

南开大学

本科生毕业论文（设计）

中文题目： 高频回归分析在资本资产定价问题中的应用

外文题目： Application of High-frequency Regression Analysis
in Capital Asset Pricing Problems

学 号： 1810175

姓 名： 刘欣雨

年 级： 18 级

专 业： 数学与应用数学

系 别： 数理金融与精算系

学 院： 数学科学学院

指导教师： 陈大川老师

完成日期： 2022 年 5 月 20 日

关于南开大学本科生毕业论文（设计） 的声明

本人郑重声明：所呈交的学位论文，是本人在指导教师指导下，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，本学位论文的研究成果不包含任何他人创作的、已公开发表或没有公开发表的作品内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。本学位论文原创性声明的法律责任由本人承担。

学位论文作者签名：

刘欣雨

2022 年 6 月 1 日

本人声明：该学位论文是本人指导学生完成的研究成果，已经审阅过论文的全部内容，并能够保证题目、关键词、摘要部分中英文内容的一致性和准确性。

学位论文指导教师签名：

陈大川

2022 年 6 月 1 日

摘 要

过去的金融理论通常将证券价格变化视为几何布朗运动对进行资本资产定价。本文将证券价格变化视为漂移项与波动项均为时变变量的 $Itô$ 过程,并在考虑高频数据观测值与真实值存在微观结构噪声的情况下,借助非参数模型框架与 Smoothed-TSRV 方法,得到时变参数的相合估计值。根据随机模型,进一步得到资产与市场组合的时变相关关系,并带入资本资产定价模型。本文通过排序法构造模拟资产组合、进行截面回归与时序回归证明 CAPM 在高频情况下表现良好,参数估计值较准确。并在高频数据框架下验证了存在随 β 上升减小的 Betting Against Beta 异象。

关键词: 资本资产定价; 高频数据; CAPM; 市场微观结构噪声; BAB 异象

Abstract

Prior studies regard security price to follow geometric Brownian motion. We consider security price following $Itô$ process with time-varying drift term and volatility. Dealing with the microstructure noise in high-frequency data, nonparametric method and smoothed TSRV estimator is used to get consistent estimators of time-varying parameters. According to the stochastic model, the time-varying correlation between assets and market portfolio is further obtained and brought into the capital asset pricing model. We also use ranking method, cross-sectional regression and time series regression to prove that the CAPM performs well in high frequency data and the parameter estimation is accurate. Also, Betting Against Beta anomaly is verified under the framework of high-frequency data.

Keywords: capital assets pricing; high frequency ;CAPM ;microstructure noise; betting against beta anomaly

目 录

一、 引言	1
二、 理论模型	2
(一) 传统 CAPM	2
(二) 改进 CAPM	2
1. 多因子 CAPM	3
2. Conditional CAPM	3
(三) CAPM 参数估计	4
1. 排序法估计因子收益率	4
2. 线性回归估计因子暴露	4
3. 高频数据估计因子暴露	5
(四) 高频估计理论	5
1. 模型假设	5
2. S-TSRV 法估计积分协变差	6
3. 因子暴露的估计	7
4. 统计量性质	7
(五) Betting Against Beta 异象	8
1. Betting Against Beta 异象的市场均衡理论解释	8
2. Betting Against Beta 异象的 Conditional CAPM 解释	9
三、 计量实证	9
(一) 数据来源	9
(二) 计量步骤	10
(三) 高频数据计算因子暴露	10
(四) 排序法构造市场因子价差策略	11
1. 市场因子暴露排序法	11
2. 市场因子暴露与市值双重排序法	14
(五) 回归法评估 Beta 估计量	14
(六) BAB 异象检验	15
四、 研究局限与改进方向	16
(一) 研究局限	16

(二) 改进方向	17
五、 结论	17
附 录	18
计量实验配置	18
计量实验数据	18
计量实验代码	18
参考文献	21
致 谢	23

一、引言

随机游走理论与资本资产定价模型是金融学研究两个重要方法。Louis Bachelier(1900) 在 *Theorie de la Speculation*[1] 中首次提出将随机游走模型应用于金融资产定价, 随后 Osborne(1959) 在 *Brownian Motion in the Stock Market*[2] 中提出使用几何布朗运动描述证券价格变化。而 Markowitz(1952)[3] 提出现代资产配置理论, 在平衡收益与风险的情况下得到单个资产价格与市场平均资产价格的关系。在此基础上 Treynor (1961[4], 1962[5]), Sharpe (1964) [6], Lintner (1965a[7], 1965b[8]) 以及 Mossin (1966) [9] 提出了资本资产定价模型, 用市场组合的风险和收益衡量单个资产的风险与收益, 以线性形式表示资产与市场组合收益的关系。

而上述两种方法均要求市场为有效市场 (Fama, 1995)[10]。即保证所有市场相关信息均包含于当前价格信息之中, 但是后续的实证发现有效市场假设不一定成立, 需要引入更多的信息进行资产定价。其中对于 CAPM, Fama 和 French(1993)[11] 在 CAPM 市场因子的基础上, 增加部分信息, 加入了价值因子与规模因子, 提出了三因子模型, 开创了多因子模型的先河; 而 Ravi Jagannathan 和 Zhenyu Wang(1996)[12] 提出了 Conditional CAPM, 不再将市场因子暴露 β 和市场风险溢价 ($R^M - R^f$) 视为常量, 通过这两个参数的时变引入新增信息。而对于随机游走模型, 则将几何布朗运动替换为更一般的 Itô 过程, 将漂移项与波动项都视为时变过程, 从而能够实时更新随机模型。

而对于随机游走模型中的时变参数, 可以借助高频数据, 找到具有良好统计性质的估计值, 帮助进行资产定价。此外, Zhang, Mykland 以及 Ait-Sahalia (2005)[13] 在实证中发现高频数据观测值与实际值之间存在的微观结构噪声会使得一般方法的估计结果有偏, 而借助 Smoothed-Two Scales Realized Volatility(S-TSRV) 方法 [14] 对随机过程二次协变差的估计能够消除偏差, 使得估计量具有相合性。而资产与市场组合的二次协变差估计值可以作为两者的协方差, 从而能够得到资产与市场组合的相关关系, 在高频数据的基础上应用传统 CAPM 方法, 对资本资产进行定价。

本文基于高频数据, 使用 S-TSRV 方法对随机游走模型中的时变参数进行估计。将估计结果带入 CAPM, 得到在高频情况下, 每个截面参数 β 与 ($R^M - R^f$) 都发生变化的 CAPM 模型。应用排序法构造模拟投资组合, 使用截面回归、时序回归等方法, 对参数时变的 CAPM 模型进行检验, 同时验证时变参数估计量的准确性。此外应用上述结果, 在高频情况下对 Frazzini, Pedersen(2014) 提出的 Betting Against Beta[15] 异象进行检验。

二、理论模型

(一) 传统 CAPM

1952 年, Markowitz 提出现代资产配置理论 (Modern Portfolio Theory, MPT), 将资产收益率视为随机变量, 根据不同资产收益率的期望、协方差构建资产组合, 从而使得组合超额收益期望值与标准差的比值 (夏普比率) 最大。

在 MPT 理论的基础上, Treynor (1961,1962), Sharpe (1964), Lintner (1965a, 1965b) 以及 Mossin (1966) 提出了资本资产定价模型 (Capital Asset Pricing Model, CAPM), 借助市场收益与市场风险衡量资产的收益与风险:

$$E[R_i] - R_f = \beta_i(E[R_M] - R_f) \quad (1)$$

其中 R_f 为无风险收益, R_M 为市场组合的预期收益, β_i 是资产 i 的市场因子暴露即相对市场的风险暴露。上述传统的 CAPM 包含如下假设:

- 投资者持有多种资产 (Investors hold diversified portfolios)
- 只考虑单期投资情况 (Single-period transaction horizon)
- 投资者可以以无风险利率借贷和贷款 (Investors can borrow and lend at the risk-free rate of return)
- 市场符合有效市场假设 (perfect capital market)

