R Finance 2014
Estimating Factor Models and Managing Risk with factorAnalytics

Yi-An Chen

Department of Economics
University of Washington

May 16, 2014

Road map

Overview of factorAnalytics

Estimating Factor Model

Risk Management

Background of factorAnaltyics

- R package "factorAnalytics" is designed for factor models estimation and risk management.
- It has 21 functions and 3 demo data so far, as well as an user friendly vignette.
- It talks to "PerformanceAnalytics" with portfolio risk analytics.
- It was one of the many projects of R, Statistical computing participating in Google Summer of Codes 2013.
- It is now a branch project of "ReturnAnalytics" and currently lives under it on the R-forge.
- Developers include Brian Peterson, Doug Martin, Eric Zivot, Guy Yollin, Peter Carl and Yi-An Chen.

Main Functions

- Estimating: fitTimeSeriesFactorModel.r, fitFundamentalFactorModel.r, fitStatisticalFactorModel.r
- Risk managing : factorModelSdDecomposition.r, factorModelVaRDecomposition.r, factorModelEsDecomposition.r
- Other misc. functions : factorModelCovariance.r, factorModelMonteCarlo.r
- Generic functions for fitting: plot(), predict(), summary(), print()

Introduction

Factor models for asset returns are used to

- Decompose risk and return into explainable and unexplainable components
- Generate estimates of abnormal return
- Describe the covariance structure of returns
- Predict returns in specified stress scenarios
- Provide a framework for portfolio risk analysis

Factor Model in General Form

A factor model is defined as

$$\begin{aligned} R_{it} &= \alpha_i + \beta_{1i} f_{1t} + \beta_{2i} f_{2t} + \cdots + \beta_{Ki} f_{Kt} + \epsilon_{it} \\ &= \alpha_i + \beta_i' f_t + \epsilon_{it} \end{aligned}$$

Where

- R_t is simple returns or excess returns on asset i (i $= 1 \cdots$ N) and in time period t (t$= 1 \cdots$ T).
- β_{ki} is factor exposures or factor loadings for asset i on kth factor.
- f_{kt} is common risk factor (k $= 1 \cdots$ K).
- ϵ_{it} is asset specific returns or residual returns.

Assumption

1. The common risk factors, f_t, are stationary with unconditional moments

$$\mathbb{E}[f_t] = \mu_f$$
$$cov(f_t) = \mathbb{E}[(f_t - \mu_f)(f_t - \mu_f)] = \Omega_f$$

2. Asset specific returns, ϵ_{it}, are uncorrelated to each common risk factors f_{kt}.

$$cov(f_{kt}, \epsilon_{it}) = 0, \forall\, k, i, t$$

3. Error terms, ϵ_{it}, are serially uncorrelated and contemporaneously uncorrelated across assets

$$\begin{aligned} cov(\epsilon_{it}, \epsilon_{jt}) &= \sigma_i^2, \forall\, i = j,\ t = s \\ &= 0, otherwise \end{aligned}$$

Three types of models

Factor Model:

$$R_{it} = \alpha_i + \beta_i' f_t + \epsilon_{it}$$

According to different model setup, 3 main different types of models:

- Common factor f_t is known, it is called time series factor model.
- Factor loadings β are known, from firm characteristics. It is called fundamental factor model.
- both β and f_t are unknown and extracted from asset returns, it is called statistical factor model.

Fitting Time Series Factor Model

Factor model can be written as time-series regression model for asset i:

$$R_i = \mathbf{1}_T\alpha_i + F\beta_i + \epsilon_i,\ i = 1, \cdots, N$$

where

- R_i is T $\times$ 1,
- α_i is scalar,
- F is T $\times$ K common risk factors,
- β_i is K $\times$ 1 vector,
- $\mathbb{E}[\epsilon_i^2] = \sigma_i^2$, time-series homoskedasticity.

Estimation can be done by N times time-series regression. Function "fitTimeSeriesFactorModel()" is design to do it.

fitTimeSeriesFactorModel()

```
> args(fitTimeSeriesFactorModel)
function (assets.names, factors.names, data = data, num.factor.s
    fit.method = c("OLS", "DLS", "Robust"), variable.selection =
    decay.factor = 0.95, nvmax = 8, force.in = NULL, subsets.met
        "backward", "forward", "seqrep"), lars.criteria = "Cp",
    add.up.market.returns = FALSE, add.quadratic.term = FALSE,
    excess.market.returns.name)
```

- 3 different estimation methods: OLS(lm), DLS(wls), Robust(lmRob)
- support different model selection methods :
  1. "none" for no selection,
  2. "stepwise" is forward/backward stepwise OLS regression (step() or step.lmRob()),
  3. "all subsets" is all subsets regression (regsubsets()), subsets methods are "exhaustive", "backward", "forward" and "seqrep"
  4. "lar" , "lasso" is based on package "lar"

(Go to R script.)

