Beyond Black–Litterman: *Letting the Data Speak*

GUOFU ZHOU

GUOFU ZHOU
is a professor of finance at Olin Business School, Washington University in St. Louis, MO.
zhou@wustl.edu

Since the seminal work of Markowitz [1952], the mean-variance framework has been widely used in practice. However, as is the case with any model, the true parameters (the expected returns and risks) are unknown and have to be estimated from data. Based on standard estimates, the optimized solution takes on unusually large long and short positions when no portfolio constraints are imposed, and takes on many zero positions when short-selling constraints are imposed. Given that the parameter estimation errors are nontrivial, and that the economic intuition of these extreme positions is not obvious, it is hence difficult to implement the optimized mean-variance solution in practice.

The Black–Litterman model (BLM), originated by Black and Litterman [1992], provides a novel solution to the problem. It assumes that the investor starts with views of the market, then updates them with their own views via the Bayesian rule. In a typical case, the market views are taken as the equilibrium expected returns implied by the capital asset pricing model, or CAPM. If their views are the same as the market, and if the investor treats the equilibrium expected returns as the true expected returns, he should hold the equilibrium or value-weighted market portfolio. When the investor does have a view on the relative performance of the assets, the BLM provides a way for the investor to tilt his optimal portfolio away from the market according to the strength of his view. Hence, the key insight of the BLM is to blend the investor's proprietary views with those of the market to construct a portfolio that is optimal in a certain sense. Since its publication, the BLM has generated wide applications and a number of studies. Recently, for example, Fabozzi, Focardi, and Kolm [2006] incorporated various trading strategies into the model, and Martellini and Ziemann [2007] extended it beyond the mean-variance framework and applied it to hedge funds.

The theoretical basis of the BLM is Bayesian learning. The investor starts with views of the market as represented by an equilibrium model, then updates them with his own proprietary views based on the Bayesian rule. This is different from the usual Bayesian decision making where the learning takes place in light of the data—the observed asset returns tell us how they have behaved in the past and how they might behave in the future. On the one hand, theoretically, with an infinite amount of data, we can learn about the true expected returns, rendering the equilibrium model unnecessary. On the other hand, with a limited amount of data available in practice, estimates of the model parameters can be noisy and an equilibrium model can help provide useful information to tie down the parameters. In short, the BLM makes clever use of an equilibrium model, but ignores the data or the

data-generating process of the asset returns. To our knowledge, subsequent studies of the BLM also ignore the role that the data play. The relation to data is an important missing link of the BLM, because the data are both economically and statistically crucial in understanding the dynamics of asset returns, including, in particular, their expected returns and risks, which are important for investment decisions.

In this article, we extend the BLM to allow the Bayesian learning to use all available information: the equilibrium model, the investor's views, and the data. Hence, we link the BLM to quantitative models of asset returns and to the vast academic literature on Bayesian decision making. Simple intuition explains the idea—the BLM can be understood as blending current research with long-run market equilibrium. Our approach combines these two insights to provide a useful Bayesian prior and to combine this prior with the current evolution of the market, say, a bull or bear phase, which ultimately determines the future return on the optimized portfolio.

Rather than sequentially combining the equilibrium model and the investor's views/priors, we provide in addition an alternative Bayesian model averaging approach. Under each of the priors, a Bayesian portfolio decision framework is well defined that updates the prior by data. The two priors imply, however, two Bayesian models for portfolio decisions. The Bayesian model averaging further utilizes the data to obtain data-dependent weights and averages them into a new model. It is this model that combines all the priors and data that are used in the final portfolio choice.

The article is organized as follows. We first review the portfolio choice problem and discuss the BLM, then provide a simple extension of the BLM. Next, we use a numerical example to illustrate the basic ideas supporting our approach. And before concluding, we present some relatively more complex generalizations of the BLM.

MEAN-VARIANCE PORTFOLIO

Consider the case of an investor who allocates funds among N risky assets and a riskless asset. Let R_t be an N-vector of returns on the risky assets in excess of the riskless asset at time t. The standard assumption regarding the probability distribution of R_t is that R_t is independent, identical, and normally distributed (i.i.d.) over time, with mean μ and covariance matrix Σ. We retain this assumption, which will be relaxed later when we discuss the various generalizations.

In the mean-variance framework, the investor's portfolio choice problem of maximizing the expected return for a given level of risk is equivalent to selecting portfolio weights w to maximize mean-variance utility,

$$U(w) = E[R_{T+1}] - \frac{\gamma}{2}\mathrm{Var}[R_{T+1}] = w'\mu - \frac{\gamma}{2}w'\Sigma w \quad (1)$$

where T is today's time, $T + 1$ is the future period, and γ is the investor's coefficient of relative risk aversion. The problem has the standard solution,

$$w = \frac{1}{\gamma}\Sigma^{-1}\mu \quad (2)$$

This is the optimal allocation per dollar of funds across N risky assets, with the balance invested in the riskless asset.

