

Gestão Ativa de Carteiras: Treynor-Black e Black-Litterman

Cláudio R. Lucinda
FEA/RP-USP

November 20, 2011

1 Introdução

Neste artigo, iremos discutir os modelos teóricos mais utilizados para a incorporação de estratégias ativas – entendidas como pontos de vista sobre determinados setores da economia e/ou ativos – dentro da estrutura tradicional de fronteira eficiente. Tais modelos são os de Treynor e Black (1973) e Black e Litterman (1991).

2 Treynor-Black

Estes autores propõem um modelo para construir um modelo quando os analistas são capazes de prever retornos anormais em um pequeno número de ações. Neste caso, a carteira ótima é construída por meio da combinação de um portfólio passivo – também chamado de *benchmark* – com uma carteira composta pelas ações cobertas pelos analistas. Evidentemente, a capacidade do modelo em entregar retornos excedentes depende fundamentalmente da capacidade de se prever retornos anormais. A sua implementação depende de as previsões dos analistas serem sujeitas a análises estatísticas e as propriedades das previsões sejam usadas no processo de otimização de carteira. Ou seja, as previsões dos analistas precisam ser quantificáveis e que sejam sujeitas a testes rigorosos de desempenho individual – o que vale também para a carteira toda.

O ponto de partida é uma premissa sobre o processo gerador dos retornos excedentes – que aqui assume-se ser o bom e velho CAPM:

$$(r_{it} - r_{ft}) = \beta_i(r_{Mt} - r_{ft}) + z_{it}$$

Neste caso, o termo z_{it} seria o retorno “anormal” da ação, e o α do ativo seria a esperança deste z_{it} , ou seja $E(z_{it}) = \alpha_i$, e podemos empilhar isso em um vetor coluna α . Podemos entender este negócio como uma medida de quanto o retorno excedente da ação pode se desviar na ausência de informações de analistas sobre a empresa.

Aqui vamos assumir que este retorno “anormal” seja correlacionado entre os diferentes ativos, ou seja, a matriz de variância-covariância entre os dois ativos pode ser dada por Ω , e é $N \times N$, sendo N o número de ativos analisados.

Para otimizar o retorno deste negócio vamos começar analisando o portfólio ótimo somente destas N ações, que vamos chamar de “carteira ativa”, representada por A . Os autores mostram que a carteira ótima neste caso passa por escolher um vetor $\mathbf{h} = (h_1, \dots, h_N)$, que maximiza a *razão de informação* da carteira, ou seja:

$$\max_{\mathbf{h}} \frac{\mathbf{h}^T \alpha}{\sqrt{\mathbf{h}^T \Omega \mathbf{h}}} \\ \text{sujeito a} \quad \mathbf{h}^T \mathbf{1} = 1$$

A carteira ótima para este problema é dada por:

$$\mathbf{h}^* = [\alpha^T \Omega^{-1} \mathbf{1}]^{-1} \Omega^{-1} \alpha$$

A taxa de retorno do portfólio ativo seria, neste caso, dada por $(r_A - r_f) = \sum_i h_i (r_i - r_f)$. O passo seguinte é, então, misturar de forma ótima esta carteira com uma posição passiva no índice. Neste caso, o retorno fica sendo $r_p = w_A r_A + (1 - w_A) r_M$. Vamos encontrar esta combinação pela maximização do Índice de Sharpe da carteira combinada:

$$\max_w S_p = \frac{E(r_p - r_f)}{\sqrt{\text{var}(r_p)}}$$

A solução deste problema gera o seguinte peso ótimo para a carteira ativa:

$$w_A^* = \frac{\alpha_A \sigma_m^2}{(1 - \beta_A) \alpha_A \sigma_m^2 + \mu_m \mathbf{h}^{*T} \Omega^{-1} \mathbf{h}^*}$$

Em que:

- $\alpha_A = \sum_i h_i^* \alpha_i$
- $\beta_A = \sum_i h_i^* \beta_i$
- $\mu_m = E(r_m)$
- $\sigma_m^2 = \text{var}(r_m)$

Podemos usar o índice de Sharpe da combinação ótima para determinar a contribuição dos analistas ao desempenho da carteira:

$$S(r_P^*)^2 = \mu_m \sigma_m^{-2} \mu_m + \alpha^T \Omega^{-1} \alpha$$

O primeiro termo nesta soma é a contribuição da carteira passiva ao desempenho, enquanto que o segundo termo é a contribuição da carteira ativa. Além disso, como mostra o BKM, podemos expressar também a variância do Tracking Error – a diferença entre os retornos da carteira ativa e da carteira passiva, como uma função dos parâmetros do modelo:

$$T_E = r_A - r_P$$

$$\begin{aligned}
r_P &= w_A^* \alpha_A + (1 - w_A^*(1 - \beta_A))r_M + w_A^* \epsilon_A \\
T_E &= w_A^* \alpha_A - w_A^*(1 - \beta_A)r_M + w_A^* \epsilon_A \\
\text{Var}(T_E) &= [w_A^*(1 - \beta_A)r_M] \text{Var}(r_M) + \text{Var}(w_A^* \epsilon_A) \\
&= [w_A^*(1 - \beta_A)r_M] \sigma_M^2 + [w_A \mathbf{h}^* \mathbf{T} \mathbf{\Omega}^{-1} \mathbf{h}^*]^2 \\
\sigma(T_E) &= w_A \sqrt{[(1 - \beta_A)r_M] \sigma_M^2 + [\mathbf{h}^* \mathbf{T} \mathbf{\Omega}^{-1} \mathbf{h}^*]^2}
\end{aligned}$$

O importante aqui é obter boas estimativas de α . A melhor forma é fazendo uma regressão dos alphas obtidos no passado em relação aos previstos, e usar a relação estimada para descontar as novas previsões.