在上述假设下 CAPM 能够对资本资产进行完美定价, 然而更多情况下, 上述假设并不能全部成立, 从而需要对 CAPM 进行改进。

(二) 改进 CAPM

进一步实证中发现, CAMP 作为静态单因子模型 (static one-factor model) 并不总能对资本资产进行完美定价: Banz(1981) 发现小市值效应, Basu(1997)[16] 发现盈利市值比 (EP) 效应, 以及后续的账面市值比 (BM) 效应、债务市值比 (DM) 效应等, 这些效应均无法被 CAMP 解释。学界这些传统 CAMP 无法解释的部分称为异象 (anomaly)。为了解决异象问题, 过往的文献放宽 CAPM 假设, 对传统 CAMP 进行改进, 较主流的两种方法是: 将单因子 (市场因子) 模型变为多因子模型; 将静态截面模型变为动态时变模型。

1. 多因子 CAPM

Fama 和 French(1993) 在 CAMP 市场因子的基础上, 加入了价值因子 (High-Minus-Low, HML) 与规模因子 (Small-Minus-Big, SMB), 提出了三因子模型, 增强了模型的可解释性。而后 Carhart(1997)[17] 在 Fama-French 三因子模型中加入了截面动量因子 (Momentum Factor, MOM), 解决了截面动量异象 (Jegadeesh and Titman, 1993) 问题。后续衍生出的多因子模型, 逐步将 β_i 从单一的市场因子暴露丰富为多因子的风险暴露, 捕捉了传统 CAPM 无法解释的风险。

2. Conditional CAPM

Ravi Jagannathan 和 Zhenyu Wang(1996) 提出了 Conditional CAPM, 从另一个角度对异象的来源进行解释: 市场因子暴露 β 和市场风险溢价并非常量, 而是随时间变化的时变 (time-varying) 值, 从而传统的 CAPM 模型不再有良好的解释性。时变的 CAPM 模型如下:

$$E[R_{i,t}|I_{t-1}] = \gamma_{0,t-1} + \gamma_{1,t-1}\beta_{i,t-1} \quad (2)$$

其中 $\gamma_{0,t-1} = E(R_{f,t-1}|I_{t-1})$ 是零 beta(市场因子) 组合的条件期望收益, $\gamma_{1,t-1} = E(R_{m,t-1}|I_{t-1}) - E(R_{f,t-1}|I_{t-1})$ 是条件市场溢价, 条件市场因子暴露为 $\beta_{i,t-1} = \frac{Cov(R_{i,t}, R_{m,t}|I_{t-1})}{Var(R_{m,t}|I_{t-1})}$

对 (2) 式左右求期望得到非条件下的期望收益:

$$E[R_{i,t}] = \gamma_0 + \gamma_1 \bar{\beta}_i + Cov(\gamma_{1,t-1}, \beta_{i,t-1}) \quad (3)$$

其中 $\gamma_0 = E[\gamma_{0,t-1}]$, $\gamma_1 = E[\gamma_{1,t-1}]$, $\bar{\beta}_i = E[\beta_{i,t-1}]$ 从而相比于传统的 CAPM 模型新增了 $Cov(\gamma_{1,t-1}, \beta_{i,t-1})$ 项, 如果条件 β 与条件市场溢价存在相关性则存在 $alpha = Cov(\gamma_{1,t-1}, \beta_{i,t-1})$ 异象因子。

进一步, 对 (3) 式通过差分近似进行调整, 可以将 Conditional CAPM 模型转换为双因子模型:

市场因子暴露 (beta) 对风险溢价敏感性 $\theta_i := \frac{Cov(\beta_{i,t-1}, \gamma_{1,t-1})}{Var(\gamma_{1,t-1})}$, 记: $\eta_{i,t-1} = \beta_{i,t-1} - \bar{\beta}_i - \theta_i(\gamma_{1,t-1} - \gamma_1)$ 从而能够将条件 beta 表示为无条件 beta、条件市场风险溢价与无条件市场风险溢价的差值和正交无关项: $\beta_{i,t-1} = \bar{\beta}_i + \theta_i(\gamma_{1,t-1} - \gamma_1) + \eta_{i,t-1}$ 其中 $E[\eta_{i,t-1}] = 0, E[\eta_{i,t-1}\gamma_{1,t-1}] = 0$

将上式带入期望收益中, 得到

$$E[R_{i,t}] = \gamma_0 + \gamma_1 \beta_i + \gamma_2 \beta_i^\gamma \quad (4)$$

其中: $\beta_i = \frac{Cov(R_{i,t}, R_{m,t})}{Var(R_{m,t})}$, $\beta_i^\gamma = \frac{Cov(R_{i,t}, \gamma_{1,t-1})}{Var(\gamma_{1,t-1})}$, $\gamma_2 = Var(\gamma_{1,t-1})$, 分别对应市场因子暴露和市场溢价因子暴露。

(三) CAPM 参数估计

1. 排序法估计因子收益率

多数情况下因子的收益是未知的，在实证研究中通过排序法构造因子模拟投资组合排除其它因子影响，得到目标因子的收益率。

因子模拟投资组合是满足：投资组合在目标因子的暴露（目标因子的 β 值）大于零，在其它因子的暴露等于零，且特质性风险最小的投资组合。在实证中一般通过用排序法 (portfolio sort) 构建因子组合：

- 将所有样本股票在同一截面的变量根据目标变量取值进行排序
- 根据排序结果，将样本股票分为 L 组，构造价差组合 (spread portfolio): 做空排名最低的一组股票，做多排名最高的一组股票
- 此多空组合的收益率即能代表目标因子的收益率

2. 线性回归估计因子暴露

对于反映公司具体情况的风格因子 (例如公司规模、市值), 根据公司信息的披露时间, 将最新的公司特征数据作为后续时间段的因子暴露, 能够得到显著的估计结果 (Jegadeesh, 2019)。而对于市场因子 β_M 却很难找到直接的变量替代, 实证中一般使用时序回归, 将过往 $t = 1, 2, \dots, T$ 时刻的市场风险溢价 $(R_M - R_f)_t$ 以及对应时刻的资产收益率 R_t^i 视为自变量与因变量进行线性回归:

$$R_t^i = \alpha_i + \beta_M^i (R_M - R_f)_t + \varepsilon, \quad t = 1, 2, \dots, T \quad (5)$$

通过最小二乘法 (OLS), 得到资产 i 的市场因子暴露 R_M^i 的最佳线性无偏估计量 (BLUE) 作为 T 时刻因子暴露 $R_{M,T}^i$ 的估计量。

而 Conditional CAPM 将 CAPM 的异象归结为因子暴露 β 并非常量, 而是随时间变化的随机变量。因此为解决异象, 需要得到时变的因子暴露的估计值。根据式 (6), 能得到 T 时刻因子暴露而要得到实时更新的因子暴露时, 则一般采用滑窗数据, 使用 $t = T - N, \dots, T$ 的数据回归计算 $R_{M,T}^i$, 例如 Bloomberg 数据库中的市场因子暴露是过去两年每周价格回归得到的。然而对于这种方法, 由于进行回归需要的样本较多, 采用的数据最大滞后期 N 也较大, 可能导致得到的因子暴露对当前市场变化不够灵敏; 而如果对高频数据使用上述回归方法计算参数, 由于高频情况下数据扰动性强, 趋势性不明显, 同样可能使得回归结果不显著。因此需要使用高频数据, 借助高频数据分析方法进行估计。

3. 高频数据估计因子暴露

过往的论文基于上述 CAPM 理论，使用金融市场数据进行研究，但是多数的数据都是季度、月度、日频等中低频数据，考虑的时间跨度较大，通常仅用于对模型进行实证检验或者进行低频资产配置。对于信息传递速度快、流动性强的金融市场，根据上述数据得到建立的资产定价模型无法用于更高频率的应用场景中。

对于高频情况下的因子暴露估计问题，借助高频数据分析理论方法，使用随机游走模型描述资产收益率的变化，借助对随机参数模型参数的估计，从而得到因子暴露的估计量是一种可行的方法。

