathbf{u} = \mathbf{Y} - \mathbf{X}_* \boldsymbol{\beta}_* - \mathbf{Z}^{(1)} \mathbf{b} - \mathbf{Z}^{(2)} \mathbf{a}$ 。

不难发现上述结果是逆 Gamma 分布(简记为  $IG$ )的核函数, 所以  $\sigma^2$  的后验分布为

$$\sigma^2|\mathbf{Y}, \boldsymbol{\beta}_*, \mathbf{b}, \mathbf{a} \sim IG \left( \frac{n}{2}, \frac{1}{2} \mathbf{u}^T \mathbf{u} \right) \quad (12)$$

注意, 逆 Gamma 分布是一个易于处理的后验分布, 可以有效简化计算步骤。

3.5 贝叶斯推断

记  $\{(\boldsymbol{\beta}^{(\varphi)}, \mathbf{D}^{(\varphi)}, \tau^{(\varphi)}, \sigma^{2(\varphi)}) : \varphi = l, \dots, M\}$  是利用 MCMC 方法生成的  $\{\boldsymbol{\beta}, \sigma^2, \mathbf{D}, \tau\}$  的观测值。同

时, 令  $L = M - l$ ,  $l$  表示迭代过程中的迭代次数, 则  $\boldsymbol{\beta}, \mathbf{D}, \tau, \sigma^2$  的贝叶斯估计分别为

$$\hat{\boldsymbol{\beta}} = \frac{1}{L} \sum_{l=l}^M \boldsymbol{\beta}^{(\varphi)}, \hat{\mathbf{D}} = \frac{1}{L} \sum_{l=l}^M \mathbf{D}^{(\varphi)}$$

$$\hat{\tau} = \frac{1}{L} \sum_{l=l}^M \tau^{(\varphi)}, \hat{\sigma}^2 = \frac{1}{L} \sum_{l=l}^M (\sigma^2)^{(\varphi)}$$

此外, 通过  $\{\boldsymbol{\beta}^{(\varphi)} : \varphi = l, \dots, M\}$  也可以得到后验协方差  $Var(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, \mathbf{Z})$  的估计

$$\widehat{Var}(\boldsymbol{\beta}|\mathbf{Y}, \mathbf{X}, \mathbf{Z}) = \frac{1}{L-1} \sum_{l=m}^M (\boldsymbol{\beta}^{(\varphi)} - \hat{\boldsymbol{\beta}})(\boldsymbol{\beta}^{(\varphi)} - \hat{\boldsymbol{\beta}})^T。$$

4 实证分析

将 50 家房地产上市公司每年的净资产收益率作为响应变量  $Y$ , 以流动资产率 ( $X_1$ )、资产负债率 ( $X_2$ )、总资产的对数 ( $X_3$ )、主营业务收入的对数 ( $X_4$ )、主营业务收入增长率 ( $X_5$ ) 作为解释变量建立半参数部分线性混合效应模型:

$$Y_{ij} = \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \beta_5 X_{5ij} + f(t_{ij}) + b_i + \varepsilon_{ij} \quad (i = 1, \dots, 50, j = 1, \dots, 10) \quad (13)$$

其中,  $f(t)$  是一个光滑函数;  $b_i$  是相互独立的随机截距, 服从正态分布  $N(0, \theta)$ 。

将数据代入模型 (13), 通过贝叶斯分析和 MC-MC 方法进行模拟可得模型 (13) 的参数估计, 具体结果见表 2、3。

表 2  $\boldsymbol{\beta}$  的估计值及其标准差

Table 2 Posterior means and standard deviations of $\boldsymbol{\beta}$		
参数	估计值	标准差
$\beta_1$	0.1466	$4.1 \times 10^{-4}$
$\beta_2$	0.0246	$9.3 \times 10^{-4}$
$\beta_3$	0.0412	$1.0 \times 10^{-3}$
$\beta_4$	0.0003	$1.9 \times 10^{-5}$
$\beta_5$	0.0004	$4.2 \times 10^{-5}$

表 3  $\theta, \sigma^2$  的估计值及标准差

Table 3 Posterior means and standard deviations of $\theta, \sigma^2$		
参数	估计值	标准差
$\theta$	0.1173	$3.9 \times 10^{-4}$
$\sigma^2$	0.3051	$1.5 \times 10^{-3}$

由表 2 可以看出, 代表公司资产结构的流动资产率的系数 0.1466 为正值, 说明流动资产率正面影响房地产上市公司的盈利能力, 即公司的流



动资产率越高,其盈利能力越好,这与房地产行业盈利能力强的公司都保持着良好的流动资产率的情况是相符的;资产负债率的系数为 0.0246,表明资产负债率可以促进房地产公司的盈利能力,即资产负债率越高公司盈利能力越强,这与现状也是一致的。总资产对数的系数为 0.0412,说明公司总资产越大,净资产收益率越高,即房地产公司规模越大,盈利能力越强,这与实际相符;主营业务收入收入的系数为 0.0003,说明主营业务收入的与数对房地产上市公司的盈利能力呈正相关,较高的主营业务收入体现着房地产公司较好的盈利能力;代表公司发展潜力的主营业务收入增长率与净资产收益率存在正相关关系,说明具有发展潜力的公司受市场和投资者的认可,市场前景好,因此利润率相对较高。

