

# ALPHA因子专题

## 因子合成方法& 宏观指标因子化& 因子换手率优化

---

科大财经

大纲:

1. Stepwise Selection算法
2. Stepwise最大化ICIR最大化合成
3. Stepwise最大化多头组收益最大化合成
4. EPU经济不稳定性宏观指标因子化
5. 课后作业 **【'GDP','CPI','SF','M2','PMI'】**
6. 因子换手率优化

# STEPWISE SELECTION

**Stepwise Selection** 是一种变量选择的方法，用于在给定的一组预测变量中选择最佳的子集来建立一个模型。它基于统计学的原理，通过逐步添加或删除变量来选择最佳模型。

**Stepwise Selection** 有两种常见的方法：前向选择和后向选择。

1. 前向选择 (**Forward Selection**)：从空模型开始，逐步选择最佳的变量添加到模型中。具体步骤如下：
  - a. 从所有可能的预测变量中选择一个变量，该变量与响应变量具有最高的相关性。
  - b. 逐步添加其他变量，每次选择与当前模型的残差最相关的变量，直到达到预设的停止准则。
2. 后向选择 (**Backward Selection**)：从包含所有预测变量的完整模型开始，逐步删除最不相关的变量，直到达到预设的停止准则。具体步骤如下：
  - a. 从包含所有预测变量的完整模型开始。
  - b. 逐步删除与当前模型的残差最不相关的变量，直到达到预设的停止准则。

**Stepwise Selection** 是一种常见的特征选择方法，用于在建立模型时选择最相关的变量。然而，需要注意的是，它可能存在一些问题，例如过拟合和选择偏差等。因此，在使用 **Stepwise Selection** 时需要谨慎，并结合领域知识和模型评估指标来进行决策。

以下是一些关于 **Stepwise Selection** 的链接，你可以进一步了解相关的内容和实现方法：

- [Stepwise Selection in R](https://www.statology.org/stepwise-selection-in-r/)
- [Stepwise Regression in Python](https://www.kaggle.com/hamelg/python-for-data-science-index#stepwise-regression)
- [Stepwise Selection on Wikipedia](https://en.wikipedia.org/wiki/Stepwise\_regression)

# 经济不确定因子 ECONOMIC POLICY UNCERTAINTY

经济政策不确定的衡量：[http://www.policyuncertainty.com/china\\_monthly.html](http://www.policyuncertainty.com/china_monthly.html)

经济政策：货币政策、财政政策 ——旨在调节总需求（刺激/抑制有效需求）

经济政策不确定性是指未来与经济相关的政策变动中包含的各类无法预知的成分 eg.政策变迁

经济政策不确定性上升，企业降低投资和劳动力需求，消费者收入降低、需求减少，进一步加剧企业需求萎靡

风险偏好降低，对股票要求更高的风险溢价，导致股价或估值降低

综上，构建公司对经济政策不确定的敏感度

使用模型：Fama-French 三因子、Fama-French 五因子

利用文本分析法搜索中国大陆具有影响力的报刊 “经济”、“政策”和“不确定性”等关键词出现的频次来计算中国经济政策不确定性指数（EPU），

EPU Indices	
All Country-Level Data	
Global	USA
Australia	Belgium
Brazil	Canada
Chile	China
Colombia	Croatia
Denmark	France
Germany	Greece
Hong Kong	India
Ireland	Italy

## China Policy Uncertainty Indices Based on Mainland Papers

[Download Data](#)

[Download Annotated Index](#)

We are pleased to host newspaper-based indices of policy uncertainty in China developed by Steven J. Davis, Dingqian Liu and Xuguang S. Sheng, based on their working paper, "Economic Policy Uncertainty in China Since 1949: The View from Mainland Newspapers." They quantify uncertainty-related concepts from October 1949 onwards using two mainland Chinese newspapers: the Renmin Daily and the Guangming Daily.

To construct an Economic Policy Uncertainty Index for China, they first obtain monthly counts of articles that contain at least one term in each of three term sets: Economics, Policy, and Uncertainty. Table 1 reports the terms in each set using Chinese characters and the corresponding English translations. In a second step, they scale the raw monthly EPU counts by the number of total articles for the same newspaper and month.

