Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Generalized Black—Litterman Model

Shea D. Chen, Andrew E. B. Lim

To cite this article:
Shea D. Chen, Andrew E. B. Lim (2020) A Generalized Black—Litterman Model. Operations Research 68(2):381-410.
https://doi.org/10.1287/opre.2019.1893

http://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 68, No. 2, March–April 2020, pp. 381–410
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Contextual Areas

A Generalized Black–Litterman Model

Shea D. Chen,[a] Andrew E. B. Lim[b]

[a] Eyon Holding Group, Taipei 10483, Taiwan; [b] Department of Analytics and Operations, Department of Finance, and Institute for Operations Research and Analytics, National University of Singapore, 119245 Singapore

Contact: sheadaniel@eyon.com.tw, https://orcid.org/0000-0003-1080-446X (SDC); andrewlim@nus.edu.sg, https://orcid.org/0000-0001-7189-3406 (AEBL)

Received: August 12, 2014
Revised: October 6, 2015; February 4, 2017; July 5, 2018; February 13, 2019
Accepted: May 14, 2019
Published Online in Articles in Advance: January 6, 2020

Subject Classifications: statistics: Bayesian, estimation; finance: investment, portfolio; simulation: statistical analysis
Area of Review: Financial Engineering

https://doi.org/10.1287/opre.2019.1893

Abstract. The Black–Litterman model provides a framework for combining the forecasts of a backward-looking equilibrium model with the views of (several) forward-looking experts in a portfolio allocation decision. The classical version uses the capital asset pricing model to specify expected returns, and assumes that expert views are unbiased noisy observations of future returns. It combines the two using Bayes' rule and the portfolio allocation decision is made on the basis of the updated forecast. The classical Black–Litterman model assumes that the equilibrium and expert models are accurately specified. This is generally not the case, however, and there may be substantial efficiency loss if misspecification is ignored. In this paper, we formulate a generalized Black–Litterman model that accounts for both misspecification and bias in the equilibrium and expert models. We show how to calibrate this model using historical view and return data, and study the value of our generalized model for portfolio construction. More generally, this paper shows how the views of multiple experts can be modeled as a Bayesian graphical model and estimated using historical data, which may be of interest in applications that involve the aggregation of expert opinions for the purpose of decision making.

Funding: This work was supported by the Singapore Ministry of Education Academic Research Fund, Tier 2 [MOE2016-T2-1-086].

Keywords: Black–Litterman model • portfolio selection • graphical models • Bayesian methods • Gibbs sampling • central limit theorem

1. Introduction

The Black and Litterman (1991, 1992) model provides a framework for combining the forecasts of a backward-looking equilibrium model with the views of (several) forward-looking experts in a portfolio allocation decision. The classical version of this model uses the capital asset pricing model (CAPM) as the equilibrium model and assumes that expert views are unbiased forecasts of future returns. These forecasts are combined using Bayes' rule, and the allocation is made using the updated return distribution. The classical Black–Litterman model (BL) assumes that the equilibrium and expert models are accurately specified, which is generally not the case. For example, the equilibrium model may ignore relevant factors that impact expected returns, whereas expert views may be biased, and efficiency loss from model misspecification can be substantial. In this paper, we formulate a generalization of the Black–Litterman model that accounts for misspecification and bias in both the equilibrium and view models and show how it can be calibrated using historical data. More generally, this paper shows how the views of multiple experts can be modeled as a Bayesian hierarchical model and estimated using a history of forecast data and outcomes that may be of interest in applications that involve the aggregation of expert opinions for the purpose of decision making.

The Black–Litterman model was introduced in Black and Litterman (1991, 1992), using the mixed estimation approach of Theil (1971) to combine the equilibrium and expert forecasts. Qian and Gorman (2001) provide an interpretation of the update equations from a Bayesian perspective, which is the framework we adopt. A comprehensive literature review can be found in Walters (2011).

Although concerns about the impact of model misspecification in the classical Black–Litterman model have been raised, many of the refinements are basically strategies for tuning parameters, and relatively little has been done to address limitations of the classical model by moving beyond the original framework. For example, the parameter τ, which scales the covariance matrix in Black and Litterman (1992), was included as a way to account for uncertainty in the

market equilibrium estimate of the expected return (see Walters 2013). One limitation of this approach is that it downweights the forecasting information that is available in the CAPM estimate and ignores the possibility that systematic errors such as forecast bias can potentially be estimated and removed. A natural approach to address the acknowledged limitations of the CAPM is explored in Krishnan and Mains (2005), which includes additional factors in the equilibrium model. Similar concerns about misspecification arise in the context of the expert model, where much of the focus is centered around how confidence levels for experts can be chosen (see, e.g., Walters 2011, where several methods are discussed).

In this paper, we propose a substantial generalization of the traditional model that accounts for its shortcomings directly. Specifically, we propose a Bayesian graphical model that accounts for systematic biases and unmodeled factors/dynamics in both the expert and equilibrium models and show how it can be estimated in a Bayesian framework using a history of expert forecasts and realized returns. We characterize the optimal mean-variance portfolio in terms of the mean and variance of returns under the posterior and show how it can be estimated using Gibbs sampling. We also show that this estimate is consistent, asymptotically normal, and converges at rate $1/\sqrt{m}$ (where m is the number of Gibbs samples), generalizing analogous results for sample average approximation (Shapiro et al. 2014) to the case when samples are dependent and there is a risk function in the objective. Empirical tests with simulated and historical data are also provided.

As we have noted, much of the literature on the Black–Litterman model stays close to the original framework, and relatively little has been done to address its shortcomings by generalizing the model, which is one contribution of this paper. One notable exception is the paper by Bertsimas et al. (2012), which combines forecasts from the equilibrium model and experts using ideas from inverse optimization. The advantage of their framework is that it allows for nonstandard expert opinions (e.g., on volatility) and the use of risk measures such as conditional value at risk (CVaR). It can also handle uncertainty about the covariance matrix using ideas from robust optimization. We note, however, that it does not address concerns about misspecification in either the equilibrium or expert models (and hence the forecast of asset returns that is generated by combining these models). Our paper is different in that we develop equilibrium and view models with both learnable bias and unresolvable estimation uncertainty using probabilistic and Bayesian techniques and show how sampling methods can be used to compute optimal portfolios. Although we focus on mean-variance optimization, we believe that mean-CVaR problems can be addressed by combining the posterior sampling methods used in this paper with the data-driven method from Rockafellar and Uryasev (2000). In summary, both papers address different limitations of the classical Black–Litterman model with different tools.

The remainder of this paper is as follows: Section 2 covers the background of the Black–Litterman model and its graphical representation. Section 3 introduces the generalized Black–Litterman model (GBL) and the associated Bayesian graphical model. Section 4 discusses the problem of estimation and shows how Gibbs sampling can be used to calibrate the model and generate samples from the updated posterior distribution of returns. The optimal portfolio is characterized in terms of the mean and variance of returns of the posterior distribution in Section 5 and can be estimated using Gibbs sampling. We show that this estimate is consistent and asymptotically normal and converges at rate $1/\sqrt{m}$ (the number of Gibbs samples). Empirical tests on simulated and real data are provided in Section 6. We conclude in Section 7.

2. The Black–Litterman Model

The Black–Litterman model estimates future returns by combining a backward-looking equilibrium model with forward-looking expert views. We briefly summarize the Bayesian perspective of the classical model (see also Qian and Gorman 2001) and provide a graphical representation of this model.

2.1. Description

A central concern in a portfolio choice problem is modeling the distribution of asset returns. The classical Black–Litterman model constructs this distribution by combining a backward-looking equilibrium model with forward-looking expert views where the following sequence of events is assumed: (i) at the start of the investment period, the investor specifies a "prior" distribution for the vector of asset returns that is calibrated using a backward-looking equilibrium model, (ii) the investor receives expert views which are *modeled* as noisy observations of the future asset returns, (iii) expert views are combined with the equilibrium model using Bayes' rule to generate an updated return distribution, (iv) a portfolio allocation is made on the basis of the updated distribution, and (v) asset returns are realized.

2.1.1. Equilibrium Model and Prior Distribution for Returns. Our financial market consists of N risky assets and a risk-free asset. Let r denote the N-dimensional vector of returns for the risky assets. We assume without loss of generality that the risk-free rate is zero. (If the risk-free rate is nonzero, simply replace r with the vector of *excess returns* and our

analysis carries through.) The classical Black–Litterman model assumes at stage (i) that r is normally distributed:

$$r \sim N(\mu, \Sigma), \tag{1}$$

where the mean μ is set equal to the expected returns obtained from a backward-looking equilibrium model such as the CAPM. There are various methods for specifying the covariance matrix Σ (see, e.g., He 2002, Idzorek 2002, Meucci 2006, Beach and Orlov 2007, Walters 2013). We can think of the distribution (1) as a "prior" for the latent return vector r whose value will only be revealed at the end of the investment period.

2.1.2. Expert Views. Expert views are solicited to refine the (prior) distribution of returns (1). Views are forward-looking forecasts of the latent asset returns which are combined with the prior model using Bayes' rule.

To illustrate the idea, suppose that our investment universe consists of three assets, Apple (AAPL), Amazon (AMZN), and IBM, where the vector of returns is given by

$$r = \begin{bmatrix} r_{AAPL} \\ r_{AMZN} \\ r_{IBM} \end{bmatrix}.$$

At the start of the investment period, prior to the realization of r, the investor receives expert forecasts related to r. Forecasts could be *absolute views* about the return of a single asset, for example, "Apple's return will be 10%," or *relative views* about the difference in return between two or more assets, for example, "Amazon will outperform IBM by 2%." We write this as $V_{AAPL} = 10\%$ and $V_{AMZN-IBM} = 2\%$, which we interpret as noisy forecasts of the (latent) returns r_{AAPL} and $r_{AMZN} - r_{IBM}$, respectively. Alternatively, we have a vector of forecasts

$$V = \begin{pmatrix} V_{AAPL} \\ V_{AMZN-IBM} \end{pmatrix} = \begin{pmatrix} 10\% \\ 2\% \end{pmatrix}$$

of a linear mapping of returns

$$Pr = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} r_{AAPL} \\ r_{AMZN} \\ r_{IBM} \end{bmatrix} = \begin{bmatrix} r_{AAPL} \\ r_{AMZN} - r_{IBM} \end{bmatrix}.$$

More generally, experts express views about the returns of K different asset portfolios represented through a $K \times N$ "pick"/"link" matrix P. Each row of P selects a linear combination/portfolio of assets and the vector of views is a forecast of the returns of each of these portfolios. Observe that rows of the link matrix representing absolute views sum to 1, whereas rows representing relative views sum to 0. (For more discussion on specifying P, see He (2002), Idzorek (2002), and Satchell and Scowcroft (2000).)

