9. Advanced Black–Litterman Theory

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

Definition 9.1. 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 9.2.

Definition 9.2. Given a Black-Litterman-Bayes (BLB) model as per Definition 9.1, 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

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

Definition 9.3. Given a benchmark portfolio with holdings $\boldsymbol{h}_B$ (eg. the market portfolio), and given a Black-Litterman-Bayes model (Def. 9.1), 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

$$\boldsymbol{h}_B = \underset{\boldsymbol{h}}{\operatorname{argmax}} \int u(\boldsymbol{h}'\boldsymbol{r}) p(\boldsymbol{r} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{9.1}$$

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: $p(\boldsymbol{r} \mid \boldsymbol{\theta})$ and $f(\boldsymbol{q} \mid \boldsymbol{\theta})$. 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.

Even if $\boldsymbol{\theta}$ simply represents the mean vector of asset returns, such returns are widely recognized to be non-normal. Replacing (**??**) with a Laplace distribution may fit empirical asset returns more accurately. This corresponds to Lasso regression, a special case of Bayesian regression, in the same sense that the original Black–Litterman model is similar to ridge regression. Giacometti et al. (2007) also investigated heavy-tailed distributions in the context of Black–Litterman optimization.

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.

9.2. **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

$$\boldsymbol{r} = \mathbf{X}\boldsymbol{f} + \boldsymbol{\epsilon}, \quad \mathbb{E}[\boldsymbol{\epsilon}] = 0, \quad \mathbb{V}[\boldsymbol{\epsilon}] = \mathbf{D} \tag{9.2}$$

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

$$\mathbf{D} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2) \;\text{ with all }\; \sigma_i^2 > 0. \tag{9.3}$$

The variable $\boldsymbol{f}$ in (9.2) 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

$$\mathbb{E}[\boldsymbol{f}] = \boldsymbol{\mu}_f, \quad \text{and} \quad \mathbb{V}[\boldsymbol{f}] = \boldsymbol{F}. \tag{9.4}$$

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 (9.2), (9.3) and (9.4) entails associated reductions of the first and second moments of the asset returns:

$$\mathbb{E}[\boldsymbol{r}] = \mathbf{X}\boldsymbol{\mu}_f, \quad \text{and} \quad \boldsymbol{\Sigma} := \mathbb{V}[\boldsymbol{r}] = \mathbf{D} + \mathbf{X}\boldsymbol{F}\mathbf{X}' \tag{9.5}$$

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 (9.3) implies that $\boldsymbol{\Sigma}^{-1}$ exists.

For simplicity, we treat $\mathbf{X}$ as non-stochastic and assume $k \ll n$. Then (9.5) 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 (9.2) 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. 9.1, we are free to choose $\boldsymbol{\theta}$ as any vector of parameters appearing in a parametric statistical model for asset returns; (9.2)-(9.4) 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

$$f(\boldsymbol{q} \,|\, \boldsymbol{\theta}) = \prod_{i=1}^{k} \exp[-\frac{1}{2\omega_i^2}(\theta_i - q_i)^2] \tag{9.6}$$

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. 9.3.

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 (9.1). The first step in evaluating (9.1) is to compute the a priori density on returns, $\int p(\boldsymbol{r} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$. Since $\pi(\boldsymbol{\theta})$ and $p(\boldsymbol{r} \mid \boldsymbol{\theta})$ 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:

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

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

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,

$$\begin{aligned}
&\int p(\boldsymbol{r} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} = (2\pi)^{k/2} \left|\boldsymbol{H}\right|^{-1} \exp[-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\eta}'\boldsymbol{H}^{-1}\boldsymbol{\eta})] \\
&= \frac{(2\pi)^{k/2}}{\det \boldsymbol{H}} \exp\Big[-\frac{1}{2}\Big\{\boldsymbol{r}'\boldsymbol{\Sigma}^{-1}\boldsymbol{r} + \boldsymbol{\xi}'\boldsymbol{V}^{-1}\boldsymbol{\xi} \\
&\qquad - (\boldsymbol{\xi}'\boldsymbol{V}^{-1} + \boldsymbol{r}'\boldsymbol{\Sigma}^{-1}\mathbf{X})\boldsymbol{H}^{-1}(\boldsymbol{V}^{-1}\boldsymbol{\xi} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{r})\Big\}\Big]
\end{aligned}$$

