

An alternative approach to Black Litterman

Yixiang Wang*, Yue Wang^{†‡}

December 24, 2016

Abstract

In this report, we provided a generalized form of Black Litterman by showing explicitly how one would combine multiple views of factor returns on each single factor under APT model framework. Along with the form, we applied a variety of techniques to forecast factor returns and assessed its impact in this framework.

1 Introduction

In this study, we conclude that the simplest EWMA-like Holt winter approach is the most appealing single filter to forecast the factor returns that yield an IR of 1.92 within Black Litterman framework applied in this dataset. The remainder of the paper is organized as follow: section 2 discusses the theoretical foundation for the Black Litterman model construction and associated mathematics. Section 3 describes the data and assumption and methodologies used to generate the result. Section 4 show the main results. Section 5 explores the behavior of the associated results and concludes.

*yixiang.wang.ny@gmail.com

[†]jasonyuewang@gmail.com

[‡]Special thanks to our professor Gordon Ritter.

2 Theoretical Foundations

2.1 Overview of APT Model

Multi-factor models assume a linear functional form

$$R_{t+1} = X_t f_t + \epsilon_t, \mathbb{E}[R_t] = 0, \mathbb{V}[R_t] = D \quad (2.1)$$

where R_{t+1} is an n -dimensional random vector containing the cross-section of returns in excess of the risk-free rate over some time interval $[t, t + 1]$, X_t is a (non-random) $n \times p$ matrix that is known before time t . Also ϵ_t is assumed to follow a mean-zero distribution with diagonal variance- covariance matrix

$$D := \text{diag}(\sigma_1^2, \dots, \sigma_n^2) \text{ with all } \sigma_i^2 > 0. \quad (2.2)$$

Eq(1.2) entails that all significant sources of correlation are already captured by factors, represented as columns of X_t . We henceforth suppress the implicit time index since most of our discussions concern a single time interval. Note, for later use, that D^{-1} exists and can be computed in $O(n)$ time. The variable f in (2.1) denotes a p -dimensional random vector process which cannot be observed directly; information about the f -process must be obtained via statistical inference. We assume that the f -process has finite first and second moments given by

$$E[f] = \mu_f, \text{ and } V[f] = F \quad (2.3)$$

The primary outputs of a statistical inference process are the parameters f and F mentioned in (3.3), and other outputs one might be interested in include estimates of the daily realizations \hat{f}_t .

2.2 Overview of Generalized Black Litterman Framework

A Black-Litterman-Bayes model consists of:

- A parametric statistical model for asset returns $p(r \mid \theta)$ with finite-dimensional parameter vector.
- A prior π on the parameter space.
- A likelihood function $f(q \mid \theta)$ where θ is any parameter vector appearing in a parametric statistical model for asset returns, and q is a vector supplied by portfolio managers or economists.
- A utility function $u(w)$ of final wealth in the sense of Arrow(1971) and Pratt(1964).

Given a Black-Litterman-Bayes (BLB) model, the associated BLB optimal portfolio is defined to be

$$h^* \in \arg \max \mathbb{E}[u(\mathbf{h}'\mathbf{r} \mid \mathbf{q})] \quad (2.4)$$

where $\mathbb{E}[\cdot \mid q]$ denotes the expectation with respect to the posterior predictive density for the random variable \mathbf{r} . In other words, h^* maximizes posterior expected utility.

$$p(r \mid q) = \int p(r \mid \theta) p(\theta \mid q) d\theta \quad \text{where} \quad (2.5)$$

$$p(\theta \mid q) = \frac{p(q \mid \theta) \pi(\theta)}{\int p(q \mid \theta) \pi(\theta) d\theta} \quad (2.6)$$

Given a benchmark portfolio with holdings \mathbf{h}_B (eg. the market portfolio), and given a Black-Litterman-Bayes model (given above), the prior $\pi(\theta)$ is said to be benchmark-optimal if \mathbf{h}_B maximizes expected utility of wealth, where the expectation is taken with respect to the a priori distribution on asset returns $p(r) = \int p(r \mid \theta) \pi(\theta) d\theta$ so

$$h \in \arg \max_h \int u(\mathbf{h}'\mathbf{r}) p(\mathbf{r} \mid \theta) d\theta \quad (2.7)$$

In summary, The model of Black and Litterman (1991) is the special case in which $r \mid \theta$ is multivariate normal with mean θ and $f(\cdot \mid \cdot)$ is the normal likelihood for a regression of the portfolio manager's views, the utility of final wealth is the CARA function $u(w) = -e^{-\delta w}$, and the prior is the unique normal distribution which is benchmark-optimal with respect to the market portfolio.