Consistent with the idea that views are forward-looking forecasts of asset returns, views (conditional on r) are modeled as normal random variables with mean Pr and covariance matrix Ω, namely,

$$V|r \sim N(Pr, \Omega). \tag{2}$$

Here, Ω models the accuracy of the forecast V of Pr and is often interpreted as a measure of the level of confidence in the expert's opinions. There are several approaches to defining Ω, but generally this is determined at the discretion of the practitioner.

2.1.3. Updating the Return Distribution. With the prior model for returns (1), the model of views (2), and numerical values of expert forecasts $V = v$ (e.g., $V_{AAPL} = 10\%$ and $V_{AMZN-IBM} = 2\%$ in our example), the distribution of the return r is updated to a posterior $r|V = v$ using Bayes' rule. Under these assumptions, the updated distribution of returns is also normal:

$$r|v \sim N(\mu_{BL}, \Sigma_{BL}), \tag{3}$$

where

$$\begin{aligned} \mu_{BL} &= (\Sigma^{-1} + P'\Omega^{-1}P)^{-1}[\Sigma^{-1}\mu + P'\Omega^{-1}v], \\ \Sigma_{BL} &= (\Sigma^{-1} + P'\Omega^{-1}P)^{-1}. \end{aligned} \tag{4}$$

This shows that the mean return is an average of the prior mean and the view weighted by the confidence in each, as determined by the inverse of the covariance matrices (the multivariate analog of the "precision" of a normal posterior distribution). The precision of the return distribution increases from Σ^{-1} in (1) to

$$(\Sigma_{BL})^{-1} = \Sigma^{-1} + P'\Omega^{-1}P,$$

as determined by the link matrix P and the observation noise Ω. Portfolio allocations are obtained by solving a mean-variance problem with return distribution (3)–(4).

Compared with Black and Litterman (1992), we have dropped the scalar τ, which represents uncertainty in the market equilibrium vector, in favor of incorporating it into Σ (for discussion on the role of τ, see Walters 2013).

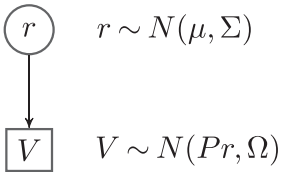
2.1.4. Optimal Portfolio. The optimal portfolio is obtained by solving a mean-variance problem with the updated return distribution (3). With risk coefficient $\gamma \in [0, \infty)$, we solve

$$\max_{\pi} \mathbb{E}[r]'\pi - \frac{\gamma}{2}\pi'\mathrm{Var}(r)\pi \equiv (\mu_{BL})'\pi - \frac{\gamma}{2}\pi'(\Sigma_{BL})\pi.$$

The optimal portfolio is given by

$$\pi^* = \frac{1}{\gamma}(\Sigma_{BL})^{-1}\mu_{BL}. \tag{5}$$

Figure 1. Graphical Representation of the Classical Black–Litterman Model



2.2. Graphical Representation

Bayesian graphical models are acyclic graphs where nodes represent random variables and directed arcs represent conditional dependence (Gilks et al. 1996, Gelman et al. 2004). These are useful representations of what may otherwise be complex models and allow for straightforward applications of probabilistic inference. In accordance with standard practice, we depict the nodes of unobserved random variables with circles and the nodes of observed random variables with squares.

The Bayesian graphical model of the classical Black–Litterman model is very simple and shown in Figure 1. Here, the view V is observed data, so it appears in the square node, whereas the return r is unobserved and appears in the circle.

3. Generalized Black–Litterman Model

We describe the generalized Black–Litterman model and show how it can be estimated using a history of expert views and returns.

3.1. Overview

Generally speaking, both the classical and generalized models have the goal of estimating the return, but differ in modeling assumptions and the data used. A comparison of both models is provided in Table 1, and we now elaborate on the details. From a high level, the generalized model accounts for the possibility that the assumptions about returns and expert views in the classical model may be incorrect and improves the model by accounting for these issues in order to better estimate the return.

Consider first the model of the views. In the classical model, views are *assumed* to be unbiased noisy observations of the realized future return (2). In the generalized model, we account for the possibility that views are biased and use the history of views and returns to estimate this bias in order to improve the forecast of the next period's return.

One feature of a set of views is that they can be influenced by multiple factors in ways that are not

Table 1. Summary of the Classical and Generalized Black–Litterman Models

Model traits	Black–Litterman model	Generalized Black–Litterman model
Expected return	Expected returns equal the CAPM estimate: Expected returns (μ) are constant Expected returns are assumed to equal the CAPM estimate ($\mu = \alpha$)	Expected returns do not equal the CAPM estimate: Expected returns μ_t vary every period The CAPM estimate α contains systemic bias b^μ Expected return will have mean centered around the CAPM with the bias offset ($\mu_t \vert b^\mu \sim N(\alpha + b^\mu, \beta)$)
Returns	Returns are normally distributed with mean μ and known covariance Σ	Returns are normally distributed with mean μ_t, which varies every period, and known covariance Σ
Expert views	Views (V) are unbiased Views are "centered" around next period's realized returns ($V \vert r \sim N(Pr, \Omega)$)	Views (V_t) may be biased (b_t^V) Views are "centered" around next period's realized returns plus the bias ($V_t \vert r_t, b_t^V \sim N(Pr_t + b_t^V, \Omega)$) Bias is normally distributed with constant mean and covariance ($b_t^V \vert \delta^V, \Theta^V \sim N(\delta^V, \Theta^V)$)
Unknown variables	r: return	r_t: return $\mu_1, \ldots, \mu_t$: expected returns b^μ: error in the CAPM estimate $b_1^V, \ldots, b_t^V$: bias in expert views δ^V, Θ^V: mean and covariance of expert bias
Exogenous variables		
Historical data	Link matrix P V: reported expert view	Link matrix P $V_1, \ldots, V_{t-1}$ and V_t: historical and current expert views $r_1, \ldots, r_{t-1}$: history of returns
User-specified parameters	Σ: covariance of returns Ω: covariance of expert views α: CAPM estimate	Σ: covariance of returns Ω: covariance of expert views α: CAPM estimate β: covariance of expected return $(\delta_0^\mu, \Theta_0^\mu)$: parameters of prior for bias (b^μ) in the CAPM η_0^V, κ_0^V, ν_0^V, Λ_0^V: parameters of prior for mean and variance of view-bias (δ^V, Θ^V)

necessarily understood. One implication is that biases for different views change randomly from period to period and may be dependent. (For example, different components may correspond to views of different experts, who may (or may not) interact personally, read research from the same sources, etc.) We model this by assuming that the view bias b_t^V at time t is an independent draw of a normal random variable with parameters (δ^V, Θ^V). We note, however, that the bias is not observed. Rather, the decision maker only sees the view V_t and the subsequent return r_t, where the view is normal with a mean $Pr_t + b_t^V$. (In contrast, the view in the classical model (2) is unbiased and has mean Pr_t.) The parameters of the bias distribution (δ^V, Θ^V) are estimated using historical view and return data and used to correct for the bias at time t.

The graphical model Figure 2 generalizes the classical model Figure 1 and illustrates the multiperiod nature of the data being used in the generalized model, whereas Figure 3 extends this further by including the model of view bias. Faced with the situation depicted in Figure 3, the plan is to use the history of views and returns, $(V_1, r_1), \ldots, (V_{t-1}, r_{t-1})$ to estimate the statistical properties (δ^V, Θ^V) of the bias, and to use this information to correct for the bias b_t^V in the view V_t so as to improve the forecast of the future return r_t. As before, the history of views and returns $(V_1, r_1), \ldots, (V_{t-1}, r_{t-1})$ and the view V_t at the start of period t are observed and appear in boxes. The other random variables, including the future return r_t, are not observed and are depicted in circles.

Consider now the model for equilibrium returns. In the classical model (1), returns are random variables where the mean μ_t is constant and assumed to equal the estimate α of expected returns from the CAPM, $\mu_t = \alpha$. In reality, however, the actual expected return is likely to differ from the CAPM forecast because of breakdowns in the assumptions underlying the CAPM or the noninclusion of relevant factors and unmodeled return dynamics (Fama and French 2004). In particular, there is concern about the impact of using a misspecified equilibrium model on the portfolio allocations that come from the classical Black–Litterman model, and various methods have been proposed to account for this error (e.g., Black and Litterman (1991) uses the scalar τ to scale the covariance matrix of the expected returns vector).

Figure 2. Graphical Representation of Black–Litterman After Multiple Periods

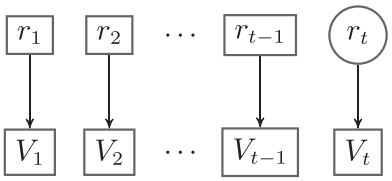
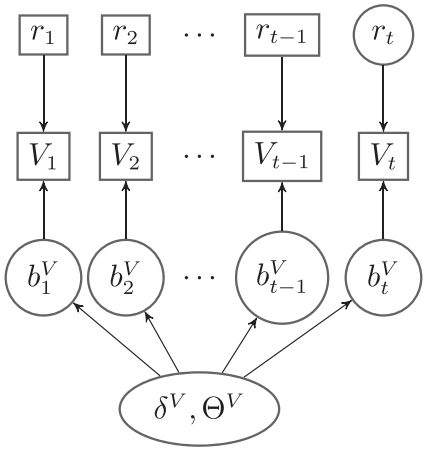


Figure 3. Graphical Representation of Generalized Black–Litterman Accounting for Expert View Bias



We account for errors in the CAPM forecast of expected returns (model misspecification, unmodeled factors, etc.) by dropping the assumption that μ_t is constant, observable, and equal to α, but allow μ_t to change from period to period according to independent samples of a normal random variable with a mean $\alpha + b^\mu$. Here, b^μ is a constant correction term that models bias in the CAPM forecast of expected returns, whereas randomness in μ_t models deviations from the mean $\alpha + b^\mu$ due to unmodeled stochastic factors that influence the expected returns from one period to the next. These modeling assumptions extend Figure 2 to the graphical model found in Figure 4. We emphasize, however, that the expected return μ_t and the correction b^μ are not observable but need to be estimated using knowledge of the CAPM forecast α and the history of returns.

