Project Introduction

This is a rough research survey about alpha trading. In this project, I built up a pipeline of alpha trading in cluding data processing, factor analysis, and model calibration.

Codes

- rqdata_utils.py: Utils dealing with the rice quant platform data
- FactorAnalysis.ipynb: Factor returns profile visulization
- BARRA.ipynb: BARRA's risk model with three calibration schemes: cross-sectional regression, global gradient descend, and Kalman filter.

Dataset

The dataset is not available as it is too large, here, I used Chinese A-stocks as an example (hard for free US equities' data). The data frame is multi-indexed similar to Quantopian's format. However, feel free to apply your own dataset.

Goal

- **Equity Return Forecasting**
- **Portfolio Risk Estimation**
    - APT
        - Risk Exposure: $\beta_{i,k}$
        - Risk Premium: P_k
        - Contribution of Risk Factor to Long Term Excess Return:

$$E[r_i] - TB = \sum_k \beta_{i,k} P_k$$

    - BARRA
        - Factor Return Covariance: V
        - Portfolio Risk: σ_p

- Portfolio Risk Exposures:

$$x_p = X^T h_p$$

- Marginal Contribution for Total Risk:

$$MCTR = \frac{Vh_p}{\sigma_p}$$

- Portfolio Risk-Adjusted Expected Return:

$$U = h_p^T r - \lambda \cdot h_p^T V h_p$$

Model Classification

- CAPM
    - a kind of sigle-factor model
    - usually, a validity benchmark for other models
- APT
    - factor returns are assumed to be known
    - factor exposure can be regressed from factor returns
    - aimed at forecasting
    - how to fit: Fama-Macbeth Algorithm
- Multi-Index Models
    - statistical indogeneous model using factor analysis
    - useful at factors parsimouny and decouple
- Multi-Factor Risk Models(BARRA)
    - factor exposures are assumed to be known (can be derived as the rescaled factor value)
    - factor return can be regressed from factor exposures
    - aimed at risk management
    - how to fit: cross-sectional regression

Calibration Algorithms

Here I used 2 traditional way add a novel Kalman filter technique (see KalmanFilter.ipynb or

MultiFactorModel.ipynb)

- Time-series regression (fix equity)
- Cross-sectional regression (fix time-stamp)
- Kalmn filter (APT model allowing risk exposure and risk premium to vary over time. In another word, a dynamic model with gaussian noise)

Notes on Factor Models

CAPM

- Author: Markovitz(1959)
- single-factor:
- explain: security returns

APT

- Author: Stephen A. Ross(1976)
- multi-factor
- explain: security returns

Postulates:

- The linear model

$$r_i(t) - \alpha_i = \sum_{k=1}^{K} \beta_{ik} \cdot f_k(t) + \epsilon_i(t)$$

where $f_k(t)$ is the realization(value) of risk factor at time t

- No pure arbitrage profit

Conclusion

- Exposure of each security on each factor
- Risk premium on each factor

$$(Mean[r_i(t)])_i = P_0 + \sum_{k=1}^{K} \beta_{ik} \cdot P_k$$

or make $\beta_{0,k}$ equals 1 for each k,

$$(Mean[r_i(t)])_i = \sum_{k=0}^{K} \bar{\beta}_{i,k} \cdot P_k$$

where P_0 is the risk free return

- Portfolio exposure to each factor

$$Portfolio_{it} = \beta_0 + \beta_k \cdot f_{kit}$$

Three alternative calibration methods

- **statistical techniques** such as factor analysis, principle analysis
  - **Goodness**: good for determining the number of relevent risk factors
  - **Undesirable**: hard to interpret
- **portfolios**: K different well-diversified portfolios as substitutions
  - **Goodness**: lead to insights
  - **Fama-Macbeth regression**
- **economic theory** (highly developed art)
  - **Goodness**: Intuitively appealing set of factors that admit economic interpretation of risk exposures
  - **Goodness**: Using economic information in addition to stock return. Avoid using stock return to explain stock return
  - **factors**:
    1. confidence risk
    2. time horizon risk
    3. inflation risk
    4. bussiness cycle risk
    5. market-timing risk

Generalizations

The simplicity of APT framework is a great virtue. It is helpful to understand the true sources of stock returns.

The basic APT model can be enhanced in many ways.

- Allow risk prices P_k to vary over time
- Allow risk exposures $\beta_{i,k}$ to vary over time
- Use Bayesian mothods to produce optimal out-of-sample forcasts for the risk exposures and hence for the expected returns
- Introduce additional factor with zero-risk prices. Although do not contribute to expected return, help to explain the volatility.

Multi-Index Models (Factor Analysis & PCA)

Goal

Using historical return extract the factors

$$r_{it} = \alpha_i + \sum_k \beta_{ik} \cdot f_{kt}$$

where

$$E[\epsilon_{it}\epsilon_{jt}] = 0$$

$$E[\epsilon_{it} f_{kt}] = 0$$

f_{kt} : the return on index k inperiod t

β: sensitivities

Estimation

Either exposure or factor return can be asserted on a priori grounds with the other identified empirically, or both can be identified empirically.