However, in practice, μ and Σ have to be estimated to compute w. The standard approach is to estimate them using the sample mean and sample covariance matrix of the asset returns, based on data available from period 1 to T. The problem is that when the estimates are treated as if they are the true parameters and are plugged into Equation (2), the resulting optimal portfolio weights—the optimized solution—often generate unusually large long or short positions when no portfolio constraints are imposed and many zero positions when short-selling constraints are imposed. This is documented widely in the literature and demonstrated here by the example provided later. Alternative estimates of μ and Σ have also been developed over the years. Kan and Zhou [2007] provided some recent developments and references.

BLACK–LITTERMAN MODEL

Black and Litterman [1992] provided an elegant solution to the extreme position problem arising from using the plug-in portfolio weights. Their idea was to start from the value-weighted index of risky assets, which we refer to, for simplicity, as stocks. In equilibrium, if the CAPM is true, and if all investors hold the same view and have the same risk aversion, the demand for risky assets should be exactly equal to the outstanding supply. Hence, the portfolio weights should be given by the optimal portfolio weights formula, Equation (2). This implies that the equilibrium expected excess returns

(or the equilibrium risk premium, as it is often referred to in the literature) must satisfy

$$\mu^e = \gamma \Sigma w_e \tag{3}$$

where γ is the market risk-aversion coefficient; w_e is the equilibrium portfolio weights, or weights of the value-weighted stock index; and Σ is the covariance matrix of asset returns often estimated by the sample covariance matrix or by a risk model. In practice, the equilibrium expected returns are easily evaluated from Equation (3).

Black and Litterman [1992] took a Bayesian approach, referred to as a *mixed estimation strategy*, to learn the value of the true expected excess return μ, which is the critical input for the mean-variance portfolio choice decision. Recall that a Bayesian regards all parameters as random variables. Hence, μ is naturally assumed to be normally distributed with mean μ^e,

$$\mu = \mu^e + \epsilon^e, \quad \epsilon^e \sim N(0, \tau\Sigma) \tag{4}$$

where ϵ^e is the deviation of μ from μ^e that is normally distributed with zero mean and covariance matrix $\tau\Sigma$, and τ is a scalar indicating the degree of belief in how close μ is to the equilibrium value. In the absence of any views on future stock returns, and in the special case of $\tau = 0$, the investor's portfolio weights must be equal to w_e, the weights of the value-weighted index.

However, an active portfolio manager or an informed investor is likely to have views on μ that are different from μ^e in a substantial way. Black and Litterman [1992] showed that views on the relative performance of stocks can be represented mathematically by a single vector equation,

$$P\mu = \mu^v + \epsilon^v, \quad \epsilon^v \sim N(0, \Omega) \tag{5}$$

where P is a $K \times N$ matrix summarizing K views, μ^v is a K-vector summarizing the prior means of the portfolio views, and ϵ^v is the residual vector. The covariance matrix of the residuals, Ω, measures the degree of confidence the investor has in his views.

Applying the mixed estimation to blend both the equilibrium relationship and the investor's views, Equations (4) and (5), Black and Litterman [1992] obtained the Bayesian updated expected returns and risks as

$$\bar{\mu}_{BL} = [(\tau\Sigma)^{-1} + P'\Omega^{-1}P]^{-1}[(\tau\Sigma)^{-1}\mu^e + P'\Omega^{-1}\mu^v] \tag{6}$$

$$\bar{\Sigma}_{BL} = \Sigma + [(\tau\Sigma)^{-1} + P'\Omega^{-1}P]^{-1} \tag{7}$$

Plugging these two updated estimates into the optimal portfolio formula given earlier, or Equation (2), we obtain the BLM solution to the portfolio choice problem.

Note that the Black–Litterman expected return, $\bar{\mu}_{BL}$, is a weighted average of the equilibrium expected return and the investor's views. Intuitively, the less confident the investor is in his views, the closer $\bar{\mu}_{BL}$ is to the equilibrium value, and the closer the Black–Litterman portfolio is to w_e. He and Litterman [1999] showed mathematically that the BLM portfolio weights are a weighted average of a scaled w_e and the weights implied by the investor's views. Hence, the BLM tilts the investor's optimal portfolio away from the market portfolio according to the strength of the investor's views. Because the market portfolio is without any extreme positions, any suitably controlled tilt should also yield a portfolio without any extreme positions. This is perhaps one of the major reasons the BLM is so popular in practice.