Market Timing Model

Henriksson-Merton Model (1981)

$$R_{p,t} - R_f = \alpha + \beta(R_{m,t} - R_f) + \gamma D + \epsilon_t$$

where

- R_t is asset returns, R_m is market returns, R_f is risk-free rate.
- D $= \max(0, R_{m,t} - R_f)$ mimicking payoff to an put option.
- γ is market timing ability. Numbers of free puts generated by manager's strategy.

Treynor-Mazuy Model (1966)

$$R_{p,t} - R_f = \alpha + \beta(R_{m,t} - R_f) + \gamma(R_{m,t} - R_f)^2 + \epsilon_t$$

If γ is positive, it indicates the manager's investment strategy having market timing ability, making more when market is up and loss less when it is down.

Fitting Statistical Factor Model

- In statistical factor models, the factor realizations f_t are not directly observable and must be extracted from the observable returns R_t using statistical methods. The primary methods are factor analysis and principal components analysis.
- Traditional factor analysis and principal component analysis are usually applied to extract the factor realizations if the number of time series observations, T, is greater than the number of assets, N.
- If $N > T$, then the sample covariance matrix of returns becomes singular which complicates traditional factor and principal components analysis. In this case, the method of "asymptotic principal component analysis" is more appropriate.

Factor Analysis

$$\underbrace{R_t}_{N\times 1} = \underbrace{\mu}_{N\times K} + \underbrace{B}_{N\times K}\underbrace{f_t}_{K\times 1} + \underbrace{\epsilon_t}_{N\times 1}, \; t = 1, \cdots, T$$

- $cov(f_t, \epsilon_t)$=0, for all t,s
- $\mathbb{E}[f_t] = \mathbb{E}[\epsilon_t] = 0$
- $var(f_t) = \mathbf{1}_K$
- $var(\epsilon_t) = D$, where D is diagonal matrix with σ_i^2 along the diagonal.
- Return covariance matrix Ω can be decomposed as

$$\Omega = BB' + D$$

- If T $>$ N, it is a traditional factor model. Error term ϵ_{it} is assumed no serial and cross-sectional correlation for all i,t.

Traditional Factor Model Estimation

When T > N, Sample principal components are computed from the spectral decomposition of the N $\times$ N sample covariance matrix $\hat{\Omega}_N$:

$$\hat{\Omega}_N = \hat{P}\hat{\Lambda}\hat{P}'$$

$$\hat{P} = [\hat{p_1} \vdots \hat{p_2} \vdots \cdots \vdots \hat{p_N}],\ \hat{P}' = \hat{P}^{-1}$$
$$\hat{\Lambda} = diag(\hat{\lambda}_1, \cdots, \hat{\lambda}_N),\ \hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_N$$

where P is matrix of orthonormal eigenvectors and Λ are corresponding eigenvalues. The estimated factor realizations are simply the first K sample principle components

$$\hat{f}_{kt} = \hat{p}_k' R_t,\ k = 1, \cdots, K,$$
$$\hat{f}_t = (\hat{f}_{1t}, \cdots, \hat{f}_{Kt})$$

The factor loading for each assets, β_i and the residual variance σ_i^2 can be estimated via OLS from the time series regression

$$R_{it} = \alpha_i + \beta_i' \hat{f}_t + \epsilon_{it},\ t = 1, \cdots, T$$

Asymptotic Principle Components

- When N >T, Asymptotic principal component analysis (APCA) by Conner and Korajczyk (1986) will be used.
- APCA is based on eigenvector analysis of the T $\times$ T matrix $\hat{\Omega}_T$. APCA estimates of factor f_k are the first K eigenvectors of $\hat{\Omega}_T$.

  That is $\hat{F} = [f_1 \vdots \cdots \vdots f_K]$ is an orthonormal T $\times$ K matrix.
- It works easier with T $\times$ T matrix $\hat{\Omega}_T$ rather than N $\times$ N matrix $\hat{\Omega}_N$ when N grows large.
- It allows approximate factor structure of returns. ϵ_{it} are allowed weakly cross-sectional correlation which is more realistic assumption.
- Conner and Korajczyk (1986) and Bai and Ng (2002) methods can be used to determine numbers of factors when N $>$ T.