2.3 Mathematics behind Black Litterman Portfolio Construction

Starting with some views:

$$\mathbb{E}[\mathbf{Pr}] = \mathbf{q} \quad (2.8)$$

views serve as the noisy observations about \mathbf{r} that statistical inference would be performed against

$$\mathbf{r} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \quad (2.9)$$

To represent its noisy property, we specify uncertainty associated with the view that expressed

$$\mathbf{P}\boldsymbol{\theta} = \mathbf{q} + \boldsymbol{\epsilon}^{(v)}, \quad \boldsymbol{\epsilon}^{(v)} \sim N(0, \boldsymbol{\Omega}), \quad \boldsymbol{\Omega} = \text{diag}(w_1, w_2, \dots, w_k) \quad (2.10)$$

The noise is represented by $\epsilon^{(v)}$. The intuition behind this is if any random variable r comes from a density $p(r | \theta)$ with parameter θ , and if one were given a set of noisy observations of realizations of r , then one could infer something about θ by statistical inference. To conduct statistical inference, not only one would need observations, but also a fully specified statistical model, which includes both prior and likelihood. based on the above definition, the likelihood is given as follows:

$$f(\mathbf{q} | \boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2}(\mathbf{P}\boldsymbol{\theta} - \mathbf{Q})' \boldsymbol{\Omega}^{-1}(\mathbf{P}\boldsymbol{\theta} - \mathbf{Q})\right] \quad (2.11)$$

which essentially is a standard normal for a multiple linear regression with dependant variables \mathbf{q} and design matrix \mathbf{P} . In the absence of any sort of information/views which could constitute alpha over the benchmark, the optimization procedure should simply return the global CAPM equilibrium portfolio, with holdings denoted \mathbf{h}_{eq}

$$\mathbf{r} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \text{ and } \boldsymbol{\theta} \sim N(\boldsymbol{\Pi}, \mathbf{F}) \quad (2.12)$$

$$\mathbb{E}[\mathbf{p}'\mathbf{r}] = \mathbf{P}'\boldsymbol{\Pi} \text{ and } \mathbb{V}[\mathbf{p}'\mathbf{r}] = \mathbf{p}'(\boldsymbol{\Sigma} + \mathbf{C})\mathbf{p} \quad (2.13)$$

We also need to make a choice whether to use the conditional or unconditional variance in optimization: equally:

$$\mathbb{V}[\mathbf{r} | \boldsymbol{\theta}] = \boldsymbol{\Sigma} \text{ v.s. } \mathbb{V}[\mathbf{r}] = \boldsymbol{\Sigma} + \mathbf{C} \quad (2.14)$$

In this case, we are concerned with unconditional variance i.e. $\mathbb{V}[\mathbf{r}] = \boldsymbol{\Sigma} + \mathbf{C}$. Denote $\mathbf{h} \in \mathbb{R}^n$ as the vector of portfolio holdings, which has **unit of dollars**. Mean Variance optimization with a risk-aversion parameter $\epsilon > 0$ leads to

$$\mathbf{h}_{eq} = \epsilon^{-1}(\boldsymbol{\Sigma} + \mathbf{C})^{-1}\boldsymbol{\Pi} \quad (2.15)$$

In a special case where $\mathbf{C} = \tau\boldsymbol{\Sigma}$ with some scalar τ we will have

$$\mathbf{h}_{eq} = \epsilon^{-1}(1 + \tau)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi} \quad (2.16)$$

we obtain the log posterior as following:

$$(\mathbf{P}\boldsymbol{\theta} - \mathbf{Q})' \boldsymbol{\Omega}^{-1}(\mathbf{P}\boldsymbol{\theta} - \mathbf{Q}) + (\boldsymbol{\theta} - \boldsymbol{\Pi})' \mathbf{C}^{-1}(\boldsymbol{\theta} - \boldsymbol{\Pi}) \quad (2.17)$$

$$= \boldsymbol{\theta}'[\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1}]\boldsymbol{\theta} - 2(\mathbf{q}'\boldsymbol{\Omega}^{-1}\mathbf{P} + \boldsymbol{\Pi}'\mathbf{C}^{-1}) \quad (2.18)$$