The generalized Black–Litterman model combines both the expert (Figure 3) and equilibrium (Figure 4) models and is shown in Figure 5. With this blueprint in mind, we can now discuss specifics.

Figure 4. Graphical Representation of Generalized Black–Litterman Model Accounting for Equilibrium Errors

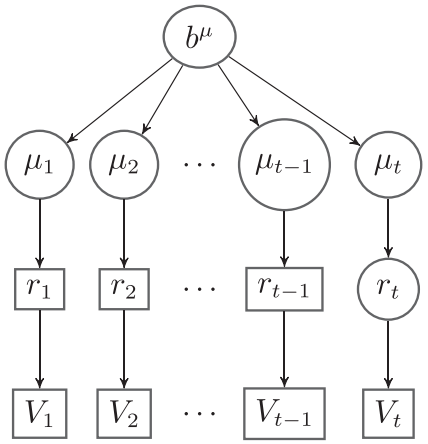
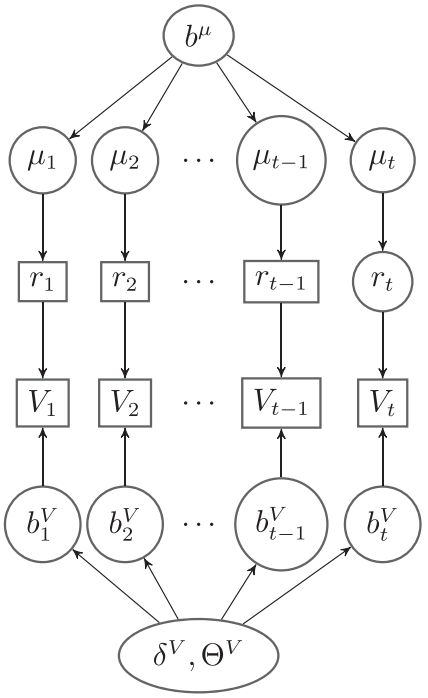


Figure 5. Graphical Representation of the Generalized Black–Litterman Model that Accounts for Violations in Black and Litterman's (1991, 1992) Assumptions Regarding the Market Equilibrium and Expert Views



Note. The terms in circles are unobserved factors at the start of time t that need to be estimated, whereas the terms in the rectangles are data.

3.2. Graphical Model

Our market consists of N risky assets and a risk-free asset. The vector of expert views observed at the beginning of period $t = 1, 2, \ldots$ is denoted by $V_t \in \mathbb{R}^K$, and the vector of returns observed at the end of time period is denoted by $r_t \in \mathbb{R}^N$. We denote the history of observed returns and expert views up to and including period t by $\mathcal{R}_t \equiv \{r_1, r_2, \ldots, r_t\}$ and $\mathcal{V}_t \equiv \{V_1, \ldots, V_t\}$, respectively. At the start of period t, the investor has a history of forecasts $\mathcal{V}_{t-1} = \{V_1, \ldots, V_{t-1}\}$ and returns $\mathcal{R}_{t-1} = \{r_1, \ldots, r_{t-1}\}$ and a forecast V_t of the return r_t (which will be observed only at the end of the period). The objective is to use these data, $\{\mathcal{V}_{t-1}, V_t\}$ and $\mathcal{R}_{t-1}$, to estimate the parameters and biases in the equilibrium and forecast model and to generate a forecast of the return r_t, which is to be used to optimize the time t asset allocation. We now provide details of the components of the equilibrium return and forecast model.

3.2.1. Equilibrium Model. We assume throughout that asset returns are normal with mean μ_t and covariance matrix Σ:

$$r_t|\mu_t \sim N(\mu_t, \Sigma). \tag{6}$$

The classical Black–Litterman model assumes that the mean return μ_t is deterministic and equals the forecast of expected returns α from the CAPM (i.e., $\mu_t = \alpha$ for every t in (6)). In the generalized Black–Litterman model, we assume that μ_t is randomly drawn each period from a normal distribution with mean $\alpha + b^\mu$ and covariance matrix β:

$$\mu_t|b^\mu \sim N(\alpha + b^\mu, \beta). \tag{7}$$

Here α is the CAPM forecast of expected return and is known to the investor, whereas b^μ is a constant but is unknown to the investor. The term b^μ models the bias in the CAPM estimate of expected returns, whereas the randomness in μ_t accounts for unmodeled stochastic factors that change μ_t over time. The term b^μ is the average impact of these unmodeled stochastic factors on the expected return, which can be learned over time, whereas fluctuations in these unmodeled factors is captured by the assumption that μ_t is latent and generated as independent and identically distributed (i.i.d.) at the start of each period with covariance β (and, hence, there is uncertainty in μ_t that is impossible to resolve).

Empirical failings of the CAPM are discussed in Fama and French (2004) and provide some justification of our model. (Figures 2 and 3 in Fama and French 2004 are examples of annualized expected returns deviating significantly from the predictions of the CAPM.) Although many of these limitations of the CAPM have been partly addressed by the Fama and French (1993, 1996) three-factor model, it still remains the case that these improved models are still misspecified (e.g., the three-factor model is unable to account for momentum effects; Fama and French 2004).

The covariance matrix β is a parameter of the model chosen by the investor that describes the uncertainty in the latent expected return μ_t. By virtue of (7), one might choose β so that a reasonable range of values for the expected return μ_t is covered by the spread associated with this normal distribution. For example, $\beta = 0.2^2 I$ means that the 95% confidence region of μ_t is $\pm 1.96 \times 0.2 = \pm 0.392$ ($\approx \pm 40\%$). Alternatively, one can choose Σ and β so that the covariance of returns under GBL is consistent with historical returns. Both approaches, together with experimental results, are discussed in Section 6.3.

We model uncertainty in the constant b^μ by specifying a prior $N(\delta_0^\mu, \Theta_0^\mu)$, where the hyperparameters $(\delta_0^\mu, \Theta_0^\mu)$ are user specified and are updated using Bayes' rule conditional on the history of forecasts $\mathcal{V}_{t-1}$ and returns $\mathcal{R}_{t-1}$ available at the start of period t.

3.2.2. Expert Views. Views in the classical Black–Litterman model are unbiased noisy observations of (linear functions of) returns. In the generalized Black–Litterman model, we allow for the possibility that views are biased by assuming

$$V_t|r_t, b_t^V \sim N(Pr_t + b_t^V, \Omega), \tag{8}$$

where biases $b_1^V, b_2^V, \ldots, b_t^V$ are not observed by the investor and drawn independently from a normal distribution

$$b_t^V | \delta^V, \Theta^V \sim N(\delta^V, \Theta^V). \tag{9}$$

The parameters (δ^V, Θ^V) of the bias distribution are constant but are not known by the investor. We assume a *normal inverse Wishart* prior,

$$(\delta^V, \Theta^V) \sim NIW(\eta_0^V, \kappa_0^V, \nu_0^V, \Lambda_0^V);$$

that is, the prior on the covariance matrix Θ^V is an inverse Wishart distribution with parameters $\nu_0^V \in \mathbb{R}$ and $\Lambda_0^V \in \mathbb{R}^{K \times K}$, whereas δ^V, given Θ^V, is normal:

$$\delta^V \sim N\left(\eta_0^V, \frac{1}{\kappa_0^V}\Theta^V\right),$$

where $\eta_0^V \in \mathbb{R}^K$ and $\kappa_0^V \in \mathbb{R}$. Under this specification, the mean of the prior on the covariance matrix $\mathbb{E}[\Theta^V] = \Lambda_0^V/(\nu_0^V - K - 1)$. If $\eta_0^V = 0$ and $\Lambda_0^V = 0$, then $\Theta^V = 0$, $\delta^V = 0$, and $b_t^V = 0$ (with probability 1), and the view model (9) degenerates to that of classical Black–Litterman where view bias is zero. The parameters $(\eta_0^V, \kappa_0^V, \nu_0^V, \Lambda_0^V)$ are updated using Bayes' rule and the history of views and returns.

Observe from (9) that Θ^V is not necessarily diagonal, so view biases at time t can be correlated. (This can happen, for example, if experts have the same source of information, or discuss views among themselves before coming up with a final opinion.) This model can handle data sets with asynchronous views (e.g., when different experts release views at different frequencies, or when the portfolios about which views are being expressed, as given in the rows of the link matrix P, can change from one period to the next). This would involve changing the link matrix P to match the views being expressed and introducing a bias term and associated hyperparameters for each view type. Bias parameters are updated only when the associated view is expressed. To ease notation, we will assume throughout that the link matrix does not change from period to period.

3.2.3. Asymmetric Treatment of β and Θ^V.

Our treatment of the equilibrium and view models is not symmetric in that the variance of expected returns (β) in the equilibrium model is assumed to be known, whereas that of the view bias (Θ^V) needs to be learned. This is a modeling choice that is justified in part by us being generally more familiar with returns than views (and hence, more confident about selecting ballpark values for β than Θ^V), as well as for computational reasons.

Specifically, the assets we are investing in are typically "long lived," so reasonable values for β can be chosen on the grounds that it is a measure the spread/uncertainty in the expected return μ_t, and we know what reasonable values of returns are. (Methods for specifying β are discussed in Section 6.3.) For the expert model, however, the statistical characteristics of the views will change depending on the group of experts providing them. Because track records are short, there is generally more uncertainty about the parameters describing the views and reasonable ranges for the view biases (which is described by Θ^V), though this can be reduced with data.

Second, the computational burden of parameter estimation increases with the number of parameters that need to be learned, so although it *is* possible to learn the pair (b^μ, β) using a normal inverse Wishart prior as we have done for (δ^V, Θ^V) instead of treating β as an input, we would then need to sample an $N \times N$ covariance matrix from the joint posterior in each round of Gibbs sampling, which corresponds to $N(N-1)/2$ random variables. When β is an input parameter, the number of samples grows linearly in N. (See Section 4.3 for discussion of how computational demands in each round of Gibbs sampling scale with the number of parameters in the model.)

