ARTICLE IN PRESS
JID: EOR [m5G;November 3, 2016;16:42]

European Journal of Operational Research 000 (2016) 1–9

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 xxx

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

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

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,

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

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

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

Please cite this article as: P. Kolm, G. Ritter, On the Bayesian interpretation of Black–Litterman, European Journal of Operational Research (2016), http://dx.doi.org/10.1016/j.ejor.2016.10.027

and covariance

$$\boldsymbol{H}^{-1} = [\boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P} + \boldsymbol{C}^{-1}]^{-1}. \tag{11}$$

Part of the beauty of this derivation is its simplicity: going from (7) to (11) requires just a few lines of algebra.

Investors with CARA utility of final wealth will want to solve

$$\boldsymbol{h}^* = \text{argmax}_{\boldsymbol{h}} \left\{ \mathbb{E}[\boldsymbol{h}'\boldsymbol{r}] - (\delta/2)\mathbb{V}[\boldsymbol{h}'\boldsymbol{r}] \right\}$$

where $\mathbb{E}[\boldsymbol{r}]$ and $\mathbb{V}[\boldsymbol{r}]$ denote, respectively, the unconditional mean and covariance of $\boldsymbol{r}$ under the posterior. The unconditional covariance is a sum of variance due to parameter uncertainty, and variance due to the randomness in $\boldsymbol{r}$. In other words,

$$\mathbb{V}[\boldsymbol{h}'\boldsymbol{r}] = \boldsymbol{h}'[\boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P} + \boldsymbol{C}^{-1}]^{-1}\boldsymbol{h} + \boldsymbol{h}'\boldsymbol{\Sigma}\boldsymbol{h}$$

The optimal portfolio accounting for both types of variance is then

$$\boldsymbol{h}^* = \delta^{-1}[\boldsymbol{H}^{-1} + \boldsymbol{\Sigma}]^{-1}\boldsymbol{H}^{-1}[\boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{q} + \boldsymbol{C}^{-1}\boldsymbol{\Pi}].$$

3. The most general Black–Litterman–Bayes model

3.1. Definitions

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

Definition 1. A *Black–Litterman–Bayes* model consists of:

(a) A parametric statistical model for asset returns $p(\boldsymbol{r}\,|\,\boldsymbol{\theta})$ with finite-dimensional parameter vector $\boldsymbol{\theta}$,
(b) A prior $\pi(\boldsymbol{\theta})$ on the parameter space,
(c) A likelihood function $f(\boldsymbol{q}\,|\,\boldsymbol{\theta})$ 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) and (b) simply state that we have a Bayesian statistical model, as defined in Section 1, for asset returns. Under such a model, Decision Theory (see Robert (2007, Chapter 2 and references) teaches us that the optimal decision is the one maximizing posterior expected utility. This leads us to Definition 2.

Definition 2. Given a Black–Litterman–Bayes (BLB) model as per Definition 1, the associated BLB optimal portfolio is defined to be

$$\boldsymbol{h}^* \in \text{argmax}_{\boldsymbol{h}}\, \mathbb{E}[u(\boldsymbol{h}'\boldsymbol{r})\,|\,\boldsymbol{q}]$$

where $\mathbb{E}[\,\cdot\,|\,\boldsymbol{q}]$ 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

$$p(\boldsymbol{r}\,|\,\boldsymbol{q}) = \int p(\boldsymbol{r}\,|\,\boldsymbol{\theta})p(\boldsymbol{\theta}\,|\,\boldsymbol{q})d\boldsymbol{\theta} \quad \text{where}$$

$$p(\boldsymbol{\theta}\,|\,\boldsymbol{q}) = \frac{f(\boldsymbol{q}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{q}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

Definition 3. Given a benchmark portfolio with holdings $\boldsymbol{h}_B$ (e.g. the market portfolio), and given a Black–Litterman–Bayes model (Definition 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 $p(\boldsymbol{r}) = \int p(\boldsymbol{r}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$, so

$$\boldsymbol{h}_B \in \text{argmax}_{\boldsymbol{h}} \int u(\boldsymbol{h}'\boldsymbol{r})p(\boldsymbol{r}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{12}$$

Many existing approaches are special cases of the above. The model of Black and Litterman (1991) is the special case in which $\boldsymbol{r}\,|\,\boldsymbol{\theta}$ is multivariate normal with mean $\boldsymbol{\theta}$ and $f(\,\cdot\,|\,\cdot\,)$ 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.

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.

We find that phrasing things in this way inspires the imagination. Even if $\boldsymbol{\theta}$ simply represents the mean vector of asset returns, such returns are widely recognized to be non-normal. Replacing (3) with a Laplace distribution may fit empirical asset returns more accurately. This corresponds to Least absolute deviation (LAD) regression. Giacometti, Bertocchi, Rachev, and Fabozzi (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.