Characteristics

- Have f(indexes) represents separate influence
- The structure must be parsimonious: the returns can be described in terms of limited indexes

Statistical Solutions

Let the data design the model

- PCA
- Factor Analysis: better in heteroscedastic series

Design Issue

- **The Choice of Data**: Individul stocks vs portfolio
- **The number of Index**:
    - Stactical techniques: Factor analysis, PCA
    - Common sense and economic significance play a major role in deciding on the number of factors
- **The nonuniqueness of Factors**: The researcher should realize the resulting structure is not unique. Some researchers will examine alternative structures in an atempt to understand what influences are affecting security returns and to convince themself the overall separation make an intuitive sense
- **Computational Problems**:
    - Roll and Ross: Multisample approach
    - Chen: Portfolio approach

Applications

- **Identify the Indexes set**
- **Determine the number of factors**: PCA / Factor Analysis
    - Single-group tests for each sample
        - Factor Analysis on return-generating process
        - Criteria: Chi2, AIC, **BIC**
    - Multiple-group tests for all stocks
        - Canonical Correlation (CCA):

          take two sets of variables and see what is common amongst the two sets (can be two noncorresponding variables either on index or dimension)

$$X_{N\times K}, Y_{N\times K'}$$

$$\text{x\_weights}_{K,n}$$

$$\text{y\_weights}_{K',n}$$

          Use CCA / PLS:

$$\text{X\_score}_{N\times n} = \text{Normalized}[X]_{N\times K}\text{x\_weights}_{K,n}$$

$$\text{Y\_score}_{N\times n} = \text{Normalized}[Y]_{N\times K'}\text{y\_weights}_{K',n}$$

- Determin the number:
  - r-value for $n = 10$
  - correlation matrix pattern for each number of components: $n \times n$ for $n = 1, \cdots, 10$

- **Generate Factors**

- **Calibrate sensitivities**:
  - Portfolio exposure to each factor
  - $AdjustedR^2$ (Should be stable)
  - Explanatory power: Compare these results with those for the single-index model (Should depend on the market cap)

- **Explanatory Power** of the Model for Each Stock: R2>0.7 excellent

Conclusions

- Goodness: simultaneously estimate the indexes and sensitivities in a multi-index model
- Defect: Data Minning: Using return to explain return

Multi-Factor Models for Portfolio Risk (BARRA)

$$r_{i,t} = a_{i,t} + X_{i,k,t} \cdot f_{k,t}$$

where $X_{i,k,t}$: the exposure of asset i to factor k known at time t $f_{k,t}$: the factor return to factor k during the period from time t to time $t+1$ $a_{i,t}$: the stock i's specific return during period from time t to time $t+1$ $r_{i,t}$: the excess return (return above the risk-free return) on stock i during the period from time t to time $t+1$

The risk structure

$$V_{i,j} = X_{i,k1}F_{k1,k2}X_{j,k2}^T + \Delta_{i,j}$$

$$V = X^T F X + \Delta$$

where

$F_{k1,k2}$ is the K by K covariance matrix for factor returns

$\Delta_{i,j}$ is the N by N diagonal matrix of specific variance

A portfolio described by an N-element vector h_i

- portfolio exposure: $x_p = X^T h_p$
- portfolio variance: $\sigma_p^2 = x_p^T F x_p + h_p^T \Delta h_p = h_p^T V h_p$
- Marginal Contribution for Total Risk

$$MCTR = \frac{Vh_p}{\sigma_p}$$

- Risk-adjusted expected return:

$$U = h_p^T r_p - \lambda \cdot h_p^T V h_p$$

Choosing the Factors

- External influences --> BARRA Model
  - Return in bond market (bond beta)
  - Unexpected changes in inflation
  - Change in oil price
  - Change in exchange rate

- Cross-sectional comparisons
  - Fundamental
  - Market
    - volatility
    - price
    - share turnover

- Purely internal or statistical factors
  - see multi-index model

Exposures

- Industry Exposures
  - 1/0 variable

- Risk Index Exposures
  - Volatility: beta, daily return vol, option implied vol
  - Momentum
  - Size

- Liquidity
- Growth
- Value(Fundamentals)
- Earning volatility
- Financial leverage: debt-to-equity ratios

Applications

- Rescale the Exposures
- Regress the Factor Returns Against Exposures via Cross-sectional Regression

$$f = (X^T W X)^{-1}(X^T W r)$$
$$= \sum_{i=1}^{N} C_{k,i} r_i$$

  Here factor return can be interpreted as the return to a portfolio with weights $C_{k,i}$. So factor returns are the returns to factor portfolios. This portfolio has unit exposure to the particular factor

- Factor Covariance and Specific
  - Stock returns
  - Factor exposures
  - Stock dividends, splits, and other adjustment

Model Validation

- Model Setting:
  - 50 factors
  - 1000 assets

- Measures:

  - R^2 : 30-40%. It can vary quite significantly from month to month. And depends on the market return level.
  - root mean square error: 6% roughly against 10% volatility
  - Portfolio Risk

- Goal:

  - Expain the portfolio risk
  - Forecast variances and covariances of factors and specific returns
  - Providing incisive, intuitive and interesting risk analysis

You can think of this as slicing through the other direction from the APT analysis, as now the factor returns are unknowns to be solved for, whereas originally the coefficients b were the unknowns. Another way to think about it is that you're determining how predictive of returns the factor was on that day, and therefore how much return you could have squeezed out of that factor.