3.3. Summary

The generalized Black–Litterman model is given by (6)–(9). It is assumed that the expected return estimate α is from the CAPM and known to the investor, whereas the covariance of returns Σ from (6), the uncertainty in the CAPM forecast β from (7), and the expert covariance matrix Ω for (8) are user specified. The goal at the start of period t is to determine the distribution of the future return r_t, conditional on the history of views $\mathcal{V}_{t-1}$ and returns $\mathcal{R}_{t-1}$, and the view V_t at the start of period t.

The posterior distribution for r_t is obtained from the joint posterior

$$P\big(r_t, (\mu_1, \ldots, \mu_{t-1}, \mu_t), b^\mu, (b_1^V, \ldots, b_{t-1}^V, b_t^V), (\delta^V, \Omega^V) \,|\, \mathcal{V}_{t-1}, \mathcal{R}_{t-1}, V_t\big),$$

where

- $\mu_1, \ldots, \mu_{t-1}$ is the history of expected returns and μ_t is the time t expected return,
- b^μ is the CAPM bias,
- $(b_1^V, \ldots, b_{t-1}^V)$ are the expert biases for previous periods, and b_t^V is the expert bias for the current period, and
- (δ^V, Θ^V) is the mean and covariance of the expert bias.

There is a normal prior on the CAPM bias $b^\mu \sim N(\delta_0^\mu, \Theta_0^\mu)$. The expert bias parameters have a normal inverse Wishart prior $(\delta^V, \Theta^V) \sim NIW(\eta_0^V, \kappa_0^V, \nu_0^V, \Lambda_0^V)$.

4. Estimation

Because of the complexity of the generalized Black–Litterman model, sampling methods must be used

to compute parameter estimates under the posterior. In this section, we describe Gibbs sampling, which is a Markov chain Monte Carlo method for generating samples from posterior distributions for models with graphical structure, and show how it can be applied to our model.

4.1. Gibbs Sampling

Suppose we would like to obtain samples of the multivariate random variable $X = (x_1, x_2, \ldots, x_n)$ with joint distribution $\mathbb{Q}(x_1, x_2, \ldots, x_n)$. The Gibbs sampling algorithm generates samples $X^{(1)}, X^{(2)}, \ldots$, where $X^{(i)} = (x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)})$ denotes the ith sample in the following way:

Initialize: Choose initial values for $X^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})$.

At the start of iteration $i+1$, we are given $X^{(i)} = (x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)})$. Construct $X^{(i+1)}$ as follows:

Sample: Sample $x_1^{(i+1)}$ from the conditional distribution

$$\mathbb{Q}\left(x_1 | x_2^{(i)}, x_3^{(i)}, \ldots, x_n^{(i)}\right)$$

and $x_j^{(i+1)}$, for $j = 2, \ldots, n$, from

$$\mathbb{Q}\left(x_j | x_1^{(i+1)}, x_2^{(i+1)}, \ldots, x_{j-1}^{(i+1)}, x_{j+1}^{(i)}, \ldots, x_n^{(i)}\right).$$

Repeat: Increase the index $i \to i+1$ and repeat the sampling step until the required number of samples $X^{(j)}$ have been generated.

It is well known (see, e.g., Gelman et al. 2004) that Gibbs samples have distribution $\mathbb{Q}$ as desired. Moreover, the algorithm defines a Markov chain $\{X^{(i)}, i = 0, 1, 2, \ldots\}$ with stationary distribution $\mathbb{Q}$, so convergence properties of these samples can be characterized using stability results for Markov chains (e.g., Meyn and Tweedie 1993, Gilks et al. 1996, Roberts and Rosenthal 2004). It is clear that implementability of this algorithm depends on how easily samples from the full conditional distributions $\mathbb{Q}(x_j | X_{-j})$ can be generated, which we now consider in the context of the generalized Black–Litterman model.

4.2. Gibbs Sampling for the Generalized Black–Litterman Model

In the notation of the previous section, we wish to sample the random variable

$$X = \left(r_t, b^\mu, \mu_1, \ldots, \mu_t, b_1^V, \ldots, b_t^V, \delta^V, \Theta^V\right)$$

from the posterior distribution

$$\mathbb{Q}(X) \equiv P\big(r_t, b^\mu, \mu_1, \ldots, \mu_t, b_1^V, \ldots, b_t^V, \delta^V, \Theta^V \,|\, r_1, \ldots, r_{t-1}, V_1, \ldots, V_{t-1}, V_t\big). \quad (10)$$

Priors on the bias parameters b^μ and (δ^V, Θ^V) are as specified in Section 3.3, and the parameters $\{\alpha, \beta, \Sigma, P, \Omega\}$ are also assumed to be known (Table 1). Data

$$\mathcal{D}_t = \{(V_1, r_1), (V_2, r_2), \ldots, (V_{t-1}, r_{t-1}), V_t\}$$

consists of the history of views and returns up to the previous time period $t-1$ and the view V_t of the future return r_t. The conditional distribution

$$P(r_t \,|\, r_1, \ldots, r_{t-1}, V_1, \ldots, V_{t-1}, V_t) \quad (11)$$

is our forecast of the return r_t. In contrast to the classical model, explicit expressions for the mean and variance of returns under the posterior (11) are not available. We now show how samples from the posterior can be generated using the Gibbs sampling algorithm described in Section 4.1.

Let

$$X^{(i)} \equiv \left(b^{\mu(i)}, \boldsymbol{\mu}_t^{(i)}, r_t^{(i)}, \mathcal{B}_t^{V(i)}, \delta^{V(i)}, \Theta^{V(i)}\right)$$

denote the sample generated in step i of the Gibbs sampling algorithm, where

$$\boldsymbol{\mu}_t^{(i)} \equiv \left\{\mu_1^{(i)}, \mu_2^{(i)}, \ldots, \mu_t^{(i)}\right\},$$
$$\mathcal{B}_t^{V(i)} \equiv \left\{b_1^{V(i)}, b_2^{V(i)}, \ldots, b_t^{V(i)}\right\}.$$

Then, $X^{(i+1)}$ is generated by sampling its components from the full conditional distributions of the posterior:

$$b^{\mu(i+1)} \sim P\left(b^\mu \,|\, \boldsymbol{\mu}_t^{(i)}, r_t^{(i)}, \mathcal{B}_t^{V(i)}, \left(\delta^{V(i)}, \Theta^{V(i)}\right), \mathcal{D}_t\right),$$

$$\mu_j^{(i+1)} \sim P\left(\mu_j \,|\, b^{\mu(i+1)}, \boldsymbol{\mu}_{-j}^{(i)}, r_t^{(i)}, \mathcal{B}_t^{V(i)}, \left(\delta^{V(i)}, \Theta^{V(i)}\right), \mathcal{D}_t\right), \quad j = 1, \ldots, t,$$

$$r_t^{(i+1)} \sim P\left(r_t \,|\, b^{\mu(i+1)}, \boldsymbol{\mu}_t^{(i+1)}, \mathcal{B}_t^{V(i)}, \left(\delta^{V(i)}, \Theta^{V(i)}\right), \mathcal{D}_t\right),$$

$$b_j^{V(i+1)} \sim P\left(b_j^V \,|\, b^{\mu(i+1)}, \boldsymbol{\mu}_t^{(i+1)}, r_t^{(i+1)}, \mathcal{B}_{-j}^{V(i)}, \left(\delta^{V(i)}, \Theta^{V(i)}\right), \mathcal{D}_t\right), \quad j = 1, \ldots, t,$$

$$\left(\delta^{V(i+1)}, \Theta^{V(i+1)}\right) \sim P\left((\delta^V, \Theta^V) \,|\, b^{\mu(i+1)}, \boldsymbol{\mu}_t^{(i+1)}, r_t^{(i+1)}, \mathcal{B}_t^{V(i+1)}, \mathcal{D}_t\right),$$

where $\boldsymbol{\mu}_{-j}^{(i)}$ denotes $\{\mu_1^{(i)}, \ldots, \mu_{j-1}^{(i)}, \mu_{j+1}^{(i)}, \ldots, \mu_t^{(i)}\}$ (and similarly for $\mathcal{B}_{-j}^{V(i)}$).

Consider first the problem of sampling $b^{\mu(i+1)}$. Figure 5 shows that, conditional on $\mu_1, \ldots, \mu_t, b^\mu$ is independent of the data $\mathcal{D}_t$ and the remaining parameters, so

$$b^{\mu(i+1)} \sim P\left(b^\mu \,|\, \boldsymbol{\mu}_t^{(i)}, r_t^{(i)}, \mathcal{B}_t^{V(i)}, \delta^{V(i)}, \Theta^{V(i)}, \mathcal{D}_t\right) \equiv P\left(b^\mu | \boldsymbol{\mu}_t^{(i)}\right).$$

Because b^μ has a normal prior $P(b^\mu) \equiv N(\delta_0^\mu, \Theta_0^\mu)$ and a normal likelihood (7), it follows from Bayes' rule that

$$P\left(b^\mu | \boldsymbol{\mu}_t^{(i)}\right) \equiv P\left(b^\mu \,|\, \mu_1^{(i)}, \ldots, \mu_t^{(i)}\right) \propto P(b^\mu) \times P\left(\mu_1^{(i)}, \ldots, \mu_t^{(i)} \,|\, b^\mu\right),$$

where $\boldsymbol{\mu}_t^{(i)}, = (\mu_1{}^{(i)}, \ldots, \mu_t{}^{(i)})$ plays the role of "data" in this update of the prior of b^μ to $P(b^\mu|\boldsymbol{\mu}_t^{(i)})$. It is well known that this distribution is normal, so $b^{\mu(i+1)}$ is easy to sample (we provide details in Step 1 of the algorithm below). More generally, the dependence structure implied by the graphical model and the choice of conditional distributions relating the different random quantities in this model imply that sampling from the full conditional distribution for $b^{\mu(i+1)}$ is easy.