3.2. APT and factor models

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

$$\boldsymbol{r} = \boldsymbol{X}\boldsymbol{f} + \boldsymbol{\epsilon}, \quad \mathbb{E}[\boldsymbol{\epsilon}] = 0, \quad \mathbb{V}[\boldsymbol{\epsilon}] = \boldsymbol{D} \tag{13}$$

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 $\boldsymbol{X}$ 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

$$\boldsymbol{D} = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2) \quad \text{with all} \quad \sigma_i^2 > 0. \tag{14}$$

The variable $\boldsymbol{f}$ in (13) 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

$$\mathbb{E}[\boldsymbol{f}] = \boldsymbol{\mu}_f, \quad \text{and} \quad \mathbb{V}[\boldsymbol{f}] = \boldsymbol{F}. \tag{15}$$

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 (13)–(15) entails associated reductions of the first and second moments of the asset returns:

$$\mathbb{E}[\boldsymbol{r}] = \boldsymbol{X}\boldsymbol{\mu}_f, \quad \text{and} \quad \boldsymbol{\Sigma} := \mathbb{V}[\boldsymbol{r}] = \boldsymbol{D} + \boldsymbol{X}\boldsymbol{F}\boldsymbol{X}' \tag{16}$$

where $\boldsymbol{X}'$ denotes the transpose of $\boldsymbol{X}$. The elements of $\boldsymbol{\mu}_f$ are called *factor risk premia*. We will continue to use $\boldsymbol{\Sigma}$ to denote $\boldsymbol{D} + \boldsymbol{X}\boldsymbol{F}\boldsymbol{X}'$ throughout this section, and (14) implies that $\boldsymbol{\Sigma}^{-1}$ exists.

For simplicity, we treat $\boldsymbol{X}$ as non-stochastic and assume $k \ll n$. Then (16) 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 $\boldsymbol{D}$, for a total of $n+k(k+3)/2$. Models of the form (13) 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), Connor, Goldberg, and Korajczyk (2010) for more discussion and examples.

In the language of Definition 1, we are free to choose $\boldsymbol{\theta}$ as any vector of parameters appearing in a parametric statistical model for asset returns; (13)–(15) 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), Fabozzi, Focardi, and Kolm (2006), Fabozzi et al. (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

$$f(\boldsymbol{q}\,|\,\boldsymbol{\theta}) = \prod_{i=1}^{k} \exp\left[-\frac{1}{2\omega_i^2}(\theta_i - q_i)^2\right] \tag{17}$$

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 Definition 3.

If the random process model driving the unobservable factor returns $\boldsymbol{f}_t$ is stationary, i.e. $\boldsymbol{\mu}_f$, $\boldsymbol{F}$ 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 $\hat{\boldsymbol{f}}_t = (\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t'\boldsymbol{r}_{t+1}$ 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 $\boldsymbol{r}_{t+1} \sim N(\boldsymbol{X}_t\boldsymbol{f}_t, \boldsymbol{D})$ and the various $\boldsymbol{f}_t$ are modeled as i.i.d. draws $\boldsymbol{f}_t \sim N(\boldsymbol{\mu}_f, \boldsymbol{F})$. 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, Carlin, Stern, and Rubin (2003, Chapter 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 is 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 (12). The first step in evaluating (12) 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 $\boldsymbol{\Sigma} = \boldsymbol{D} + \boldsymbol{X}\boldsymbol{F}\boldsymbol{X}'$ as above, since this is the asset-level covariance in an APT model. Straightforward calculations then show:

$$\begin{aligned}-2\log[p(\boldsymbol{r}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})] &= -2\log N(\boldsymbol{r};\ \boldsymbol{X}\boldsymbol{\theta}, \boldsymbol{\Sigma}) - 2\log N(\boldsymbol{\theta};\ \boldsymbol{\xi}, \boldsymbol{V})\\ &= (\boldsymbol{r}-\boldsymbol{X}\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{r}-\boldsymbol{X}\boldsymbol{\theta}) + (\boldsymbol{\theta}-\boldsymbol{\xi})'\boldsymbol{V}^{-1}(\boldsymbol{\theta}-\boldsymbol{\xi})\\ &= \boldsymbol{\theta}'\boldsymbol{H}\boldsymbol{\theta} - 2\boldsymbol{\eta}'\boldsymbol{\theta} + \boldsymbol{z}\end{aligned}$$

where for notational simplicity, we have introduced the auxiliary variables

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

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,

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

Let's multiply out the second quadratic term:

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

Note that $\int p(\boldsymbol{r}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ 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:

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

Similarly, referring to Lemma 1, the mean is

$$\mathbb{E}_\pi[\boldsymbol{r}] = \mathbb{V}_\pi[\boldsymbol{r}]\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{H}^{-1}\boldsymbol{V}^{-1}\boldsymbol{\xi} \tag{18}$$

$$= (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{H}^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1})^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{H}^{-1}\boldsymbol{V}^{-1}\boldsymbol{\xi} \tag{19}$$

The *a priori* optimal portfolio under CARA utility is of course

$$(\delta\mathbb{V}_\pi[\boldsymbol{r}])^{-1}\mathbb{E}_\pi[\boldsymbol{r}] = \delta^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{H}^{-1}\boldsymbol{V}^{-1}\boldsymbol{\xi} \tag{20}$$

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, (20) gives a very simple characterization of those that can: they are precisely of the form $\delta^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi}$ where $\boldsymbol{\Pi}$ is some linear combination of the columns of $\boldsymbol{X}$. 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 is not 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 $\boldsymbol{X}$ contains the CAPM betas, then the individual asset risk premia will be proportional to that column of $\boldsymbol{X}$, and then the market portfolio *will* be realizable as *a priori* optimal, as per (20).

