Original Article

Explicit coupling of informative prior and likelihood functions in a Bayesian multivariate framework and application to a new non-orthogonal formulation of the Black–Litterman model

Received (in revised form): 24th January 2011

François Ogliaro

is a Pricing and Algorithmic Trading Developer at BNP Paribas and was formerly Mathematical Developer at OCCAM Financial Technology. He is détaché from the Theoretical Chemistry and Biochemistry Laboratory, Université Henri-Poincaré, Nancy, having previously held posts at Imperial College, London, King's College London, the Hebrew University of Jerusalem and the University of Liverpool. He holds a PhD in Chemical Sciences from the Université de Rennes and a Gaussian development licence, and has published over 30 articles on theoretical and quantum chemistry.

Robert K. Rice

is a Chief Executive of OCCAM Financial Technology, which he founded in 1993 to provide flexible, transparent software to financial practitioners wishing to build their own risk models. After graduating from Oxford, he pioneered the use of computers in corporate finance, becoming an advisor to SAMA in 1981, Managing Director of Baring Quantitative Management in 1987 and Director of Investment Technolgy at QUANTEC in 1989. Current research interests include Real Estate, and tools for assessing the sanity of market forecasts.

Stewart Becker

is a Quantitative Developer with Citigroup. He was formerly lead developer at OCCAM Financial Technology. He holds an MA in Computer Science from the University of Cambridge.

Raul Leote de Carvalho

is a Head of Quantitative Strategies and Research at BNP-Paribas Asset Management, Paris. Before joining BNP Paribas, he was a research associate in physics at University College London, Ecole Normale Supérieure de Lyon and Wuppertal Universität. He holds a PhD in Theoretical Physics from the University of Bristol, and an MSc in Condensed Matter Physics from the University of Lisbon.

Correspondence: Robert K. Rice, OCCAM Financial Technology, 220 Tower Bridge Road, London, SE1 2UP, London, United Kingdom
E-mail: robert.rice@occamsrazor.com

ABSTRACT Under an assumption of normality, we explore a non-orthogonal Bayesian technique in which redundant information can in principle be filtered out of the posterior distribution by the explicit coupling of the prior and likelihood functions. The Black–Litterman forecasting model widely used by investment practitioners in various forms is revisited in the light cast by the new technique, and implications for the posterior mean and overall posterior density are examined. A numerical backtest experiment conducted on a portfolio of MSCI sector indices invested using a total return acceleration strategy over the 2003–2007 period sheds some light on the possible benefits of the non-orthogonal approach. Non-orthogonal coupling is found to improve both the future expected returns and the risk model. The resulting

competitive advantage to an investor applying the technique to portfolio construction is then investigated in terms of relative performance within the mean-variance framework. With the present simplified backtest settings, the annual outperformance ranges from 13 to 98 basis points after 36 rebalancing periods, depending on the accuracy of the original forecasts.
Journal of Asset Management advance online publication, 15 September 2011; doi:10.1057/jam.2011.19

Keywords: Bayesian inference; information filtering; forecast; Black–Litterman model; backtesting; portfolio optimisation

INTRODUCTION

Black and Litterman (1990, 1991) (BL) suggest the application of the 'known covariance unknown mean' Bayesian solution to statistical inference to generate expected asset returns $r_\perp \in \mathbb{R}^n$ that are suitable for use in mean-variance Markowitz-type portfolio allocation. In essence, the approach consists of generating one-step-ahead *posterior* returns on the assets by means of a precision-matrix-weighted combination of the investor's *prior* views of their future returns $q_m \in \mathbb{R}^m$ with the distribution of their *implied* returns $\hat{\pi}_n \in \mathbb{R}^n$ obtained by reverse optimisation from the historic covariance and the current market capitalisation of the assets.

Implied returns are preferred in the original BL formulation to the *historic* returns that might have been expected, as the posterior returns are used as an input to a Markowitz-type portfolio allocation exercise, where overweight and underweight positions relative to the benchmark are desirable only to the extent that the investor actually has prior views. In this initial report, this refinement is ignored as it complicates the definition and discussion of the forecasting error at the core of the study, and historic returns are used instead.[1]

Assume an m- and n-variate multinormal distribution, respectively, for the prior $N_m(q_m, \Omega_{m,m})$ and historic $N_n(\hat{\pi}_n, \tau\hat{\Sigma}_{n,n})$ densities with $\tau \in \mathbb{R}$ a user-specified scaling factor, and define a $P_{m,n} \in \mathbb{R}^{m,n}$ matrix that relates priors on combinations of assets to the individual assets eligible for inclusion in the portfolio. The original formulation of the BL model states that the posterior $n \times 1$ vector of expected returns follows a multivariate normal distribution $N_n(r_\perp, c_\perp)$ with parameters

$$c_\perp = \left(\tau^{-1}\hat{\Sigma}^{-1} + P'\Omega^{-1}P\right)^{-1} \qquad (1)$$

and

$$r_\perp = c_\perp\left(\tau^{-1}\hat{\Sigma}^{-1}\hat{\pi} + P'\Omega^{-1}q\right). \qquad (2)$$

The BL approach is attractive because it makes it easy to incorporate prior information into a gently updated vector of one-step-ahead returns, leading the optimiser away from a corner solution and, to the extent that the benchmark is itself diversified, towards a similarly diversified allocation. But it has yet to be adopted as a standard weapon in the asset management arsenal, perhaps because less quantitative practitioners have been reluctant to adopt a procedure whose full specification is still a matter of debate. A brief literature overview will make this clear.

Fusai and Meucci (2003) equate the prior covariance matrix with a scaled historic counterpart, whereas most other studies restrict Ω to its diagonal elements ϖ_{ii}. Bevan and Winkelman (1998) calibrate these ϖ_{ii} elements on the basis of the Z-score of the prior views,[2] whereas, for example, He and Litterman (1999), Idzorek (2007) and Martellini and Zieman (2007) assume that a more or less elaborate functional form of the ϖ_{ii}/τ ratio can be derived from the (diagonal) historical volatility. Idzorek (2007) reports that values of τ vary from 0.01 to 1 from one practitioner to another: compare the value of the deterministic τ in

BL and in Beach and Orlov (2007), but also note that Satchell and Scowcroft (2000) use a stochastic scaling factor. Finally, and at a more fundamental level, Bagasheva *et al* (2006) note that 'there is no consensus *[in the literature]* as to which one of the distributions *[in Eq. 2]* defines the prior and which one the sampling density'.