Table 1. Term Sets for Index of Economic Policy Uncertainty in China

Category	English Terms	In Chinese Characters
Uncertainty	uncertain/uncertainty/ not certain/unsure/ not sure/hard to tell/ unpredictable/unknown	不确定/不明确/不明朗/未明/ 难料/难以预计/难以估计/ 难以预测/难以预料/未知
Economics	economy/business	经济/商业
Policy	fiscal/monetary/ China Securities Regulatory Commission/ China Banking Regulatory Commission/ Ministry of Finance/ The People's Bank of China/ National Development and Reform Commission / Opening-up/ Reform/ Ministry of Commerce/ legislation/tax/ national bonds/government debt/ central bank/ Ministry of Commerce/ tariff/governmental deficit	财政/货币/证监会/银监会/财 政部/人民银行/国家发改委/ 开放/改革 /商务部/法律/法规 /税收/国债/政府债务/央行/外 经贸部/关税/政府赤字

## 2.2.1 时间序列回归

时间序列回归 (time-series regression) 简单直接, Black et al. (1972) 最早使用它来检验CAPM。这种方法在回归时使用因子收益率作为自变量 (independent variable) 或解释变量 (explanatory variable), 以资产的超额收益率作为因变量 (dependent variable) 或被解释变量 (explained variable)。

此方法更适合分析由风格因子构成的多因子模型, 这是因为人们可以使用2.1节介绍的排序法构建风格因子的因子模拟投资组合, 并计算其收益率作为解释变量。对于其他类别的因子, 比如GDP等宏观经济因子, 由于难以应用排序法构建它的因子模拟投资组合以及计算收益率, 所以这种方法就无法使用, 举例来说, Fama and French (1993) 中的价值 (HML) 和规模 (SMB) 因子均是风格因子。该文使用独立双重排序构建了这两个因子的投资组合, 并计算了它们的收益率时间序列, 使用它们就可以方便地进行时间序列回归检验。

令  $\lambda_t$  表示  $t$  期因子收益率向量,  $R_{it}^e$  为资产  $i$  在  $t$  期的超额收益率, 这二者在时序上满足如下线性关系:

$$R_{it}^e = \alpha_i + \beta_i' \lambda_t + \varepsilon_{it}, \quad t = 1, 2, \dots, T \quad (2.13)$$

$$R_p - R_f = \beta_0 * EPU + \beta_1 * (Rm - Rf) + \beta_2 * HML + \beta_3 * SMB$$

对每个资产  $i=1, 2, \dots, N$ , 使用简单最小二乘 (Ordinary Least Squares, OLS) 对模型 (2.13) 进行参数估计。在时间序列回归中, 回归方程右侧自变量是因子收益率  $\lambda_t$ , 左侧的因变量是  $R_{it}^e$ , 回归得到资产  $i$  在因子上的暴露  $\hat{\beta}_i$  向量, 截距  $\hat{\alpha}_i$ , 以及残差  $\hat{\varepsilon}_{it}$ 。一旦有了  $\hat{\alpha}_i$  和  $\hat{\beta}_i$ , 将  $R_{it}^e$  和  $\lambda_{it}$  在时序上取均值就可得:

$$E_T[R_i^e] = \hat{\alpha}_i + \hat{\beta}_i' \hat{\lambda}, \quad i = 1, 2, \dots, N \quad (2.14)$$

式中  $E_T[\cdot] \equiv \frac{1}{T} \sum (\cdot)$  表示对样本数据在时序上取均值;  $\hat{\lambda} = ET[\hat{\lambda}_t]$  是资产预期收益率和因子暴露在截面上的关系式。时间序列回归中的截距  $\hat{\alpha}_i$  正是资产  $i$  的定价误差的估计。由 (2.14) 可知, 时间序列回归的好处是可以方便地估计每个因子的预期收益率。对于任意因子  $k$ , 其收益率序列  $\lambda_{kt}$  在时序上的均值就是因子  $k$  预期收益率的估计:

$$\hat{\lambda}_k = ET[\lambda_{kt}], \quad k = 1, 2, \dots, K \quad (2.15)$$