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 (17) 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

$$p(\boldsymbol{\theta} \mid \boldsymbol{q}) = \frac{f(\boldsymbol{q} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{q} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

2. Calculate the *a posteriori* distribution of asset returns (also called the posterior predictive density), given by

$$p(\boldsymbol{r} \mid \boldsymbol{q}) = \int p(\boldsymbol{r} \mid \boldsymbol{\theta})p(\boldsymbol{\theta} \mid \boldsymbol{q})d\boldsymbol{\theta} \tag{21}$$

3. Calculate the mean–variance optimal portfolio under (21).

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 do not 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

$$\mathbb{V}[\boldsymbol{r} \mid \boldsymbol{q}] = (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\tilde{\boldsymbol{V}}^{-1} + \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1})^{-1}. \tag{22}$$

$$\mathbb{E}[\boldsymbol{r} \mid \boldsymbol{q}] = \mathbb{V}[\boldsymbol{r} \mid \boldsymbol{q}]\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\tilde{\boldsymbol{V}}^{-1} + \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\tilde{\boldsymbol{V}}^{-1}\tilde{\boldsymbol{\xi}} \tag{23}$$

Step 3 is then a completely standard calculation of a mean–variance optimal portfolio from (22) and (23):

$$\boldsymbol{h}^* = \delta^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi} \tag{24}$$

$$\boldsymbol{\Pi} := \boldsymbol{X}\tilde{\boldsymbol{\mu}}_f \tag{25}$$

$$\tilde{\boldsymbol{\mu}}_f := (\boldsymbol{V}^{-1} + \boldsymbol{\Omega}^{-1} + \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{V}^{-1}\boldsymbol{\xi} + \boldsymbol{\Omega}^{-1}\boldsymbol{q}) \tag{26}$$

Eqs. (24)–(26) represent the solution to Black–Litterman optimization in the context of APT. They are written in a suggestive form: the asset-level risk premia $\boldsymbol{\Pi} = \boldsymbol{X}\tilde{\boldsymbol{\mu}}_f$ are linear combinations of the factors which form the columns of $\boldsymbol{X}$. 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.

3.3. An empirical example for APT

In this section, we present a detailed empirical example of the application of Eqs. (24)–(26) and other machinery of Section 3.2. We study the United States equity market over the period 1992–2015 through the lens of an APT model (13)–(15). 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. For further discussions of the five risk premia mentioned here, see Fama and French (1993), Connor et al. (2010), Menchero, Morozov, and Shepard (2008), Asness et al. (2013). There are likely other sources of systematic risk which could be considered in a more complete model; our goal is only to illustrate the techniques of Section 3.2, and this model will do nicely for that purpose. With the exception of the volatility premium, our model is intentionally very similar to that of Fama and French (1992), which is one of the most-cited papers in finance. The classic paper on the volatility premium (Ang, Hodrick, Xing, & Zhang, 2006) was not until 2006, but all other factors had been discussed in the academic literature prior to the beginning of our sample period.

As before we let k denote the number of factors. In this example, $k \approx 75$ due to the 5 risk premia and about 70 industries. Our industry classification is based on the "major group" of the Standard Industrial Classification (SIC) system. The model is

$$\boldsymbol{r}_{t+1} = \boldsymbol{X}_t\boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \quad \mathbb{E}[\boldsymbol{\epsilon}] = 0, \quad \mathbb{V}[\boldsymbol{\epsilon}] = \boldsymbol{D}$$

where $\boldsymbol{r}_{t+1} \in \mathbb{R}^n$ denotes a cross-section of asset returns over the interval $[t, t+1]$, and $\boldsymbol{X}_t$ is calculated using data knowable before day t. Specifically, $\boldsymbol{r}_t$ are close-to-close total returns, and t will denote a day, or when more precision is required, the exact time of the US equity market close on day t.

The first 5 columns of $\boldsymbol{X}_t$ represent *exposures* to the five risk premia mentioned above. With the exception of market beta, which we do not transform in any way, the exposures are calculated as a "raw" value which is then "gaussianized" in the cross section. Gaussianization refers to a robustification procedure for extremely fat-tailed data by which the data is converted to ranks and passed through the inverse CDF of a normal. The latter preserves the order, but reshapes the data to appear normally distributed.

(1) Market beta: each asset's daily excess return time series is winsorized and regressed against the S&P 500 excess return time series, over a trailing two year window with an intercept. The beta is the slope coefficient in this regression. The results are further improved using the Vasicek (1973) Bayesian adjustment in the cross section.
(2) Size: market capitalization.
(3) Volatility: as in the market beta calculation, each asset's daily excess return time series is winsorized and regressed against the S&P 500 excess return time series, over a trailing two year window. The mean-square error (MSE) from this regression is the raw exposure.
(4) Momentum: trailing compound returns over the last 12 months with the last 1 month excluded. For this and the next, see Asness et al. (2013) and references therein.
(5) Value: the raw exposure is e_t/p_t where e_t represents the sum of the trailing 4 quarters' earnings-per-share (EPS), adjusted for any splits which occurred between the earnings announcement date and date t, and p_t is the close price on day t.