3 Black-Litterman

O modelo Black-Litterman permite que os investidores combinem seus pontos de vista sobre os desempenhos dos diferentes ativos com o equilíbrio de mercado de uma forma que resulta em carteiras que são intuitivas e diversificadas. Além disso, alguns insights ajudam a encontrar carteiras mais diversificadas do que as que encontraríamos caso utilizássemos somente dados históricos, bem como sensibilidade aos inputs do modelo.

O modelo utiliza uma abordagem Bayesiana para combinar os pontos de vista subjetivos de um investidor acerca dos retornos esperados de um ou mais ativos com o vetor de equilíbrio de mercado de retornos esperados (que seria a prior) para, posteriormente, encontrar uma nova estimativa combinada de retornos esperados. O vetor resultante de retornos (o posterior), leva a carteiras intuitivas com pesos razoáveis de carteira.

O ponto de partida do modelo Black-Litterman são os retornos de “equilíbrio”, que não necessariamente são os mesmos que observamos com dados históricos. Estes retornos de equilíbrio saem de uma função utilidade média-variância quadrática:

$$U = \mathbf{w}^T \mathbf{\Pi} - \left(\frac{\lambda}{2}\right) \mathbf{w}^T \mathbf{\Omega} \mathbf{w}$$

Em que \mathbf{w} é um vetor de pesos dos diferentes ativos, $\mathbf{\Pi}$ o vetor de retornos esperados de equilíbrio, e $\mathbf{\Omega}$ a matriz de variância-covariância dos retornos dos ativos. Tirando a derivada desta função com relação a \mathbf{w} , temos:

$$\frac{\partial U}{\partial \mathbf{w}} = \mathbf{\Pi} - \lambda \mathbf{\Omega} \mathbf{w} = 0$$

Reorganizando, temos as estimativas de retornos de equilíbrio:

$$\mathbf{\Pi} = \lambda \mathbf{\Omega} \mathbf{w}$$

Neste caso o termo λ , uma medida do peso relativo de retorno esperado versus variância, serve como um fator de normalização – mais retorno esperado por unidade de risco (maior λ), maiores retornos esperados de equilíbrio. Usualmente se utiliza para calibrar o λ um índice de Sharpe médio¹. A partir desta equação, podemos derivar um

¹Por exemplo, multiplicando os dois lados por \mathbf{w}^T , temos que $E(r_p) - r_f = \lambda \sigma^2$, que é uma aproximação do Índice de Sharpe.

vetor de pesos dos ativos ótimo para servir de ponto de partida:

$$\mathbf{w}^* = (\lambda\Omega)^{-1}\Pi$$

Agora, para definirmos uma distribuição prior, precisamos de uma matriz de variância-covariância dos retornos de equilíbrio. Black e Litterman fazem a premissa que esta variância, Ω_Π , é proporcional à matriz Ω original, ou seja, $\Omega_\Pi = \tau\Omega^2$. Essas duas coisas formam a distribuição “prior” dos retornos $P(A) \sim \mathbb{N}(\Pi, \Omega_\Pi) = \mathbb{N}(\Pi, \tau\Omega)$. O passo seguinte é quantificar os pontos de vista que temos sobre os diferentes ativos, para combinarmos com este prior.

3.1 “Pontos de Vista” sobre os diferentes ativos

Nesta seção, iremos detalhar melhor o que consideramos os “pontos de vista” sobre os retornos médios estimados anteriormente. De uma perspectiva Bayesiana, vamos descrever estes pontos de vista como sendo a distribuição condicional.

Em primeiro lugar, por construção iremos assumir que cada ponto de vista seja único e não correlacionado com outros pontos de vista, de forma a gerar uma distribuição condicional com matriz de variância-covariância com elementos fora da diagonal iguais a zero. A principal razão para isso é para dar estabilidade numérica aos resultados obtidos anteriormente. Vamos imaginar que tenhamos k pontos de vista sobre os N ativos, a partir das seguintes matrizes:

- P , uma matriz $k \times N$ de pontos de vista sobre os diferentes ativos. Para um ponto de vista relativo, a soma dos pesos é zero, para uma visão absoluta a soma dos pesos é 1.
- Q uma matriz $k \times 1$ de retornos associados com cada ponto de vista.
- Σ , uma matriz $k \times k$ de covariância entre os diferentes pontos de vista. Como dissemos anteriormente, esta é uma matriz diagonal, e Σ^{-1} pode ser representada pelo grau de confiança destes pontos de vista.