Jacod (1994)[18] 与 Jacod, Protter (1998)[19] 首先从概率理论角度提出高频数据分析相关理论，Foster and Nelson (1996)[20] 与 Comte, Renault (1998)[21] 从计量经济学的角度对高频数据进行研究。本文沿用 Mykland 与 Zhang 汇总发展的理论框架 [22]，基于随机分析理论模型，对 CAPM 模型中的因子暴露进行估计：即将资产价格对数 Ξ_t 的变化视为连续的 $Itô$ 过程：

$$\Xi_t = \Xi_0 + \int_0^t \mu_u du + \int_0^t \sigma_u dW_t \quad (6)$$

从而资产 i 在 t 时刻的因子暴露 $\beta_i^t = \frac{\langle \Xi_i, \Xi^M \rangle_t}{\langle \Xi^M, \Xi^M \rangle_t} = \frac{(\sigma^i \sigma^{M'})_t}{(\sigma^M \sigma^{M'})_t}$ ，从而通过估计 $(\sigma^i \sigma^{j'})_t$ 得到相应的 beta 估计量。具体理论如下节所示。

(四) 高频估计理论

传统的随机游走模型用几何布朗运动对资产价格进行描述，而金融资产收益率通常使用对数差分进行计算。同时高频数据分析假设漂移 (drift) 项 μ_t 与二次变差 σ_t^2 也是服从一定随机过程的变量，从而能够从中得到的时变的因子暴露。进一步高频数据实证中发现 microstructure noise 的存在会对价格观测产生误差，为修正这一误差，Zhang, Mykland 以及 Ait-Sahalia (2005) 提出了 Two Scales Realized Volatility 估计的方法，本文采用 Smoothed Two Scales Realized Volatility (S-TSRV) 借助高频数据对证券的实时因子暴露 (spot beta) 进行估计：

1. 模型假设

假设资产价格过程：资产 ($i=1,2,\dots,q$) 价格的对数 $(\Xi_t)_{0 \leq t \leq \tau} = (\Xi_t^{(1)}, \Xi_t^{(2)}, \dots, \Xi_t^{(q)})$ 是 q 维连续 $Itô$ 过程：

$$\Xi_t = \Xi_0 + \int_0^t \mu_u du + \int_0^t \sigma_u dW_t \quad (7)$$

其中 W 是 q 维标准 $(\mathcal{F}_t)_{0 \leq t \leq \tau}$ -布朗运动, X_0 是 \mathcal{F}_0 -可测的, 系数 μ_u, σ_u 是可料的, 且 $c_t = (\sigma \sigma^T)_t, c_t, \mu_t$ 在 $\|\cdot\|_{\max}$ 度量下局部有限。从而 Ξ_t 的积分协方差矩阵 (integrated covariance matrix) 可以被表示为: $\langle \Xi, \Xi \rangle_t = \int_0^t c_u du$ 。

假设协方差过程: $(\Xi)_t$ 其中两个分量 r, s 的 $1 \leq r, s \leq q$ 的协方差 $c_t^{(r,s)}$ 也是 $It\hat{o}$ 过程。

假设观测过程: (observed process) : z 对于 $1 \leq r \leq q, (\Xi_t^{(r)})_{0 \leq t \leq \tau}$ 在时间点 $\mathcal{G}^{(r)} = \{0 = t_0^{(r)} < t_1^{(r)} < \dots < t_{n^{(r)}}^{(r)} = T\}$ 被观测。

微观结构噪声: 真实的价格数据在受到微观结构噪声 (microstructure noise) $\varepsilon_{t_j^{(r)}}^{(r)}$ 异象后产生了新的观测值:

$$\Xi_{t_j^{(r)}}^{*,(r)} = \Xi_{t_j^{(r)}}^{(r)} + \varepsilon_{t_j^{(r)}}^{(r)}, \quad 1 \leq r \leq q \quad (8)$$

其中随机过程 Ξ_t, μ_t, σ_t 适应于滤子 (filtration) (\mathcal{F}_t) , 观测时间 $t_{n,j}$ 是 (\mathcal{F}_t) -停时, 对于每一个 (n, j) 噪声 $\varepsilon_{n,t_{n,j}}$ 于 $t_{n,j}$ 时刻被观测到, 从而是 $\mathcal{F}_{t_{n,j}}$ 可测的, 且 $\sup_{n,j} E[\varepsilon_{n,t_{n,j}}^2] < \infty, E[\varepsilon_{n,t_{n,j}}] = 0$

2. S-TSRV 法估计积分协变差

通常对于 $It\hat{o}$ 过程 $(X_t^i), (X_t^j)$ 可以通过对 Bracket 过程 $[X^i, X^j]_T = \int_0^T \sigma_t^i \sigma_t^j dt$ 的变差估计: $[X^i, X^j]_T^G = \sum_{t_{i+1} \leq T} (X_{t_{i+1}}^i - X_{t_i}^i)(X_{t_{i+1}}^j - X_{t_i}^j)$ 对相应的协方差在 0 到 T 时间的积分进行估计。然而由于观测存在微观结构噪声: $X_t^* = X_t + \varepsilon_t$, 导致观测到的 $(X^*)_t$ 不再是半鞅, 上述估计不再具有相合性。

对此 S-TSRV 方法提出间隔 K, J 期对观测值进行一系列下采样 (subsample), 分别计算两种下采样方法得到的协方差后做差消除噪声的干扰, 从而对积分协方差矩阵 (integrated covariance matrix) 进行一致估计:

同步观测格点 $\{0 = \tau_{n,0} < \tau_{n,1} < \dots < \tau_{n,N} = T\}$ 下的光滑 TSRV (S-TSRV) 统计量的计算方式如下 $\widetilde{\langle \Xi, \Xi \rangle}_t$:

对时间 $0 \leq t \leq \tau$, 资产 $1 \leq r, s \leq q$, 与估计参数 (J, K)

$$K[\widetilde{\langle \Xi^{(r)}, \Xi^{(s)} \rangle}_t]^{(K)} = \left(\frac{1}{2} \sum_{i=1}^J + \sum_{i=J+1}^{N^*(t)-b} + \frac{1}{2} \sum_{i=N^*(t)-(K+J)+1}^{N^*(t)-K} \right) (\tilde{\Xi}_{i+K}^{(r)} - \tilde{\Xi}_i^{(r)}) (\tilde{\Xi}_{i+K}^{(s)} - \tilde{\Xi}_i^{(s)}) \quad (9)$$

其中 $N^*(t) = \max\{1 \leq i \leq N : \tau_{n,i} \leq t\}$, $\tilde{\Xi}_i^{(r)}$ 是前平均价格 (the pre-averaged price): $\tilde{\Xi}_i^{(r)} :=$

$$\frac{1}{M_{n,i}^{(r)}} \sum_{\tau_{n,i-1} \leq t_j^{(r)} \leq \tau_{n,i}} \Xi_{t_j^{(r)}}^{*,(r)}$$

以相同的方法定义 $J[\widetilde{\langle \Xi^{(r)}, \Xi^{(s)} \rangle}_t]^{(J)}$

从而 Smoothed-TSRV 对积分协变差的估计量为:

$$\langle \widehat{\Xi^{(r)}}, \widehat{\Xi^{(s)}} \rangle_t = \frac{1}{(1-b/N)(K-J)} \{K[\widetilde{\Xi^{(r)}}, \widetilde{\Xi^{(s)}}]_t^{(K)} - J[\widetilde{\Xi^{(r)}}, \widetilde{\Xi^{(s)}}]_t^{(J)}\} \quad (10)$$

3. 因子暴露的估计

积分协变差对时间差分得到当前协变差的值, 在计算中通过差分对求导过程进行近似, 得到实时协变差 (spot covariance) 估计量为:

$$\hat{c}_{\Delta T_n, t}^{(r,s)} = \frac{1}{\Delta T_n} (\langle \widehat{\Xi^{(r)}}, \widehat{\Xi^{(s)}} \rangle_{t+\Delta T_n} - \langle \widehat{\Xi^{(r)}}, \widehat{\Xi^{(s)}} \rangle_t) \quad (11)$$

进一步通过 (Ξ_t) 的两个分量 $X := \Xi^{(1)}, Y := \Xi^{(2)}$ 相互表示: $dY_t = \beta_t dX_t + dZ_t, \forall t : \langle X, Z \rangle_t = 0$, 得到 $\beta_t = \frac{c_t^{X,Y}}{c_t^{X,X}}$

从而实时 beta 统计量 (spot beta estimator) 为:

$$\hat{\beta}_{\Delta T_n, t} = \frac{\hat{c}_{\Delta T_n, t}^{X,Y}}{\hat{c}_{\Delta T_n, t}^{X,X}} \quad (12)$$