We define a stock's exposure to an industry as 1 if the stock is classified into that industry by SIC, and zero otherwise. Hence the remaining 70 columns of $\boldsymbol{X}_t$ are populated by 1s and 0s, with one column per industry.

Next we calculate the OLS factor returns

$$\hat{\boldsymbol{f}}_t = (\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t'\boldsymbol{r}_{t+1}. \tag{27}$$

Since our example concerns industry-neutral and market-neutral portfolios, we will not ascribe any persistent risk premia to the market-beta factor, nor to any of the industry factors. The cumulative factor returns of the other risk premia is shown in Fig. 1.

Fig. 1 corresponds well with our intuition and established consensus from the academic literature: the two positive risk premia are momentum and value; the two negative risk premia are size and volatility. For example, a negative premium to the size factor indicates that low market capitalization stocks tend to outperform

Fig. 1. Cumulative factor returns to risk premia $\hat{\boldsymbol{f}}_t$ as computed by (27). The two positive-drift risk premia are momentum and value; the two negative-drift risk premia are size and volatility.

high market capitalization stocks after controlling for market beta, industry, and other systematic sources of risk.

As above we continue to assume that the random process model driving the unobservable factor returns $\boldsymbol{f}_t$ is stationary, i.e. $\boldsymbol{\mu}_f$, $\boldsymbol{F}$ are approximately constant over time. None of what follows depends on having very precise forecasts of $\boldsymbol{\mu}_f$, $\boldsymbol{F}$. We take $\boldsymbol{\mu}_f$ to be zero except for the four risk premia that we want exposure to, in which case we set the corresponding components of $\boldsymbol{\mu}_f$ as: Momentum $=$ 1.0 basis points, Size $= -0.5$ basis points, Volatility $= -1.5$ basis points, and Value $=$ 0.5 basis points, in units of daily return.

We take $\boldsymbol{F}$ to be diagonal, with all industry variances set to 10 basis points/day-squared; the resulting portfolios do not really depend on the industry block of $\boldsymbol{F}$ since they are not taking any meaningful industry risk. The risk premia variances are estimated using long-range, slowly moving windows.

We estimate $\boldsymbol{D} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ via simple time-series regressions of each stock's return against the S&P 500 return over a rolling two-year window; then σ_i^2 is the residual variance from that regression. We also winsorize low values of σ_i^2 since they will be inverted when we compute $\boldsymbol{D}^{-1}$ and could lead to large portfolio weights.

At this point, we have enough information to compute the Markowitz portfolio, which is definitionally given by

$$\boldsymbol{h}_{\mathrm{mar}} := (\delta\boldsymbol{\Sigma}_t)^{-1}\boldsymbol{X}_t\boldsymbol{\mu}_f, \text{ where } \boldsymbol{\Sigma}_t = \boldsymbol{X}_t\boldsymbol{F}\boldsymbol{X}_t' + \boldsymbol{D}_t \tag{28}$$

To speed up the computation, instead of (28) we use the linear-time formula[1] of Ritter (2016). Note that (28) does not use any form of Black–Litterman technology, and it does not include "views" in the terminology used above. Hence we regard (28) as a baseline which a supposedly more sophisticated method should be able to outperform (in terms of information ratio).

To progress further, we need to compute Eqs. (24)–(26), for which we need to choose a prior $\pi(\boldsymbol{\theta};\ \boldsymbol{\xi},\ \boldsymbol{V})$ and some form of views on factor risk premia as embodied by $\boldsymbol{q}_t$ and $\boldsymbol{\Omega}_t$. For simplicity we take $\boldsymbol{\xi} = \boldsymbol{\mu}_f$ and $\boldsymbol{V} = \boldsymbol{F}$, although real-world applications may wish to build more detailed data-driven priors. Furthermore, we take $\omega_i^2 = v_{ii}$ where v_{ii} is the ith diagonal element of $\boldsymbol{V}$. This amounts to assuming our uncertainty in each view is proportional to the actual market risk of the factor that the view is about. This is reasonable, but one should also consider using for ω_i the statistical estimation error in q_i, when available.

We stress that a Black–Litterman–Bayes model is only as good as its inputs, which in this case are really the forecasts $q_{i,\,t}$. We take $q_{i,\,t}$ to be the one-period-ahead forecast of the ith risk premium's factor return from a univariate ARIMA model fit on an expanding window (i.e. the entire history up to t). In other words, we are forecasting the time-series depicted in Fig. 1 and using those forecasts for $\boldsymbol{q}$. The precise ARIMA model used for forecasting $\mathbb{E}[\boldsymbol{f}_{t+1}]$ is selected using AIC$_c$ on a rolling basis, using only data knowable as of time t.