AN EXTENSION

The BLM combines both the equilibrium model and the investor's views, but ignores the data-generating process of the asset returns. This point can be understood in two ways. First, the data-generating process is an econometric model of the data, which is generally a nonlinear system of equations that govern the evolution of the asset returns over time,

$$g(R_T, F_T, Z_T) = 0 \tag{8}$$

where g is a vector function, F_T is a vector of factors, and Z_T is a vector of random noises. In this case, the views as expressed by the simple structure of Equation (5) clearly do not capture the salient feature of the data. Second, consider the simplest possible data-generating process,

$$R_T = \mu + \epsilon, \quad \epsilon \sim N(0, \Sigma) \tag{9}$$

where ϵ is normally distributed with zero mean and covariance matrix Σ. This is the normality assumption on R_t that underlies the mean-variance analysis. In this case, how the data-generating process is used depends on how views are formed. Clearly if the views are obtained from

an auxiliary analysis, such as a political or industry assessment on certain stocks or sectors, then the data-generating process is ignored. Even if information from the past return history is used, as long as Equation (9) is not used in its entirety, the data-generating process can be regarded as being ignored; in true Bayesian learning, the prior should be updated by the law of motion of the data or by the data-generating process. Because K is not equal to N, in general, the views in the BLM are not based entirely on the data-generating process. Only in the trivial case that $K = N$ and Equation (5) is derived by Equation (9) can we consider that the views are based on the data-generating process. But this is not the way views are usually formed in practice, and hence this scenario will be excluded in our analysis.

The BLM reshapes the equilibrium allocation based on the investor's views without use of the data-generating process. The question is whether this approach is optimal. In fact, there are two potential problems. First, from an asset pricing perspective, the equilibrium relationship is subject to specification errors for which the data have a lot to say. For example, in terms of the stock market, the equilibrium relation hinges on the validity of the CAPM. But the CAPM is strongly rejected by data, as shown by Gibbons, Ross, and Shanken [1989] and Zhou [1991], among others. It is now well accepted that the Fama and French [1993] three-factor model is a much better choice, and including the momentum factor, as shown by Carhart [1997], can make the model even better. But these later models are also rejected by data. Hence, the equilibrium expected returns that rely on holding a portfolio proportional to the market capitalization are likely to be incorrect. This fact does not mean, however, that the equilibrium models are not useful. On the contrary, they provide valuable insights into the problem, but relying on them completely, even if updated by the investor's views, is not adequate. Certain adjustments based on the return data or past history should be able to improve portfolio selection, a point that this article will soon make clear.

Second, in the simplest normal data-generating process, the sample means of the asset returns play an important role in determining future stock returns, but they are not used in the BLM ad hoc Bayesian updating. To understand the importance of the data, imagine a stock that has an almost zero correlation with the rest of the market. Assume $\gamma = 3$ and the stock has an annual standard deviation of 20% with 1% market share. The annual equilibrium expected return is 1.2%. If the sample mean is 1% for the past 50 years, the equilibrium expected return of 1.2% is likely to be reasonable. But if the sample mean is 10% for the past 50 years, the equilibrium expected return of 1.2% is likely a substantial underestimate of the true expected return. When views are not strong, the Black–Litterman expected return is close to the equilibrium expected return, and will underestimate the true expected return. When the views about the stock's outperformance relative to the market are strong, errors in estimating the expected returns of other stocks can cause the relative-performance view to be quite different from what is intended. For example, a view that stock A outperforms stock B by 5% can be exaggerated to 9% if the expected return of A is overestimated by 4%. In short, under normality, the historical average returns contain important information on future expected returns and should not be almost entirely ignored.

According to statistical decision theory, the data and the associated model for the data (data-generating process) are fundamental in Bayesian learning. Three elements are available for decision making: the equilibrium model, the investor's views, and the data. The BLM combines only the first two. Mathematically, an additional combination with the third can only improve performance. The issue is how to combine all of the elements and if the new combination improves the model's performance.

A typical Bayesian inference is to update a given prior by the data, as demonstrated by Shanken [1987] and Harvey and Zhou [1990] in the context of analyzing the validity of the CAPM. What is the prior in the BLM? Although not explicitly stated in Black and Litterman [1992] or elsewhere, we can interpret that the mixed estimation strategy is an update of the prior of the equilibrium relationship with the investor's views, as we have done thus far. Because the residuals in the belief equations of the BLM are independent and normally distributed, reversing the order should yield the same result. It seems more natural, however, to interpret the equilibrium relation as the prior.

In order to link the equilibrium prior, the investor's views, and the data together, we have to regard the first two as priors. In contrast to standard Bayesian updating, instead of a single prior, we have two priors—the equilibrium relationship and the views. The Bayesian learning mechanism does allow a sequence of priors that can be updated sequentially (see, e.g., Geweke [2005]). Armed with these observations, we can extend the BLM to account for information from the data.

For simplicity, in this section we maintain the same assumption that underlies the BLM, that Σ is a known constant matrix; this assumption will be relaxed later. Assuming a multivariate regression for the asset returns, we let $\hat{\mu}_m$ be the expected return vector in the regression. Normality is a special case of the multivariate regression with only the constant regressor; in this case, $\hat{\mu}_m$ will be the historical mean of the asset returns. In general, either factors or predictable variables can be in the regression, and $\hat{\mu}_m$ will be the regression forecasts of future expected returns.