(Go to R script)

Fitting Fundamental Factor Model

Fundamental factor models use observable asset specific characteristics (fundamentals) like industry classification, market capitalization, style classification (value, growth) etc. to determine the common risk factors.

- Factor loading β are from observable firm characteristics.
- Common risk factor f_t are estimated at each t give β
- BARRA approach: Estimate f_t at time t given firm characteristics β. That is to say, running T cross-section regressions
- "fitFundamentalFactorModel()" can estimate fundamental factor model.

Example: BARRA type single factor model

A simple model in the form of a cross-sectional regression at time t :

$$\underbrace{R_t}_{N\times 1} = \underbrace{\beta}_{N\times 1}\underbrace{f_t}_{1\times 1} + \underbrace{\epsilon_t}_{N\times 1},\ t = 1, \cdots, T$$

- β is N $\times$ 1 observed firm characteristics.
- f_t is unobserved factor realization.
- $cov(f_t, \epsilon_{it}) = 0$, $var(f_t) = \sigma_f^2$ for all i, t
- Heteroskedasticity in error term : $var(\epsilon_{it}) = \sigma_i^2$ for all i $= 1, \cdots,$N
- Estimation $\hat{f}_t$ can be done by Weighted Least Square (WLS) cross-sectionally at each time t.

fitFundamentalFactorModel()

```
> args(fitFundamentalFactorModel)
function (data, exposure.names, datevar, returnsvar, assetvar,
    wls = TRUE, regression = "classic", covariance = "classic",
    full.resid.cov = FALSE, robust.scale = FALSE,
    standardized.factor.exposure = FALSE, weight.var)
```

- data is panel like $(T \times N) \times K$ data.frame. exposure.names is colnames of factor loadings.
- datevar indicates date variables and should be xts identifiable
- returnsvar indicates returns variables. In order to talk to PerformanceAnalytics. Do not use %.
- assetvar indicates assets variables.

(Go to R script)

Cont.

```
> args(fitFundamentalFactorModel)
function (data, exposure.names, datevar, returnsvar, assetvar,
    wls = TRUE, regression = "classic", covariance = "classic",
    full.resid.cov = FALSE, robust.scale = FALSE,
    standardized.factor.exposure = FALSE, weight.var)
```

- wls can be chosen to be TRUE or FASLE
- regression method can be chosen as classic or robust (lmRob())
- covariance method can be chosen as classic or robust (covRob())
- If full.resid.cov controls is TRUE, residual covariance matrix is not diagonal but full.
- If standardized.factor.exposure logical is TRUE. Factor exposure will be standardized to regression weighted mean 0 and standardized deviation to 1.

(Go to R script)

BARRA type Industry Model

The industry model with K industries is defined as :

$$
\begin{aligned}
R_{it} &= \beta_{1i} f_{1t} + \beta_{2i} f_{2t} + \cdots + \beta_{Ki} f_{Kt} + \epsilon_{it},\ i = 1, \cdots, N,\ t = 1, \cdots, T \\
var(\epsilon_{it}) &= \sigma_i^2,\ i = 1, \cdots, N \\
cov(\epsilon_{it}, f_{jt}) &= 0,\ j = 1, \cdots K,\ i = 1, \cdots, N \\
cov(f_{it}, f_{jt}) &= \sigma_{ij}^2,\ i, j = 1, \cdots K
\end{aligned}
$$

where

$$
\begin{aligned}
\beta_{ik} &= 1, \text{ if asset i is in industry k} \\
&= 0, \text{ otherwise}
\end{aligned}
$$

(Go to R script)

Factor Risk Budgeting

- Additively decompose (slice and dice) individual asset or portfolio return risk measures into factor contributions
- Allow portfolio manager to know sources of factor risk for allocation and hedging purposes
- Allow risk manager to evaluate portfolio from factor risk perspective

Factor Risk Decompositions Assume asset or portfolio return R_t can be explained by a factor model

$$R_t = \alpha + \beta' f_t + \epsilon_t$$

$$f_t \sim iid(\mu_f, \Omega_f), \epsilon_t \sim iid(0, \sigma_\epsilon^2), cov(f_{k,t}, \epsilon_s) = 0, \forall k, t, s$$