## 5 结束语

本文根据贝叶斯理论对半参数部分线性混合效应模型进行了统计推断,可以发现模型固定效应、随机效应及其方差以及随机误差在先验分布下的后验分布均为标准分布,有效简便了模型估计的统计计算。同时,纵向数据下利用半参数部分线性混合效应模型对房地产上市公司盈利能力相关影响因素的实证结果表明这些因素对于房地产公司盈利能力有着不同程度的影响,实证分析结论也与房地产行业的实际情况相符,说明所提模型在实际应用中是可行的。但是在实际情况中,房地产上市公司的盈利能力不仅受到文中所述的解释变量影响,还受到国家政策、宏观经济状况等外部因素的影响,从而造成模型的精确度降低,如果可以量化这些因素并纳入模型中将会有更加精准的结果,这也是本文后续进一步研究的内容。

## 参考文献:

- [1] Lin X H, Carroll R J. Semiparametric regression for clustered data[J]. *Biometrika*, 2001, 88(4): 1179-1185.
- [2] Fan J Q, Li R Z. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis[J]. *Journal of the American Statistical Association*, 2004, 99(467): 710-723.
- [3] Wang N, Carroll R J, Lin X H. Efficient semiparametric marginal estimation for longitudinal/clustered data[J]. *Journal of the American Statistical Association*, 2005, 100(469): 147-157.
- [4] Lin X H, Carroll R J. Semiparametric regression for clustered data using generalized estimating equations[J]. *Journal of the American Statistical Association*, 2001, 96(455): 1045-1056.
- [5] Zhang D W, Lin X H, Raz J, et al. Semiparametric stochastic mixed models for longitudinal data[J]. *Journal of the American Statistical Association*, 1998, 93(442): 710-719.
- [6] Verbyla A P, Cullis B R, Kenward M G, et al. The analysis of designed experiments and longitudinal data by using smoothing splines[J]. *Journal of the Royal Statistical Society, Series C*, 1999, 48(3): 269-311.
- [7] Ruppert D, Wand M P, Carroll R J. Semiparametric regression[M]. New York, USA: Cambridge University Press, 2003.
- [8] 王佐仁, 杨琳. 贝叶斯统计推断及其主要进展[J]. *统计与信息论坛*, 2012, 27(12): 3-8.  
Wang Z R, Yang L. Bayesian inference and main progress[J]. *Statistics & Information Forum*, 2012, 27(12): 3-8. (in Chinese)
- [9] Li Y S, Lin X H, Müller P. Bayesian inference in semiparametric mixed models for longitudinal data[J]. *Biometrics*, 2010, 66(1): 70-78.
- [10] 胡梦娜. 上市公司盈利能力衡量指标的选取[J]. *合作经济与科技*, 2016(21): 167-169.  
Hu M N. Selection of the profitability analysis index of list real estate corporations[J]. *Co-Operative Economy & Science*, 2016(21): 167-169. (in Chinese)
- [11] 赵琪. 基于 MCMC 方法的贝叶斯统计推断[J]. *中国科技信息*, 2012(10): 64-65.  
Zhao Q. Bayesian inference based on MCMC method[J]. *China Science and Technology Information*, 2012(10): 64-65. (in Chinese)
- [12] Wahba G. Improper priors, spline smoothing and the problem of guarding against model errors in regression[J]. *Journal of the Royal Statistical Society, Series B*, 1978, 40(3): 364-372.
- [13] Green P J. Penalized likelihood for general semi-parametric regression models[J]. *International Statistical Review*, 1987, 55(3): 245-259.
- [14] Leonard T, Hsu J S J. Bayesian methods[M]. New York, USA: Cambridge University Press, 1999: 245.

# A study of the factors affecting the profitability of real estate listed companies based on semiparametric partially linear mixed models

YUAN ChaoJie SUN Yu LI ZhiQiang\*

(Faculty of Science, Beijing University of Chemical Technology, Beijing 100029, China)

**Abstract:** Based on the longitudinal data for fifty Chinese real estate listed companies from 2006 to 2015, this paper analyzes the factors affecting their profitability from the viewpoints of asset structure, debt structure, firm size, company risk and development potential. In addition, we set up a semiparametric partially linear mixed model between the return on equity, current assets ratio and asset – liability ratio and use empirical analysis to propose the estimators of the model via Bayesian inference. Finally, we compare the empirical results with the actual situation for the real estate listed companies and find the results are in good agreement, which indicates that the model is effective.

**Key words:** longitudinal data; semiparametric mixed effects model; Bayes analysis; real estate listed company; profitability analysis

(责任编辑:汪 琴)