”completing the square” Lemma: If a multivariate normal random variable $\boldsymbol{\theta}$ has density $p(\boldsymbol{\theta})$ and

$$-2\log p(\boldsymbol{\theta}) = \boldsymbol{\theta}'\mathbf{H}\boldsymbol{\theta} - 2\boldsymbol{\eta}'\boldsymbol{\theta} + \text{term}(\text{ without } \boldsymbol{\theta}) \quad (2.19)$$

then $\mathbb{V}[\boldsymbol{\theta}] = \mathbf{H}^{-1}$ and $\mathbb{E}[\boldsymbol{\theta}] = \mathbf{H}^{-1}\boldsymbol{\eta}$. By above lemma: with $\mathbf{H} = \mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1}$ the posterior has **mean**

$$\mathbf{v} = [\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1}]^{-1}[\mathbf{P}'\boldsymbol{\Omega}^{-1}\mathbf{q} + \mathbf{C}^{-1}\boldsymbol{\Pi}] \quad (2.20)$$

and **covariance**

$$\mathbf{H}^{-1} = [\mathbf{P}'\mathbf{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1}]^{-1} \quad (2.21)$$

Based on CARA utility, equivalently to solve the following:

$$\mathbf{h}^* \in \arg \max_{\mathbf{h}} \{\mathbb{E}[\mathbf{h}'\mathbf{r}] - (\delta/2)\mathbb{V}[\mathbf{h}'\mathbf{r}]\} \quad (2.22)$$

where $\mathbb{E}[\mathbf{r}]$ and $\mathbb{V}[\mathbf{r}]$ represent **unconditional mean** and **covariance** of \mathbf{r} under the posterior. we have the **variance** term can be represented as following:

$$\mathbb{V}[\mathbf{h}^{-1}\mathbf{r}] = \mathbf{h}'[\mathbf{P}'\mathbf{\Omega}^{-1}\mathbf{P} + \mathbf{C}^{-1}]^{-1}\mathbf{h} + \mathbf{h}'\mathbf{\Sigma}\mathbf{h} \quad (2.23)$$

and optimal portfolio holding is then

$$\mathbf{h}^* = \delta^{-1}[\mathbf{H}^{-1} + \mathbf{\Sigma}]^{-1}\mathbf{H}^{-1}[\mathbf{P}'\mathbf{\Omega}^{-1}\mathbf{q} + \mathbf{C}^{-1}\mathbf{\Pi}] \quad (2.24)$$

where $[\mathbf{H}^{-1} + \mathbf{\Sigma}]^{-1}$ is the updated **covariance** and $\mathbf{H}^{-1}[\mathbf{P}'\mathbf{\Omega}^{-1}\mathbf{q} + \mathbf{C}^{-1}\mathbf{\Pi}]$ is the updated **mean**

2.4 Step by Step Black Litterman Portfolio Construction

Step 1 involves of calculate the posterior distribution of $\boldsymbol{\theta}$, after the view being taken into account, based on the assumption, the posterior distribution is of the same distributional family as the prior (again normal) - but with different parameters If original parameters are $\boldsymbol{\xi}, \mathbf{V}$ i.e. $\pi(\boldsymbol{\theta}) \propto N(\boldsymbol{\xi}, \mathbf{V})$ Then the updated parameters are:

$$\tilde{\boldsymbol{\xi}} = (\mathbf{V}^{-1} + \mathbf{\Omega}^{-1})^{-1}(\mathbf{V}^{-1}\boldsymbol{\xi} + \mathbf{\Omega}^{-1}\mathbf{q}) \quad (2.25)$$

$$\tilde{\mathbf{V}} = (\mathbf{V}^{-1} + \mathbf{\Omega}^{-1})^{-1} \quad (2.26)$$