4. 统计量性质

式 (10), 式 (12) 中的统计量在满足下列约束下具有相合性:

1. 观测格点在 0 到 T 之间均匀取得, 观测间隔 $\Delta \tau_n = \frac{T}{N}$
2. 微观结构噪声最多影响 J 期观测:

$$E(\bar{\epsilon}_{n,i} | \mathcal{F}_{\tau_{i-J}}) = 0 \quad (13)$$

3. 微观结构噪声二阶矩约束:

$$E[\sup_i E(\bar{\epsilon}_{n,i}^2 | \mathcal{F}_{\tau_{i-J}})] = o_p(\Delta \tau_n^+ (K-J)^{1/2}) \quad (14)$$

$$\sup_i \text{Var}(\hat{\epsilon}_i^{(s_1)} \hat{\epsilon}_i^{(s_2)}) = O_p(n^2) \quad (15)$$

4. 差分时间间隔 $\{\Delta T_n\}_{n \geq 1}$ 足够小:

$$a_n^{-2} \Delta T_n \rightarrow \infty, \Delta T_n \rightarrow 0, \quad n \rightarrow \infty \quad (16)$$

其中 $a_n = [(K_n - J_n) \Delta \tau_n]^{\frac{1}{2}}$

从上述条件 1、2、3 可以得到: 积分协变差矩阵的估计量是一致统计量:

$$|\langle \widehat{\Xi^{(r)}}, \widehat{\Xi^{(s)}} \rangle_t - \langle \Xi^{(r)}, \Xi^{(s)} \rangle_t| = O_p(a_n) \quad (17)$$

因子暴露统计量是相合的:

$$\sup_t |\hat{\beta}_{\Delta T_n, t} - \beta_t| = o_p(1) \quad (18)$$

其中: $\beta_t = \frac{c_t^{X,Y}}{c_t^{X,X}}$

(五) Betting Against Beta 异象

根据上述理论基础, 本文在高频情况下研究 Betting Against Beta (BAB) 异象。Betting Against Beta 异象由 Andrea Frazzini 以及 Lasse Heje Pedersen 在 2014 年 *Betting Against Beta* 一文中提出。BAB 异象因子 alpha 与市场因子 beta 负相关, 从而可以构造低 beta 股票多头、高 beta 股票空头的组合获取超额收益。

1. Betting Against Beta 异象的市场均衡理论解释

Frazzini 与 Lasse 首先通过解决证券市场一般均衡问题从理论上说明了 BAB 异象的存在性:

模型假设市场上共有 I 位代理人 (agents) $i = 1, 2, \dots, I$, 每位代理人持有财富 W_t^i 并持有两期, 在 t 时刻可选证券组合 $x = (x^1, \dots, x^S)'$, 市场上存在证券 $s = 1, 2, \dots, S$, 对应分红为 δ_t^s 。

根据上述假设, 对市场均衡情况进行求解: 对于每位代理, 在风险厌恶固定的情况下最大化效用函数:

$$x'(E_t(P_{t+1} + \delta_{t+1}) - (1 + r^f)P_t) - \frac{\gamma^2}{2} x' \Omega_t x \quad (19)$$

且满足资金约束: $m_t^i \sum_s x^s P_t^s \leq W_t^i$, 以及供需平衡约束: $\sum_i x^i = x^*$, 其中 m_t^i 代表资金杠杆 (贷款杠杆/准备金率), $\gamma = \sum_i \frac{1}{\gamma^i}$ 表示风险厌恶系数, Ω_t 代表收益率协方差矩阵, 对应收益率按照价格 P_t^i 计算: $r_{t+1}^i = \frac{P_{t+1}^i + \delta_{t+1}^i}{P_t^i} - 1$ 。

根据一阶条件以及资金、供需约束得到均衡价格: $P_t = \frac{E_t(P_{t+1} + \delta_{t+1}) - \gamma \Omega_t x^*}{1 + r^f + \psi_t}$, 其中 $\psi_t = \sum_i \frac{\gamma}{\gamma^i} \psi_t^i$, ψ_t^i : 为资金约束对应的拉格朗日乘子, 能够衡量资金杠杆的松紧程度。根据均衡价格得到均衡收益率:

$$E_t(r_{t+1}^s) = r^f + \psi_t + \beta_t^s [E_t(r_{t+1}^M) - r^f - \psi_t] \quad (20)$$

从而根据 (20) 式, 存在异象 $\alpha_t^s = \psi_t(1 - \beta_t^s)$ 且 α 随 β 升高降低。

该论文对 BAB 因子构造多空组合进行异象检验: 做多低 beta 资产 (收益率 r^L), 做空高 beta 资产 (收益率 r^H) 构造市场中性组合 (组合 $\beta = 0$) 得到对应收益率:

$$r_{t+1}^{BAB} = \frac{1}{\beta_t^L}(r_{t+1}^L - r^f) - \frac{1}{\beta_t^H}(r_{t+1}^H - r^f) \quad (21)$$

$$E_t(r_{t+1}^{BAB}) = \frac{\beta_t^H - \beta_t^L}{\beta_t^L \beta_t^H} \psi_t \geq 0 \quad (22)$$

$$(23)$$

从而低 Beta-高 Beta 组合具有正的超额收益。

同时该文对 20 个国家的 55600 只股票和其它资产进行实证检验, 证明了 BAB 异象的存在性。

2. Betting Against Beta 异象的 Conditional CAPM 解释

Scott 和 Michael(2015)[23] 使用 Conditional CAPM 对 BAB 因子异象进行解释: 根据 Conditional CAPM, 如果证券的条件市场因子暴露 (β) 与市场溢价 (γ) 存在相关性, 则存在异象 $\alpha = Cov(\gamma_{t-1}, \beta_{t-1})$ 。而 Scott 和 Michael 证实了 High-minus-low beta 组合的条件 beta 与市场溢价负相关, 与市场波动正相关。从而无条件的 alpha 是对真实 alpha 的下偏估计, 因此如果考虑到条件市场风险, 就能解决 Betting Against Beta 异象。

三、计量实证

(一) 数据来源

本文选择 2007 年, 标准普尔 100 指数 (SP100) 成分股中流动性最好的 70 支股票以及对应指数基金 OEF(iShares SP 100 ETF) 作为研究对象。从纽约证券交易所 (NYSE) 的 Trade and Quote(TAQ) 数据库中, 得到 2007 全年 251 个交易日的高频价格数据。其中美股每日交易时间 6.5 小时, 以 5 秒为频率对证券价格进行监测, 得到每日 4680 个时间节点的交易价格。同时从 Bloomberg 数据库得到 2007 年 70 支股票市值数据。

(二) 计量步骤

上述数据中资产 i 在 t 时刻的价格为 P_t^i , 对数价格为 $\Xi_t^i = \ln(P_t^i)$ 。记 $\Delta T = 1 \text{ day}$, 第 k 天时间为 $T = k\Delta T, k = 1, 2, \dots, 251$ 。根据资产的每日收盘价 $P_{T, \text{close}}^i$ 计算日收益率 $R_T^i = \ln(P_T^i) - \ln(P_{T-1}^i)$ 。将 OEF 收益率 R_t^{OEF} 作为市场组合收益率 R_T^M , 由于每日无风险利率远低于资产日收益, 从而忽略不计 $R^{\text{risk-free}} = 0$ 。

根据回归与高频数据分析理论, 得到进一步实证结果:

1. 根据每 $\Delta\tau = 5s$ 均匀观测的高频数据, 计算每日市场因子暴露 β_T^i 。
2. 在每个 $T = k\Delta T$ 截面, 根据第一步得到的 70 只股票的 $\beta_T^i, i = 1, 2, \dots, 70$, 使用排序法计算市场因子收益率, 并与 OEF 收益率进行比较, 检验 beta 估计效果。
3. 在每个 $T = k\Delta T$ 截面, 将第一步得到的 β_T^i 视为自变量, 将收益率 R_T^i 视为因变量进行截面回归, 得到 T 时刻对应的市场因子收益率 \hat{R}_T^M , 与 OEF 收益率比较, 检验 beta 估计效果。
4. 借助第一步计算得到的 β_T^i 对 BAB(betting against beta) 异象进行检验。