Based on recursive updating logic, the Black–Litterman estimates, which are updates of the equilibrium relationship by the views, can be further updated based on the likelihood function of the data to obtain the following:

$$\hat{\mu}_T = [\Delta^{-1} + (\Sigma/T)^{-1}]^{-1}[\Delta^{-1}\bar{\mu}_{BL} + (\Sigma/T)^{-1}\hat{\mu}_m] \quad (10)$$

$$\hat{\Sigma}_T = \Sigma + [\Delta^{-1} + (\Sigma/T)^{-1}]^{-1} \quad (11)$$

where T is the sample size, and $\Delta = [(\tau\Sigma)^{-1} + P'\Omega^{-1}P]^{-1}$ is the covariance matrix of the Black–Litterman estimate $\bar{\mu}_{BL}$.

The data-updated expected return $\hat{\mu}_T$ is a weighted average of the Black–Litterman expected return and the model's expected return/forecast based on the data. (The weighting will be more complex, as we will address later, if the data-generating process is beyond a multivariate regression.) If we have a lot of data so that T is large, it will be determined more by $\hat{\mu}_m$. In the limit when we have an infinite amount of data, it must converge to the true expected returns of the data-generating process. This is the desired consistency property of Bayesian decision making. With ample data, we should be able to make the right decisions. Non-dogmatic priors should have no impact on the inference. But if T is small or the belief in either the equilibrium relationship or the views is strong, $\hat{\mu}_T$ will be determined primarily by $\bar{\mu}_{BL}$. In this case, $\bar{\mu}_{BL}$ offers additional information that should help tie down the correct parameter values.

Because the BLM itself can be treated as a prior, we can further replace Δ in Equations (10) and (11) by $\eta\Delta$ for obtaining $\hat{\mu}_T$, where η is a hyperparameter to reflect the degree of belief in the BLM, which earlier had a value of one. The smaller the η, the greater the belief in the BLM. In the extreme case of $\eta = 0$, we believe in the BLM completely and the result will not be altered by data. Therefore, our approach is strictly an extension of the BLM, and it reduces to the BLM when $\eta = 0$. The interesting question is how to empirically determine the value of η. Note that the BLM says nothing theoretically about how to choose τ and Ω, which represent the beliefs or confidence in the equilibrium relationship or the views. Similiar to these hyperparameters, the choice of η is a matter of empirical calibration and portfolio management experience.

AN EXAMPLE

In this section, for simplicity we assume normality. The argument is that if there is value in blending the BLM with the data even in the simplest data-generating process, there must be value for complex dynamic processes, such as GARCH and stochastic volatility models with time-varying risk premiums. In the normality model, consider an allocation of funds into the seven major industrial countries—Australia, Canada, France, Germany, Japan, the U.K., and the U.S.—with monthly dollar returns available from Global Financial Data for the period January 1970 to December 2007. The second column of Exhibit 1 reports the equilibrium expected excess returns (annualized) of the seven countries, computed based on Equation (3) with a γ value of 2.5. The fifth column reports the equilibrium/value-weighted weights (in percentage points). Compared to the years studied by He and Litterman [1999], the stock market in France now has a market share of 7.9% versus 5.2% and in Germany of 6.5% versus 5.5%. So, France surpasses Germany in terms of stock market value, though it has only about half of Germany's GDP. The third column of Exhibit 1 reports the country sample average excess returns, the $\hat{\mu}_T$. Compared to the equilibrium returns, two interesting observations can be made. First, the cross-country differences

EXHIBIT 1

Asset Returns and Portfolio

	μ^e	$\hat{\mu}$	μ_{BL}	w_e	$\hat{w}$	w_{BL}
Australia	5.1	8.2	6.1	4.0	27.2	3.7
Canada	5.2	6.1	6.5	6.8	–19.1	43.0
France	5.5	8.3	6.5	7.9	34.5	21.8
Germany	4.5	5.1	4.5	6.5	–23.3	–8.6
Japan	4.2	6.7	4.7	13.5	25.6	12.2
U.K.	5.9	8.7	6.4	12.0	23.4	10.9
U.S.	4.8	5.9	4.6	49.2	60.1	7.9

are much greater. Second, the rankings of the countries in terms of the returns are quite different. For example, Japan has the lowest return of 4.2%, but is now ranked fourth with a return of 6.7%. Clearly, the optimal portfolio weights based on the sample averages, reported in the sixth column under $\hat{w}$ and computed from Equation (2), are totally different from the equilibrium weights with huge short positions of 23.3% and 19.1% on Germany and Canada, respectively. Interestingly, the long position in the U.S. is 60.1%, not too far from the equilibrium value of 49.2%.