Similarly, it can be seen from Figure 5 that

$$\mu_j^{(i+1)} \sim P\left(\mu_j \,|\, \boldsymbol{\mu}_{-j}^{(i)}, b^{\mu(i+1)}, r_t^{(i)}, \mathcal{B}_t^{V(i)}, \delta^{V(i)}, \Theta^{V(i)}, \mathcal{D}_t\right) \equiv P\left(\mu_j \,|\, b^{\mu(i+1)}, r_j\right),$$

and Bayes' rule together with Equations (6) and (7) imply that

$$P\left(\mu_j \,|\, b^{\mu(i+1)}, r_j\right) \propto P\left(r_j \,|\, \mu_j\right) P\left(\mu_j \,|\, b^{\mu(i+1)}\right),$$

where $P(\mu_j \,|\, b^{\mu(i+1)}) \equiv N(\alpha + b^{\mu(i+1)}, \beta)$ plays the role of a normal prior on μ_j (recall that α, the CAPM forecast, is assumed to be known, and the covariance matrix β is a user specified matrix), whereas $P(r_j \,|\, \mu_j)$ is a normal likelihood (6) for μ_j. It follows that $\mu_j^{(i+1)}$ is generated by sampling from a normal distribution, the parameters of which are provided in Step 2 of the algorithm below.

Observing the general rule that parameters associated with "parent" nodes in the graph in Figure 5 characterize the distribution that plays the role of the prior, whereas the parameters of the "children" play the role of data, the conditional distributions associated for $r_t{}^{(i+1)}$, $b_j^{V(i+1)}$ and $(\delta^{V(i+1)}, \Theta^{V(i+1)})$ can be similarly derived. As we have already seen, the full conditional distributions are easy to simulate if the conditional distributions relating parents to children form a conjugate pair.

Full conditional distributions for the remaining parameters $r_t, b_1^V, \ldots, b_t^V$, and (δ^V, Θ^V) can be derived similarly and the algorithm summarized as follows.

4.2.1. Gibbs Sampling Algorithm for GBL.

For every $i = 1, 2, 3, \ldots$, let the ith Gibbs sample be denoted by

$$X^{(i)} \equiv \left(b^{\mu(i)}, \boldsymbol{\mu}_t^{(i)}, r_t^{(i)}, \mathcal{B}_t^{V(i)}, \delta^{V(i)}, \Theta^{V(i)}\right),$$

where

$$\boldsymbol{\mu}_t^{(i)} \equiv \left\{\mu_1^{(i)}, \mu_2^{(i)}, \ldots, \mu_t^{(i)}\right\},$$
$$\mathcal{B}_t^{V(i)} \equiv \left\{b_1^{V(i)}, b_2^{V(i)}, \ldots, b_t^{V(i)}\right\}.$$

Given $\{\alpha, \beta, \Sigma, P, \Omega\}$ and prior hyperparameters $\{\delta_0^\mu, \Theta_0^\mu, \eta_0^V, \kappa_0^V, \nu_0^V, \Lambda_0^V\}$, the Gibbs sampling algorithm is as follows:

Initialize: Choose initial $X^{(0)}$.

For iteration $i+1$ in the loop, we are given $X^{(i)}$ and construct $X^{(i+1)}$ in the following manner:

Step 1. Sample $b^{\mu(i+1)} \sim N(\hat{\alpha}_{b^\mu}, \hat{\beta}_{b^\mu})$, where

$$\hat{\alpha}_{b^\mu} = \left(t\beta^{-1} + \{\Theta_0^\mu\}^{-1}\right)^{-1}\left[t\beta^{-1}\left(\overline{\mu}_t^{(i)} - \alpha\right) + \{\Theta_0^\mu\}^{-1}\delta_0^\mu\right],$$
$$\hat{\beta}_{b^\mu} = \left(t\beta^{-1} + \{\Theta_0^\mu\}^{-1}\right)^{-1},$$

and $\overline{\mu}_t^{(i)} \stackrel{\text{def}}{=} \frac{1}{t}\sum_{j=1}^t \mu_j^{(i)}$ is the sample mean of $\mu_1^{(i)}, \ldots, \mu_t^{(i)}$.

Step 2. Sample each component of $\boldsymbol{\mu}_t^{(i+1)}$ as $\mu_j^{(i+1)} \sim N(\hat{\alpha}_{\mu,j}, \hat{\beta}_{\mu,j})$ for $j = 1, 2, \ldots, t$, where

$$\hat{\alpha}_{\mu,j} = \left(\Sigma^{-1} + \beta^{-1}\right)^{-1}\left[\Sigma^{-1}r^* + \beta^{-1}\left(\alpha + b^{\mu(i+1)}\right)\right],$$
$$\hat{\beta}_{\mu,j} = \left(\Sigma^{-1} + \beta^{-1}\right)^{-1}$$

and

$$r^* = \begin{cases} r_j & \text{if } j = 1, 2, \ldots, t-1, \\ r_t^{(i)} & \text{if } j = t. \end{cases}$$

Step 3. Sample $r_t^{(i+1)} \sim N(\hat{\alpha}_r, \hat{\beta}_r)$, where

$$\hat{\alpha}_r = \left(P'\Omega^{-1}P + \Sigma^{-1}\right)^{-1}\left[P'\Omega^{-1}\left(V_t - b_t^{V(i)}\right) + \Sigma^{-1}\mu_t^{(i+1)}\right],$$
$$\hat{\beta}_r = \left(P'\Omega^{-1}P + \Sigma^{-1}\right)^{-1}.$$

Step 4. Sample each component of $\mathcal{B}_t^{V(i+1)}$ as $b_j^{V(i+1)} \sim N(\alpha_{b^V,j}, \beta_{b^V,j})$ for $j = 1, 2, \ldots, t$, where

$$\hat{\alpha}_{b^V,j} = \left(\Omega^{-1} + \Theta^{V(i)^{-1}}\right)^{-1}\left[\Omega^{-1}\left(V_j - Pr^*\right) + \Theta^{V(i)^{-1}}\delta^{V(i)}\right],$$
$$\hat{\beta}_{b^V,j} = \left(\Omega^{-1} + \Theta^{V(i)^{-1}}\right)^{-1}$$

and

$$r^* = \begin{cases} r_j & \text{if } j = 1, 2, \ldots, t-1, \\ r_t^{(i+1)} & \text{if } j = t. \end{cases}$$

Step 5. Sample $\delta^{V(i+1)}, \Theta^{V(i+1)} \sim NIW(\eta_t^V, \kappa_t^V, \nu_t^V, \Lambda_t^V)$, where

$$\eta_t^V = \frac{\kappa_0^V}{\kappa_0^V + t}\eta_0^V + \frac{t}{\kappa_0^V + t}\overline{b_t^V}^{(i+1)},$$
$$\kappa_t^V = \kappa_0^V + t,$$
$$\nu_t^V = \nu_0^V + t,$$
$$\Lambda_t^V = \Lambda_0^V + \sum_{j=1}^t \left(b_j^{V(i+1)} - \overline{b_t^V}^{(i+1)}\right)\left(b_j^{V(i+1)} - \overline{b_t^V}^{(i+1)}\right)' + \frac{\kappa_0^V t}{\kappa_0^V + t}\left(\overline{b_t^V}^{(i+1)} - \eta_0^V\right)\left(\overline{b_t^V}^{(i+1)} - \eta_0^V\right)',$$

and $\overline{b_t^V}^{(i+1)} = \frac{1}{t}\sum_{j=1}^t b_j^{V(i+1)}$.

Repeat: Repeat Steps 1–5 to generate as many samples of $X^{(j)}$ ($j \geq 1$) as necessary.

The samples $X^{(i)}$ are distributed according to the posterior distribution

$$\mathbb{Q}(X) \equiv P\big(r_t, b^\mu, \mu_1, \ldots, \mu_t, b_1^V, \ldots, b_t^V, \delta^V, \Theta^V \,|\, r_1, \ldots, r_{t-1}, V_1, \ldots, V_{t-1}, V_t\big),$$

and the samples $r_t^{(i)}$ have distribution $P(r_t|r_1, \ldots, r_{t-1}, V_1, \ldots, V_{t-1}, V_t)$. Convergence properties of estimates of the optimal portfolio generated using Gibbs samples (law of large numbers and central limit theorem) are discussed in Section 5.

4.2.2. Example. We simulate 50 periods of view and return data and the views for period 51 from a model of the form Figure 5 and estimate the generalized Black–Litterman model with these data using Gibbs sampling. The purpose of this experiment is to illustrate the output generated by the Gibbs sampling algorithm and its ability to identify errors/biases in the underlying data-generating model (DGM).

For the data-generating model, we assume a return covariance matrix

$$\Sigma = \begin{pmatrix} 0.015838413 & -0.01307935 & 0.001936589 \\ -0.013079347 & 0.03004989 & -0.012119245 \\ 0.001936589 & -0.01211924 & 0.007054950 \end{pmatrix}$$

and equilibrium expected return $\alpha = (0.09, 0.05, 0.04)$. We choose $b^\mu = 0$ and $\beta = 0$ so the true expected return is indeed α (there is no equilibrium error). The link matrix P is equal to the identity (absolute views only). The mean and variance of the view bias are $\delta^V = -0.4$ and $\Theta^V = 0.015^2 I$, respectively. The variance of the observed views is $\Omega = 0.01I$.

In setting $b^\mu = 0$ and $\beta = 0$, the assumptions of the classical Black–Litterman model concerning the equilibrium are satisfied in the data-generating model. The classical assumptions concerning the views are violated, however, and the plots show how the generalized model estimates the view bias and corrects for it in the posterior return distribution. A more comprehensive test comparing the performance of portfolios computed on the basis of the classical and generalized models, with real and simulated data, is presented in Section 6.

Figure 6 shows the prior and posterior distributions of δ^V, demonstrating that GBL is able to detect the negative bias in the simulation and be fairly accurate in its estimation, even with a very uninformative prior. Figure 7 shows the prior and posterior distributions for BL and GBL when the view biases b_{51}^V for each of the assets happened to be negative (the value of b_{51}^V for each of the assets is shown in the plots). This plot demonstrates the effect of view bias on the posterior BL return distributions and the correction of this bias in the GBL distribution. In particular, the GBL posterior return distribution for each of the assets has a higher mean due to its ability to correct the negative view bias. Consistent with (33), the GBL posterior variance is higher because of the additional uncertainty in the model.

4.3. Computational Demands

The computational demands in each round of the Gibbs sampling algorithm can be summarized as follows:

- b^μ, μ_1, $\mu_2, \ldots, \mu_t$, and r_t are N dimensional and multivariate normal, so the number of samples generated by the algorithm grows linearly in N;
- $b_1^V, b_2^V, \ldots, b_t^V$ and δ^V are K dimensional multivariate normal, whereas Θ^V is quadratic in K and has an inverse Wishart distribution, and the number of samples grows quadratically in K;
- there are $2t$ multivariate normal random variables μ_1, $\mu_2, \ldots, \mu_t$ and $b_1^V, b_2^V, \ldots, b_t^V$, so the number of samples is linear in t.