Step 2 follows similar calculation by using the updated values $\boldsymbol{\xi}, \mathbf{V}$ we can find explicitly find the $\mathbb{V}[\mathbf{r} \mid \boldsymbol{\theta}]$ and $\mathbb{E}[\mathbf{r} \mid \boldsymbol{\theta}]$ (The explicit formula is omitted here) Step 3 involves solve it in mean variance framework, but also substitution we can obtain the updated portfolio as following:

$$\mathbf{h} = \delta^{-1}\mathbf{\Sigma}^{-1}\mathbf{\Pi} \quad (2.27)$$

$$\mathbf{\Pi} := \mathbf{X}\tilde{\mathbf{u}}_f \quad (2.28)$$

$$\tilde{\mathbf{u}}_f := (\mathbf{V}^{-1} + \mathbf{\Omega}^{-1} + \mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{V}^{-1}\boldsymbol{\delta} + \mathbf{\Sigma}^{-1}\mathbf{q}) \quad (2.29)$$

3 Main result

3.1 Data

For this projcet, we study the United States equity market over the period 2006-2015 through the lens of an APT model. For each day t in our sample, we take n =

2000 and select the top n stocks in the US market, sorted by market capitalization. We chose this value because stocks falling below the top 2000 by market cap tend to be illiquid and have wide spreads, making them difficult to trade for institutional investors. We restrict the study to common stock, hence excluding closed-end funds, REITs, ETFs, unit trusts, depository receipts, warrants, etc. Our only data sources for this study were CRSP and IBES which we access via the Wharton Research Data Service.

We construct our model to contain five of the most commonly-studied and well-known sources of systematic risk: market beta, size, value, momentum, and volatility, as well as a classification of the stocks into industries. some interpretation of the risk factors:

- Market beta: each asset's daily excess return time series is winsorized and regressed against the SPX, over a 2-year window w intercept. The beta is the slope coefficient. The results are further improved using the Vasicek (1973) Bayesian adjustment in the cross section.
- Size: market capitalization.
- Volatility: as in the market beta calculation, each asset's daily excess return time series is winsorized and regressed against the SP 500 excess return time series, over a trailing two year window. The mean-square error (MSE) from this regression is the raw exposure.
- Momentum: trailing compound returns over the last 12 months with the last 1 month excluded.
- Value: the raw exposure is et/pt where et represents the sum of the trailing 4 quarters' earnings-per-share (EPS), adjusted for splits, and pt is the close price on day t .

3.2 Assumption

The study presented in this paper heavily relies on the prior ground-breaking work that has been done by Peter Kolm and Gordon Ritter in their paper.

Within APT, the factor f could be well defined by its first and second moment, (when f has a finite first and second moment) we could induce the following:

$$\mathbb{E}[\mathbf{f}] = \boldsymbol{\mu}_f, V[\mathbf{f}] = \mathbf{F} \quad (3.1)$$

which leads to a further reduction of the following:

$$\mathbb{E}[\mathbf{f}] = \boldsymbol{\mu}_f, \boldsymbol{\Sigma} := V[\mathbf{f}] = \mathbf{D} + \mathbf{X}\mathbf{F}\mathbf{X}' \quad (3.2)$$

Back to the black litterman, we are free to choose $\boldsymbol{\theta}$ (the distribution of prior $\boldsymbol{\theta}$) as any vector of parameters appearing in a parametric statistical model for asset returns; It seems to be natural to choose $\boldsymbol{\theta} = \mathbf{u}_f$, the \mathbf{k} parameters describing the factor risk premia. For simplicity we treat \mathbf{F} as a constant matrix, just as the original Black-Litterman model treats $\boldsymbol{\Sigma}$ as a constant matrix.

