European Journal of Operational Research 258 (2017) 564–572

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

On the Bayesian interpretation of Black–Litterman

Petter Kolm[a], Gordon Ritter[a,b,*]

[a] Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, NY 10012, United States
[b] Rutgers University and Baruch College, United States

ARTICLE INFO

Article history:
Received 21 April 2016
Accepted 15 October 2016
Available online 20 October 2016

Keywords:
Finance
Investment analysis
Bayesian statistics
Black–Litterman
Portfolio optimization

ABSTRACT

We present the most general model of the type considered by Black and Litterman (1991) after fully clarifying the duality between Black–Litterman optimization and Bayesian regression. Our generalization is itself a special case of a Bayesian network or graphical model. As an example, we work out in full detail the treatment of views on factor risk premia in the context of APT. We also consider a more speculative example in which the portfolio manager specifies a view on realized volatility by trading a variance swap.

1. Introduction

The topic of portfolio optimization in the style of Black and Litterman (1992, 1991) seems to have generated more than its share of confusion over the years, as evidenced by articles with titles such as *"A demystification of the Black–Litterman model"* (Satchell & Scowcroft, 2000), etc. The method itself is often described as "Bayesian" but the original authors do not elaborate directly on connections with Bayesian statistics.

In language universally familiar to statisticians (Robert, 2007), a *Bayesian statistical model* consists of:

1. A vector-valued random variable $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^d$ distributed according to $f(\boldsymbol{x} \,|\, \boldsymbol{\theta})$, where realizations of $\boldsymbol{x}$ have been observed and only the parameter $\boldsymbol{\theta}$ (which belongs to a real vector space $\Theta \subseteq \mathbb{R}^\ell$) is unknown, and
2. A prior density $\pi(\boldsymbol{\theta})$ on Θ.

The function $f(\boldsymbol{x} \,|\, \boldsymbol{\theta})$ is called the *likelihood* and, after conditioning on $\boldsymbol{\theta}$, forms a density on the *data space* $\mathcal{X} \subseteq \mathbb{R}^d$. The *posterior* is the density on Θ proportional to $f(\boldsymbol{x} \,|\, \boldsymbol{\theta})\pi(\boldsymbol{\theta})$, and the normalization factor drops out of certain calculations. In Bayesian statistics, all statistical inference is based on the posterior.

The paper by Litterman and He (1999) contains many references to a "prior" but only one mention of a "posterior" without details, and no mention of a "likelihood."

In the present note, we clarify the exact nature of the Bayesian statistical model to which Black–Litterman optimization corresponds, in terms of the prior, likelihood, and posterior. In the process we also lay out the full set of assumptions made, some of which are glossed over in other treatments.

2. Black, Litterman, and Bayes

Consider a view such as "the German equity market will outperform a capitalization-weighted basket of the rest of the European equity markets by 5%," which is an example presented in Litterman and He (1999). Let $\boldsymbol{p} \in \mathbb{R}^n$ denote a portfolio which is long one unit of the DAX index, and short a one-unit basket which holds each of the other major European indices (UKX, CAC40, AEX, etc.) in proportion to their respective aggregate market capitalizations, so that $\sum_i p_i = 0$. Let $q = 0.05$ in this example. This view may be equivalently expressed as

$$\mathbb{E}[\boldsymbol{p}'\boldsymbol{r}] = q \in \mathbb{R} \tag{1}$$

where $\boldsymbol{r}$ is the random vector of asset returns over some subsequent interval, and q denotes the expected return, according to the view. If there are multiple such views, say

$$\mathbb{E}[\boldsymbol{p}_i'\boldsymbol{r}] = q_i, \quad i = 1 \dots k$$

then the portfolios $\boldsymbol{p}_i$ are more conveniently arranged as rows of a matrix $\boldsymbol{P}$, and the statement of views becomes

$$\mathbb{E}[\boldsymbol{P}\boldsymbol{r}] = \boldsymbol{q} \text{ for } \boldsymbol{q} \in \mathbb{R}^k. \tag{2}$$

In the language of statistics, the core idea of Black and Litterman (1991) is to treat the portfolio manager's views as noisy observations which are useful for performing statistical inference concerning the parameters in some underlying model for $\boldsymbol{r}$. For example, if

$$\boldsymbol{r} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \tag{3}$$

* Corresponding author.
E-mail addresses: ritter@post.harvard.edu, gordon.ritter@gsacapital.com (G. Ritter).

http://dx.doi.org/10.1016/j.ejor.2016.10.027
0377-2217/

with $\boldsymbol{\Sigma}$ a known positive-definite $n \times n$ matrix, then the views (2) can be recast as "observations" relevant for inference on the parameter $\boldsymbol{\theta}$.