Now, assume that we have two views, similar to the case examined by He and Litterman [1999]. The first view is that France will outperform Germany by 2% with a standard error of 2%. The second view is that Canada will outperform the U.S. by 2% with a standard error of 4%. The BLM estimates of the expected returns are obtained from Equations (6) and (7), and reported in the fourth column of Exhibit 1. The expected return differences between France and Germany and between Canada and the U.S. are updated from 1% to 2% and from 0.4% to 1.9%, respectively. Correspondingly, the relative portfolio weights, reported in the last column, are substantially adjusted according to the new expected returns. For example, the equilibrium holding of 6.5% in Germany now becomes a short position of 8.6%. Both France and Canada have increased their long positions substantially from 7.9% to 21.8% and from 6.8% to 43.0%, respectively.

To understand the role that data plays, imagine that the sample mean and covariance matrix are obtained from a large sample size and are thus reliable estimates of the true expected returns and risks. Because they are quite different from the equilibrium returns, the latter must be inaccurate; hence, a combination of the equilibrium returns with the supposed correct views will not necessarily yield outstanding performance. Nevertheless, even when the data are not plentiful, Bayesian learning suggests that they provide useful information on the moments of the returns as well as on their dynamic behavior. It is thus worthwhile to blend the BLM solution with the data. Therefore, suppose that the previous sample mean and covariance matrix were obtained with hypothetical sample sizes of $T = 60$, 120, and 240, respectively, which correspond to 5, 10, and 20 years of monthly data. We will examine how the expected returns and weights will be updated with these typical sample sizes. Exhibit 2 reports the results. The expected blended returns, $\hat{\mu}_T s$,

EXHIBIT 2
Black–Litterman Model Blended with Data

	$\hat{\mu}_{60}$	$\hat{\mu}_{120}$	$\hat{\mu}_{240}$	$\hat{w}_{60}$	$\hat{w}_{120}$	$\hat{w}_{240}$
Australia	7.2	7.6	7.9	17.7	21.3	23.8
Canada	6.1	6.1	6.1	–0.1	–7.1	–12.2
France	7.7	7.9	8.1	32.6	33.3	33.8
Germany	4.7	4.9	5.0	–20.1	–21.3	–22.1
Japan	5.9	6.2	6.4	20.6	22.5	23.8
U.K.	7.8	8.1	8.3	18.7	20.4	21.7
U.S.	5.4	5.6	5.7	46.6	51.6	55.2

are computed from Equation (10), the portfolio weights, $\hat{w}_T s$, are computed from Equation (2) based on $\hat{\mu}_T s$, and the covariance matrix is computed from Equation (11).

$\hat{\mu}_T$ approaches $\hat{\mu}$ quickly as T increases. Even with $T = 60$, the expected returns are already close to $\hat{\mu}$. The most striking impact is on the portfolio weights. The 43.0% position in Canada now becomes a short position of 0.1%. This change is drastic, but nevertheless it is a move in the right direction because the sample weight is a short position of 19.1%. With $T = 240$, the portfolio weights are no longer much different from $\hat{w}$. The data-generating process here is quite simple, so that the speed of the sample size in being able to adjust the prior is high. But in a more realistic situation when the data-generating process is complex, as discussed in the next section, the speed can be much slower. In addition, if the confidence in the BLM is high as reflected by a small Δ matrix (or by a small hyperparameter as discussed in the next section), the speed will be lower too. Regardless of the sample size, the blending or updating is consistent with Bayesian decision theory, because it will be optimal to update the BLM given the data at hand and the assumed data-generating process.

A simple way to see the effectiveness of the updating is to assume that the sample estimates are the true parameters. This is, of course, biased against the BLM, but is still useful in order to see how fast the BLM, assumed incorrect, can go to the true model. In this case, the certainty-equivalent return based on the true weights $\hat{w}$ is 5.00%, and the certainty-equivalent return of the w_{BL} is only 3.89%. The updated certainty-equivalent returns are 4.89%, 4.95%, and 4.99% with the sample sizes 60, 120, and 240, respectively. Interestingly, although $\hat{w}_{60}$ is half way from w_{BL} to $\hat{w}$ and is not nearly as extreme as $\hat{w}$, it obtains almost all of the optimal expected portfolio return. The implication is that, if the data represent the

truth, the blending of the BLM even with a small amount of data can quickly approach the level of the optimal expected portfolio return. In general, given enough data and given the true data-generating process, there is no guarantee that the BLM solution will converge to any limit, but the Bayesian blending must converge to the true optimal solution. Of course, the true data-generating process might be a complex and chaotic nonlinear system which may never be fully known. But to the extent that today's sophisticated quantitative and econometric approaches can be used to produce a model that captures much of the behavior of the data, the Bayesian updating of the BLM with such a data-generating process should clearly add value to the investment process.