Rewrite the factor model as

$$\begin{aligned} R_t &= \alpha + \beta' f_t + \epsilon_t = \alpha + \beta' f_t + \sigma_\epsilon \times z_t \\ &= \alpha + \bar{\beta}' \bar{f}_t \\ \bar{\beta} &= (\beta', \sigma_\epsilon)', \ \bar{f}_t = (f_t, z_t)', \ z_t = \frac{\epsilon_t}{\sigma_\epsilon} \sim iid(0, 1) \\ \sigma_{FM}^2 &= \bar{\beta}' \Omega_{\bar{f}} \bar{\beta}, \ \Omega_{\bar{f}} = \begin{pmatrix} \Omega_f & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Euler's Theorem and Risk Decomposition

Let $RM(\bar{\beta})$ denote the risk measure σ_{FM}, VaR_{α}^{FM} and ETL_{α}^{FM} as function of β. We have following risk decomposition:

$$
\begin{aligned}
RM(\bar{\beta}) &= \sum_{j=1}^{K+1} \bar{\beta}_j \frac{\partial RM(\bar{\beta})}{\partial \bar{\beta}_j} \\
&= \bar{\beta}_1 \frac{\partial RM(\bar{\beta})}{\partial \bar{\beta}_1} + \bar{\beta}_2 \frac{\partial RM(\bar{\beta})}{\partial \bar{\beta}_2} + \cdots + \bar{\beta}_{K+1} \frac{\partial RM(\bar{\beta})}{\partial \bar{\beta}_{K+1}} \\
&= \beta_1 \frac{\partial RM(\bar{\beta})}{\partial \beta_1} + \cdots + \beta_K \frac{\partial RM(\bar{\beta})}{\partial \beta_K} + \sigma_{\epsilon} \frac{\partial RM(\bar{\beta})}{\partial \sigma_{\epsilon}}
\end{aligned}
$$

Terminology

Factor j marginal contribution to risk

$$\frac{\partial RM(\bar{\beta})}{\partial \bar{\beta}_j}$$

Factor j contribution to risk

$$\bar{\beta}_j \frac{\partial RM(\bar{\beta})}{\partial \bar{\beta}_j}$$

Factor j percentage contribution to risk

$$\frac{\bar{\beta}_1 \frac{\partial RM(\bar{\beta})}{\partial \bar{\beta}_1}}{RM(\bar{\beta})}$$

It can be shown that:

$$\frac{\partial VaR_{\alpha}^{FM}(\bar{\beta})}{\partial \bar{\beta}_j} = \mathbb{E}[\bar{f}_t | R_t = VaR_{\alpha}^{FM}(\bar{\beta})], \, j = 1, \cdots, K+1$$
$$\frac{\partial ETL_{\alpha}^{FM}(\bar{\beta})}{\partial \bar{\beta}_j} = \mathbb{E}[\bar{f}_t | R_t \leq VaR_{\alpha}^{FM}(\bar{\beta})], \, j = 1, \cdots, K+1$$

Remarks:

- Intuitive interpretation as stress loss scenario.
- Analytic results are available under normality.

(Go to R script)

Factor Model Monte Carlo

Problem: Short history and incomplete data limits applicability of historical simulation, and risk budgeting calculations are extremely difficult for nonnormal distributions

- Use fitted factor model to simulate pseudo assets return data preserving empirical characteristics
- Use full history for factors (t=1,$\cdots$,T) and observed history for asset returns (t=$t_i, \cdots, T$). Data structure is unequal.
- Do not assume full parametric distributions for assets returns and risk factor returns
- Estimate tail risk and related measures non-parametrically from simulated return data

Simulation Algorithm

Estimate factor models for each asset using partial history for assets and risk factors.

$$R_{it} = \hat{\alpha}_{it} + \hat{\beta}_i' f_t + \hat{\epsilon}_{it},\ t = t_i, \cdots, T$$

Simulate B values of risk factors by resampling with replacement from full history of risk factor:

$$f_1^*, \cdots, f_B^*$$

Simulate B values of the factor model residuals from empirical distribution or fitted non-normal distribution

$$\hat{\epsilon}_1^*, \cdots, \hat{\epsilon}_B^*$$

Create pseudo factor model returns from fitted factor model parameters, simulated factor variables and simulated residuals:

$$R_{it}^* = \hat{\beta}_i' f_t^* + \hat{\epsilon}_{it}^*, t = 1, \cdots, B$$

Check out factorAnalytics

It is downloadable on R-forge

https://r-forge.r-project.org/R/?group_id=579

Thanks You for Coming

Comments Welcome

Yi-An Chen

Email: chenyian@uw.edu