Let's multiply out the second quadratic term:

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

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

$$\begin{aligned}
\mathbb{E}_\pi[\boldsymbol{r}] &= \mathbb{V}_\pi[\boldsymbol{r}]\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{H}^{-1}\boldsymbol{V}^{-1}\boldsymbol{\xi} \\
&= (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{H}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1})^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{H}^{-1}\boldsymbol{V}^{-1}\boldsymbol{\xi}
\end{aligned}$$

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

$$(\kappa\mathbb{V}_\pi[\boldsymbol{r}])^{-1}\mathbb{E}_\pi[\boldsymbol{r}] = \kappa^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{H}^{-1}\boldsymbol{V}^{-1}\boldsymbol{\xi} \tag{9.7}$$

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, (9.7) 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 (9.7).

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 (9.6) 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
$$p(\boldsymbol{r} \,|\, \boldsymbol{q}) = \int p(\boldsymbol{r} \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, \boldsymbol{q}) d\boldsymbol{\theta} \tag{9.8}$$
3. calculate the mean-variance optimal portfolio under (9.8).

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

$$\begin{aligned} \mathbb{V}[\boldsymbol{r} \,|\, \boldsymbol{q}] &= (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\mathbf{X}(\tilde{\boldsymbol{V}}^{-1} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1})^{-1}. \\ \mathbb{E}[\boldsymbol{r} \,|\, \boldsymbol{q}] &= \mathbb{V}[\boldsymbol{r} \,|\, \boldsymbol{q}]\boldsymbol{\Sigma}^{-1}\mathbf{X}(\tilde{\boldsymbol{V}}^{-1} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\tilde{\boldsymbol{V}}^{-1}\tilde{\boldsymbol{\xi}} \end{aligned} \tag{9.9}$$

Step 3 is then a completely standard calculation of a mean-variance optimal portfolio from (9.9):

$$\begin{aligned} \boldsymbol{h}^* &= \kappa^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi} \\ \boldsymbol{\Pi} &:= \mathbf{X}\tilde{\boldsymbol{\mu}}_f \\ \tilde{\boldsymbol{\mu}}_f &:= (\boldsymbol{V}^{-1} + \boldsymbol{\Omega}^{-1} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\boldsymbol{V}^{-1}\boldsymbol{\xi} + \boldsymbol{\Omega}^{-1}\boldsymbol{q}) \end{aligned} \tag{9.10}$$

Eqns. (9.10) 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 ξ_i/V_{ii} and/or high expected return-uncertainty ratios Q_i/ω_i^2.

9.3. **Multi-Period Views.** Consider an investor who has n-step ahead forecasts of asset return for all n up to some maximal horizon; this is called an *alpha term structure*. Typically, these forecasts are increasingly uncertain as n increases. In other words, the 10-step ahead forecast typically has larger "error bars around the prediction" than the 1-step ahead forecast. This is *not* a statement that the asset's return volatility is expected to increase in the future; rather, it reflects the fact that, usually, we can predict less accurately about the distant future than we can about the near future. Even for an asset with constant volatility the term structure of forecast error variance can be expected to be upward-sloping.

In this context, a trade that is executed more quickly incurs more market impact, but also enjoys the benefit of exposure to the early part of the alpha term structure where the forecast is more certain. Generally speaking, in the presence of trading costs, the fact that our uncertainty increases as we look further ahead must be balanced against the market impact costs associated to trading too quickly.

Suppose one wishes to use an n-period optimization to optimize utility of final wealth in n days. If a naive optimization method is employed, then profit made on the n-th day is viewed by the optimizer as equivalent to profit made on the first day – both increase final wealth. Hence the optimizer is willing to trade off a slightly worse portfolio on the first day for a modestly better portfolio on the n-th day. The naive approach is suboptimal because it ignores the fact that the n-step-ahead forecasts have much larger forecast error variance than the one-step-ahead forecasts. In the extreme version of this conundrum, one may extend the term structure far enough into the future that one knows almost nothing about the end of the term structure, yet the naive approach treats the end as if it were of equal importance to the beginning.