A key aspect of the model is that the practitioner must also specify a level of uncertainty or "error bar" for each view, which is assumed to be an independent source of noise from the volatility already accounted for in a model such as (3). This is expressed as the following more precise restatement of (2):

$$\boldsymbol{P\theta} = \boldsymbol{q} + \boldsymbol{\epsilon}^{(v)}, \quad \boldsymbol{\epsilon}^{(v)} \sim N(0, \boldsymbol{\Omega}), \quad \boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \ldots, \omega_k) \tag{4}$$

Portfolio managers in this model specify *noisy, partial, indirect* information about $\boldsymbol{\theta}$, through their views. The information is *partial and indirect* because the views are on portfolio returns, i.e. linear transformations of returns, rather than on the asset returns directly. The information is *noisy*, with the noise modeled by $\boldsymbol{\epsilon}^{(v)}$, because the future is always uncertain.

A subjective, uncertain view about what will happen to a certain portfolio in the future is conceptually distinct from a noisy experimental observation such as an attempt to measure some physical constant with imperfect laboratory equipment. Nonetheless, for building intuition, we suggest thinking of a portfolio manager's forecast as an "observation of the future" in which the measuring device is a rather murky and unreliable crystal ball. Only in this way is it analogous to the noisy measurements in experimental design which much of statistics is designed to model.

Quite generally, 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. This would be the predicament of a physicist with a noisy measuring device, measuring a quantity that is itself random, and we suppose the physicist wants to know about the underlying data-generating process. Black and Litterman essentially say that the portfolio manager's view, if it is worth anything, should contain some (noisy) information about the future, so the view is, mathematically, no different from a noisy observation of a realization of (a linear transformation of) future returns.

As noted above, to perform statistical inference, observations alone are not sufficient; one needs to fully specify the statistical model, which includes a likelihood and a prior. In fact (4) specifies the likelihood as

$$f(\boldsymbol{q}\,|\,\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{P\theta} - \boldsymbol{q})'\boldsymbol{\Omega}^{-1}(\boldsymbol{P\theta} - \boldsymbol{q})\right] \tag{5}$$

which is the standard normal likelihood for a multiple linear regression problem with dependent variable $\boldsymbol{q}$ and design matrix $\boldsymbol{P}$.

A feature of Bayesian statistics that is dissimilar from frequentist statistics is the ability to perform inference in data-scarce situations. In Bayesian statistics, even a single observation can lead to valid inferences for multi-parameter models due to the presence of a prior. In essence, when less information is available, more weight is given to the prior.

The classic regression problem has the number of variables much less than the number of observations, and is therefore identifiable. However, the need to perform inference in models with many more variables than observations also arises in many applications. Notably, this arises in the analysis of gene expression arrays, and is typically handled by Bayesian methods such as ridge and the lasso (Tibshirani, 1996).

In a Black–Litterman model with one single view, there is one observation and still n parameters to serve as the subjects for statistical inference: $\boldsymbol{\theta} \in \mathbb{R}^n$ are the unobservable means of the asset returns. More generally, we may be presented with no views, one, or very many. When views are collected from many diverse portfolio managers or economists, they may contain internal contradictions; i.e. it may be impossible that they all come true exactly. Bayesian regression is the ideal tool to deal with all such cases. Internal contradictions in the views simply mean that there is no exact (zero-residual) solution to the regression equations, which in fact is the typical situation in classic (identifiable) linear regression.

We have not yet specified the prior, but Black and Litterman were motivated by the guiding principle that, 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 $\boldsymbol{h}_{eq}$. Hence in the absence of any views, and with prior mean equal to $\boldsymbol{\Pi}$, the investor's model of the world is that

$$\boldsymbol{r} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}), \quad \text{and} \quad \boldsymbol{\theta} \sim N(\boldsymbol{\Pi}, \boldsymbol{C}) \tag{6}$$

for some covariance $\boldsymbol{C}$ whose inverse represents the amount of precision in the prior. For any portfolio $\boldsymbol{p}$, then, according to (6) we have

$$\mathbb{E}[\boldsymbol{p}'\boldsymbol{r}] = \boldsymbol{p}'\boldsymbol{\Pi} \quad \mathbb{V}[\boldsymbol{p}'\boldsymbol{r}] = \boldsymbol{p}'(\boldsymbol{\Sigma} + \boldsymbol{C})\boldsymbol{p}.$$

In fact we must make a choice whether to use the conditional or unconditional variance in optimization: $\mathbb{V}(\boldsymbol{r}\,|\,\boldsymbol{\theta}) = \boldsymbol{\Sigma}$ but $\mathbb{V}(\boldsymbol{r}) = \boldsymbol{\Sigma} + \boldsymbol{C}$. Since investors are presumably concerned with unconditional variance of wealth, the unconditional variance form is preferable.