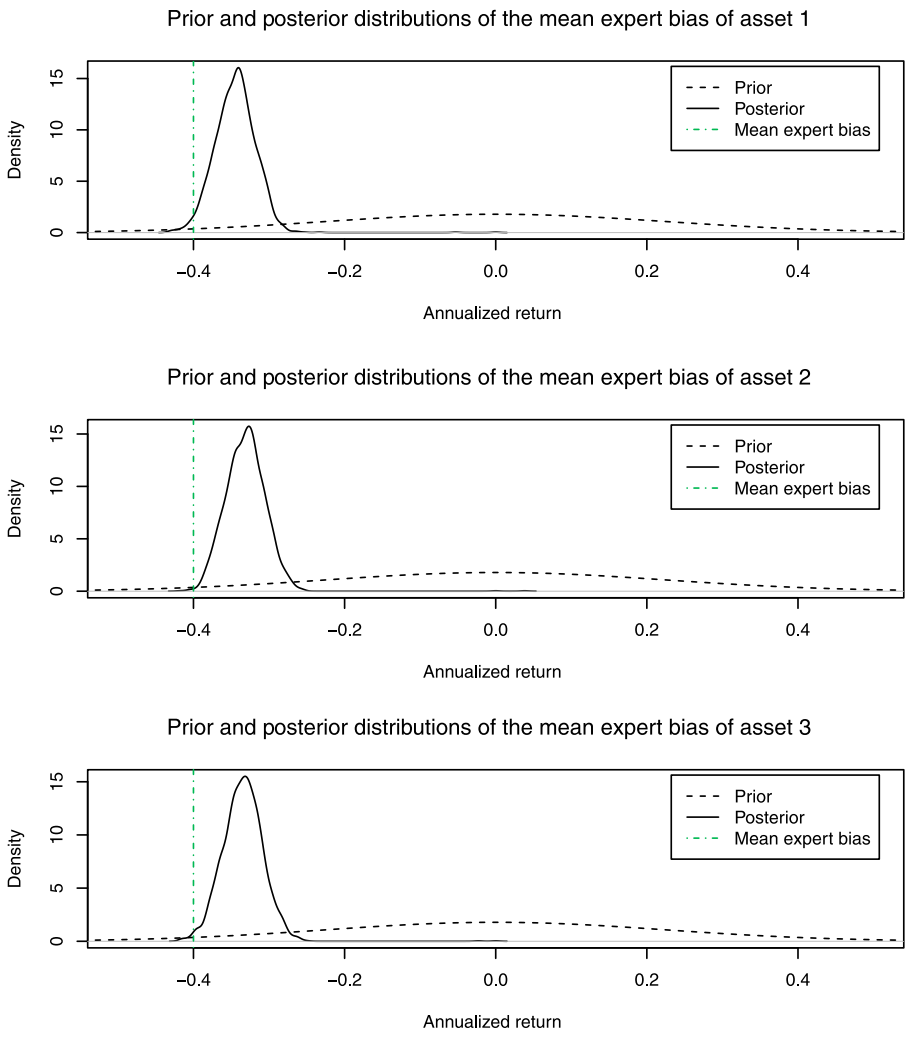
Whereas the number of samples grows linearly in t and N and quadratically in K, the random variables themselves are from well-known distributional families (multivariate normal, inverse Wishart). Of course, this is not the complete story, as it is the convergence rate of estimates, which are given by sample averages of the samples, that determines the number of rounds required to obtain estimates of sufficient accuracy. We discuss this issue in the next section.

As noted in Section 3.2, our current model assumes that the variance of the expected return β is known, which differs from our treatment of the view model where Θ^V needs to be learned. Although it is possible to learn (b^μ, β) jointly by adopting a normal inverse Wishart prior, the number of random variables that need to be generated in each round of Gibbs sampling then grows quadratically in N. Quadratic growth in K (in the case of Θ^V) is less of a concern because the number of views is typically smaller than the number of assets N. The impact of the number of assets, views, and historical samples on computational times is illustrated in Section 6.

5. Optimal Portfolio Choice

We present conditions that guarantee consistency and asymptotic normality of the estimate of the optimal portfolio computed using Gibbs samples and show how to determine the number of samples required to obtain a predefined level of accuracy in the estimate. Structural properties of the GBL portfolio are also discussed, including natural regularization properties of the GBL model.

Figure 6. (Color online) The Histograms Correspond to the Prior and Posterior Distributions of the Mean Expert Bias δ^V for Data Simulated from a Model Where There Is a Negative View Bias and No Equilibrium Error



Note. Although the expected value of the bias $\delta^V = -0.4$ in the data-generating model is underestimated, the direction of the bias is substantial and clearly captured.

5.1. Approximating the Optimal Portfolio

Consider the portfolio choice problem

$$\max_{\pi} \mathbb{E}_{\mathbb{Q}}[r_t]'\pi - \frac{\gamma}{2}\pi' \mathrm{Var}_{\mathbb{Q}}(r_t)\pi, \tag{12}$$

where the expectation and variance of returns are computed under the posterior

$$\mathbb{Q} \equiv \mathbb{P}(\cdot \,|\, (V_1, r_1), \ldots, (V_{t-1}, r_{t-1}), V_t).$$

The optimal portfolio

$$\pi^* = \frac{1}{\gamma}\mathrm{Var}_{\mathbb{Q}}(r_t)^{-1}\mathbb{E}_{\mathbb{Q}}[r_t]$$

is approximated by

$$\begin{aligned}\pi(m) &= \frac{1}{\gamma}\Lambda_t(m)^{-1}\bar{r}_t(m) \\ &\equiv \arg\max_{\pi}\left\{\bar{r}(m)'\pi - \frac{\gamma}{2}\pi'\Lambda(m)\pi\right\},\end{aligned} \tag{13}$$

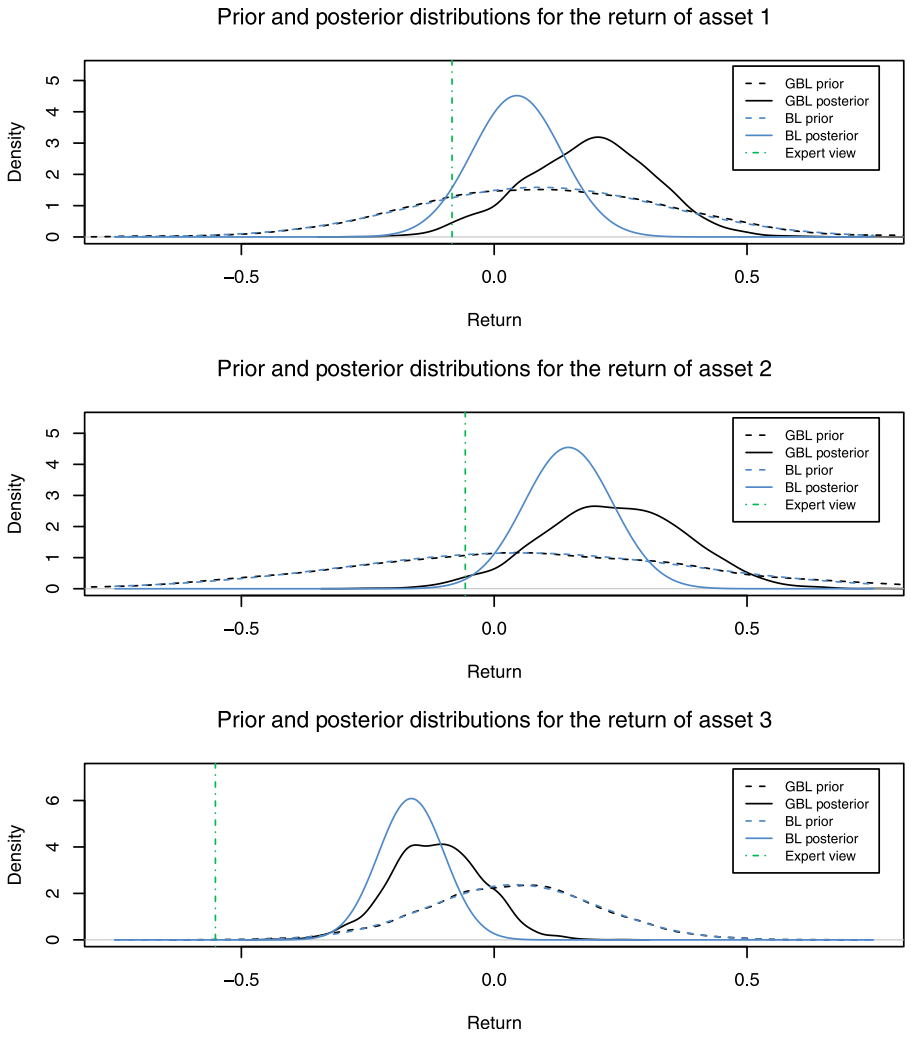
where estimates of the posterior mean and variance of returns

$$\begin{aligned}\bar{r}_t(m) &= \frac{1}{m}\sum_{i=1}^{m} r_t^{(i)}, \\ \Lambda_t(m) &= \frac{1}{m}\sum_{i=1}^{m}\left(r_t^{(i)} - \bar{r}_t\right)\left(r_t^{(i)} - \bar{r}_t\right)'\end{aligned} \tag{14}$$

are obtained using Gibbs samples. If $\mathbb{E}_{\mathbb{Q}}|r_t|^2 < \infty$, the ergodic theorem implies that

$$\begin{aligned}\lim_{m\to\infty}\bar{r}_t(m) &= \mathbb{E}_{\mathbb{Q}}[r_t], \\ \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} r_t^{(i)} r_t^{(i)\prime} &= \mathbb{E}_{\mathbb{Q}}[r_t r_t'] \Longrightarrow \lim_{m\to\infty}\Lambda_t(m) \\ &= \mathrm{Var}_{\mathbb{Q}}(r_t), \text{ a.s.}\end{aligned} \tag{15}$$

Figure 7. The Histograms Show the Prior and Posterior Distributions of Returns for Data Simulated from a Model Where There Is a Negative View Bias and No Equilibrium Error



Note. Here, the expected value of the bias in the data-generating model is negative ($\delta^V = -0.4$), and the realized bias of the expert views in period 51, b_{51}^V, is also negative. The generalized Black–Litterman model attempts to remove this bias by estimating δ^V. This results in a return distribution that is shifted to the right relative to that of the classical Black–Litterman model (which assumes that there is no view bias).