In a single-period setting, if trading costs are negligible, then methods for incorporating forecast uncertainty into portfolio construction are fairly well-known. In a pair of classic papers, Black and Litterman (1991) and Black and Litterman (1992) showed how to construct optimal portfolios with uncertain forecasts in a single-period setting, by introducing forecast uncertainty as an explicit numerical parameter. Mathematical details of the model were fleshed out in He and Litterman (1999) and extended by Fabozzi, Focardi, and Kolm (2006) and Kolm and Ritter (2016). In order to correctly balance the increasing uncertainty of information against the costs of trading too quickly, we need to generalize the classic work of Black and Litterman (1991) to a multi-period setting with realistic trading costs.

The inputs to the Black–Litterman model are views, which are expected returns on given portfolios, and uncertainties in those views. Supposing the current time is t, a set of k distinct views about the future n periods from now consists of a matrix P_{t+n} with k rows, a vector $q_{t+n} \in \mathbb{R}^k$, and a matrix Ω_{t+n} where

$$P_{t+n} r_{t+n} = q_{t+n} + \epsilon^{(v)}, \quad \epsilon^{(v)} \sim N(0, \Omega_{t+n}). \tag{9.11}$$

with r_{t+n} denoting the random vector of n-period-ahead asset returns. In practice, often a diagonal form

$$\Omega_{t+n} = \mathrm{diag}(\omega^2_{t+n,i} : i = 1, \ldots, k) \tag{9.12}$$

is used, indicating that the uncertainty term in each view is independent of the other views' uncertainty terms. It is perfectly reasonable that the uncertainty terms may be statistically independent, even when the underlying view portfolios are correlated. Often the portfolio P will be fixed as of time t, while the forecasts q_{t+n} and their standard errors ω_{t+n} will meaningfully depend on n.

We suppress the time subscripts when they aren't relevant. For any view, scaling P, q, and ω by the same constant gives an equivalent view containing the same information. Similarly, multiplying P and q both by -1 expresses an equivalent view. Hence all economically meaningful expressions, such as expressions for optimal portfolios, must be invariant under the same set of symmetries.

Note that (9.11) is well-suited to describe relative-value pricing relationships or approximate arbitrage relationships: if asset i is mis-priced relative to asset j, this means that one row of P has nonzero elements of the same magnitude and opposite sign in the i-th and j-th positions, and zero elsewhere.

As we attempt to look further into the future, our uncertainty will typically increase. This is especially true where asset return forecasts are concerned. If the result of a clinical trial will be released in exactly 10 days from now, we may be more certain that the volatility on that day will be high, but less certain of the direction of the return than we would be on a normal day. This leads us to a property that elements of Ω typically have along the time axis:

$$\omega_{t+n,i} < \omega_{t+n+k,i} \quad \text{for all } k > 0. \tag{9.13}$$

It is to be emphasized that that the $\omega_{t,i}$ are not asset return variances. There is no reason, in general, that an asset's return variance should increase in the future in the manner suggested by (9.13).

9.4. **Black-Litterman Optimization With Statistical Forecasts.** Assume, quite generally, that the returns on some asset or some portfolio of assets are contemporaneously related to some discrete-time stochastic process Z whose value at time t is z_t. The most straightforward application is simply that z_t is the return observed in period t on some fixed portfolio of assets, but at the highest level of generality, all one needs is some statistical model connecting z_t with returns on some known portfolios. We assume that the time-series z_t is amenable to forecasting via any structural time-series model such as those described in the classic textbooks by Hamilton or Tsay.

Any such structural time-series model (for example, a vector autoregression/-VAR) produces *n-step-ahead forecasts* for any desired integer $n > 0$. In other words, the model allows us to calculate the *predictive density* $p(z_{t+n} \,|\, z_t, z_{t-1}, \ldots, z_1)$. If we assume that z_t is a Markov process, then the predictive density does not depend on the full history, and can then be written more simply as $p(z_{t+n} \,|\, z_t)$. The predictive distribution is often summarized by its mean and 95% quantiles, as in the following well-known airline passenger model:

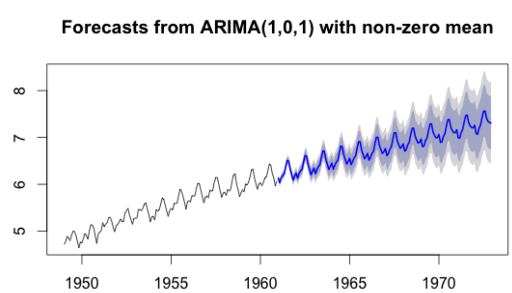

Figure 1: airline passenger data (black, log space) and ARIMA(1,0,1) n-step ahead predictive distributions $p(z_{t+n} \,|\, z_t)$, summarized by their moments and quantiles (blue).