Para entender como podemos traduzir diferentes pontos de vista em elementos destas matrizes, vamos pegar um exemplo. Considere que existam quatro ativos e dois pontos de vista. O primeiro deles é um ponto de vista relativo, em que o ativo 1 vai superar o ativo 3 em 2% com um grau de confiança σ_1 , e um ponto de vista absoluto em que o investidor considera que o ativo 2 vai gerar um retorno de 3% com um grau de confiança de σ_2 . Note que o investidor não tem nenhuma opinião formada sobre o ativo 4, significando que o seu retorno não deve ser ajustado. Esses pontos de vista implicam a seguinte configuração para as matrizes:

$$P = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}; Q = \begin{bmatrix} 2 \\ 3 \end{bmatrix}; \Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

²Usualmente existem duas formas de calibrar este parâmetro, a primeira seria $\tau = 1/T$, e a segunda seria $\tau = 1/T-k$. A primeira alternativa é a chamada “estimativa de máxima verossimilhança” e a segunda “melhor estimativa em sentido quadrático”, sendo T o tamanho da amostra e k o número de ativos. Se estivermos usando 5 anos de dados mensais, a primeira abordagem implica $\tau = 1/60$.

Podemos representar esta distribuição condicional em termos de retornos como sendo $P(B|A) \sim \mathbb{N}(P^{-1}Q, [P^T \Sigma^{-1} P]^{-1})$. Evidentemente, nem sempre é possível inverter P , mas dificilmente vamos precisar disso mais adiante. Mas algo mais difícil é determinar os termos da matriz Σ . Existem alguns mecanismos para isso:

- Proporcional à variância do prior
- Usando um intervalo de confiança
- Usando a variância dos resíduos em um modelo de fatores

Sobre o primeiro dos pontos, podemos assumir que a variância dos pontos de vista é proporcional à variância dos retornos dos ativos, como fizemos com o prior. Neste caso, determinaríamos outro parâmetro de calibragem, τ_2 , e $\Sigma = \tau_2 \Omega$, além disso, se $\tau_2 = \tau$, a solução final vai ficar independente do parâmetro, o que facilita bastante as coisas

A segunda opção passa por estimar diretamente as variâncias. A forma mais fácil de se estimar isso é definindo um intervalo de confiança em torno do retorno médio estimado, por exemplo, o ativo 2 tem um retorno médio de 3%, com a expectativa de 68% entre o intervalo de 2 a 4%. Sabendo que 68% da distribuição normal está entre um desvio padrão, isto implica em $\sigma_2^2 = 0,01^2$.

A terceira forma é usar os elementos diagonais dos resíduos de um modelo de fator único. Podemos escrever esta matriz variância-covariância como sendo:

$$V = BV(F)B^T + V(\epsilon)$$

Em que a matriz B é os betas dos diferentes fatores e $V(F)$ a matriz de variância-Covariância dos retornos dos diferentes fatores, e $V(\epsilon)$ a variância-covariância dos resíduos dos ativos. Esta matriz $V(\epsilon)$ pode ser usada – pelo menos os termos diagonais – para calibrar Σ .

3.2 Combinando as duas coisas – Teoria de Bayes

Uma das grandes vantagens do modelo Black-Litterman é que permite que possamos combinar as duas distribuições anteriormente colocadas de uma forma estatisticamente correta, com a ajuda do Teorema de Bayes. A idéia aqui é encontrar a distribuição dos retornos condicional aos pontos de vista, dada a distribuição dos dados – a prior – e um ponto de vista condicional à esta distribuição dos dados. Pelo teorema de Bayes temos a seguinte fórmula:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Em que $P(A|B)$ é o que queremos saber – a distribuição dos dados condicional aos pontos de vista – e $P(B|A)$ e $P(A)$ já foram discutidos anteriormente. $P(B)$, a distribuição não condicional, não precisa ser explicitada e vai sumir nas constantes de integração da derivação que se segue.

Pode-se mostrar que a distribuição posterior que queremos pode ser expressa como:

$$P(A|B) \sim \mathbb{N}([(\tau \Omega)^{-1} \Pi + P^T \Sigma^{-1} Q] [(\tau \Omega)^{-1} + P^T \Sigma^{-1} P]^{-1}, ((\tau \Sigma)^{-1} + P^T \Sigma^{-1} P)^{-1})$$

Podemos representar a média e a variância posterior alternativamente como sendo:

$$\begin{aligned}\hat{\Pi} &= \Pi + \tau\Omega P^T[(P\tau\Omega P^T) + \Sigma]^{-1}[Q - P\Pi] \\ M &= ((\tau\Omega)^{-1} + P^T\Omega P)^{-1}\end{aligned}$$

Só note que M , a variância posterior, é a variância da média posterior $\hat{\Pi}$ em torno da média observada. Ou seja, é a incerteza na média posterior e não a variância dos retornos. Para calcularmos a variância dos retornos – se quisermos rodar um otimizador de carteiras, por exemplo – precisamos somar este M na variância observada dos retornos:

$$\hat{\Sigma}_p = \Sigma + M$$