With these data in hand, we then compute for each day t, the Black–Litterman–Bayes portfolio holdings as per (24):

$$\boldsymbol{h}_{\mathrm{blb}} = \delta^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{V}^{-1} + \boldsymbol{\Omega}^{-1} + \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{V}^{-1}\boldsymbol{\xi} + \boldsymbol{\Omega}^{-1}\boldsymbol{q}) \tag{29}$$

The computation of (29) can be sped up considerably by means of the Woodbury matrix inversion lemma: one has $\boldsymbol{\Sigma}^{-1} = \boldsymbol{D}^{-1} - \boldsymbol{Z}[\boldsymbol{F}^{-1} + \boldsymbol{X}'\boldsymbol{Z}]^{-1}\boldsymbol{Z}'$ where $\boldsymbol{Z} = \boldsymbol{D}^{-1}\boldsymbol{X}$, which only requires calculating the inverses of diagonal or low-dimensional matrices.

We then directly compare the two time-series of transaction-cost-free profits: $\boldsymbol{h}_{\mathrm{blb},t}'\boldsymbol{r}_{t+1}$ and $\boldsymbol{h}_{\mathrm{mar},t}'\boldsymbol{r}_{t+1}$. The cumulative (pre-trading cost) profit from both portfolios is shown in Fig. 2. They have been scaled to have the same volatility.

The respective information ratios (IR) are 1.8 for the BLB portfolios and 1.3 for the Markowitz portfolios. Hence we conclude that the Black–Litterman–Bayes optimization procedure has done what it was designed to do: it has successfully used the one-period-ahead predictions of risk premia returns as "views," and incorporated these views into the optimization, leading to an improved information ratio and less severe drawdowns. In some cases the BLB portfolio actually benefits when the Markowitz portfolio has a drawdown because the negative trend is picked up quickly enough by the ARIMA model generating the views, and the factor having the drawdown is given a negative view.

3.4. Simultaneous views on drift and volatility

Consider a fixed time interval $[t, T]$ and let $r = (S_T - S_t)/S_t$ denote an asset's return over the interval. Suppose that a portfolio manager has a view, that $r > 0$ and also that the volatility over

[1] The linear-time formula of Ritter (2016) is implemented in R as follows: `Z <- t(X) %*% sqrt(D.inv) (1.0/delta) * sqrt(D.inv) %*% ginv(Z) %*% solve(V + ginv(Z %*% t(Z)), mu_f).`

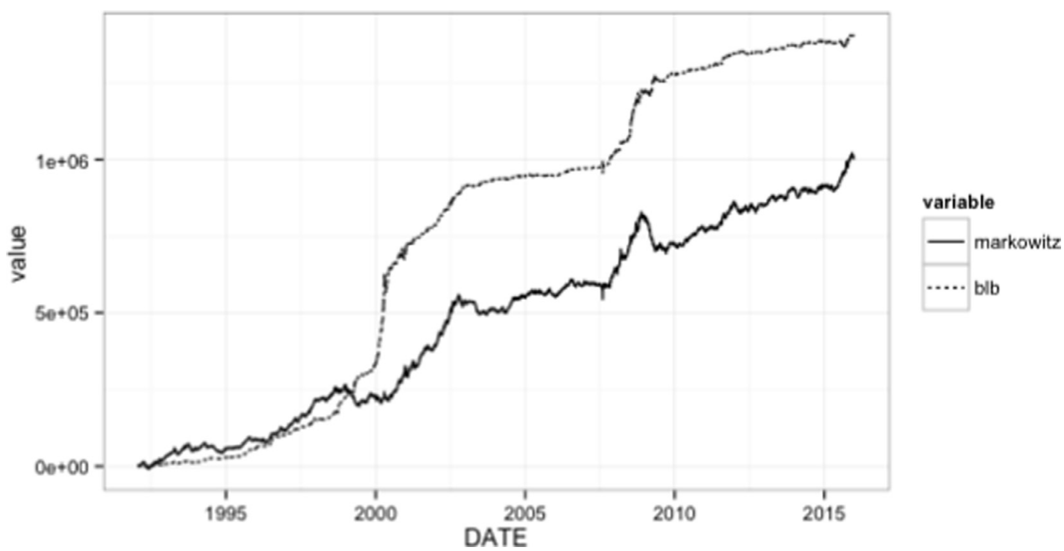

Fig. 2. Cumulative returns to Markowitz portfolio (solid line, IR = 1.3) and Black–Litterman–Bayes portfolio (dotted line, IR = 1.8).

$[t, T]$ will be higher than the options market is currently anticipating. One relatively pure representation of the market's view on what the realized variance will be is the strike of a variance swap. The payoff of a variance swap is

$$n_{\text{var}}(\sigma_R^2 - \sigma_{\text{str}}^2)$$

where n_{var} is the notional, σ_R^2 is realized variance of the underlying over the sampling period, and σ_{str}^2 is the "strike variance."