Define the likelihood function to be (Normal likelihood again)

$$f(q \mid \theta) = \prod_{i=1}^k \exp\left[-\frac{1}{2w_i^2}(\theta_i - q_i)^2\right] \quad (3.3)$$

Conventionally, there are mainly two types of priors: one driven by historical data, and one driven by the desire to have some specific benchmark turn out to be optimal under the model of the prior, in this paper we will solely address the first approach. Residing on the nature of APT model, we applied an OLS approach to estimate the factor return cross-sectionally at any given t by introducing the data driven approach

- random process driving f is stationary implies $\boldsymbol{\mu}_f$ and \mathbf{f} constant overtime
- use simple OLS estimates $f_t = (\mathbf{X}_t' \mathbf{X}_t)^{-1} \mathbf{X}_t' \mathbf{r}_{t+1}$ to be prior mean
- resulting model treat each time period as a "group" with $\mathbf{r}_{t+1} \sim N(\mathbf{X}_t \mathbf{f}_t, \mathbf{D})$ and \mathbf{f}_t modeled as i.i.d. draws $\mathbf{f}_t \sim N(\mathbf{u}_t, \mathbf{F})$
- The statistical inference problem is then to infer $\boldsymbol{\theta} = \mathbf{u}_f$ from obs of $\mathbf{D}t$

Note: The "data-driven" approach to prior selection that we have just described has the advantage of not requiring a benchmark portfolio. In particular, this makes sense for absolute return strategies where the elective benchmark is cash.

3.3 Methodology and our approach

3.3.1 Estimation of factor returns and covariances

Define the prior:

- 6-month lookback window
- \mathbf{u}_f to be the mean of looking back window factor return
- \mathbf{F} to be the covariance of looking back window factor return
- $\mathbf{D} := \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ with all $\sigma_i^2 > 0$ via time series of cross sectional regression residual at each time t and average over based on initial looking back window

3.3.2 Market Portfolio and Black Litterman Portfolio

Based on the above assumptions, we will be able to derive the following results:
Market Portfolio:

$$\mathbf{F}_{mar} := (\delta \mathbf{\Sigma})^{-1} \mathbf{X}_t \mathbf{u}_f, \text{ where } \mathbf{\Sigma}_t = \mathbf{X}_t \mathbf{F} \mathbf{X}_t' + \mathbf{D}_t \quad (3.4)$$

Black Litterman Portfolio:

$$\mathbf{h}_{blb} = \delta^{-1} \mathbf{\Sigma}^{-1} \mathbf{X} (\mathbf{V}^{-1} + \mathbf{\Omega}^{-1} + \mathbf{X}' \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} (\mathbf{V}^{-1} \boldsymbol{\delta} + \mathbf{\Sigma}^{-1} \mathbf{q}) \quad (3.5)$$

3.3.3 Generalized Black Litterman Portfolio

Instead of presenting single view for each single factor (recall the likelihood define previously)

$$f(q \mid \theta) = \prod_{i=1}^k \exp\left[-\frac{1}{2w_i^2}(\theta_i - q_i)^2\right] \quad (3.6)$$

where specifically each i corresponding to a single view to a single factor with total of \mathbf{k} factors.

we proposed a transform of likelihood to be

$$f(q \mid \theta) = \prod_{i=1}^k \prod_{j=1}^n \exp\left[-\frac{1}{2w_{ij}^2}(\theta_i - q_{ij})^2\right] \quad (3.7)$$

where n is the number of views for each single factor

In black litterman, the optimal holding based on this construct will be instead

$$\mathbf{h}_{blb} = \delta^{-1} \mathbf{\Sigma}^{-1} \mathbf{X} (\mathbf{V}^{-1} + \mathbf{M}' \mathbf{\Omega}^{-1} \mathbf{M} + \mathbf{X}' \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} (\mathbf{V}^{-1} \boldsymbol{\delta} + \mathbf{M}' \mathbf{\Sigma}^{-1} \mathbf{q}) \quad (3.8)$$

Where \mathbf{M} are specified as following: we can think originally \mathbf{M} is an identity matrix $\mathbf{1}_{n,n}$, now updated to a matrix for each factor, we have equal number of methods to

forecast, say \mathbf{n} then its \mathbf{n} by \mathbf{k} matrix defined as following

$$\mathbf{M} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \dots & x_{1,k} \\ x_{2,1} & x_{2,2} & x_{2,3} & \dots & x_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & x_{n,3} & \dots & x_{n,k} \\ x_{n+1,1} & x_{n+1,2} & x_{n+1,3} & \dots & x_{n+1,k} \\ x_{n+2,1} & x_{n+2,2} & x_{n+2,3} & \dots & x_{n+2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{2n,1} & x_{2n,2} & x_{2n,3} & \dots & x_{2n,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{k(n-1)+1,1} & x_{k(n-1)+1,2} & x_{k(n-1)+1,3} & \dots & x_{k(n-1)+1,k} \\ x_{k(n-1)+2,1} & x_{k(n-1)+2,2} & x_{k(n-1)+2,3} & \dots & x_{k(n-1)+2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{kn,1} & x_{kn,2} & x_{kn,3} & \dots & x_{kn,k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \quad (3.9)$$