(三) 高频数据计算因子暴露

假设证券对数价格 Ξ 变化是 $It\hat{o}$ 过程:

$$\Xi_t = \Xi_0 + \int_0^t \mu_u du + \int_0^t \sigma_u dW_t \quad (24)$$

使用高频数据分析方法, 用对数价格二阶协变差 (quadratic covariation) 表示收益率协方差, 从而计算个股每天 ($\Delta T = 1 \text{ day} / 6.5h$) 的市场因子暴露 $\beta_{\Delta T, t}$:

1. 根据股票 i 在第 t 天的与市场组合 (OEF) 的变差、协变差瞬时值计算 beta:

$$\begin{aligned} \hat{\beta}_{\Delta T, t} &= \frac{d\langle \Xi^{(i)}, \Xi^{(\text{OEF})} \rangle_t}{d\langle \Xi^{(\text{OEF})}, \Xi^{(\text{OEF})} \rangle_t} \\ &= \frac{(\sigma^{(i)} \sigma^{(\text{OEF})'})_t}{(\sigma^{(\text{OEF})} \sigma^{(\text{OEF})'})_t} \\ &:= \frac{\hat{c}_{\Delta T, t}^{i, \text{OEF}}}{\hat{c}_{\Delta T, t}^{\text{OEF}, \text{OEF}}} \end{aligned} \quad (25)$$

2. 第 t 天的协变差 $c_{\Delta T, t}^{(i, \text{OEF})}$ 通过对 $\int_0^t c_{\Delta T, \tau}^{(i, \text{OEF})} d\tau$ 的差分估计得到:

$$\hat{c}_t^{(i, \text{OEF})} = \frac{1}{\Delta T} \int_t^{t+\Delta T} \hat{c}_\tau^{(i, \text{OEF})} d\tau \quad (26)$$

同样的方法计算市场因子二阶变差 $c_{\Delta T, t}^{(\text{OEF}, \text{OEF})}$

3. 借助 Smoothed-TSRV 方法估计第 t 日协变差 $\int_t^{t+\Delta T} c_\tau^{(i, OEF)} d\tau$:

$$\begin{aligned} \int_t^{t+\Delta T} \widehat{c_\tau^{(i, OEF)}} d\tau &= \langle \widehat{\Xi^{(i)}}, \widehat{\Xi^{(OEF)}} \rangle_t^{t+\Delta T} \\ &= \frac{1}{(1-b/N)(K-J)} \{K[\widehat{\Xi^{(i)}}, \widehat{\Xi^{(OEF)}}]_t^{(K)} - J[\widehat{\Xi^{(i)}}, \widehat{\Xi^{(OEF)}}]_t^{(J)}\} \end{aligned}$$

其中 $K[\widehat{\Xi^{(i)}}, \widehat{\Xi^{(OEF)}}]_t^{(K)}$ 计算方法为:

$$K[\widehat{\Xi^{(i)}}, \widehat{\Xi^{(OEF)}}]_t^{(K)} = \left(\frac{1}{2} \sum_{l=1}^J + \sum_{l=J+1}^{N-b} + \frac{1}{2} \sum_{l=N-(K+J)+1}^{N-K} \right) (\widehat{\Xi}_{l+K}^{(i)} - \widehat{\Xi}_l^{(i)}) (\widehat{\Xi}_{l+K}^{(OEF)} - \widehat{\Xi}_l^{(OEF)}) \quad (27)$$

以相同的方法定义 $J[\widehat{\Xi^{(r)}}, \widehat{\Xi^{(s)}}]_t^{(J)}$ 。

其中 (J, K) 为 S-TSRV 方法参数, $\widehat{\Xi}_t^{(i)}$ 表示第 1 个五秒内证券的对数平均价格, 根据每日高频数据 (每日交易时间 6.5 小时, 数据频率为 5s, 从而每日共 $N=4680$ 组数据) 对当日因子暴露 beta 进行估计。

参数 (J, K) 需要满足 $1 \leq J \leq K, K-J = O_p(N^{\frac{2}{3}})$, 从而在 $n \rightarrow \infty$ 时对于任何满足条件的 (J, K) , 上述算法得到的因子暴露估计量都是相合的。选择样例, 根据不同的 (J, K) 组合计算当日市场因子暴露, 得到 J, K 取值影响 Beta 的表面图。根据表面图结果, 选择市场因子暴露 (beta) 估计结果较平稳的区域, 本文取 $J=400, K=800$ 。

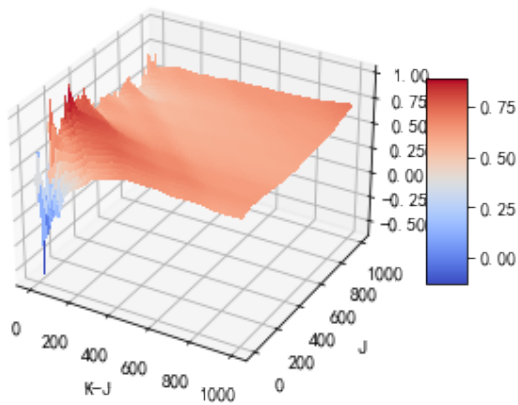
(四) 排序法构造市场因子价差策略

1. 市场因子暴露排序法

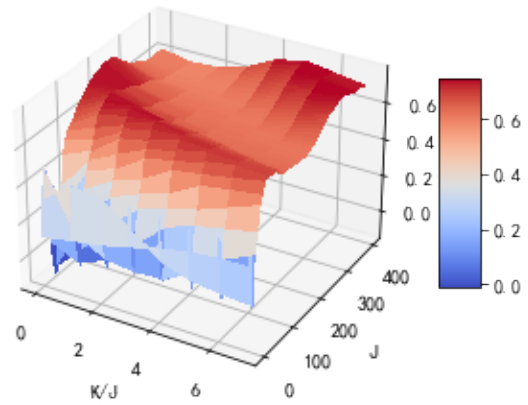
2007 年共 251 个交易日, 按照上述高频数据分析方法估计资产 i 第 t 天的市场因子暴露为 β_t^i 。在第 $t+1$ 日根据第 t 日 $\beta_t^i, i=1, 2, \dots, 70$ 排序从低到高平均分为 10 组, 组内股票等权做多, 构造 10 组投资组合, 并根据两日收盘价计算当日每组持仓收益率 $\ln(P_t^{close}) - \ln(P_{t-1}^{close})$ 。得到从第二天开始 250 天 10 种资产组合的每日收益率。

3 图中显示, 资产组合收益率随 beta 值增大而升高。对每组资产组合全年收益率总和数据进行 Spearman 秩相关系数检验来验证这一点, 上述组合收益率由低到高排序为 X_R , 上述组合按照市场因子暴露 beta 由低到高排序为 X_{beta} , 计算秩相关系数:

$$\rho_{Rank} = \frac{Cov(X_R, X_{beta})}{\sqrt{Var(X_R)Var(X_{beta})}} \quad (28)$$



(K-J,J,beta) 表面图



(K/J,J,beta) 表面图

图 1 参数 (J,K) 对 beta 估计量的影响。富国银行 (WFC)2007 年 1 月 3 日高频股价数据得到的当日市场因子暴露随 S-TSRV 参数 (J,K) 变化表面图，y 轴均为参数 J 的取值，x 轴分别为 K 与 J 的差值与比值。