In addition to these ambiguities in the model specification, two other important issues remain unresolved:

(i) The BL model embodies an empirical Bayesian technique and therefore, as Press (1989) points out, becomes increasingly flawed as the information shared between the historic and informative prior signals increases.
(ii) Although in the original formulation of the model Black and Litterman (1990) restrict themselves to estimating the mean return alone, Satchell and Scowcroft (2000) point out that one can, and in principle should, also use the posterior covariance matrix as a model of risk. Fund managers attempting to do so have found that this extended formulation (BLSS) can result in a dramatic underestimation of the latent risk of individual assets, with a corresponding effect when they are aggregated into portfolios, especially if the prior covariance matrix is diagonal.

In this contribution, we establish that the risk underestimation (ii) originates from overlooking the non-orthogonality of the historic and prior signals to each other (i), and explore a possible solution to the problem. The basic idea consists of recasting the inverse probability theorem as a joint multivariate distribution of the prior and historic likelihood functions so that the double counting of shared information is explicitly taken into account by introducing an off-diagonal *coupling*[3] term. This, we hope, will reduce the strain on the purely orthogonal BLSS model when it is applied to two strongly non-orthogonal signals. This may in turn diminish the need to resort to the *ad hoc* assumptions illustrated above to produce 'sensible' hyperparameters for Equations (1) and (2), and ultimately help the BL framework move towards a more widely accepted specification.

This contribution is organised as follows: Part I introduces a possible new tool for information filtering from a general Bayesian viewpoint and then briefly examines its implications for the first and second moment of the posterior density. Part II presents a first numerical exploration (based on the 10 MSCI global sector indices) of the benefits of a non-orthogonal Bayesian approach for the posterior density. First, the merits of coupling the historic and prior returns are examined by monitoring the forecasting power of the posterior one-step-ahead expected returns. Then the benefits of the coupling for the risk model are examined by backtesting the performance of a minimum variance portfolio invested according to a naïve acceleration strategy. This two-step approach allows the impact of the coupling on the posterior mean to be examined separately from that on the posterior risk model. A more applied empirical study of the benefits of introducing off-diagonal coupling to the BL and BLSS models is forthcoming.

NON-ORTHOGONAL BAYESIAN INFERENCE

The problem to be solved

Consider the likelihood function

$$L(\hat{\pi}_n|\mu_n,\ I_{\mathrm{H}}) \sim N(\mu_n,\ \hat{\Sigma}_{n,n}|I_{\mathrm{H}}) \qquad (3)$$

of an observable $\pi_n \in \mathbb{R}^n$ random vector that follows a normal distribution centred on an unobservable mean μ_n. The sample mean vector $\hat{\pi}_n$ and covariance matrix $\hat{\Sigma}_{n,n}$ together encompass all the *Historical* information I_H contained in the observed historic distribution. Let

$$g(\mu_n) = P_{m,n}\mu_n \sim N(q_m|I_P, \Omega_{m,m}|I_P) \quad (4)$$

be the conjugate prior whose location and dispersion parameters rely solely on the *Prior* information I_P. In (4), $P_{m,n}$ $(m \leqslant n)$ is a projection matrix whose coefficients are known and deterministic.

Assuming (3) and (4) are uncorrelated, Bayes's theorem expresses the posterior distribution $(\mu\,|\,\hat{\pi})$ as the product of the prior and likelihood distributions. As Press 1989 points out, this independence requirement is not fully satisfied in an empirical Bayesian framework when the hyperparameters of the conjugate prior are fitted using information from the observed distribution (I_H). As the $I_H \cap I_P$ intersection expands, the assumption underlying the conditional probability theorem is increasingly violated and $(\mu\,|\,\hat{\pi})$ becomes pathological.

A possible solution

Assume that we know the second-moment $\Gamma_{m,n}$ matrix correlating the L and g vectors.[4]

Theorem *For a joint normal distribution of a likelihood function (3) and a prior (4) correlated by Γ, given $\hat{\pi}$, P, q, and two symmetric positive definite matrices Ω and $\hat{\Sigma}$, the posterior density $(\mu/\hat{\pi})$ when it exists (depending on Γ) is a normal distribution N(r, c) with parameters*

$$c = \left(\hat{\Sigma}^{-1} + (P - \Gamma\hat{\Sigma}^{-1})'(\Omega - \Gamma\hat{\Sigma}^{-1}\Gamma')^{-1} \times (P - \Gamma\hat{\Sigma}^{-1})\right)^{-1}, \quad (5)$$

$$r = c\left(\hat{\Sigma}^{-1}\hat{\pi} + (P - \Gamma\hat{\Sigma}^{-1})'(\Omega - \Gamma\hat{\Sigma}^{-1}\Gamma')^{-1} \times (q - \Gamma\hat{\Sigma}^{-1}\hat{\pi})\right). \quad (6)$$

Proof The proof is a generalisation of the proof in Satchell and Scowcroft (2000) to cover cases where L and g are explicitly correlated by Γ.