The portfolio manager may be more certain about the volatility view than the drift view, as in the example of an upcoming event: one can be sure that once the event occurs, the volatility will spike, but the direction is harder to predict. Presumably this portfolio manager would like to hold a portfolio consisting of the underlying asset and a variance swap. The final wealth of such a portfolio (neglecting the initial wealth) will be

$$w = hr + n_{\text{var}}(\sigma_R^2 - \sigma_{\text{str}}^2) \tag{30}$$

To construct a Black–Litterman–Bayes model, we will need a Bayesian statistical model for asset returns. More precisely, for the bivariate joint distribution of the two terms in (30). We take the parameters to be

$$\boldsymbol{\theta} = (\mu, \sigma^2),$$

the mean and variance of the true underlying distribution of r, which is assumed to be normal. Concretely, we assume r is driven by a Brownian motion with drift, in which $S_T - S_t$ is the sum of $n+1$ normally-distributed i.i.d. increments at the sample points of the variance swap. The theorem of Cochran (1934) shows that σ_R^2 follows a scaled chi-squared distribution:

$$n\frac{\sigma_R^2}{\sigma^2} \sim \chi_n^2.$$

We write this as $\sigma_R^2 \sim \chi^2_{\text{scaled}}(\sigma_R^2; \sigma^2)$ where

$$\chi^2_{\text{scaled}}(x; \nu) = \frac{2^{-n/2} n e^{-nx/(2\nu)} (nx/\nu)^{\frac{n}{2}-1}}{\nu\Gamma(n/2)}$$

In particular,

$$\mathbb{E}[\sigma_R^2 \mid \sigma^2] = \sigma^2 \text{ and } \mathbb{V}[\sigma_R^2 \mid \sigma^2] = 2\sigma^4/n. \tag{31}$$

The portfolio manager must express views via a likelihood function $f(q \mid \boldsymbol{\theta})$ for some as-yet undetermined data q. Views on the asset's return and variance can be expressed as

$$\mu \sim N(q, \omega^2), \quad \text{and} \quad \sigma^2 \sim \text{IG}(\alpha, \beta) \tag{32}$$

where $\text{IG}(\alpha, \beta)$ is the inverse gamma distribution with p.d.f.

$$g_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right)$$

In other words the portfolio manager specifies (q, ω^2) as in the classical Black–Litterman model for expected return and uncertainty in the view. Additionally – and this is the new ingredient – the portfolio manager expresses views on volatility by giving (α, β) such that $\sigma^2 \sim \text{IG}(\alpha, \beta)$. Reasonable views have $\alpha > 2$ so that $\text{IG}(\alpha, \beta)$ has finite first two moments.

Perhaps the most "boring" class of views on volatility would be characterized by those having $\mathbb{E}[\sigma^2] = \beta/(\alpha - 1) = \sigma_{\text{str}}^2$, because σ_{str}^2 represents the consensus view on variance already implied by the prevailing volatility surface in the market. This may be a useful reference point. More generally, any model of a random process that can be calibrated to the volatility surface could be used to set parameters or otherwise specify the prior and the views.

The likelihood function which encodes these views is then

$$f(q \mid \boldsymbol{\theta}) = g_{\alpha,\beta}(\sigma^2) N(\mu; q, \omega^2), \quad \text{where} \quad \boldsymbol{\theta} = (\mu, \sigma^2).$$

To complete the specification of the model, we take the prior $\pi(\mu, \sigma^2)$ to be a special case of the Normal-Inverse-Gamma distribution $\text{NIG}(\mu_0, \alpha_0, \beta_0)$. In other words,

$$\mu \mid \sigma^2 \sim N(\mu_0, \sigma^2), \quad \sigma^2 \sim \text{IG}(\alpha_0, \beta_0) \tag{33}$$

where $\alpha_0 > 2$. A zero subscript indicates a hyper-parameter appearing in the prior. [2]

It is interesting that σ_{str}^2 is essentially quoted by the market, hence the determination of a market-implied prior is perhaps easier than for portfolios of non-derivative instruments, which was the case for the original examples of Black–Litterman optimization. In the present case, the market gives us information directly about the parameters of the model via the quoted strike of a variance swap. Therefore we feel that any reasonable definition of *market-implied prior* $\pi(\sigma^2)$ should have the property that

$$\mathbb{E}_\pi[\sigma^2] = \sigma_{\text{str}}^2.$$

Priors not satisfying this property are of course possible, but they cannot claim to be in agreement with the options market. If the

[2] This is a reasonable prior to use in cases where some data may have already been observed; the inverse gamma arises as the marginal posterior distribution for the unknown variance of a normal distribution if an uninformative prior is used.

prior is of the inverse gamma family, $\mathrm{IG}(\alpha_0, \beta_0)$ then $\mathbb{E}_\pi[\sigma^2] = \beta_0/(\alpha_0 - 1)$, and hence $\beta_0 = \sigma^2_{\mathrm{str}}(\alpha_0 - 1)$. One can match both moments of the sampling distribution (31) by setting $\alpha_0 = 2 + n/2$ and $\beta_0 = \sigma^2_{\mathrm{str}}(1 + n/2)$.