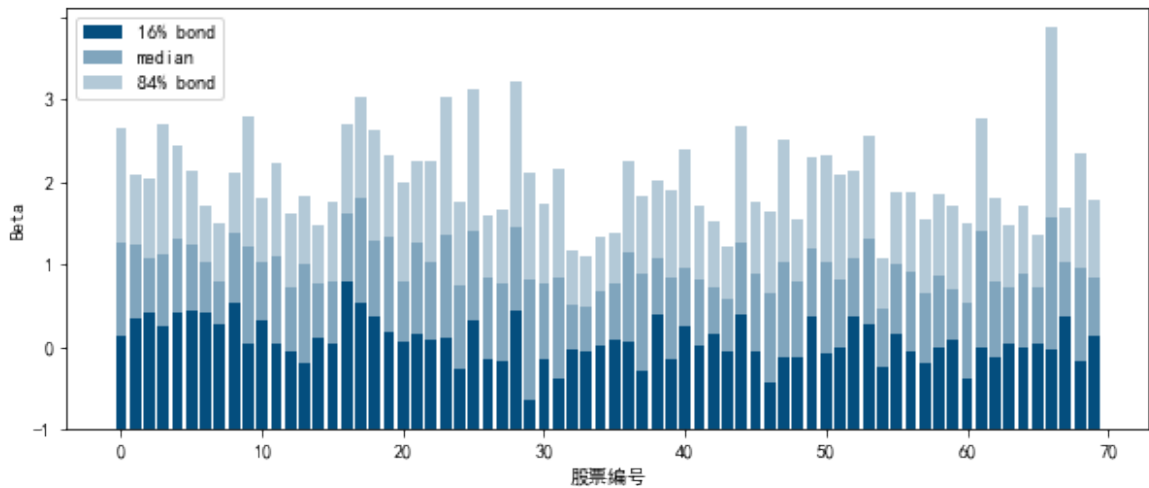


图 2 70 只股票 2007 年 beta 值的中位数及 72% 取值区间。根据高频数据计算得到 SP100 成分股中流动最好的 70 值股票在 2007 年每个交易日的市场因子暴露 beta 值，图中展示每只股票全年 251 个交易日 beta 的 16%,50%,84% 分位数

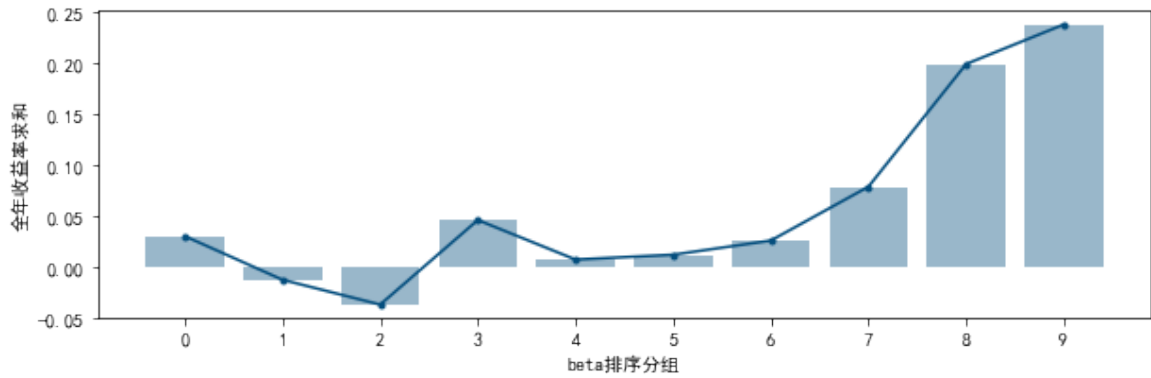


图3 收益率与beta值的关系。按照beta取值将股票池70只股票从小到大分为10组，每组7只股票，分别计算每组每日收益率全年总和

为0.0395，对应P-value为0.0251，表示收益率与beta值之间存在相关关系，且是正相关关系。而2007年全年OEF每日收益率求和为0.0395，上述结果表明，跟随市场组合，选择beta较大的投资组合能获得更多的beta收益。

此外，根据上述方法构造做多高beta资产，做空低beta资产的资产中性(money neutral)每日轮动策略：在第t+1日根据第t日 β 对70只股票进行排序，做多beta位于前70%的股票，做空beta位于后30%的股票，对所有选择的股票进行等权配置，计算每日收益率 $R_t^{HML}, t = 2, \dots, 251$ 。对该组合250日收益率序列进行T检验： $H_0: \bar{R}_t^{HML} = 0 \leftrightarrow H_1: \bar{R}_t^{HML} \neq 0$ 计算统计量 $t = \frac{\bar{R}_t^{HML}}{\sqrt{\text{Var}(R_t^{HML})}}$ 为1.6739，对应p-value为0.0954，即在低于10%错误概率情况下可以认为该组合具有非零收益。

同时可将上述组合每日收益率视为市场因子收益率的估计值(\hat{R}_t^M)，并使用该估计值对市场因子收益率代理(R_t^{OEF})使用最小二乘法进行线性回归，回归结果显示通过排序法构造投资组合得到的市场因子收益率对真实市场因子收益率有可解释性，进一步证明通过高频数据得到的市场因子暴露beta较准确。

表1 排序法组合收益率对OEF收益率OLS回归结果

	1	2	3	4	5	6
	系数	标准误	t统计量	P> t	[0.025]	[0.975]
组合收益率	0.5872	0.077	7.607	0.000	0.435	0.739
常数项	-0.0003	0.001	-0.500	0.618	-0.001	0.001
Log-Likelihood	847.09		R ²	0.189		
样本数量	250		Adj-R ²	0.186		

2. 市场因子暴露与市值双重排序法

上述对市场因子暴露 β 的单排序并没有考虑其它因素的影响。为了排除股票市值对组合收益率的干扰，下面引入市值变量，和 β 同时进行条件双重排序：

1. 每天对 70 只股票当日市值进行排序，按照市值从低到高划分为三组： $CAP1, CAP2, CAP3$ ，每组分别包含 23, 24, 23 只股票。
2. 在 $CAP1, CAP2, CAP3$ 每组中分别选择市场因子暴露 β 最大的 30% 股票与每组中 β 最小的 30% 股票，记为 $CAP1^H, CAP1^L, CAP2^H, CAP2^L, CAP3^H, CAP3^L$ 。
3. 取所有 β 最大的组的并集为控制市值情况下的高 β 组合： $G^H = CAP1^H \cup CAP2^H \cup CAP3^H$ ，取所有 β 最小的组的并集为控制市值情况下的低 β 组合： $G^L = CAP1^L \cup CAP2^L \cup CAP3^L$ 。
4. 构造做多高 β 组合，做空低 β 组合 $G^{HML} = G^H - G^L$ ，计算此组合的收益率。

对上述 G^{HML} 组合收益率进行平均收益是否为 0 的 T 检验，得到 t 统计量为 1.7606，对应 p-value 为 0.0795。说明根据市值与市场因子暴露双重排序构造的投资组合具有显著的正收益。

进一步使用上述组合收益率对 OEF 收益率进行线性回归，得到结果组合收益率依然显著，但是系数相对仅使用 Beta 排序构造组合收益率的回归系数有所降低，且拟合优度也相对减小，可能是由于先进行市值分组后构造的高低 β 组合 β 值的差距没有单一 β 分组明显。说明与市场因子暴露 β 相比，市值对收益率的影响有限。

表2 β 与市值双重排序法组合收益率对OEF收益率OLS回归结果

	1	2	3	4	5	6
	系数	标准误	t统计量	$P > t $	[0.025]	[0.975]
组合收益率	0.1910	0.027	6.983	0.000	0.137	0.245
常数项	-0.0003	0.001	-0.477	0.634	-0.001	0.001
Log-Likelihood	843.32		R^2	0.164		
样本数量	250		Adj- R^2	0.161		

(五) 回归法评估 Beta 估计量

上述排序法构造因子模拟投资组合验证了，根据高频数据估计的市场因子暴露 β 较为准确，且能够帮助构造组合获得 β 收益。下面进一步通过回归验证 β

估计的准确性。

由于传统的 CAPM 模型只能 $E(R_t^i) = \beta_i E(R_t^M) + \alpha_i$ 通过对不同时间 $t = 1, 2, \dots$ 进行时序回归, 得到一个 β_i 值, 这样显然是无法直接对使用高频数据估计的 β_t^i 进行比较, 进行评估的。因此我们进一步使用 (β_t^i) 估计市场因子收益率, 与真实市场因子收益率相比较, 间接说明市场因子暴露 beta 估计的准确性。

对每个 $t = 2, 3, \dots, 251$, 根据高频数据估计得到的当日各股票的市场因子暴露 $\beta_t^i, (i = 1, 2, \dots, 70)$, 以及当日各股收益率 $R_t^i, i = 1, 2, \dots, 70$ 进行截面回归:

$$R_t^i = \beta_t^i R_t^M + \alpha_t, \quad i = 1, 2, \dots, 70 \quad (29)$$

得到当日市场因子收益估计量 \hat{R}_t^M , 对 $t = 2, 3, \dots, 251$ 共 250 个交易日进行 250 此截面回归, 得到每日市场因子收益率估计量序列: $(\hat{R}_t^M), t = 2, 3, \dots, 251$ 。

