## 8. Advanced Black–Litterman Theory

8.1. **Generalizing the Model.** The observations in the previous section now allow us to easily formulate the most general model of this type.

*Definition* 8.1. A *Black-Litterman-Bayes* model consists of:

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

Items (a)-(b) simply state that we have a Bayesian statistical model, as defined above, for asset returns. Under such a model, Decision Theory (see Robert (2007, Ch. 2) and references) teaches us that the optimal decision is the one maximizing posterior expected utility. This leads us to Definition 8.2.

*Definition* 8.2. Given a Black-Litterman-Bayes (BLB) model as per Definition 8.1, the associated BLB optimal portfolio is defined to be

$$\boldsymbol{h}^* = \underset{\boldsymbol{h}}{\operatorname{argmax}}\, \mathbb{E}[u(\boldsymbol{h}'\boldsymbol{r}) \mid \boldsymbol{q}]$$

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

$$\begin{aligned} p(\boldsymbol{r} \mid \boldsymbol{q}) &= \int p(\boldsymbol{r} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \boldsymbol{q}) d\boldsymbol{\theta} \quad \text{where} \\ p(\boldsymbol{\theta} \mid \boldsymbol{q}) &= \frac{f(\boldsymbol{q} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{q} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta}} \end{aligned}$$

$\rightarrow$ density on parameter space.

*Definition* 8.3. Given a benchmark portfolio with holdings $\boldsymbol{h}_B$ (eg. the market portfolio), and given a Black-Litterman-Bayes model (Def. 8.1), the prior $\pi(\boldsymbol{\theta})$ is said to be *benchmark-optimal* if $\boldsymbol{h}_B$ maximizes expected utility of wealth, where the expectation is taken with respect to the *a priori* distribution on asset returns $p(\boldsymbol{r}) = \int p(\boldsymbol{r} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$, so

$$\boldsymbol{h}_B = \underset{\boldsymbol{h}}{\operatorname{argmax}} \int u(\boldsymbol{h}'\boldsymbol{r}) p(\boldsymbol{r} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{8.1}$$

Many existing approaches are special cases of the above. The model of Black and Litterman (1991) is the special case in which $\boldsymbol{r} \mid \boldsymbol{\theta}$ is multivariate normal with mean $\boldsymbol{\theta}$ and $f(\cdot \mid \cdot)$ is the normal likelihood function of the portfolio manager's views, the utility of final wealth is any increasing, concave utility function, and the prior is the unique normal distribution which is benchmark-optimal with respect to the market portfolio.

An interesting feature of the model is that there are *two* functions which both play the role of likelihood functions: $p(\boldsymbol{r} \,|\, \boldsymbol{\theta})$ and $f(\boldsymbol{q} \,|\, \boldsymbol{\theta})$. Equivalently, we have a triple of random vectors: $(\boldsymbol{r}, \boldsymbol{q}, \boldsymbol{\theta})$ which are not pairwise independent, but $\boldsymbol{r}$ and $\boldsymbol{q}$ are *conditionally independent* given $\boldsymbol{\theta}$. In Bayesian statistics, such situations are commonplace. A *Bayesian network* (or "graphical model") is, intuitively, an arbitrary collection of random variables whose conditional independence structure is specified by a (typically directed and acyclic) graph, so this system could be considered a Bayesian network with three nodes. We refer the reader to Pearl (2014) for the authoritative treatise on Bayesian networks, but suffice it to say that inference with much larger networks than the $(\boldsymbol{r}, \boldsymbol{q}, \boldsymbol{\theta})$ network is now commonplace.

More generally, $\boldsymbol{\theta}$ is allowed to be any set of parameters appearing in a parametric statistical model for asset returns, not necessarily their means. We explore this class of generalizations in the next sections.