To progress further we need to understand the joint posterior of r and σ_R^2, which are the two random variables driving Eq. (30),

$$\int p(r, \sigma_R^2 \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, q) d\boldsymbol{\theta} = \int \chi^2_{\mathrm{scaled}}(\sigma_R^2; \sigma^2) \left\{ \int p_N(r; \mu, \sigma^2) p(\mu, \sigma^2 \,|\, q) d\mu \right\} d\sigma^2$$

where p_N denotes the pdf of a normal. We will now show that the inner integral over μ can be done in closed form, for which we need the posterior $p(\mu, \sigma^2 \,|\, q)$:

$$p(\mu, \sigma^2 \,|\, q) \propto \underbrace{g_{\alpha,\beta}(\sigma^2) p_N(\mu; q, \omega^2)}_{\text{likelihood}} \times \underbrace{p_N(\mu; \mu_0, \sigma^2) g_{\alpha_0,\beta_0}(\sigma^2)}_{\text{prior}}$$

Neglecting proportionality constants, the previous line is proportional to

$$p_N\big(\mu; \mu', (\sigma')^2\big) g_{\alpha',\beta'}(\sigma^2)$$

where

$$\mu' := \frac{q\sigma^2 + \mu_0\omega^2}{\sigma^2 + \omega^2}, \quad (\sigma')^2 := \frac{\sigma^2\omega^2}{\sigma^2 + \omega^2}$$
$$\alpha' := \alpha + \alpha_0 + 1, \quad \beta' = \beta + \beta_0$$

Then, we obtain

$$\int p_N(r; \mu, \sigma^2) p(\mu, \sigma^2 \,|\, q) d\mu = g_{\alpha',\beta'}(\sigma^2) p_N(r; \mu', \sigma^2 + (\sigma')^2)$$

The bivariate posterior predictive density $p(r, \sigma_R^2)$ can thus be reduced to

$$\int \chi^2_{\mathrm{scaled}}(\sigma_R^2; \sigma^2) g_{\alpha',\beta'}(\sigma^2) p_N \times \left(r; \frac{q\,\sigma^2 + \mu_0\,\omega^2}{\sigma^2 + \omega^2}, \sigma^2 + (\sigma^{-2} + \omega^{-2})^{-1} \right) d\sigma^2 \tag{34}$$

As an aside, it is worth noting that the marginals of this distribution are simpler than the original due to the structure of the integrand; e.g. the marginal distribution of σ_R^2 is then

$$\int p(r, \sigma_R^2) dr = \int \chi^2_{\mathrm{scaled}}(\sigma_R^2; \sigma^2) g_{\alpha',\beta'}(\sigma^2) d\sigma^2$$

For any choice of h, n_{var} the wealth random variable is, as above, $w = hr + n_{\mathrm{var}}(\sigma_R^2 - \sigma^2_{\mathrm{str}})$. Hence

$$\mathbb{E}[w] = h\mathbb{E}[r] + n_{\mathrm{var}}(\mathbb{E}[\sigma_R^2] - \sigma^2_{\mathrm{str}})$$

and $\mathbb{E}[R]$ and $\mathbb{E}[\sigma_R^2]$ can be computed (from the marginals) before optimizing over h, n_{var}. Similarly,

$$\mathbb{V}[w] = h^2\mathbb{V}[r] + n_{\mathrm{var}}^2\mathbb{V}[\sigma_R^2] + 2\ hn_{\mathrm{var}}\,\mathrm{cov}(r, \sigma_R^2)$$

Hence for the purposes of optimizing $\mathbb{E}[w] - (\delta/2)\mathbb{V}[w]$ over h, n_{var}, we have only to compute (h, n_{var}-independent) estimates of the first and second moments of the joint posterior (34). Various numerical methods can be used to compute such estimates to any desired accuracy. Given the low dimension of the problem, a multidimensional numerical integration procedure will work well. Specifically, we suggest[3] writing a one-dimensional adaptive quadrature routine to compute $p(r, \sigma_R^2)$ as in (34) for any specific values of (r, σ_R^2) and then using this within a two-dimensional lattice integration to compute moments such as $\int rp(r, \sigma_R^2)drd\sigma_R^2$, $\int(r - \mathbb{E}[r])(\sigma_R^2 - \mathbb{E}[\sigma_R^2])p(r, \sigma_R^2)drd\sigma_R^2$ etc.

[3] Source code for performing this procedure is available by emailing the author(s).

We present this example in the spirit of writing down one of the simplest nontrivial models that illustrates the procedure. One weakness of the above structure, which is really a weakness of the Normal-Inverse Gamma family, is that the model does not properly capture the negative correlation which exists between return and volatility; higher volatility states of the world are empirically observed to coincide with states of the world in which risk assets such as equities experience negative returns.

4. Conclusions

We live in an exciting era characterized by rapid advances in statistical inference techniques, often driven by corresponding advances in machine learning and the need to analyze ever larger and more complex data sets. Many of the associated inference and model selection techniques can be naturally understood in a Bayesian context. For example, the Lasso (Tibshirani, 1996) is equivalent to Bayesian regression with a Laplace prior. In light of these developments, we believe that the coming decades will, increasingly, see instances of Bayesian statistical models being used to model asset returns in empirical finance. See Rachev, Hsu, Bagasheva, and Fabozzi (2008) and Kolm, Tütüncü, and Fabozzi (2014) for surveys of the field.