(see Meyn and Tweedie 1993, theorem 17.0.1; Roberts and Rosenthal 2004, theorem 4, fact 5), and the portfolio estimate is consistent:

$$\lim_{m\to\infty} \pi(m) = \pi^* \text{ (a.s.)}.$$

5.2. Central Limit Theorem for the Approximate Solution

Although consistency is a desirable property, the rate at which $\pi(m)$ converges to π^* is also of interest as this determines the number of samples m required to obtain a sufficiently accurate approximation. Similar issues are addressed in Shapiro et al. (2014) for the sample average approximation

$$\pi(m) \triangleq \arg\max_{\pi} \frac{1}{m}\sum_{i=1}^{m} f\left(\pi, r^{(i)}\right) \tag{16}$$

of the stochastic optimization problem

$$\pi^* \triangleq \arg\max_{\pi} \mathbb{E}_{\mathbb{Q}}[f(\pi, r)] \tag{17}$$

under the assumption that $r^{(i)}$ $(i = 1, \ldots, m)$ are generated i.i.d. from the distribution $\mathbb{Q}$ (typically a simulation model). These results do not apply in our setting because the approximate objective (13) is not a sample average of i.i.d. terms due to Gibbs samples being dependent and the variance term in (13) not being in the form (16). We now develop analogous results that can be applied to (12) and (13).

To account for the variance term in (12), consider the optimization problem

$$\max_{\pi,c}\left\{\mathbb{E}_{\mathbb{Q}}[r_t]'\pi - \frac{\gamma}{2}\mathbb{E}_{\mathbb{Q}}\left[\left(r_t'\pi - c\right)^2\right]\right\}. \tag{18}$$

It is easy to show that (π^*, c^*) is a solution of (18) if and only if π^* solves the original problem (12) and $c^* = \mathbb{E}_{\mathbb{Q}}[r_t]'\pi^*$, so both problems are equivalent. Likewise, $(\pi(m), c(m))$ is a solution of

$$\max_{\pi, c}\left\{\bar{r}_t(m)'\pi - \frac{\gamma}{2}\frac{1}{m}\sum_{i=1}^{m}\left(r_t^{(i)\prime}\pi - c\right)^2\right\} \tag{19}$$

if and only if $\pi(m)$ solves (13) and $c(m) = \bar{r}_t(m)'\pi(m)$. This allows us to study the asymptotics of $\pi(m)$ through (18) and (19) instead of (12) and (13), the advantage being that it no longer involves the variance of returns, and hence can be approximated by a sample average approximation.

Defining $\psi^* = \gamma\pi^* \in \mathbb{R}^N$ and $\eta^* = \gamma c^* \in \mathbb{R}$, the solution (π^*, c^*) of (18) is uniquely characterized by the first order conditions

$$\begin{bmatrix} \mathbb{E}_{\mathbb{Q}}(r_t) - \left(\mathbb{E}_{\mathbb{Q}}(r_t r_t')\psi^* - \mathbb{E}_{\mathbb{Q}}(r_t)\eta^*\right) \\ \left(\mathbb{E}_{\mathbb{Q}}(r_t)'\psi^* - \eta^*\right) \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{20}$$

(observe that this equation and (ψ^*, η^*) are independent of γ). Likewise, $(\pi(m), c(m))$ are uniquely determine by the first order conditions for the approximate problem (19),

$$\begin{bmatrix} \frac{1}{m}\sum_{i=1}^{m} r_t^{(i)} - \gamma\frac{1}{m}\sum_{i=1}^{m}\left(r_t^{(i)\prime}\pi(m) - c(m)\right)r_t^{(i)} \\ \gamma\frac{1}{m}\sum_{i=1}^{m}\left(r_t^{(i)\prime}\pi(m) - c(m)\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which can be written

$$A_m - \gamma B_m\begin{pmatrix} \pi(m) - \pi^* \\ c(m) - c^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{21}$$

where

$$A_m = \frac{1}{m}\sum_{i=1}^{m}\begin{bmatrix} r_t^{(i)} - \left(r_t^{(i)}r_t^{(i)\prime}\psi^* - r_t^{(i)}\eta^*\right) \\ \left(r_t^{(i)\prime}\psi^* - \eta^*\right) \end{bmatrix} = \frac{1}{m}\sum_{i=1}^{m} h\left(r_t^{(i)}\right),$$

$$B_m = \frac{1}{m}\sum_{i=1}^{m}\begin{bmatrix} r_t^{(i)}r_t^{(i)\prime} & -r_t^{(i)} \\ -r_t^{(i)\prime} & 1 \end{bmatrix},$$

with $h : \mathbb{R}^N \to \mathbb{R}^{N+1}$ defined by

$$h(r) = \begin{bmatrix} r - \left(rr'\psi^* - r\eta^*\right) \\ \left(r'\psi^* - \eta^*\right) \end{bmatrix}. \tag{22}$$

To characterize the asymptotic properties of $\pi(m)$, central limit theorems for Markov chains are needed. Toward this end, let $\{X^{(i)}, i = 0, 1, 2, \ldots\}$ be the Markov chain defined by the Gibbs sampling algorithm, $\mathbb{Q}$ be its stationary distribution (i.e., the GBL posterior at the start of period t), and

$$P^i(x, A) \triangleq \mathbb{P}\left[X^{(i)} \in A \mid X^{(0)} = x\right]$$

the associated i-step transition distribution of this Markov chain. We say that $\{X^{(i)}, i = 1, 2, 3, \ldots\}$ is *uniformly ergodic* (Roberts and Rosenthal 2004) if

$$\|P^i(x, \cdot) - \mathbb{Q}(\cdot)\| \triangleq \sup_A |P^i(x, \cdot) - \mathbb{Q}(\cdot)| \le M\rho^i, \quad i = 1, 2, 3, \ldots$$

for some $\rho < 1$ and $M < \infty$, where the supremum is taken over all measurable sets.

If $\{X^{(i)}, i = 0, 1, 2, \ldots\}$ is uniformly ergodic and $\mathbb{E}_{\mathbb{Q}}(|r_t|^4) < \infty$, the central limit theorem for Markov chains (see Roberts and Rosenthal 2004, theorem 23) implies

$$\sqrt{m}A_m \rightsquigarrow N(0, H), \tag{23}$$

where $\rightsquigarrow$ denotes weak convergence (convergence in distribution) and the covariance matrix $H \in \mathbb{R}^{(N+1)\times(N+1)}$ is given by

$$H \equiv \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \mathbb{E}_{\mathbb{Q}}\left[h\left(r_t^{(0)}\right)h\left(r_t^{(0)}\right)'\right] + 2\sum_{i=1}^{\infty}\mathbb{E}_{\mathbb{Q}}\left[h\left(r_t^{(0)}\right)h\left(r_t^{(i)}\right)'\right] \tag{24}$$

$(H_{11} \in \mathbb{R}^{N\times N}, H_{12}', H_{21} \in \mathbb{R}^N, H_{22} \in \mathbb{R})$. The condition $\mathbb{E}_{\mathbb{Q}}|r_t|^4 < \infty$ also implies

$$\lim_{m\to\infty} B_m = \begin{bmatrix} \mathbb{E}_{\mathbb{Q}}[r_t r_t'] & -\mathbb{E}_{\mathbb{Q}}[r_t] \\ -\mathbb{E}_{\mathbb{Q}}[r_t]' & 1 \end{bmatrix} \equiv L, \text{ a.s.} \tag{25}$$

(see (15)) and, hence, $B_m \rightsquigarrow L$. The following result characterizes the asymptotic properties of the approximate solution $\pi(m)$.

Theorem 1. *Suppose that the Markov chain $\{X^{(i)}, i = 0, 1, 2, \ldots\}$ generated by Gibbs sampling is uniformly ergodic and $\mathbb{E}_{\mathbb{Q}}(|r_t|^4) < \infty$. Then*

$$\sqrt{m}(\pi(m) - \pi^*) \rightsquigarrow N(0, V),$$

where

$$V = \frac{1}{\gamma^2}\mathrm{Var}_{\mathbb{Q}}(r_t)^{-1}\begin{bmatrix} I & \mathbb{E}_{\mathbb{Q}}(r_t) \end{bmatrix}\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \cdot \begin{bmatrix} I \\ \mathbb{E}_{\mathbb{Q}}(r_t)' \end{bmatrix}\mathrm{Var}_{\mathbb{Q}}(r_t)^{-1} \tag{26}$$

and H is defined by (24).

Observe that B_m and L are invertible, with

$$L^{-1} = \begin{bmatrix} \mathbb{E}_{\mathbb{Q}}[r_t r_t'] & -\mathbb{E}_{\mathbb{Q}}[r_t] \\ -\mathbb{E}_{\mathbb{Q}}[r_t]' & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} I \\ \mathbb{E}_{\mathbb{Q}}(r_t)' \end{bmatrix} \mathrm{Var}_{\mathbb{Q}}(r_t)^{-1} \begin{bmatrix} I & \mathbb{E}_{\mathbb{Q}}(r_t) \end{bmatrix}$$

and

$$B_m^{-1} = \left(\frac{1}{m} \sum_{i=1}^{m} \begin{bmatrix} r_t^{(i)} r^{(i)\prime} & -r_t^{(i)} \\ -r_t^{(i)\prime} & 1 \end{bmatrix} \right)^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} I \\ \bar{r}_t(m)' \end{bmatrix} \Lambda_t(m)^{-1} \begin{bmatrix} I & \bar{r}_t(m) \end{bmatrix},$$

where $\bar{r}_t(m)$ and $\Lambda_t(m)$ are defined as in (14). This implies that (21) can be written

$$\sqrt{m} \begin{pmatrix} \pi(m) - \pi^* \\ c(m) - c^* \end{pmatrix} = \frac{1}{\gamma} B_m^{-1} (\sqrt{m} A_m). \tag{27}$$

Because L is deterministic, B_m and L are invertible, $B_m \rightsquigarrow L$, and $\sqrt{m} A_m \rightsquigarrow N(0, H)$, it follows from Slutsky's theorem that

$$\sqrt{m} \begin{pmatrix} \pi(m) - \pi^* \\ c(m) - c^* \end{pmatrix} \rightsquigarrow N\left(0, \frac{1}{\gamma^2} L^{-1} H (L^{-1})'\right).$$

The term V is obtained by substituting the expression for L^{-1} and extracting the portion of the covariance matrix $L^{-1}H(L^{-1})'$ associated with $\pi(m) - \pi^*$.