8.2. **APT and Factor Models.** Generalizing further, the parameter vector $\boldsymbol{\theta}$ could represent means (and covariances) of unobservable latent factors in an APT model (Ross, 1976; Roll and Ross, 1980). Such models assume a linear functional form

$$\boldsymbol{r} = \mathbf{X}\boldsymbol{f} + \boldsymbol{\epsilon}, \quad \mathbb{E}[\boldsymbol{\epsilon}] = 0, \quad \mathbb{V}[\boldsymbol{\epsilon}] = \mathbf{D} \tag{8.2}$$

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

$$\mathbf{D} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2) \ \text{ with all } \ \sigma_i^2 > 0. \tag{8.3}$$

The variable $\boldsymbol{f}$ in (8.2) denotes a $k$-dimensional random vector process which cannot be observed directly; information about the $\boldsymbol{f}$-process must be obtained via statistical inference.

Specifically, we assume that the $\boldsymbol{f}$-process has finite first and second moments given by

$$\mathbb{E}[\boldsymbol{f}] = \boldsymbol{\mu}_f, \quad \text{and} \quad \mathbb{V}[\boldsymbol{f}] = \boldsymbol{F}. \tag{8.4}$$

若有factor的factor Return持续为正，则可能是这是一个未发现的risk、有一些人想get rid of、有些人想take这个risk。

When necessary, we will use $\boldsymbol{f}_t$ to denote a realization of the $\boldsymbol{f}$-process on day $t$, but we will usually suppress the implicit time subscript.

The model (8.2), (8.3) and (8.4) entails associated reductions of the first and second moments of the asset returns:

$$\mathbb{E}[\boldsymbol{r}] = \mathbf{X}\boldsymbol{\mu}_f, \quad \text{and} \quad \boldsymbol{\Sigma} := \mathbb{V}[\boldsymbol{r}] = \mathbf{D} + \mathbf{X}\boldsymbol{F}\mathbf{X}' \tag{8.5}$$

where $\mathbf{X}'$ denotes the transpose of $\mathbf{X}$. The elements of $\boldsymbol{\mu}_f$ are called *factor risk premia*. We will continue to use $\boldsymbol{\Sigma}$ to denote $\mathbf{D} + \mathbf{X}\boldsymbol{F}\mathbf{X}'$ throughout this section, and (8.3) implies that $\boldsymbol{\Sigma}^{-1}$ exists.

For simplicity, we treat $\mathbf{X}$ as non-stochastic and assume $k \ll n$. Then (8.5) reduces the number of parameters necessary to describe the density $p(\boldsymbol{r})$ from $O(n^2)$ down to the $k$ parameters in $\boldsymbol{\mu}_f$, the $k(k+1)/2$ parameters in $\boldsymbol{F}$, and $n$ parameters in $\mathbf{D}$, for a total of $n + k(k+3)/2$. Models of the form (8.2) are ubiquitous in practice, and for good reason: in equity markets $n$ is too large to allow direct estimation of $\boldsymbol{\Sigma}$. See Fabozzi, Focardi, and Kolm (2010) and Connor, Goldberg, and Korajczyk (2010) for more discussion and examples.

In the language of Def. 8.1, we are free to choose $\boldsymbol{\theta}$ as any vector of parameters appearing in a parametric statistical model for asset returns; (8.2)-(8.4) is such a model, so as a starting point, choose $\boldsymbol{\theta} = \boldsymbol{\mu}_f$, the $k$ parameters describing the factor risk premia. For simplicity we treat $\boldsymbol{F}$ as a constant matrix, just as the original Black-Litterman model treats $\boldsymbol{\Sigma}$ as a constant matrix.

What kinds of views on factor risk premia do we expect portfolio managers to have? The simplest and most parsimonious scenario is that we have a view on each factor risk premium that is independent of our views on other factors. For example, consider value and momentum, as discussed at length by Asness, Moskowitz, and Pedersen (2013), and Fabozzi, Focardi, and Kolm (2006) and Fabozzi, Focardi, and Kolm (2010) going back to work of Fama and French (1993) and Carhart (1997).