Assuming that (3) and (4) follow a joint multivariate normal distribution, the posterior density can be expressed as[5]

$$(\mu|\hat{\pi}) \propto \exp -\frac{1}{2} \times \left(\begin{pmatrix} \mu - \hat{\pi} \\ P\mu - q \end{pmatrix}' \begin{pmatrix} \hat{\Sigma} & \Gamma' \\ \Gamma & \Omega \end{pmatrix}^{-1} \begin{pmatrix} \mu - \hat{\pi} \\ P\mu - q \end{pmatrix}\right). \quad (7)$$

If we now write

$$c^{-1} = \begin{pmatrix} I \\ P \end{pmatrix}' \begin{pmatrix} \hat{\Sigma} & \Gamma' \\ \Gamma & \Omega \end{pmatrix}^{-1} \begin{pmatrix} I \\ P \end{pmatrix}, \quad (8)$$

$$r = c^{-1} \begin{pmatrix} I \\ P \end{pmatrix}' \begin{pmatrix} \hat{\Sigma} & \Gamma' \\ \Gamma & \Omega \end{pmatrix}^{-1} \begin{pmatrix} \hat{\pi} \\ q \end{pmatrix}, \quad (9)$$

and retain just the three terms of (7) in μ while introducing the $\exp - 1/2(r'c^{-1}r)$ constant, we have

$$(\mu|\hat{\pi}) \propto \exp -\frac{1}{2} \times \left((\mu - r)'c^{-1}(\mu - r)\right). \quad (10)$$

Equations (8) and (9) take the form given in *Theorem* when the (symmetric) inner matrix of c^{-1} is expressed using the matrix inversion lemma:

$$c^{-1} = \begin{pmatrix} I \\ P \end{pmatrix}' \begin{pmatrix} \hat{\Sigma}^{-1} + \hat{\Sigma}^{-1}\Gamma'\Phi\Gamma\hat{\Sigma}^{-1} & -\hat{\Sigma}^{-1}\Gamma'\Phi \\ -\Phi\Gamma\hat{\Sigma}^{-1} & \Phi \end{pmatrix} \times \begin{pmatrix} I \\ P \end{pmatrix}$$

writing $\Phi = (\Omega - \Gamma\hat{\Sigma}^{-1}\Gamma')^{-1}$.

Let us now examine the key mechanistic features of (5) and (6).

Fact 1: When $P = I$, $\Omega = \hat{\Sigma}$ are positive definite $n \times n$ matrices and $\Gamma = \gamma\sqrt{\hat{\Sigma}\Omega}$ with $\gamma \in \mathbb{R}$

$$c(\gamma) = \frac{(1+\gamma)}{2}\hat{\Sigma}, \quad (11)$$

$$r = \frac{\hat{\pi} + q}{2}. \quad (12)$$

Fact 1 is a direct application of *Theorem*. It indicates that when the prior and historic signals have identical covariance matrices, the Γ-coupling has no effect on the location vector but scales linearly the posterior covariance. Note that $c(\gamma)$ is positive definite solely on the $\gamma \in [-1, +\infty]$ interval. In finite-precision arithmetic, a posterior density evaluated via (5) will become ill-defined in $\gamma = \pm 1$ as $(\Omega - \Gamma\hat{\Sigma}^{-1}\Gamma')^{-1} \to \infty$.

Fact 2: When $P = I$, $\hat{\Sigma}$ and $\Omega = \xi^2\hat{\Sigma}$ are symmetric positive definite $n \times n$ matrices with $\xi \in \mathbb{R}^+$, and $\Gamma = \gamma\sqrt{\hat{\Sigma}\Omega}$ with $\gamma \in \mathbb{R}$

$$c(\gamma,\ \xi) = \xi^2 \frac{1 - \gamma^2}{1 + \xi^2 - 2\gamma\xi}\hat{\Sigma} \qquad (13)$$

$$r(\gamma,\ \xi) = \frac{1}{1 + \xi^2 - 2\gamma\xi} \times \left(\xi(\xi - \gamma)\hat{\pi} + (1 - \gamma\xi)q\right). \qquad (14)$$

Fact 2 is a direct application of *Theorem*. It shows that $c(\gamma, \xi)$ adopts a bell-shaped curvature for a large part of the definition interval, whereas the location vector is now affected by the strength of Γ-coupling, although it remains independent of the covariance matrices. $c(\gamma, \xi)$ and $r(\gamma, \xi)$ diverge in $\gamma \to (1 + \xi^2)/2\xi$. $c(\gamma, \xi)$ is monotonically increasing on $\gamma \in]-\infty, \min(\xi, 1/\xi)]$ and $\gamma \in [\max(\xi, 1/\xi), +\infty[$ intervals and monotonically decreasing in $\gamma \in [\min(\xi, 1/\xi), (1 + /\xi^2)/2\xi[$ and $\gamma \in\ ](1 + \xi^2)/2\xi, \max(\xi, 1/\xi)]$ intervals. Clearly, $c(\gamma,\ \xi)$ is always positive definite on the $\gamma \in [-1, 1]$ interval. Finally, note the smooth transition of the posterior density between the historic pole at $\gamma = 1/\xi$ (that is $c(1/\xi, \xi) = \hat{\Sigma}$; $r(1/\xi, \xi) = \hat{\pi}$) and the prior pole at $\gamma = \xi$ (that is $c(\xi, \xi) = \xi^2\hat{\Sigma} = \Omega$; $r(\xi, \xi) = q$). From this point of view, the Γ-coupling plays a role similar to the scaling factor τ in the original BL model (1) and (2).

Fact 3: When $P = I$, Ω and $\hat{\Sigma}$ are two symmetric positive definite $n \times n$ matrices such that $[\Omega, \hat{\Sigma}] = 0$, with $\hat{\Sigma} = \Theta\hat{\Delta}\Theta'$, $\Omega = \Theta\Psi\Theta'$ and $\Gamma = \gamma\sqrt{\hat{\Sigma}\Omega}$ with $\gamma \in \mathbb{R}$, then

$$\Gamma = \gamma\sqrt{\hat{\Sigma}}\sqrt{\Omega} = \gamma\sqrt{\Omega}\sqrt{\hat{\Sigma}}, \qquad (15)$$

$$c(\gamma,\ \hat{\Sigma},\ \Omega) = \Theta\Phi\Theta'. \qquad (16)$$

$$\Phi = \left(1 - \gamma^2\right) \times \left(\hat{\Delta}^{-1} + \Psi^{-1} - 2\gamma\sqrt{\hat{\Delta}}^{-1}\sqrt{\Psi}^{-1}\right)^{-1}. \qquad (17)$$