使用上述每日市场因子收益率估计真实市场因子收益率:

$$\hat{R}_t^M = R_t^{OEF} * coef + \alpha_M, \quad t = 2, 3, \dots, 251 \quad (30)$$

上述时序回归结果中, 截面回归得到的市场因子收益率系数显著, 且为 0.9554, 而常数项 α_M 不显著, 回归拟合优度 $R^2 = 0.553$, 说明市场因子收益率估计值是对真实市场因子收益的较好拟合值, 进一步说明高频数据估计得到的市场因子暴露 beta 较为准确。

表3 截面回归估计市场因子收益率对OEF收益率OLS回归结果

	1	2	3	4	5	6
	系数	标准误	t统计量	P> t	[0.025]	[0.975]
市场因子收益率估计值	0.9554	0.055	17.508	0.000	0.848	1.063
常数项	0.0004	0.000	0.935	0.351	0.000	0.000
Log-Likelihood	921.46		R ²	0.553		
样本数量	250		Adj-R ²	0.551		

(六) BAB 异象检验

上述实证已经通过排序法和回归方法证明使用高频数据得到的市场因子暴露 beta 值较准确, 下面使用每日市场因子暴露 beta 估计值, 对 BAB(betting against beta) 异象进行检验:

根据 Frazzini, Lasse(2014) 对 BAB(betting against beta) 异象的理论推断 $\alpha_t^i = \psi_t(1 - \beta_t^i)$, 即 BAB 异象因子 alpha 与市场因子暴露 beta 有负相关关系。首先对 alpha 与 beta 进行回归来证明这一点:

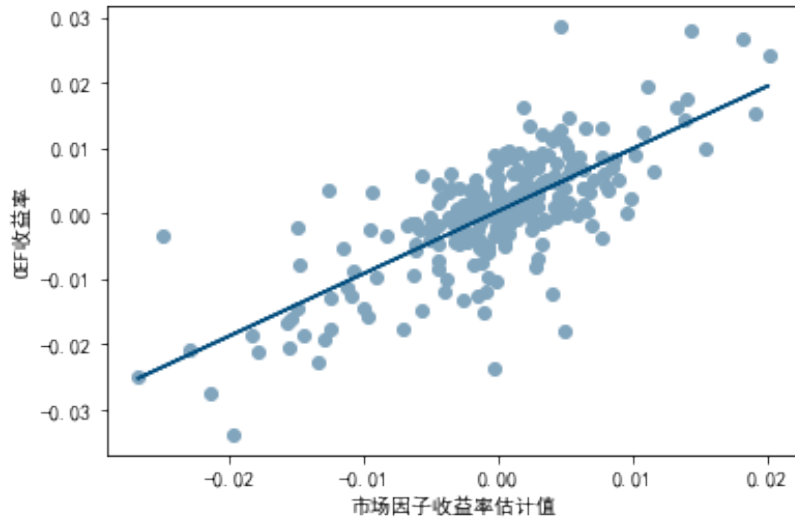


图 4 市场因子收益率估计值与真实值通过高频数据估计的市场因子暴露 beta 进行截面回归得到的市场因子收益率估计值，与真实市场因子收益率:OEF 每日收益率的散点图；以及根据两者线性回归结果拟合的回归曲线。

股票 i 第 t 天的收益率为 R_t^i ，市场因子暴露为 $\hat{\beta}_t^i$ ，OEF 第 t 天的收益率为 R_t^{OEF} 。由 CAPM 知： $R_t^i = \hat{\beta}_t^i R_t^{OEF} + \alpha_t^i + \varepsilon_t^i, i = 1, 2, \dots, 70, t = 2, 3, \dots, 251$ ，其中 ε_t^i 为随机扰动项 0。从而得到估计值 $\hat{\alpha}_t^i = \alpha_t^i + \varepsilon_t^i = R_t^i - \hat{\beta}_t^i * R_t^{OEF}$ 。将所有 $\hat{\beta}_t^i$ (共 70*250 个) 对所有 $\hat{\alpha}_t^i$ 进行回归： $\hat{\alpha}_t^i = \hat{\beta}_t^i k + const$ 得到回归系数为显著的负值，且回归拟合优度为 0.603，符合 BAB 异象因子 alpha 与市场因子暴露 beta 有负相关关系的结论。

表4 市场因子暴露Beta对BAB异象alpha的OLS回归结果

	1	2	3	4	5	6
	系数	标准误	t统计量	P> t	[0.025]	[0.975]
beta	-0.0029	1.800E-05	-163.164	0.000	-0.003	-0.003
常数项	0.0032	0.000	20.294	0.000	0.003	0.004
Log-Likelihood	42887.00		R ²	0.603		
样本数量	17500		Adj-R ²	0.603		

四、研究局限与改进方向

(一) 研究局限

- 本文研究 SP100 成分股最活跃的 70 只股票，研究对象都为大市值股票，不能代表所有上市股票。另一方面，大市值公司存在市场势力 (market power)，可能干扰市场的均衡情况，从而对 CAPM 与随机模型产生影响。

- 本文构造因子模拟投资组合时只考虑价格收益，没有考虑市场摩擦。在高频交易情况下手续费的支出占比较大，而本文计算收益率时并没有将手续费加入计算。

(二) 改进方向

- 本文只考虑了大市值股票，后续需要在研究样本中加入小市值股票进行分析。
- 需要考虑市场真实情况，研究存在交易手续费的情况下的高频交易。
- 本文只考虑了市场因子与市值影响的情况，后续可以加入更多因子，研究多因子模型在高频数据下的表现。
- 在研究高频 Betting Against Beta 异象时，只进行了简单回归验证，后续可以加入更多信息条件，对高频数据下的异象进行解释。

五、结论

过去的金融理论通常将证券价格变化视为几何布朗运动，从而通过对随机过程的分析进行资本资产定价。进一步的理论表明，可以将证券价格变化视为漂移项与波动项均为时变变量的 $Itô$ 过程，并可以借助高频数据，找到具有良好统计性质的时变参数估计值，帮助进行资产定价。其中高频数据观测值与实际值之间存在的微观结构噪声会使得一般方法的估计结果有偏，而借助 Smoothed-Two Scales Realized Volatility 方法对随机过程二次协变差的估计能够消除偏差，使得估计量具有相合性。

根据高频数据得到股票与市场组合的二次协变差估计值，将二次协变差视为变量间的协方差，从而能够在高频数据的基础上应用传统 CAPM 方法，对资本资产进行定价。

本文使用 2007 年标普 100 最活跃的 70 只成分股作为研究对象，对 5s 频率的高频价格数据进行分析。得到股票市场因子暴露的相合估计。进一步借助单一 beta 排序法构造模拟市场组合，市值、beta 双重排序法构造模拟市场组合，验证了市场因子暴露估计量的准确性。且根据高低 beta 构造的投资组合在该年市场组合整体收益为正的情况下具有显著正收益。并通过截面回归以及市场因子收益实际值与估计值的时序回归，间接说明高频数据估计的市场因子暴露 beta 准确。最后通过 CAPM 残差与 beta 的回归验证了存在随 beta 上升减小的 Betting Against Beta 异象。

附 录

计量实验配置

- Windows
- Python3.8.8
- Pytorch 1.9
- NVIDIA GPU + CUDA CuDNN

计量实验数据

本文选择 2007 年，标准普尔 100 指数 (SP100) 成分股中流动性最好的 70 支股票以及对应指数基金 OEF(iShares SP 100 ETF) 作为研究对象。从纽约证券交易所 (NYSE) 的 Trade and Quote(TAQ) 数据库中，得到 2007 全年 251 个交易日的高频价格数据。其中美股每日交易时间 6.5 小时，以 5 秒为频率对证券价格进行监测，得到每日 4680 个时间节点的交易价格。同时从 Bloomberg 数据库得到 2007 年 70 支股票市值数据。