A quantitative portfolio manager might have two views: (1) a view on the value premium, and, separately from that, (2) a view on the momentum premium. It would be atypical for portfolio managers to have views on, say, the sum or difference of the value and momentum premia, or more generally on "portfolios of factors." Hence to keep things simple but still useful, we take the likelihood function to be

$$f(\boldsymbol{q} \,|\, \boldsymbol{\theta}) = \prod_{i=1}^{k} \exp[-\frac{1}{2\omega_i^2}(\theta_i - q_i)^2] \tag{8.6}$$

parameter of factor returns.

The choice of prior $\pi(\boldsymbol{\theta})$ is very interesting. We discuss 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 as in in Def. 8.3.

If the random process model driving the unobservable factor returns $\boldsymbol{f}_t$ is stationary, ie. $\boldsymbol{\mu}_f, \boldsymbol{F}$ are approximately constant over time, then we could obtain a prior for $\boldsymbol{\theta} = \boldsymbol{\mu}_f$ by taking the posterior from a simple Bayesian time-series model for the factor returns $\boldsymbol{f}_t$. In particular, the historical mean of the OLS estimates $\hat{\boldsymbol{f}}_{t+1} = (\mathbf{X}_t'\mathbf{X}_t)^{-1}\mathbf{X}_t'\boldsymbol{r}_{t+1}$ could be taken as the prior mean. More generally, this problem lends itself well to a hierarchical (or mixed-effects model) approach. Each time period is a "group" and one has models for $\boldsymbol{r}_{t+1} \sim N(\mathbf{X}_t\boldsymbol{f}_{t+1}, \mathbf{D})$ and the various $\boldsymbol{f}_{t+1}$ are modeled as i.i.d. draws $\boldsymbol{f}_{t+1} \sim N(\boldsymbol{\mu}_f, \boldsymbol{F})$. The statistical inference problem is then to infer $\boldsymbol{\theta} = \boldsymbol{\mu}_f$ from observations of $\boldsymbol{r}_t$, a special case of the hierarchical approach discussed in Gelman et al. (2003, Ch. 15). The posterior from this procedure is a possible prior for use in the Black-Litterman procedure.

The "data-driven" approach to prior selection that we have just described has the advantage of not requiring a benchmark portfolio. This makes sense for absolute return strategies where the effective benchmark is cash. It's very common in Bayesian statistics for the posterior from one analysis to become the prior for subsequent analysis.

Alternatively, if there is a benchmark portfolio $\boldsymbol{h}_B$, then closest in spirit to Black and Litterman (1991) would be to search for a benchmark-optimal prior, as defined above. To progress any further, we need to introduce notation for the hyperparameters in $\pi(\boldsymbol{\theta})$, so let's say $\pi(\boldsymbol{\theta}) \sim N(\boldsymbol{\xi}, \boldsymbol{V})$ with $\boldsymbol{\xi} \in \mathbb{R}^k$ and $\boldsymbol{V} \in S^k_{++}$, the set of symmetric positive definite $k \times k$ matrices. Choosing a prior then amounts to choosing $\boldsymbol{\xi}$ and $\boldsymbol{V}$, which are constrained by (8.1). The first step in evaluating (8.1) is to compute the a priori density on returns, $\int p(\boldsymbol{r} \,|\, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$. Since $\pi(\boldsymbol{\theta})$ and $p(\boldsymbol{r} \,|\, \boldsymbol{\theta})$ are both Gaussian, this is another completion of squares.

We continue to use the notation $\boldsymbol{\Sigma} = \mathbf{D} + \mathbf{X}\boldsymbol{F}\mathbf{X}'$ as above, since this is the asset-level covariance in an APT model. Straightforward calculations then show:

$$-2\log[p(\boldsymbol{r} \,|\, \boldsymbol{\theta})\pi(\boldsymbol{\theta})] = -2\log N(\boldsymbol{r};\ \mathbf{X}\boldsymbol{\theta}, \boldsymbol{\Sigma}) - 2\log N(\boldsymbol{\theta};\ \boldsymbol{\xi}, \boldsymbol{V})$$