Equation (15) is obvious as, for any given function f, $[\Omega, \hat{\Sigma}] = 0$ implies $[f(\Omega), f(\hat{\Sigma})] = 0$. Observe that a similar relationship does not exist when Ω and $\hat{\Sigma}$ do not commute, as the $\Omega\hat{\Sigma}$ matrix is no longer symmetric and therefore cannot be factored as a $\Gamma\Gamma'$ product (even if under certain conditions it has a square root). Equation (16) is obvious as all individual $\mathbb{R}^{n \times n}$ terms of equation (5) are diagonalisable by the same unitary transformation Θ, as is the posterior covariance. Equations (16) and (17) are a direct application of *Theorem*. Denoting the (necessarily positive) individual diagonal entries of $\hat{\Delta}$ and Ψ by $\hat{\delta}_{ii}$ and ψ_{ii}, it is easy to see using equation (17) that the posterior covariance is positive definite if the n conditions

$$\frac{1}{1 - \gamma^2}\left(\psi_{ii} + \hat{\delta}_{ii} - 2\gamma\sqrt{\hat{\delta}_{ii}}\sqrt{\psi_{ii}}\right) > 0 \qquad (18)$$

are all simultaneously satisfied for $i \in [1, n]$. Overall, (18) is not very informative, indicating that even for a simple case where $[\Omega, \hat{\Sigma}] = 0$ it is difficult to anticipate the domain of γ on which c is positive definite.

Note that the drastic restrictions used to derive *Facts* 1–3 will almost never be satisfied by real-life applications. However, the key results established above will facilitate the interpretation of the results of complex non-orthogonal inference problems for which, say, the prior and historic matrices are close to being equal, or a scaled version of each other, or commute.

Discussion

Let us examine the implications of *Fact* 1. *Fact* 1 analyses the effect of setting the prior covariance matrix equal to its historical (likelihood) counterpart.

For $\gamma \to 1$, the risk matrix remains unchanged. This is reasonable: when no extra information is available (that is, $I_{\mathrm{H}} \cup I_{\mathrm{P}} = \{I_{\mathrm{P}}\} = \{I_{\mathrm{H}}\}$), historic and prior are fully correlated and the output of the prediction process is identical to its input.

For $\gamma = 0$, following (11), the variance is divided by two. Alternatively stated, the classical (that is, orthogonal) Bayesian inference assumes $I_{\mathrm{H}} \cap I_{\mathrm{P}} = \{\ \}$ and uses the whole of the new information provided by the prior to reduce the dispersion of $(\mu \mid \hat{\pi})$, so that the addition of the prior exactly doubles the information available. However, if the information is incomplete, that is if $(I_{\mathrm{H}} \cup I_{\mathrm{P}})^c \neq \{\ \}$, some dispersion remains.

When $\gamma \to -1$, the risk associated with the random process simply vanishes. This is as expected in a situation where the information available would sum up to $I_{\mathrm{U}} = (I_{\mathrm{H}} \cup I_{\mathrm{H}}^c)$. The collapse of the forecast to a deterministic value (which can also be produced by iterating the standard Bayesian inference process) is clearly undesirable when trying to apply Bayesian inference to strongly stochastic processes.

For $\gamma < -1$, c cannot be interpreted as a covariance matrix as it is negative definite and the model is better abandoned. For $\gamma > 1$, the risk matrix can be made arbitrarily large and in particular larger than the original $\hat{\Sigma}$, which was not possible in the original orthogonal formulation.

In the light of these remarks, the off-diagonal Γ-coupling term can be given the following interpretations: Γ relates to

1. the extent of information sharing between historic (3) and prior (4),
2. the rate of convergence (or degree of predictability) of the posterior mean towards a deterministic value,
3. the degree of predictability of the posterior mean, depending on the size of its variance.

AN INITIAL NUMERICAL EXPLORATION OF THE NON-ORTHOGONAL FORECASTING PROCEDURE

The non-orthogonal model

The original BL/BLSS model formulated in (1) and (2) is the special case $\Gamma = 0$ of a more general non-orthogonal formulation of the Black–Litterman model (NOBL) defined by (5) and (6): it assumes that the priors are completely uninformed by history. This hypothesis rarely holds with respect to returns forecast by financial analysts. Even the more subjective or qualitative views of fund managers are typically informed to some extent by the history of the asset returns in question. The non-orthogonality of historic and prior is even more evident when the priors are estimated directly from the historic signal using more quantitative techniques such as GARCH in Beach and Orlov (2007).

Thus, a technique for the explicit filtering out of redundant information has the potential to be useful to most fund managers currently using methods based on Bayesian inference. It now remains to be seen how it performs on real market data. In particular, how does it impact the posterior mean (the BL model) and the overall posterior density (the BLSS model)? To begin to answer these questions, we carried out the simplified (proof-of-concept) backtest simulation presented in the next section.

Numerical experiment settings

The universe consists of the $N = 10$ global MSCI sector indices in 10 times-series (TS) of monthly returns for the period from 1 January 1999 to 31 December 2007.

The raw material for our numerical experiments consists of 36 monthly posterior densities $N(r_t, c_t)$ for each of the two models. For the original BLSS model, r_t and $c_t(t=1, \ldots, T=36)$ are computed using equations (1) and (2) with $\tau=1$, and for the NOBL model using (5) and (6), in both cases for the period beginning 1 January 2005. The historic parameters $\hat{\pi}_t$ and $\hat{\Sigma}_t$ appearing in (1), (2), (5) and (6) are estimated from the sample using all historical observations available up to that date (see Note 1).

There is one view per position (so P in (5) and (6) is the identity matrix). Following an acceleration strategy, the vector of the $N=10$ mean values (in fact their 'instantaneous' value, so shown here without the '$\wedge$' character) of the prior signal at time t for the period $t+1$ is denoted $q(t+1)^{(t)}$ (or $q(t+1)_t$) and computed as

$$\begin{aligned} q(t+1)^{(t)} &= q(t+1)_t \\ &= \zeta(\pi_{t-1} + \chi(\pi_t - \pi_{t-1})) \\ &\quad + (1-\zeta)\pi_{t+1} \end{aligned} \tag{19}$$

Note that π_{t-1}, π_t and π_{t+1} in (19) are three $N \times 1$(that is, 10×1) vectors of instantaneous historic returns observed at $t-1$, t and $t+1$, and $(\zeta, \chi) \in \mathbb{R}^2$ two user-specified scaling factors, so that

- when $\zeta=0$, the prior forecasts are made with perfect foresight, without reference to history;
- when $\zeta=1$, the priors are based only on history; and
- when $0<\zeta<1$, the priors are based partly on history, partly on foresight[6]

and in the last two cases:

- if $\chi>0$, they implement a naïve momentum or acceleration strategy;
- if $\chi<0$, they implement a naïve reversal strategy.