Throughout the following, we use the letter $\boldsymbol{h} \in \mathbb{R}^n$ to denote a vector of portfolio holdings; it has units of dollars, or whatever numéraire currency the investor is using. Mean–variance optimization with the moments as given above, and with risk-aversion parameter $\delta > 0$, leads to

$$\boldsymbol{h}_{\mathrm{eq}} = \delta^{-1}(\boldsymbol{\Sigma} + \boldsymbol{C})^{-1}\boldsymbol{\Pi}.$$

Any combination of $\boldsymbol{\Pi}$, $\boldsymbol{C}$ satisfying this will lead to a model with the desired property – that the optimal portfolio with only the information given in the prior is the prescribed portfolio $\boldsymbol{h}_{\mathrm{eq}}$. In particular, taking $\boldsymbol{C} = \tau\boldsymbol{\Sigma}$ with some arbitrary scalar $\tau > 0$, as did the original authors, leads to

$$\boldsymbol{\Pi} = \delta(1 + \tau)\boldsymbol{\Sigma}\boldsymbol{h}_{\mathrm{eq}}$$

We thus have the normal likelihood (5) and the normal prior (6) which is a *conjugate prior* for that likelihood, meaning that the posterior is of the same family (i.e. also normal in this example). A detailed discussion of conjugate priors is found in Robert (2007, Section 3.3).

The negative log posterior is thus proportional to (neglecting terms that do not contain $\boldsymbol{\theta}$):

$$(\boldsymbol{P\theta} - \boldsymbol{q})'\boldsymbol{\Omega}^{-1}(\boldsymbol{P\theta} - \boldsymbol{q}) + (\boldsymbol{\theta} - \boldsymbol{\Pi})'\boldsymbol{C}^{-1}(\boldsymbol{\theta} - \boldsymbol{\Pi}) \tag{7}$$

$$= \boldsymbol{\theta}'\boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P\theta} - \boldsymbol{\theta}'\boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{q} - \boldsymbol{q}'\boldsymbol{\Omega}^{-1}\boldsymbol{P\theta} \tag{8}$$

$$+ \boldsymbol{\theta}'\boldsymbol{C}^{-1}\boldsymbol{\theta} - \boldsymbol{\theta}'\boldsymbol{C}^{-1}\boldsymbol{\Pi} - \boldsymbol{\Pi}'\boldsymbol{C}^{-1}\boldsymbol{\theta}$$

$$= \boldsymbol{\theta}'[\boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P} + \boldsymbol{C}^{-1}]\boldsymbol{\theta} - 2(\boldsymbol{q}'\boldsymbol{\Omega}^{-1}\boldsymbol{P} + \boldsymbol{\Pi}'\boldsymbol{C}^{-1})\boldsymbol{\theta} \tag{9}$$

The following lemma, known colloquially as "completing the squares" will be useful:

Lemma 1. *If a multivariate normal random variable* $\boldsymbol{\theta}$ *has density* $p(\boldsymbol{\theta})$ *and*

$$-2\log p(\boldsymbol{\theta}) = \boldsymbol{\theta}'\boldsymbol{H\theta} - 2\boldsymbol{\eta}'\boldsymbol{\theta} + (\text{terms without } \boldsymbol{\theta})$$

then $\mathbb{V}[\boldsymbol{\theta}] = \boldsymbol{H}^{-1}$ *and* $\mathbb{E}\boldsymbol{\theta} = \boldsymbol{H}^{-1}\boldsymbol{\eta}$.

Lemma 1 follows directly from the fact that, for $\boldsymbol{H}$ symmetric,

$$\boldsymbol{\theta}'\boldsymbol{H\theta} - 2\boldsymbol{v}'\boldsymbol{H\theta} = (\boldsymbol{\theta} - \boldsymbol{v})'\boldsymbol{H}(\boldsymbol{\theta} - \boldsymbol{v}) - \boldsymbol{v}'\boldsymbol{Hv}$$

For the quadratic term to match (9) we must have $\boldsymbol{H} = \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P} + \boldsymbol{C}^{-1}$ and hence the posterior has mean

$$\boldsymbol{v} = [\boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P} + \boldsymbol{C}^{-1}]^{-1}[\boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{q} + \boldsymbol{C}^{-1}\boldsymbol{\Pi}] \tag{10}$$