Indeed, an ideal forecast $\hat{z}$ of z_{t+n} given the information set $\mathcal{F}_t$ should satisfy

$$\mathbb{E}[(z_{t+n} - \hat{z})^2 \,|\, \mathcal{F}_t] = \min_f \mathbb{E}[(z_{t+n} - f)^2 \,|\, \mathcal{F}_t]$$

where the minimization is over all functions of z_t. The solution is that the optimal forecast $\hat{z}$ is $\hat{z} = \mathbb{E}[z_{t+n} \,|\, z_t]$.

The *n-step ahead forecast error* is defined as

$$e_t(n) := z_{t+n} - \hat{z} \tag{9.14}$$

which is, of course, a random variable at time t since z_{t+n} is not known. The *forecast error variance* is then defined to be

$$\sigma^2_{t+n\,|\,t} := \mathbb{V}[e_t(n) \,|\, \mathcal{F}_t] \tag{9.15}$$

The truly fundamental object (in the sense that it contains all of the information we have about various periods in the future) is the collection of predictive densities:

$$\{p(z_{t+n} \,|\, z_t) : n = 1, 2, 3, \ldots\}$$

which we call *the term structure of information.*

As long as the variable being forecasted, z_t, is identifiable as the return on some portfolio of assets, then referring to (9.11), we may set a row of P to be this given portfolio, and then identify the Black-Litterman quantities q, ω in terms of forecasts produced by our statistical model:

$$q_{t+n} = \hat{z} = \mathbb{E}[z_{t+n} \,|\, z_t], \quad \omega^2_{t+n} = \mathbb{V}[e_t(n) \,|\, \mathcal{F}_t] = \sigma^2_{t+n\,|\,t}. \tag{9.16}$$

Eq. (9.16) is the basic relation connecting the uncertainty variances ω^2_{t+n}, which are inputs to the Black–Litterman procedure, to the output of multi-step-ahead statistical forecast procedures.

It is useful at this point to give a concrete class of examples. Suppose we have two assets with returns x_t, y_t in period t. Suppose for some constant β, that

$$z_t := y_t - \beta x_t$$

is the return on a portfolio which is long one unit of the asset whose return is y_t and short β units of the x_t asset. Define a matrix P whose one single row is the stationary portfolio:

$$P = [1 \quad -\beta] \tag{9.17}$$

Then by construction,

$$P r_{t+n} = y_{t+n} - \beta x_{t+n} \equiv z_{t+n}$$

We take the q_{t+n} and ω_{t+n} for all n to be given by (9.16), and assume for simplicity that z_t follows a mean-zero AR(1) model:

$$z_t - \phi z_{t-1} = a_t \text{ with } |\phi| < 1 \text{ and } a_t \sim N(0, \sigma_a^2).$$

Then, a standard calculation shows,

$$\hat{z}(t+n) = \phi^n z_t$$
$$\mathbb{V}[e_t(n)] = (1 + \phi^2 + \ldots + \phi^{2(n-1)})\sigma_a^2.$$

Note that as $n \to \infty$, we have $\hat{z}(t+n) \to 0$ which is the long-run mean, and moreover the variance of forecast error converges to $(1-\phi^2)^{-1}\sigma_a^2$, which is the variance of z_t. More generally, the long-term forecast is the marginal distribution of z_t.

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 Litterman (1992). "Global portfolio optimization". In: *Financial Analysts Journal*, pp. 28–43.

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.

Giacometti, Rosella et al. (2007). "Stable distributions in the Black–Litterman approach to asset allocation". In: *Quantitative Finance* 7.4, pp. 423–433.

He, Guangliang and Robert Litterman (1999). "The intuition behind Black-Litterman model portfolios". In: *Goldman Sachs Investment Management Series*.

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.