For the purposes of the analysis that follows, we have focused on 3 types of prior signals: '100H0F' when $\zeta=100$ per cent, '80H20F' when $\zeta=80$ per cent and '20H80F' when $\zeta=20$ per cent. In each case, we have assumed an acceleration strategy with χ set to 2. Once again, many alternative strategies are possible and a few of them will be examined in detail in the forthcoming empirical study.

The confidence levels of the priors ($\hat{\Omega}$) used to compute $N(r_t, c_t)$ are calculated as a sample covariance matrix estimated directly from the 10 prior TS generated using (18). The off-diagonal coupling term is written as

$$\hat{\Gamma} = \gamma \hat{\Gamma}_{\mathrm{H,F}} \tag{20}$$

where $\gamma \in \mathbb{R}$ is specified by the user and the (generally) asymmetrical matrix $\Gamma_{\mathrm{H,F}}$ is calculated directly from the historic and prior TS. Note that the manifold assumptions and parameters required to provide an adequate specification of the BL model discussed in the literature review above are replaced in the NOBL model by a single parameter γ, which by contrast to the original τ from the BL model can be given a clear interpretation in terms of redundant information filtering. Otherwise, the model as it stands is fully specified by the data.

The first numerical experiment (used to produce Figure 1) examines the forecasting

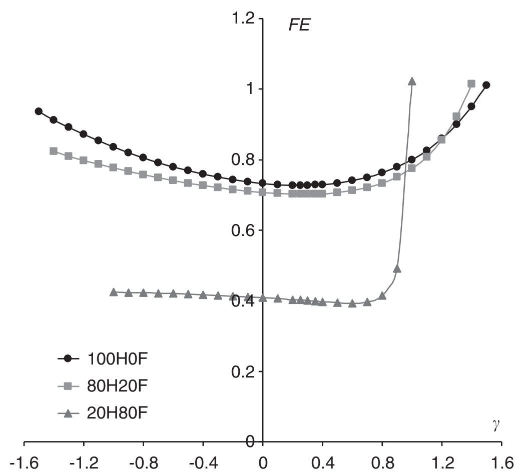

Figure 1: Forecasting error *FE* as a function of the Γ-coupling for prior returns with a decreasing historic component. The minimum *FE* for the 100H0F, 80H20F and 20H80F cases is found at $\gamma_{\underline{FE}} = 0.25$, 0.30 and 0.60, respectively.

power of the posterior one-step-ahead expected returns r_t computed for the $N = 10$ MSCI sectors. This is achieved by computing the Forecast Error (*FE*) defined as the L_2 norm of the difference between the posterior and the observed one-step-ahead returns for each of the N individual MSCI positions for each of the 36 investment periods:

$$FE = \sqrt{\sum_{t=1}^{n=36} (r_t - \pi_{t+1})'(r_t - \pi_{t+1})}. \quad (21)$$

The optimum filtering of redundant information achievable for a given problem is defined here as $\gamma_{\underline{FE}} = \arg\min_\gamma(FE)$ that is, obtained by simple application of the maximum likelihood principle. Finally, observe that *FE* is computed without any reference to a portfolio or to the mean-variance Markowitz framework: this first part of our numerical experiment is a general exercise in TS forecasting independent of any asset allocation problem.

The benefit of improved second moments is more difficult to establish as there is no natural benchmark for the one-step-ahead expected covariance matrix (c). However, we can obtain a first and rather *ad hoc*, but still useful, *indication* of the extent of dispersion of the posterior density (shown in Figure 2) by simply computing the total risk

$$\sigma_{1/N}(Post) = w'_{1/N} c w_{1/N}, \quad (22)$$

of the so-called $1/N$ portfolio that is fully diversified over the 10 MSCI sectors. The value of the off-diagonal scaling that maximises the posterior dispersion is denoted by $\gamma_{\overline{\sigma}} = \arg\max_\gamma(\sigma_{1/N})$

Another way of assessing the impact of the non-orthogonal coupling on the posterior covariance is to run a multi-period allocation backtest and measure the outperformance. Typically, we are interested in comparing a portfolio's instantaneous ($\mu(t)$) (Figure 3) and accumulated ($\mu^{acc}(t)$)returns (Figure 4) for each period of the backtesting simulation for different values of γ.

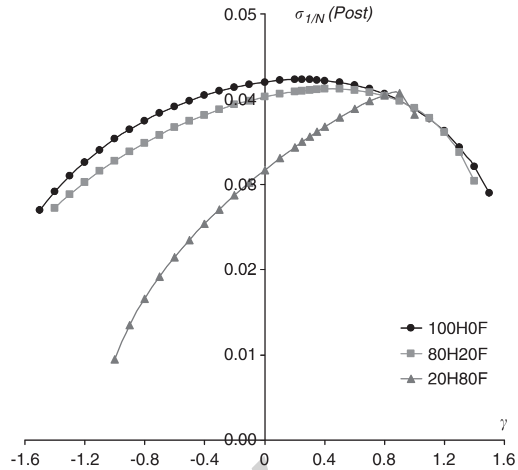

Figure 2: Risk $\sigma_{1/N}$(*Post*) of the posterior distribution as a function of the Γ-coupling for prior returns with a decreasing historic component. The maximum $\sigma_{1/N}$(*Post*) risk for the 100H0F, 80H20F and 20H80F cases is found at $\gamma_{\overline{\sigma}}$ = 0.25, 0.50 and 0.90, respectively.