and $\mathbf{\Omega}$ to be diagonal matrix \mathbf{n} by \mathbf{k} to represent the uncertainty in each specific view by (\mathbf{i}, \mathbf{j}) entry of the matrix - i -th view about j -th factor.

$$\mathbf{\Omega} = \begin{bmatrix} w_{11} & & & & \\ & w_{12} & & & \\ & & \ddots & & \\ & & & w_{nk-1} & \\ & & & & w_{nk} \end{bmatrix} \quad (3.10)$$

To make computationally efficient, we make $\forall \mathbf{j}, \mathbf{i} \in (1, n)$, in another word, each factors we have exactly n views, this generalization however works for any j such that $\mathbf{k}_{j_m} \neq \mathbf{k}_{j_n}$ where $m \neq n$

3.4 Factor Return Forecasting methods

A core aspect of bayesian method or equivalently Black Litterman framework in our case is only as good as its input, which in the cases are really the views or the forecast of the factor returns namely $\mathbf{q}_{i,t}$ we will mainly address three broad aspects of forecasting toolbox, namely simple MA filter, traditional statistical time series models as well as a particular machine learning technique named Recurrent Neural Networks

3.4.1 Exponential Moving Average

Without going into too much details (it's abundant in literature), we have three smoothed window with 5-days, 30-days and 120-days of half-life

3.4.2 Traditional Time Series

- Autoregressive integrated moving average (ARIMA)
- Holt Winter
- Vectorized AutoRegression

All the above techniques, we applied an auto model and parameter selection that embedded in R or Python that based on either AIC or BIC criteria to select the best model to do the one day ahead forecast based on looking-back window data

3.4.3 Machine learning with Recurrent Neural Network

Recurrent Neural Network handles the information loss that encountered by traditional neural network to allow information to persist (It's unclear how a traditional neural network could use its reasoning about previous events to inform later ones)

- Recurrent Neural Network (RNN)
 - networks with loops
 - can be thought as multiple networks and each passing a message to successor
 - ability to connect previous information to the present task
- Long Short Term Memory networks (LSTM Networks)
 - a special kind of RNN
 - aimed to solve long term dependences as the gap between past and present widen where RNN might underperform
- Gated Recurrent Unit (GRU)
 - a more dramatic variation - simpler than LSTM
 - combines partial loss information into update, growing popularity

From the existing studies, it's unclear which version of RNNs works best.

4 Main Results

4.1 IR of portflios

Aside from the above-mentioned nine ways forecasting factor return, we also applied our generalized framework defined in 3.8 to combine some different return forecast mechanisms.

our IR (Note: here we directly use daily PnL as proxy of \mathbf{r}) is defined as:

$$\frac{\mathbb{E}(\mathbf{r})}{\mathbb{V}(\mathbf{r})} * \text{Annualized Factor} \quad (4.1)$$

Period from 2006.05-2015	Information Ratio
ARIMA	1.65251310236
GRU	1.41910485371
Holt Winters	1.92422234646
LSTM	1.56851696395
MA5	1.893804042
Markowitz	1.74641973298
Simple RNN	1.5712844355
Simple RNN and MA5	1.90596235524
VAR	1.00522552586
Multi-EMA	1.65257029739

Below is the cumulative PnL of the abover portfolios:

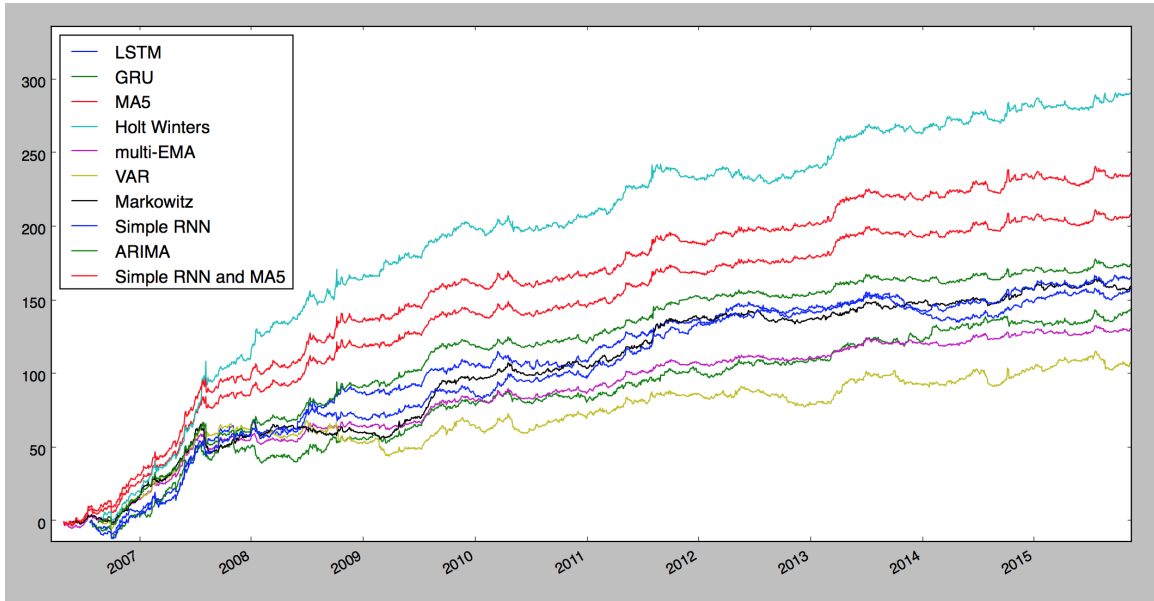


Figure 1: Cumulative PnL

4.2 Recap of proposed Generalized Framework

Remark: Given the time constraints, we applied generalized framework defined in (3.8) to the following:

- combine different Exponential Moving Average windows
- combine RNN with Exponential Moving Average short window

In practice, the generalized framework could applied any number of views, even in the case where number of views for each factor returns are different.

5 Summary and conclusion

In general, among of all the different strategies we applied in this study, all of them would be able to achieve an $IR > 1$. With the worst performer being VAR and Top performer to be Holt Winter method. It is very surprising that the simplest short window moving average filter works really well which corresponds to the second best single alone strategy among the all. By blending multi views, we managed to achieved a better risk-adjusted performance as indicated by MA5 combined with RNN achieved a better IR than both RNN alone and MA5 alone, but the outperformance over MA5 is marginal. However, this blending method applied on multiEMA doesnt really beat MA5. Even more surprisingly, without any bayesian inference, the stand-alone historical average approach seems to be working well with IR of 1.75. This seems daunting as it doesnt coincide with the result that presented by Gordon in their earlier paper, However, one needs to keep in mind that the dataset we applied is a shorter version of original study. we didnt explicitly assume a factor return and covariance, so the initial looking-back windows will impose a big impact on the performance of underlying Markowitz portfolio performance. In general, the performance of RNNs' family doesnt diff much which is relatively stable. As opposed to machine learning, the traditional time series have a divergent performance difference with ARIMA being mediocre. On the other hand, it entails us how sensitive the bayesian model would be to its posterior.

Acknowledgments

We are very grateful to Professor Gordon Ritter

References

- [1] Kolm, Petter N. and Ritter, Gordon *On the Bayesian Interpretation of Black-Litterman*. European Journal of Operational Research, Forthcoming, 2016.
 - [2] Guangliang He and Robert Litterman *The Intuition behind Black-Litterman Model Portfolios*. Goldman Sachs Asset Management (2006).
 - [3] Satchell, Stephen, and Alan Scowcroft *A demystification of the Black-Litterman model: Managing quantitative and traditional portfolio construction*. Journal of Asset Management 1.2, pp. 138-150 (2000).
- @misc RNN, author=Christopher, Olah,title = Understanding LSTM Networks, howpublished="http://colah.github.io/posts/2015-08-Understanding-LSTMs/", year = 2015, note="[Online; accessed 20-Dec-2016]"