$$\begin{aligned} &= (\boldsymbol{r} - \mathbf{X}\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{r} - \mathbf{X}\boldsymbol{\theta}) + (\boldsymbol{\theta} - \boldsymbol{\xi})'\boldsymbol{V}^{-1}(\boldsymbol{\theta} - \boldsymbol{\xi}) \\ &= \boldsymbol{\theta}'\boldsymbol{H}\boldsymbol{\theta} - 2\boldsymbol{\eta}'\boldsymbol{\theta} + \boldsymbol{z} \end{aligned}$$

where for notational simplicity, we have introduced the auxiliary variables

$$\boldsymbol{H} = \boldsymbol{V}^{-1} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X},\ \boldsymbol{\eta} = (\boldsymbol{\xi}'\boldsymbol{V}^{-1} + \boldsymbol{r}'\boldsymbol{\Sigma}^{-1}\mathbf{X})'$$

and

$$\boldsymbol{z} = \boldsymbol{r}'\boldsymbol{\Sigma}^{-1}\boldsymbol{r} + \boldsymbol{\xi}'\boldsymbol{V}^{-1}\boldsymbol{\xi}.$$

Completing the square again,

$$\boldsymbol{\theta}'\boldsymbol{H}\boldsymbol{\theta} - 2\boldsymbol{\eta}'\boldsymbol{\theta} + \boldsymbol{z} = (\boldsymbol{\theta} - \boldsymbol{v})'\boldsymbol{H}(\boldsymbol{\theta} - \boldsymbol{v}) - \boldsymbol{v}'\boldsymbol{H}\boldsymbol{v} + \boldsymbol{z}, \quad \boldsymbol{v} = \boldsymbol{H}^{-1}\boldsymbol{\eta}$$

The integral over $\boldsymbol{\theta}$ is then a standard Gaussian integral, which is performed via the formula

$$\int \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{v})'\boldsymbol{H}(\boldsymbol{\theta} - \boldsymbol{v})\right] d\boldsymbol{\theta} = \sqrt{\frac{(2\pi)^k}{\det \boldsymbol{H}}}$$

Hence,

$$\int p(\boldsymbol{r} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} = (2\pi)^{k/2}\,|\boldsymbol{H}|^{-1}\exp[-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\eta}'\boldsymbol{H}^{-1}\boldsymbol{\eta})]$$

$$= \frac{(2\pi)^{k/2}}{\det \boldsymbol{H}}\exp\Big[-\frac{1}{2}\Big\{\boldsymbol{r}'\boldsymbol{\Sigma}^{-1}\boldsymbol{r} + \boldsymbol{\xi}'\boldsymbol{V}^{-1}\boldsymbol{\xi} - (\boldsymbol{\xi}'\boldsymbol{V}^{-1} + \boldsymbol{r}'\boldsymbol{\Sigma}^{-1}\mathbf{X})\boldsymbol{H}^{-1}(\boldsymbol{V}^{-1}\boldsymbol{\xi} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{r})\Big\}\Big]$$

Let's multiply out the second quadratic term:

$$\begin{aligned}&(\boldsymbol{\xi}'\boldsymbol{V}^{-1} + \boldsymbol{r}'\boldsymbol{\Sigma}^{-1}\mathbf{X})\boldsymbol{H}^{-1}(\boldsymbol{V}^{-1}\boldsymbol{\xi} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{r})\\ &= \boldsymbol{\xi}'(\boldsymbol{V}\boldsymbol{H}\boldsymbol{V})^{-1}\boldsymbol{\xi} + 2\boldsymbol{\xi}'(\boldsymbol{H}\boldsymbol{V})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{r}\\ &\quad + \boldsymbol{r}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{H}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{r}\end{aligned}$$

Note that $\int p(\boldsymbol{r} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ is a Gaussian probability distribution for the random vector $\boldsymbol{r}$, so to find the covariance, we just collect the quadratic terms in $\boldsymbol{r}$ and take the inverse:

$$\mathbb{V}_\pi[\boldsymbol{r}] = (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{H}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1})^{-1}.$$