The backtest starts on 1 January 2005 and runs for 3 years with rebalancing taking place on the last day of each month. The backtest engine is required to propagate the minimum variance portfolio: for each of the $t = 1, \ldots, T = 36$ backtest periods, the Markowitz-type optimiser takes the non-orthogonal posterior density $N(r_t, c_t)$ as input and produces a minimum variance portfolio. The minimum variance portfolio is a good candidate for our study as (i) it always exists and (ii) does not depend on the vector of means, and thus we can study the benefits of our approach for the posterior returns separately from its benefits for the posterior risks. We opted against using a benchmark as the variable weights of the benchmark will introduce an extra source of uncertainty that will complicate our analysis. To ensure that the allocated portfolio is not too concentrated (owing to the use of historic returns in lieu of implied returns) (see Note 1), the portfolio individual weights w_i are constrained to vary within the -10 per cent $< w_i < 30$ per cent interval with $i = 1, \ldots, N$. It must be emphasised that the application in this note is simply a numerical exploration of the model and not a true empirical study, which will follow.

RESULTS

The benefits of applying Γ-coupling to the BL and BLSS models are inferred from the computation of the forecasting error *FE*, the posterior risk $\sigma_{1/N}(Post)$, and the instantaneous and accumulated returns shown in Figures 1–4.[7]

Over a relatively large domain of the parameter γ (here roughly $\gamma \in [-1; 0.8]$), the *FE* shown in Figure 1 reassuringly tends to decrease as the historic component of the prior decreases, and the foresight component increases. Irrespective of the prior actually used, there always exists some value of the Γ-coupling that improves the forecasting accuracy of the posterior. The optimum information filtering in the 100H0F, 80H20F and 20H80F cases is found at $\gamma_{\underline{FE}} = 0.25$, 0.30 and 0.60, respectively. Thus, in the present backtest experiment based on MCSI data, $\gamma_{\underline{FE}}$ is found to be an increasing function of the knowledge of the one-step-ahead realised returns. This suggests that the Γ-coupling controls the portion of historical signal filtered out of the posterior density. A similar role is played by τ in the orthogonal BL model (1) and (2).

Interestingly, $\gamma_{\underline{FE}} \neq 0$ even in the 100H0F case. This is not totally surprising and can be given two equivalent interpretations: the TS of future realised MSCI returns can be predicted from the historical prices or, alternatively, the simple acceleration strategy expressed by equation (19) is to some extent successful in capturing future market moves.

Figure 2 sketches the dispersion of the posterior density as a function of the strength of the Γ-coupling when prior returns have a decreasing historic component. The highest risk (the maximum of the parabola-like curves) for the 100H0F, 80H20F and 20H80F cases is found at $\gamma_{\underline{FE}} = 0.25$, 0.50 and 0.90, respectively. The Γ-coupling introduced in the NOBL model is therefore successful in reducing the risk underestimation that results from the orthogonal model. Note that for priors with a realistically heavy historic component, $\gamma_{\bar{\sigma}}$ and $\gamma_{\underline{FE}}$ almost coincide.

For both the 100H0F and 80H0F cases, $\sigma_{1/N}(Post)$ is a broad bell-shaped curve with a positive maximum $\gamma_{\bar{\sigma}}$. This is consistent with the analytical form of $c(\gamma, \xi)$ given in (13), which applies when the prior covariance matrix can be approximated by $\Omega \approx \xi^2 \hat{\Sigma}$. With our particular settings, it is obvious from (19) that for example, $\xi^2 \approx 4$ for the 100H0F case. In the 20H80F case, on the other hand, $\sigma_{1/N}(Post)$ scales almost linearly with $\gamma_{\bar{\sigma}}$ as described in *Fact 1*. When the prior has a small historic component, the historic and prior covariance matrices tend to be equal to each other and (11) is quasi-satisfied. Equation (12) is also quasi-satisfied, as can be seen from the rather stagnant *FE* shown in Figure 1, for a large portion of the definition interval.

Figure 3 monitors over the 36 backtest periods a variety of instantaneous (per period) *ex post* outperformance $\Delta\mu(t)$ measures for the minimum variance portfolio for each of the three cases, using $\gamma_{\underline{FE}}$, the value of the coupling constant associated with the minimum *FE* in each case. Figure 4 provides the same information on an accumulated basis.

In our notation:

- $\mu(H)$ denotes the return on a portfolio allocated using the historical covariance matrix (that is, without resorting to BL);
- $\mu(\gamma = 0)$ denotes the return on a portfolio allocated using the standard BLSS covariance matrix (that is, computed using a gamma coupling with γ set to 0);
- $\mu(\gamma_{\underline{FE}})$denotes the return of a portfolio allocated using the NOBL covariance matrix computed using the gamma coupling with the γ set to the value $\gamma_{\underline{FE}}$ which minimises the forecasting error

It follows that

- when $\Delta\mu(t) = \mu(\gamma_{\underline{FE}}) - \mu(\gamma = 0) > 0$ the NOBL model outperforms the original BLSS formulation;
- when $\Delta\mu(t) = \mu(H) - \mu(\gamma = 0) > 0$ the historic risk model outperforms its original BLSS counterpart, and;
- when $\Delta\mu(t) = \mu(\gamma_{\underline{FE}}) - \mu(H) > 0$ the NOBL model outperforms the historical risk model

and vice versa.

It can be seen from Figure 3 that the outperformance measures are random variables all of which change sign. None of the three estimators of the risk model strongly dominates its analogues in a Markowitz mean–variance sense. Interestingly, the volatility of the $\Delta\mu(t)$ measures tends to increase as the historic component of the priors decreases, as we pass from 100H0F through 80H20F to 20H80F. This is consistent with intuition: the importance of the choice of the risk model will increase with the increasing quality of the forecast signals. Alternatively stated, when the forecasts do not contain any valuable, that is, non-historical information, the three risk models should yield essentially similar results because the historical signal cannot be improved upon.

The accumulated outperformances $\Delta\mu^{acc}(t)$ are shown in Figure 4. Some comments are appropriate.