FURTHER GENERALIZATIONS

The central idea of our approach so far has been to treat the BLM as part of the recursive priors in the formal Bayesian learning mechanism based on the data. In this way, insights of the BLM can be incorporated into an arbitrary data-generating process, such as a sophisticated econometric model of asset returns that accounts for both regime changes (e.g., bull, bear, and panic) and stochastic volatilities. Our discussion thus far has been kept simple to allow our framework to developed without the hinderance of technical details. Going forward, we will allow for more complex models of the data. For brevity, however, we will only outline the major steps and provide the relevant references for implementation details.

The first extension of the earlier analysis is to allow Σ to be treated as a matrix with unknown parameters in the multivariate regression, which will account for errors associated with estimating Σ in practice. The estimation errors are not a problem when N is small, because the sample covariance matrix is usually an accurate estimate of Σ. However, they can matter substantially when N is large, as shown by Kan and Zhou [2007]. To capture the uncertainty associated in estimating Σ, we can specify an inverted Wishart prior distribution,

$$p_0(\Sigma) \propto |\Sigma|^{-\frac{\nu}{2}} \exp\left\{-\frac{1}{2} tr\Sigma^e\Sigma^{-1}\right\} \qquad (12)$$

where Σ^e is the earlier constant estimate of the covariance matrix in the equilibrium relationship of the BLM, and ν is a parameter that captures the degree of confidence in the estimate. These types of priors can be easily updated via the Bayesian rule. The greater the ν, the more weight we put on Σ^e. When ν is large enough, the updated Σ becomes virtually Σ^e, the same as the earlier case where Σ is treated as constant. However, with reasonable views, the data should inform us about the distribution of Σ. As an alternative to Equation (12), and like Qian and Gorman [2001] in the classical framework, various prior views on the volatilities and the correlations of assets can be expressed, and combined with the data. In general, almost any priors on Σ can be accommodated.

Denote by $\Phi_0(\mu)$ the normal prior density on μ with the Black–Litterman estimate as the mean and $\eta\Delta$ as the covariance matrix. Based on the two informed priors, the posterior distribution of the parameters then becomes

$$p(\mu, \Sigma) \propto p_0(\Sigma)\Phi_0(\mu)\mathcal{L}(\mu, \Sigma) \qquad (13)$$

where $\mathcal{L}(\mu, \Sigma)$ is the likelihood function of the data. It can be shown that the posterior mean of the expected return is roughly a weighted average of the prior mean and the sample mean of the data. The more data, the more we can learn in order to change our prior belief about the parameters.

Once the uncertainty in estimating Σ is modeled, the portfolio choice decision will be implemented differently, due to a couple of complexities. First, although updating on the expected returns can be accomplished based on Equation (13), by combining the Black–Litterman estimate with the likelihood function of the data, their posterior distribution will no longer be normal; it will be *t*-distributed. Furthermore, the Bayesian portfolio weights should now be determined by maximizing the expected utility under the predictive distribution of the data, rather than the simple classical mean-variance solution. Fortunately, the tools provided by Pástor [2000], for example, can be readily adapted to the implementation here.

As used in our analysis, the Bayesian approach not only accounts for parameter estimation risk, but is also applicable to an arbitrary utility, thus extending Martellini and Ziemann [2007]; letting $U(w)$ be the utility of holding a portfolio w at time $T + 1$, the optimal portfolio is obtained by maximizing the expected utility under the predictive distribution,

$$\hat{w}^{\text{Bayes}} = \text{argmax}_w \int_{R_{T+1}} U(w) p(R_{T+1} \mid \Phi_T)\, dR_{T+1}$$
$$= \text{argmax}_w \int_{R_{T+1}} \int_{\mu} \int_{\Sigma} U(w)\, p(R_{T+1} \mid \mu, \Sigma, \Phi_T)$$
$$\times\, p(\mu, \Sigma \mid \Phi_T)\, d\mu d\Sigma dR_{T+1} \tag{14}$$

where $p(R_{T+1} \mid \Phi_T)$ is the predictive density, $p(\mu, \Sigma \mid \Phi_T)$ is the posterior density of μ and Σ, and Φ_T denotes the data up to time T. Thus, the Bayesian approach maximizes the average expected utility over the distribution of the parameters and the utility function can be nonquadratic.

Our second generalization of the BLM is to allow the expected returns to be explained by a number of factors with an unknown residual covariance matrix. The return data-generating process can then be written as

$$R = \alpha + \beta_1 f_t + \cdots + \beta_q f_q + \epsilon \tag{15}$$

where $f_1, \ldots, f_q$ are the factors and q is the number of factors. Unlike the case in previous sections, the residual covariance matrix here is unknown. The factor model is the usual market model regression if the market portfolio is taken as the single factor. Under the factor model, the Bayesian portfolio choice can be made in almost exactly the same way as in the first generalization. Moreover, because the factors provide additional information on the asset returns, the estimated expected returns should now be more accurate than before.