Many practical problems now require inference in "data scarce" situations where the number of parameters may greatly exceed the number of observations. Black–Litterman optimization with a small number of views is one such problem.

Any Bayesian statistical model of asset returns, together with a utility function of final wealth in the sense of Arrow (1971) and Pratt (1964), gives rise to an associated Black–Litterman–Bayes optimization procedure. Key to the generality of this procedure is that $\boldsymbol{\theta}$ can be any vector of parameters appearing in a statistical model for asset returns, and need not be simply a parameter representing the mean return. Section 3 also shows that our Black–Litterman–Bayes generalization is itself a special case of a Bayesian network in the sense of Pearl (2014).

In the examples to Section 3 we show that the Arbitrage Pricing Theory of Ross (1976), which has become central to the entire practice of quantitative trading, has a natural Bayesian extension and we write down the associated Black–Litterman–Bayes optimal portfolio. We also discuss the main two types of priors which are relevant to this kind of model: data-driven priors and benchmark-optimal priors.

In the process we hope to have clarified the precise sense in which the original model of Black and Litterman (1992); 1991) is "Bayesian." Usually, in Bayesian statistics, the likelihood function plays the essential role of connecting the empirical data to the parameters. The structure of the experimental design must be encoded in the likelihood function (which is why the matrix in a regression is called the "design matrix.")

In Black–Litterman type models, the likelihood is not really what connects the parameters to empirical data, unless we broaden the definition of "empirical data" to include data on portfolio managers' views. We reconcile this by regarding portfolio managers' views as (possibly *very* noisy) observations of the future, and as such, operationally no different from empirical data obtained through a very unreliable measuring device.

References

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance, 61*(1), 259–299.

Arrow, K. J. (1971). Essays in the Theory of Risk-Bearing. *Markham economics series*. North-Holland.

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance, 68*(3), 929–985.

Black, F., & Litterman, R. (1992). Global portfolio optimization. *FinancialAnalysts Journal, 48*(5), 28–43.

Black, F., & Litterman, R. B. (1991). Asset allocation: Combining investor views with market equilibrium. *The Journal of Fixed Income, 1*(2), 7–18.
Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance, 52*(1), 57–82.
Cochran, W. G. (1934). The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. In *Proceedings of the cambridge philosophical society on mathematical: vol.30* (pp. 178–191). Cambridge University Press.
Connor, G., Goldberg, L. R., & Korajczyk, R. A. (2010). *Portfolio risk analysis*. Princeton University Press.
Fabozzi, F. J., Focardi, S. M., & Kolm, P. N. (2006). Incorporating trading strategies in the Black–Litterman framework. *The Journal of Trading, 1*(2), 28–37.
Fabozzi, F. J., Focardi, S. M., & Kolm, P. N. (2010). *Quantitative equity investing: Techniques and Strategies*. John Wiley & Sons.
Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance, 47*(2), 427–465.
Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3–56.
Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis* (2nd ed.). Taylor & Francis.
Giacometti, R., Bertocchi, M., Rachev, S. T., & Fabozzi, F. J. (2007). Stable distributions in the Black–Litterman approach to asset allocation. *Quantitative Finance, 7*(4), 423–433.
Kolm, P. N., Tütüncü, R., & Fabozzi, F. J. (2014). 60 years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research, 234*(2), 356–371.
Litterman, R., & He, G. (1999). The intuition behind Black–Litterman model portfolios. Goldman Sachs investment management series. Golman, Sachs & Co.
Menchero, J., Morozov, A., & Shepard, P. (2008). The Barra global equity model (gem2), MSCI Barra research notes p. 53. MSCI, Inc.
Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.
Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica: Journal of the Econometric Society, 32*(1/2), 122–136.
Rachev, S. T., Hsu, J. S., Bagasheva, B. S., & Fabozzi, F. J. (2008). *Bayesian methods in finance*: 153. John Wiley & Sons.
Ritter, G. (2016). Stable linear-time optimization in arbitrage pricing theory models. *Risk Magazine, 29*(9), 82–85.
Robert, C. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. Springer Science & Business Media.
Roll, R., & Ross, S. A. (1980). An empirical investigation of the arbitrage pricing theory. *The Journal of Finance, 35*(5), 1073–1103.
Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of economic theory, 13*(3), 341–360.
Satchell, S., & Scowcroft, A. (2000). A demystification of the Black–Litterman model: Managing quantitative and traditional portfolio construction. *Journal of Asset Management, 1*(2), 138–150.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological), 58*(1), 267–288.
Vasicek, O. A. (1973). A note on using cross-sectional information in Bayesian estimation of security betas. *The Journal of Finance, 28*(5), 1233–1239.

Please cite this article as: P. Kolm, G. Ritter, On the Bayesian interpretation of Black–Litterman, European Journal of Operational Research (2016), http://dx.doi.org/10.1016/j.ejor.2016.10.027