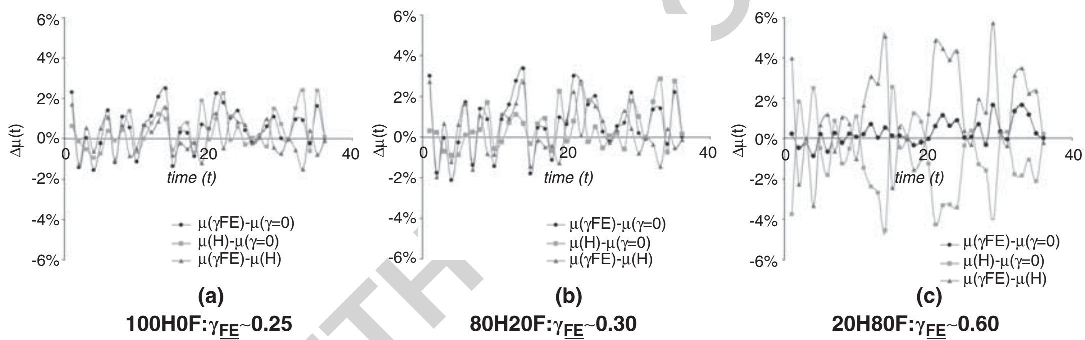

Figure 3: Difference in the monthly instantaneous and portfolio returns $\Delta\mu(t)$ computed for the 36 monthly backtest periods for each of the three cases

- A positive $\Delta\mu(t) = \mu(\gamma_{\underline{FE}}) - \mu(\gamma = 0)$ means that the NOBL model outperforms the original BLSS formulation.
- A positive $\Delta\mu(t) = \mu(H) - \mu(\gamma = 0)$ means that the historic risk model outperforms its original BLSS counterpart.
- A positive $\Delta\mu(t) = \mu(\gamma_{\underline{FE}}) - \mu(H)$ means that the NOBL model outperforms the historic risk model.

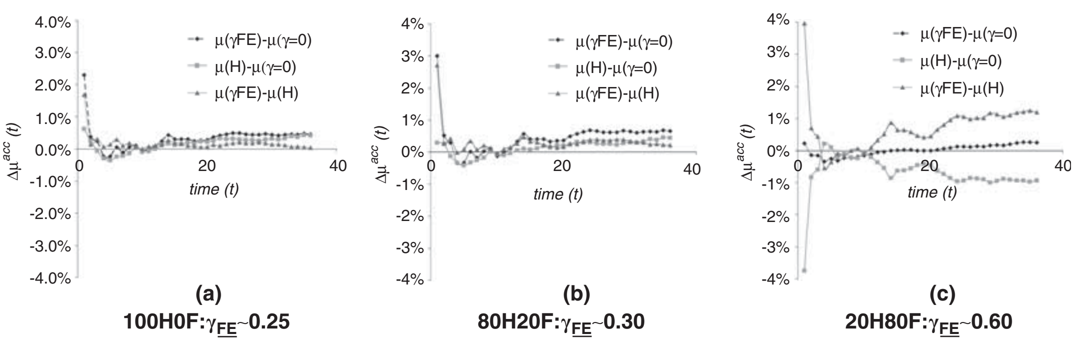

Figure 4: Difference in the annualised portfolio returns $\Delta\mu^{acc}(t)$ accumulated over the 36 monthly backtest periods

- A positive $\Delta\mu(t) = \mu(\gamma_{\underline{FE}}) - \mu(\gamma = 0)$ means that the NOBL model outperforms the original BLSS formulation.
- A positive $\Delta\mu(t) = \mu(H) - \mu(\gamma = 0)$ means that the historic risk model outperforms its original BLSS counterpart.
- A positive $\Delta\mu(t) = \mu(\gamma_{\underline{FE}}) - \mu(H)$ means that the NOBL model outperforms the historic risk model.

First, observe that when the historic component of prior returns is high, as in Figure 4(a) and (b), that $\mu(H)-\mu(\gamma=0)$ is generally positive, that is the historic model outperforms BLSS; it is only when the historic component is low and the foresight component high, as in Figure 4(c), that $\mu(H)-\mu(\gamma=0)$ is generally negative and BLSS represents an improvement. On the basis of this particular data set at least, it is therefore preferable not to use the BLSS risk model but to stick with the historical risk model unless the priors contain a high degree of genuine foresight, which, we submit, will rarely be the case. Alternatively stated, the BLSS approach is found to be useful only when the forecasts contain significantly more than redundant historical information.[8]

Second, note that when the historic component of prior returns is high, as in Figure 4(a) and (b), that $\mu(\gamma_{\underline{FE}}) - \mu(\gamma = 0)$ is generally positive, and that $\mu(\gamma_{\underline{FE}}) - \mu(H)$ is rather small, if also generally positive. That is, the non-orthogonal filtering in NOBL has improved the results obtained using BLSS and renders them of similar quality to those computed using only the historical risk model. Alternatively stated, the non-orthogonal filtering has compensated for the error made when choosing the risk model. Once again, BLSS appears to be counterproductive when the priors do not show any significant foresight. In contrast, when the priors do show significant foresight as in Figure 4(c), both $\mu(\gamma_{\underline{FE}}) - \mu(\gamma = 0)$ and $\mu(\gamma_{\underline{FE}}) - \mu(H)$ remain consistently positive after the first few periods. This suggests that, in the regrettably rarer cases when priors are of good quality, the NOBL approach may add significant value.

We now turn to the effect of using the different approaches in allocating the minimum variance portfolio. Detailed examination of the results shown in Figure 4 indicates that, in terms of this initial exploration, a NOBL fund manager might over a 3-year period outperform a BLSS competitor by anything from 10 to 63 basis points per annum and a colleague relying only on historical data by up to 98 basis points. Such yearly outperformance may appear at first sight rather small, but we found[9] that it is statistically significant and, indeed, if maintained on a consistent basis, it would be regarded by any practising fund manager as high. We should remember that:

1. the current asset allocation process may not necessarily convert the full competitive advantage of the approach into outperformance;
2. the results are based on only 36 rebalancing periods, which leaves little room for a weak stochastic dominance to express itself, as outperformance accumulates only slowly and more dramatic results could possibly be obtained by considering a longer horizon;
3. the minimum risk portfolio is relatively stable with respect to the risk model (it does not depend on the mean, which is responsible for most changes in weights);
4. the number of degrees of freedom of the asset allocation problem is small. There are only 10 positions in the universe and the weights have been constrained (see ‘Note 1’). This reduces the variation in portfolio allocation and leaves less room for possible outperformance.
5. as already noted, we monitor here only that outperformance associated with the improved risk model. We do not take into consideration the combined effect of the improving the return model as well.