Completing the squares as before, the mean is

$$\begin{aligned}\mathbb{E}_\pi[\boldsymbol{r}] &= \mathbb{V}_\pi[\boldsymbol{r}]\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{H}^{-1}\boldsymbol{V}^{-1}\boldsymbol{\xi} \quad \text{tmp4}\\ &= (\boldsymbol{\Sigma}^{-1} + \underbrace{\boxed{\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{H}^{-1}}}_{\text{tmp1}}\underbrace{\boxed{\mathbf{X}'\boldsymbol{\Sigma}^{-1}}}_{\text{tmp2}})^{-1}\overbrace{\underbrace{\boxed{\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{H}^{-1}}}_{\text{tmp1}}\underbrace{\boxed{\boldsymbol{V}^{-1}\boldsymbol{\xi}}}_{\text{tmp3}}}^{\text{tmp5}}\end{aligned}$$

These models are once again based on elliptical distributions and satisfy mean-variance equivalence for any utility function. The *a priori* optimal portfolio is then

$$(\kappa\mathbb{V}_\pi[\boldsymbol{r}])^{-1}\mathbb{E}_\pi[\boldsymbol{r}] = \kappa^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{H}^{-1}\boldsymbol{V}^{-1}\boldsymbol{\xi} \tag{8.7}$$

but unlike the classic Black-Litterman case, it is no longer true that any arbitrary benchmark portfolio can be realized as an *a priori* optimal portfolio. In fact, (8.7) gives a very simple characterization of those that can: they are precisely of the form $\kappa^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi}$ where $\boldsymbol{\Pi}$ is some linear combination of the columns of $\mathbf{X}$. That is to say, they are portfolios which are optimal with respect to a set of individual asset risk premia that come from the factor model. From the standpoint of APT, this is not a real restriction; if the original APT model isn't mis-specified, then residuals should be independent, and not additional sources of risk premia for use in forming expected returns.

Not every possible portfolio is realizable as *a priori* optimal, hence the market portfolio may not be. However, at least we can say that if the market is in a CAPM equilibrium and if one of the columns of $\mathbf{X}$ contains the CAPM betas, then the individual asset risk premia will be proportional to that column of $\mathbf{X}$, and then the market portfolio *will* be realizable as *a priori* optimal, as per (8.7).

We now leave behind the question of the prior and continue with calculating the *a posteriori* optimal portfolio, i.e. the portfolio which takes into account the views (8.6) on the factor risk premia. This calculation proceeds in three steps:

1. calculate the posterior distribution of $\boldsymbol{\theta}$, after the views are taken into account, which is given by
$$p(\boldsymbol{\theta} \,|\, \boldsymbol{q}) = \frac{f(\boldsymbol{q} \,|\, \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{q} \,|\, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$
2. calculate the *a posteriori* distribution of asset returns (also called the posterior predictive density), given by
$$p(\boldsymbol{r} \,|\, \boldsymbol{q}) = \int p(\boldsymbol{r} \,|\, \boldsymbol{\theta})p(\boldsymbol{\theta} \,|\, \boldsymbol{q})d\boldsymbol{\theta} \tag{8.8}$$
3. calculate the mean-variance optimal portfolio under (8.8).

Fortunately, Step 1 is easy since the normal prior is a *conjugate prior* for the normal likelihood, meaning that the posterior distribution is of the same distributional family as the prior (again normal), but with different values for the hyperparameters. By a straightforward calculation, if the prior is normal with hyperparameters $\boldsymbol{\xi}, \boldsymbol{V}$ then the posterior has hyperparameters $\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{V}}$ where