The factor model also permits an investor to express his prior belief on the degree of validity of a given asset pricing model. If the investor believes dogmatically in the Fama and French [1993] three-factor model, the alphas should be identically zero. In general, a normal prior can be imposed on the alphas with a multiplier parameter on the covariance matrix to reflect the degree of belief about how close the alphas are to zero. This prior can be easily combined with $\Phi_0(\mu)$. In addition, the investor can compare and assess competing asset pricing models. Pástor and Stambaugh [2000] pioneered studies of this kind. Their insights and tools can be readily applied here in conjunction with the BLM.

The factors used in the preceding analysis are assumed observable, that is, we know what they are and can collect data on them. Alternatively, we can assume that they are latent factors, which are completely unobservable and unknown, except for the value of q. In this case, the Bayesian learning is much more complex than before. The methods developed by Geweke and Zhou [1994] and Nardari and Scruggs [2007], however, are well suited for the purpose.

Our third generalization of the BLM is to allow informational variables to forecast the future expected returns, so that the portfolio decision varies with time-varying economic conditions. In this case, the data-generating process can be modeled as

$$R_t = a_0 + BZ_{t-1} + v_t \tag{16}$$

where Z_{t-1} is a vector of M predictive variables available at time t, and v_t is a vector of the residuals. The predictive variables can be allowed to follow a VAR(1) process,

$$Z_t = C_0 + C_1 Z_{t-1} + u_t \tag{17}$$

with residuals u_t. In finance, a huge literature exists on the predictability of stock returns. Avramov [2004], among others, showed how predictability, as well as asset pricing models, can be used to significantly improve portfolio selection. All of the existing tools developed in this area can be adapted to combine with the BLM in the same way as before.

Finally, much of the existing studies rely on the normality assumption of asset returns. Tu and Zhou [2004] showed that normality is rejected with P-values virtually equal to zero for U.S. stock returns. But the t-distribution can model the data well despite the skewness of the data, since the chance of observing a large skewness from a symmetric t is very high. Given a t-distribution for the data, the portfolio allocations can be drastically different from those under normality, but without sizable differences in terms of the utilities. However, beyond being stationary, the model can adapted to capture business cycles and structural breaks in the economy. In general, both the portfolio allocations and the utilities will be different. For example, Tu [2007] provided a regime-switching model to capture the changing nature of bull and bear markets and found that utility gains are substantial compared to simply ignoring the dynamics of the data. This fact, too, can be combined with the BLM based on Equations (13) and (14).

MODEL AVERAGING

The equilibrium model and the investor's views can be considered as two separate priors on the parameters in

the BLM. In general, it is possible to have J priors, arising perhaps from various forecasts by various experts. Each of the priors can be multiplied by the likelihood function of the data to yield a Bayesian model for the portfolio decision. The J priors thus generate J models. The Bayesian model averaging approach is a way to formally combine the J models into a single model (alternative methods in statistical decisions may also be used for the combining).

The approach proceeds from a set of prior probabilities on the validity of the J models,

$$p_0(M_j) = \text{prior probability for Model } j \quad j = 1, 2, \ldots J \tag{18}$$

After observing the data R, the posterior probabilities are given by

$$p(M_j \mid R) = \frac{p_0(M_j)p(R \mid M_j)}{\Sigma_{j=1}^{J} p_0(M_j)p(R \mid M_j)} \tag{19}$$

where $p(R \mid M_j)$ is the marginal likelihood computed by

$$p(R \mid M_j) = \int p_0(\theta_j \mid M_j)p(R \mid \theta_j, M_j)d\theta_j \tag{20}$$

where $p_0(\theta_j \mid M_j)$ and $p(R \mid \theta_j, M_j)$ are the usual prior and posterior densities of the parameter θ_j conditional on model j being true with θ_j as the parameters.

Then, the predictive future asset return is an average of those from the J models weighted by their posterior probabilities,

$$p(R_{T+1} \mid R) = \sum_{j=1}^{J} p(R_{T+1} \mid M_j, R)p(M_j \mid R) \tag{21}$$

This, combined with the objective function in Equation (14), provides the Bayesian optimal portfolio choice.

For example, in the case of the mean-variance quadratic utility, the portfolio choice depends only on the predictive mean and variance,

$$E_M^* = \sum_{j=1}^{J} E_j^* p(M_j \mid R) \tag{22}$$

$$V_M^* = \sum_{j=1}^{J} V_j^* p(M_j \mid R) + \sum_{j=1}^{J} (E_j^* - E_M^*)(E_j^* - E_M^*)' p(M_j \mid R) \tag{23}$$

where E_j^* and V_j^* are the corresponding predictive moments from model j. Based on these moments, the portfolio weights are computed in the usual way.