As stated in the introduction, this clearly calls for a detailed empirical study that will examine the merits of the NOBL formulation for a realistic Black Litterman investor. However, our preliminary findings are encouraging.

CONCLUDING REMARKS

In this contribution, we propose and investigate a possible solution to the problem of information redundancy facing fund managers who seek to use Bayesian methods to improve the quality of forecast inputs that, intentionally or otherwise, rely significantly on historical data. The key idea consists of recasting the inverse probability theorem into a joint multivariate distribution of the prior and historic likelihood functions, which can thus be explicitly correlated via a non-orthogonal Γ-coupling term.

We show that the strength of the Γ-coupling controls the amount of redundant information that can be filtered out of the posterior density. In particular when the prior does not contain any information the posterior density is not artificially shrunk as is the case with classical (that is orthogonal). Bayesian inference approaches.

The Black–Litterman model, particularly in the form suggested by Satchell and Scowcroft, is revisited in this light. The benefits of the Γ-coupling are examined separately for posterior returns and posterior risks. In our numerical experiment based on MSCI sector indices, we find that non-orthogonal coupling reduces the forecasting error and redresses an otherwise underestimated dispersion of the one-step-ahead expected return distribution. A simple backtest experiment using a mean-variance optimiser shows how an investor can convert this model improvement into outperformance. Overall, we find that the qualitative improvement brought about by the Γ-coupling is tangible and possibly promising.

Finally, we suggest that the diversity of current specifications of the Black–Litterman model is a reflection of the inherent incapacity of any orthogonal formulation to reconcile two non-orthogonal signals, and that the explicit treatment of the information overlap in the non-orthogonal approach may help the model towards a less *ad hoc* specification.

NOTES

1. The portfolios produced by using historic returns in place of implied returns in mean-variance optimisation tend to be under-diversified. Diversification can be easily reintroduced by imposing constraints on the weights (see Section 'Numerical experiment settings').
2. Deriving the precision from a *Z*-score calculated from the historic distribution reduces the impact of the most extreme views. This is a desirable feature of a forecasting procedure operating in slow-moving conditions, but less so in rapidly changing markets.
3. Although this is indeed a form of covariance, we prefer the term coupling throughout this note to distinguish it from the off-diagonal covariance *within* the historic matrix and its prior counterpart. To our knowledge, the only earlier mention in the literature of an explicit coupling between the prior and the historic sample information is in the Ledoit and Wolf (2003) study of the shrinkage of sample covariance matrices.
4. Dimension indices and dependence information are omitted from the notation from now on.
5. The relative sign of the two components of the input vector (that is, $\pm(\mu - \pi)$ versus $\pm(q - P\mu)$) is adjusted to ensure an (intuitive) positive $(+\Gamma)$ correlation of the prior and the likelihood. Strictly speaking, if two distributions such as $\pm(\mu - \pi)$ and $\pm(q - P\mu)$ are normal, it does not imply that they are jointly multivariate normal, unless they are also independent.
6. Of course, in addition to history and foresight, the priors may be based on other (implicitly erroneous) sources of information. This possibility is not examined in the present experiment.
7. All Bayesian inference and backtest calculations have been carried out using the *POW!* 1.22.12 (2010) suite of software produced by OCCAM Financial Technology.
8. In our present setting, we focus on the very special case where the BLSS risk model is used but not, as in the BLSS approach proper, the corresponding mean. It will be interesting to determine whether the conclusion here that the BLSS approach is counterproductive for poor quality priors continues to hold when the mean is taken into account as well. We are not aware of any empirical studies investigating this point, but we have found from exploring a variety of data sets that it may well be the case (Ogliaro and Rice, in preparation).
9. To assess the statistical significance, we used the following *ad hoc* procedure. We first computed the standard deviation σ of the instantaneous $(\Delta\mu(t))$ and cumulated $\Delta\mu^{acc}(t)$ outperformance during each of the 18 months constituting the second half of the backtest window. σ was then used in conjunction with the last expected value of the outperformance $(\Delta\mu(t=36)$ or $\Delta\mu^{acc}(t=36))$ to carry out a two-tailed *Student-t* test. We found that the null hypothesis *H0* (the outperformance was not different from zero) was consistently rejected with a confidence level superior to 99.99 per cent.

REFERENCES

Bagasheva, B., Rachev, S.Z., Hsu, J. and Fabozzi, F. (2006) Bayesian Applications to the Investment Management Process. Santa Barbara: University of California. Technical

Reports of the Department of Statistics and Applied Probability.
Beach, S. and Orlov, A. (2007) An application of the Black–Litterman model with EGARCH-M-derived views for international portfolio management. *Financial Markets and Portfolio Management* 21: 147–166.
Bevan, A. and Winkelman, K. (1998) Using the Black–Litterman Global Asset Allocation Model: Three Years of Practical Experience. *Global Fixed Income Portfolio Strategy*: Goldman Sachs.
Black, F. and Litterman, R. (1990) Asset allocation: Combining investor views with market equilibrium. *Fixed Income Research*: Goldman Sachs.
Black, F. and Litterman, R. (1991) Global Asset Allocation With Equities, Bonds, and Currencies. *Fixed Income Research*: Goldman Sachs.
Fusai, G. and Meucci, A. (2003) Assessing views. *Risk* (March 2003) 16.3: S18–S21.
He, G. and Litterman, R. (1999) The Intuition Behind Black–Litterman Model Portfolios. *Investment Management Research*: Goldman Sachs.
Idzorek, T.M. (2007) A step-by-step guide to the Black–Litterman model. In: S. Satchell (ed.) *Forecasting Expected Returns in the Financial Markets*. Amsterdam: Elsevier.
Ledoit, O. and Wolf, M. (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10: 603–621.
Martellini, L. and Zieman, V. (2007) Extending Black–Litterman analysis beyond the mean-variance framework. *Journal of Portfolio Management* 33(4): 33–44.
Press, S.J. (1989) *Bayesian Statistics: Principles, Models, and Applications*. New York: Wiley, p. 43.
Satchell, S. and Scowcroft, A. (2000) A demystification of the Black–Litterman model: Managing quantitative and traditional construction. *Journal of Asset Management* 1: 138–150.