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

Step 2 follows via the same calculation we did to find the *a priori* density, but using the posterior updated values $\tilde{\boldsymbol{V}}$ and $\tilde{\boldsymbol{\xi}}$ for the hyperparameters. We don't need to do the whole calculation again, just make the substitution $\boldsymbol{\xi} \to \tilde{\boldsymbol{\xi}}$ and $\boldsymbol{V} \to \tilde{\boldsymbol{V}}$ to find

$$\begin{aligned} \mathbb{V}[\boldsymbol{r} \,|\, \boldsymbol{q}] &= (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\mathbf{X}(\tilde{\boldsymbol{V}}^{-1} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1})^{-1}. \\ \mathbb{E}[\boldsymbol{r} \,|\, \boldsymbol{q}] &= \mathbb{V}[\boldsymbol{r} \,|\, \boldsymbol{q}]\boldsymbol{\Sigma}^{-1}\mathbf{X}(\tilde{\boldsymbol{V}}^{-1} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\tilde{\boldsymbol{V}}^{-1}\tilde{\boldsymbol{\xi}} \end{aligned} \tag{8.9}$$

Step 3 is then a completely standard calculation of a mean-variance optimal portfolio from (8.9):

$$\begin{aligned} \boldsymbol{h}^* &= \kappa^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi} \qquad (8.10) \\ \boldsymbol{\Pi} &:= \mathbf{X}\tilde{\boldsymbol{\mu}}_f \\ \tilde{\boldsymbol{\mu}}_f &:= (\boldsymbol{V}^{-1} + \boldsymbol{\Omega}^{-1} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\boldsymbol{V}^{-1}\boldsymbol{\xi} + \boldsymbol{\Omega}^{-1}\boldsymbol{q}) \end{aligned}$$

Eqns. (8.10) are due to Kolm and Ritter (2016) and they represent the solution to Black-Litterman optimization in the context of APT. They are written in a suggestive form: the asset-level risk premia $\boldsymbol{\Pi} = \mathbf{X}\tilde{\boldsymbol{\mu}}_f$ are linear combinations of the factors which form the columns of $\mathbf{X}$. One can think of $\tilde{\boldsymbol{\mu}}_f$ as a set of factor risk premia "adjusted" to take account of the views, and the adjustments tend to give more weight to factors which have high prior mean-variance ratios $\xi_i/V_{ii}$ and/or high expected return-uncertainty ratios $Q_i/\omega_i^2$.

## References


Arrow, Kenneth J (1971). "Essays in the theory of risk-bearing". In:

Asness, Clifford S, Tobias J Moskowitz, and Lasse Heje Pedersen (2013). "Value and momentum everywhere". In: *The Journal of Finance* 68.3, pp. 929–985.

Black, Fischer and Robert B Litterman (1991). "Asset allocation: combining investor views with market equilibrium". In: *The Journal of Fixed Income* 1.2, pp. 7–18.

Carhart, Mark M (1997). "On persistence in mutual fund performance". In: *The Journal of Finance* 52.1, pp. 57–82.

Connor, Gregory, Lisa R Goldberg, and Robert A Korajczyk (2010). *Portfolio risk analysis*. Princeton University Press.

Fabozzi, Frank J, Sergio M Focardi, and Petter N Kolm (2006). "Incorporating trading strategies in the Black-Litterman framework". In: *The Journal of Trading* 1.2, pp. 28–37.

– (2010). *Quantitative equity investing: Techniques and Strategies*. John Wiley & Sons.

Fama, Eugene F and Kenneth R French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of financial economics* 33.1, pp. 3–56.

Gelman, Andrew et al. (2003). *Bayesian data analysis, 2nd ed.* Taylor & Francis.

Kolm, Petter and Gordon Ritter (2016). "On the Bayesian interpretation of Black–Litterman". In: *European Journal of Operational Research*.

Pearl, Judea (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Pratt, John W (1964). "Risk aversion in the small and in the large". In: *Econometrica: Journal of the Econometric Society*, pp. 122–136.

Robert, Christian (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.

Roll, Richard and Stephen A Ross (1980). "An empirical investigation of the arbitrage pricing theory". In: *The Journal of Finance* 35.5, pp. 1073–1103.

Ross, Stephen A (1976). "The arbitrage theory of capital asset pricing". In: *Journal of economic theory* 13.3, pp. 341–360.