CONCLUSION

Black and Litterman [1992] provided an ingenious and simple solution to the optimal portfolio choice problem by blending an investor's proprietary views with the equilibrium views of the market to construct a portfolio. Their model, in particular, offers a useful framework for beating the benchmark index by over- and underweighting certain assets. However, their model, as a mix of the two priors, ignores the data-generating process whose dynamics can have a significant impact on future portfolio returns. This article provides two simple ways to further blend the Black–Litterman model with the data. Three implications follow. First, theoretically, if the empirical data-generating process is a good model for the data, and if there are enough data, there is no guarantee that the BLM solution will converge to any limit, but the Bayesian blending of the BLM with data must converge to the true optimal portfolio. Second, our framework allows practitioners to combine the insights from the Black–Litterman model with the data to potentially generate more-reliable trading strategies and more-robust portfolios. Third, many existing Bayesian learning tools, which have been developed in the academic finance literature for updating priors via the likelihood function of the data, can now be readily applied to practical portfolio selections in conjunction with the Black–Litterman model.

ENDNOTE

The author is grateful to Campbell Harvey, Nan Li, Ľuboš Pástor, Edward Qian, Farris Shuggi, Jack Strauss, Jun Tu, and Ghislain Yanou for helpful comments.

REFERENCES

Avramov, D. "Stock Predictability and Asset Pricing Models." *Review of Financial Studies*, Vol. 17, No. 3 (2004), pp. 699–738.

Black, F., and R. Litterman. "Global Portfolio Optimization." *Financial Analysts Journal*, Vol. 48, No. 5 (1992), pp. 28–43.

Carhart, M. "On the Persistence in Mutual Fund Performance." *Journal of Finance*, Vol. 52, No. 1 (1997), pp. 57–82.

Fabozzi, F.J., S.M. Focardi, and P.N. Kolm. "Incorporating Trading Strategies in the Black–Litterman Framework." *Journal of Trading*, Vol. 1, No. 2 (2006), pp. 28–37.

Fama, E.F., and K.R. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, Vol. 33, No. 1 (1993), pp. 3–56.

Geweke, J.F. *Contemporary Bayesian Econometrics and Statistics*. New York, NY: Wiley, 2005.

Geweke, J.F., and G. Zhou. "Measuring the Pricing Error of the Arbitrage Pricing Theory." *Review of Financial Studies*, Vol. 9, No. 2 (1994), pp. 553–583.

Gibbons, M.R., S.A. Ross, and J. Shanken. "A Test of the Efficiency of a Given Portfolio." *Econometrica*, Vol. 57, No. 5 (1989), pp. 1121–1152.

Harvey, C.R., and G. Zhou. "Bayesian Inference in Asset Pricing Tests." *Journal of Financial Economics*, Vol. 26, No. 2 (1990), pp. 221–254.

He, G., and R. Litterman. "The Intuition Behind the Black–Litterman Model Portfolios." *Goldman Sachs Investment Management Series*, 1999.

Kan, R., and G. Zhou. "Optimal Portfolio Choice with Parameter Uncertainty." *Journal of Financial and Quantitative Analysis*, Vol. 42, No. 3 (2007), pp. 621–656.

Markowitz, H.M. "Portfolio Selection." *Journal of Finance*, Vol. 7, No. 1 (1952), pp. 77–91.

Martellini, L., and V. Ziemann. "Extending Black–Litterman Analysis Beyond the Mean-Variance Framework: An Application to Hedge Fund Style Active Allocation Decisions." *Journal of Portfolio Management*, Vol. 33, No. 4 (2007), pp. 33–45.

Nardari, F., and J.T. Scruggs. "Bayesian Analysis of Linear Factor Models with Latent Factors, Multivariate Stochastic Volatility, and APT Pricing Restrictions." *Journal of Financial and Quantitative Analysis*, Vol. 42, No. 4 (2007), pp. 857–892.

Pástor, Ľ. "Portfolio Selection and Asset Pricing Models." *Journal of Finance*, Vol. 55, No. 1 (2000), pp. 179–223.

Pástor, Ľ., and R.F. Stambaugh. "Comparing Asset Pricing Models: An Investment Perspective." *Journal of Financial Economics*, Vol. 56, No. 3 (2000), pp. 335–381.

Qian, E., and S. Gorman. "Conditional Distribution in Portfolio Theory." *Financial Analysts Journal*, Vol. 57, No. 2 (2001), pp. 44–51.

Shanken, J. "A Bayesian Approach to Testing Portfolio Efficiency." *Journal of Financial Economics*, Vol. 19, No. 2 (1987), pp. 195–215.

Tu, J. "Are Bull and Bear Markets Economically Important?" Working Paper, Singapore Management University, 2007.

Tu, J., and G. Zhou. "Data-Generating Process Uncertainty: What Difference Does It Make in Portfolio Decisions?" *Journal of Financial Economics*, Vol. 72, No. 2 (2004), pp. 385–421.

Zhou, G. "Small Sample Tests of Portfolio Efficiency." *Journal of Financial Economics*, Vol. 30, No. 1 (1991), pp. 165–191.