样本股票池为 2007 年 SP100 流动性最好的 70 只股票：T, CL, MDLZ, IBM, C, FB, BK, GM, KMI, RTN,OEF, PYPL, PM, BAC, AMGN, WBA, KHC, NEE, HD,COST, HON,CSCO, ALL, GILD, AAPL, SPG, MET, FOX,AXP, GD, V', TGT, QCOM, JPM, AMZN, VZ, MA, TXN, AIG, DHR, GOOGL, WFC, JNJ, DOW, ORCL, COF, USB,GOOG, MDT, CELG, CAT, KO, BRKB, MS, PFE, LLY,GS, ABBV, CVS, UNH, MRK, CVX, DD, EMR, INTC,UTX, MCD, MSFT, BLK, ABT。

计量实验代码

需要所有计量代码请联系作者：Liuxinyu_loisey@126.com, 下面仅展示核心代码：

```
1 import torch
2 class Index_extract():
3     def __init__(self,OEF,SP,N,K,J):
4         self.N=N
5         self.OEF=OEF.view(self.N,-1).cuda().detach()
6         self.SP=SP.view(self.N,-1).cuda().detach()
```



```

7         self.K=K
8         self.J=J
9
10    def get_beta(self):
11        SP=self.SP
12        OEF=self.OEF
13        N=self.N
14        K=self.K
15        J=self.J
16        b=K+J
17
18        # 求 K_lag variatio
19        SP_adK=torch.zeros(SP.shape,device=SP.device)
20        SP_adK[0:N-K,:]=SP[K:N,:]
21        SP_diffK=SP_adK-SP
22
23        OEF_adK=torch.zeros(OEF.shape,device=OEF.device)
24        OEF_adK[0:N-K,:]=OEF[K:N,:]
25        OEF_diffK=OEF_adK-OEF
26
27        cov1K=SP_diffK*OEF_diffK
28        variat1K=(1/2)*cov1K[0:J].sum(0)+cov1K[J:N-b].sum(0)
29        +(1/2)*cov1K[N-b:N-K].sum(0)
30
31        cov2K=OEF_diffK*OEF_diffK
32        variat2K=(1/2)*cov2K[0:J].sum(0)+cov2K[J:N-b].sum(0)
33        +(1/2)*cov2K[N-b:N-K].sum(0)
34
35        # 求 J_lag variation
36        SP_adJ=torch.zeros(SP.shape,device=SP.device)
37        SP_adJ[0:N-J,:]=SP[J:N,:]
38        SP_diffJ=SP_adJ-SP
39
40        OEF_adJ=torch.zeros(OEF.shape,device=OEF.device)
41        OEF_adJ[0:N-J,:]=OEF[J:N,:]
42        OEF_diffJ=OEF_adJ-OEF

```

```

42         cov1J=SP_diffJ*OEF_diffJ
43         variat1J=(1/2)*cov1J[0:K].sum(0)+cov1J[K:N-b].sum(0)
+ (1/2)*cov1J[N-b:N-J].sum(0)
44
45         cov2J=OEF_diffJ*OEF_diffJ
46         variat2J=(1/2)*cov2J[0:K].sum(0)+cov2J[K:N-b].sum(0)
+ (1/2)*cov2J[N-b:N-J].sum(0)
47
48         #c
49         c1=1/((1-b/N)*(K-J))*(variat1K-variat1J)
50         c2=1/((1-b/N)*(K-J))*(variat2K-variat2J)
51
52         # beta
53         beta=c1*(1/c2)
54
55         return(beta)
56
57     def get_close(self):
58         return(self.OEF[self.N -1],self.SP[self.N-1,:])
59
60     def get_group(self):
61         betas=Index_extract.get_beta(self)
62         (_,sort_index)=torch.sort(betas)
63         return(sort_index.view(10,-1))

```

代码 1 市场因子暴露 (Spot beta) 估计

参考文献

- [1] Louis Bachelier. Théorie de la spéculation. In *Annales scientifiques de l'École normale supérieure*, volume 17, pages 21–86, 1900.
- [2] Maury FM Osborne. Brownian motion in the stock market. *Operations research*, 7(2):145–173, 1959.
- [3] Harry M Markowitz. Portfolio selection. In *Portfolio selection*. Yale university press, 1968.
- [4] Jack L Treynor. Market value, time, and risk. *Time, and Risk (August 8, 1961)*, 1961.
- [5] Jack L Treynor. *Toward a theory of market value of risky assets*. 1961.
- [6] William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.
- [7] John Lintner. Security prices, risk, and maximal gains from diversification. *The journal of finance*, 20(4):587–615, 1965.
- [8] John Lintner. Distribution of incomes of corporations among dividends, retained earnings, and taxes. *The American economic review*, 46(2):97–113, 1956.
- [9] Jan Mossin. Equilibrium in a capital asset market. *Econometrica: Journal of the econometric society*, pages 768–783, 1966.
- [10] Eugene F Fama. Random walks in stock market prices. *Financial analysts journal*, 51(1):75–80, 1995.
- [11] Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.
- [12] Ravi Jagannathan and Zhenyu Wang. The conditional capm and the cross-section of expected returns. *The Journal of finance*, 51(1):3–53, 1996.
- [13] Lan Zhang, Per A Mykland, and Yacine Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411, 2005.

- [14] Per A Mykland, Lan Zhang, and Dachuan Chen. The algebra of two scales estimation, and the s-tsrv: High frequency estimation that is robust to sampling times. *Journal of Econometrics*, 208(1):101–119, 2019.
- [15] Andrea Frazzini and Lasse Heje Pedersen. Betting against beta. *Journal of Financial Economics*, 111(1):1–25, 2014.
- [16] Sanjoy Basu. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The journal of Finance*, 32(3):663–682, 1977.
- [17] Mark M Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82, 1997.
- [18] Valentine Genon-Catalot and Jean Jacod. Estimation of the diffusion coefficient for diffusion processes: random sampling. *Scandinavian Journal of Statistics*, pages 193–221, 1994.
- [19] Jean Jacod and Philip Protter. Asymptotic error distributions for the euler method for stochastic differential equations. *The Annals of Probability*, 26(1):267–307, 1998.
- [20] Dean P Foster and Daniel B Nelson. Continuous record asymptotics for rolling sample variance estimators, 1994.
- [21] Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical finance*, 8(4):291–323, 1998.
- [22] Per A Mykland and Lan Zhang. The econometrics of high frequency data. *Statistical methods for stochastic differential equations*, 124:109, 2012.
- [23] Scott Cederburg and MICHAEL S O'DOHERTY. Does it pay to bet against beta? on the conditional performance of the beta anomaly. *The Journal of finance*, 71(2):737–774, 2016.

致 谢

从 2021 年末开始在陈大川老师的帮助下学习高频数据计量相关的知识，但是后续因为疫情无法返校进行更多数据的实验，中间又因为实习又搁置了一段时间，一路磕磕绊绊终于把论文写完了。在写论文与读文献时越来越感觉到自己知识上的不足和这篇论文的贫乏。无论是概率理论还是统计方法亦或是经济思想均没有继续深入的研究与学习，最后只能借助众多“巨人的肩膀”来完成这篇毕业论文，以至于最初无意生产学术垃圾，最后完成时读此文却如食鸡肋。

但是这一路的学习却使我收获颇丰，我逐步认识到市场的均衡映照到证券的价格，从价格的随机游走参数的扰动中窥见市场的风云变幻，并想进一步研究高频数据得到的参数波动与市场信息机制的关系，但是由于时间和能力关系这篇论文只能止步于此了。概率原理、计量方法与经济金融原理丰富又复杂，现实市场更是深不可测，仅此一生怕是难以见其全貌，只能寄希望于在未来的工作学习中不断积累、更新认知，得以在一瞬触及真理的尾巴，也算是得偿所愿了。

感谢家人亲友对我大学生活学习的支持。特别感谢指导老师陈大川老师非常认真负责的指导，让我能够完成这篇论文，极大地减小了延毕的风险。感谢好友柯凌璐在大学期间的一直陪伴；感谢好友丁一哲对我统计编程学习的帮助；感谢好友李龙剑的经济学理论支持；感谢统计学院潘斌老师，让我大三到大四在组会中进一步学习统计与机器学习，得以完成这篇论文的编程工作；感谢数院李静老师、江一鸣老师、李津竹老师、胡威老师，组合数学中心史永堂老师在学习与升学过程中的帮助。祝愿各位老师亲人朋友身体健康，工作顺利，学业有成！