5.3. Number of Gibbs Samples Required

We now show how Theorem 1 can be used to estimate the accuracy of an estimate $\pi(m)$ of π^* obtained using m Gibbs samples and the number of samples m^* required to generate an estimate with the desired level of accuracy.

It will be convenient to change the coordinate system to the one that is defined by the eigenvectors of V, (26). Let $(X_i, \lambda_i) \in \mathbb{R}^N \times \mathbb{R}$ $(i = 1, \ldots, N)$ be the N eigenvector–eigenvalue pairs of V, and let

$$\xi(m) \equiv [\xi_1(m) \ldots, \xi_N(m)]' = X'(\pi(m) - \pi^*)$$

be the decomposition of $\pi(m) - \pi^*$ into orthogonal components $\xi_1(m), \ldots, \xi_N(m)$, where

$$X = [X_1, \ldots, X_N], \ \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N).$$

We assume without loss of generality that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. We also assume that $\lambda_N > 0$. It follows from Theorem 1 that $\sqrt{m}\,\xi_i(m)$ $(i = 1, \ldots, m)$ are asymptotically independent and normal with variance λ_i:

$$\sqrt{m}\,\xi(m) \rightsquigarrow N(0, \Lambda), \quad \sqrt{m}\,\xi_i(m) \rightsquigarrow N(0, \lambda_i), \ i = 1, \ldots, N.$$

We wish to estimate the accuracy of an estimate $\pi(m)$ of π^* obtained from m samples, and the number of samples such that the error $\pi(m) - \pi^*$ in the k most variable directions $(k \leq N)$ is small with high probability:

$$\mathbb{P}(|\xi_i(m)| < \tau_i, \ i = 1, \ldots, k) > 1 - \epsilon. \tag{28}$$

By focusing on the k largest eigenvalues/eigenvectors of V, we account for the main sources of variability in the estimate of the optimal portfolio. The decision maker can set $k = N$ if he or she prefers.

Observing that

$$\begin{aligned} \mathbb{P}(|\xi_i(m)| < \tau_i, \ i = 1, \ldots, k) &= \prod_{i=1}^{k} \mathbb{P}(|\xi_i(m)| < \tau_i) \\ &= \prod_{i=1}^{k} \mathbb{P}\left(\sqrt{m} \frac{|\xi_i(m)|}{\sqrt{\lambda_i}} < \sqrt{\frac{m}{\lambda_i}} \tau_i \right) \\ &\sim \prod_{i=1}^{k} \left[2\Phi\left(\sqrt{\frac{m}{\lambda_i}} \tau_i \right) - 1 \right], \end{aligned} \tag{29}$$

where the first equality follows from the independence of $\xi_1(m), \ldots, \xi_k(m)$ and the last from asymptotic normality of $\sqrt{m}\,\xi_i(m)$, it follows that

$$m^* \triangleq \inf\left\{ m : \prod_{i=1}^{k} \left[2\Phi\left(\sqrt{\frac{m}{\lambda_i}} \tau_i \right) - 1 \right] > 1 - \epsilon \right\} \tag{30}$$

is the smallest number of samples such that (28) is satisfied. Given an estimate of V (and hence $\lambda_1, \ldots, \lambda_N$) the probability (29) and m^* are easy to compute. We will show shortly how V can be estimated using the output of Gibbs sampling.

5.3.1. Upper and Lower Bounds.

Although it is relatively easy to estimate m^* by solving (30) directly, this characterization does not show how m^* scales with the number of assets N. To show this, we derive upper and lower bounds on m^* and examine how these bounds depend on N. For this analysis, we control the error in the first k principal directions (28), so k remains fixed even when the number of assets N is changing. As in (29), we assume that m is sufficiently large for $\xi(m)$ to be asymptotically normal.

For a lower bound, observe that

$$\mathbb{P}(|\xi_1(m)| < \tau_1) \geq \mathbb{P}(|\xi_i(m)| < \tau_i, \ i = 1, \ldots, k),$$

so

$$\begin{aligned} m_L &\triangleq \inf\{m : \mathbb{P}(|\xi_1(m)| < \tau_1) > 1 - \epsilon\} \\ &= \left[\Phi^{-1}\left(1 - \frac{\epsilon}{2}\right) \right]^2 \frac{\lambda_1}{\tau_1^2} \end{aligned}$$

is a lower bound on m^*, where the second equality follows from the asymptotic normality of $\sqrt{m}\xi_i(m)$.

For an upper bound, observe that for m sufficiently large,

$$\begin{aligned}
&\mathbb{P}[|\xi_i(m)| < \tau_i,\ i = 1,\dots,k] \\
&\quad = \mathbb{P}\left[\frac{|\xi_i(m)|^2}{\lambda_i} < \frac{\tau_i^2}{\lambda_i},\ i = 1,\dots,k\right] \\
&\quad \geq \mathbb{P}\left[\frac{|\xi_i(m)|^2}{\lambda_i} < \min_{i=1,\dots,k}\{\tau_i^2/\lambda_i\}\right] \\
&\quad \geq \mathbb{P}\left[\sum_{i=1}^{k}\frac{|\xi_i(m)|^2}{\lambda_i} < \min_{i=1,\dots,k}\{\tau_i^2/\lambda_i\}\right],
\end{aligned}$$

so

$$m_U \triangleq \inf\left\{m : \mathbb{P}\left[\sum_{i=1}^{k}\frac{|\xi_i(m)|^2}{\lambda_i} < \min_{i=1,\dots,k}\{\tau_i^2/\lambda_i\}\right] > 1-\epsilon\right\} \tag{31}$$

is always larger than m^*. Because

$$m\sum_{i=1}^{k}\frac{|\xi_i(m)|^2}{\lambda_i}$$

is a χ^2-random variable with k degrees of freedom, the condition in (31) is satisfied if

$$m \times \min_{i=1,\dots,k}\left\{\frac{\tau_i^2}{\lambda_i}\right\} > F^{-1}_{\chi^2_k}(1-\epsilon),$$

so

$$m_U = \frac{k}{\min_{i=1,\dots,k}\left\{\frac{\tau_i^2}{\lambda_i}\right\}}\left[\frac{1}{k}F^{-1}_{\chi^2_k}(1-\epsilon)\right].$$

Note that

$$\frac{1}{k}F^{-1}_{\chi^2_k}(1-\epsilon) \sim O(1)$$

(indeed, the limit of this quantity is 1 as $k \to \infty$).

To see how m_L and m_U, and hence m^*, scale with the number of assets, we need to make assumptions on how the thresholds $\tau_1,\dots,\tau_k$ scale with N. Assuming

$$\tau_1 = \tau_2 = \cdots = \tau_k = \frac{\delta}{N},$$

it follows that

$$\begin{aligned}
m_L &= \left[\Phi^{-1}\left(1-\frac{\epsilon}{2}\right)\right]^2\lambda_1\left(\frac{N}{\delta}\right)^2, \\
m_U &= \left[\frac{1}{k}F^{-1}_{\chi^2_k}(1-\epsilon)\right]\lambda_1 k\left(\frac{N}{\delta}\right)^2,
\end{aligned}$$

so m^* is also $O(N^2)$. If $k = N$, then the upper bound scales like $O(N^3)$. The choice of $\tau_i \sim O(1/N)$ sets the equal weighted portfolio as a benchmark for the order of magnitude of the accuracy. In general, however, the appropriate choice of τ, δ, and k is problem specific.

The bounds are tight when $k = 1$, but the gap between m_L and m_U can be very large when $k > 1$. The lower bound m_L is typically closer to m^* than m_U because it corresponds to satisfying (30) in the direction of greatest variability (i.e., $k = 1$). Regardless, the purpose of these bounds is to understand how m^* scales with the number of assets, for which the large difference between the constants in m_L and m_U does not matter.

5.3.2. Estimating V. Clearly, an estimate of V is required, which can be obtained by estimating H using output from Gibbs sampling and V using (26). We can estimate H using any number of methods, and the following approach from chapter 3 of Gilks et al. (1996) is particularly easy to implement.

We simulate samples $X^{(i)}$ $(i = 1,\dots,m)$ from the posterior and divide them into B batches each of size S (so $m = SB$). For each batch $k = 1,\dots,B$, compute the sample average

$$Y_k = \frac{1}{S}\sum_{i=(k-1)S+1}^{kS} h\left(r_t^{(i)}\right),$$

where $h(r)$ is defined as in (22), and we approximate $\psi^* = \gamma\pi^*$ and $\eta^* = \gamma c^*$ by $\gamma\pi(m) = \Lambda_t(m)^{-1}\bar{r}_t(m)$ and $\gamma c(m) = \bar{r}_t(m)'(\gamma\pi(m)) = \bar{r}_t(m)'\Lambda_t(m)^{-1}\bar{r}_t(m)$. We choose the batch size S such that $Y_1,\dots,Y_B$ are nearly independent. (If independence is a concern, one can throw away a few samples between batches that are used to compute the Y_k's). Because $\sqrt{m}A_m \rightsquigarrow N(0, H)$, Y_k is approximately $N(0, H/S)$, so

$$H \approx \frac{S}{B}\sum_{k=1}^{B}(Y_k - \bar{Y})(Y_k - \bar{Y})',$$

where $\bar{Y}$ is the sample average of the Y_k's. An estimate of V can now be obtained from (26), approximating $\mathbb{E}_{\mathbb{Q}}(r_t)$ and $\mathrm{Var}_{\mathbb{Q}}(r_t)$ by $\bar{r}_t(m)$ and $\Lambda_t(m)$ (see (14)).

5.3.3. Example. We can estimate the eigenvalues of V for our experiment in Section 6 ($N = 50$ assets, $T = 60$ months of historical data, equilibrium bias $b^\mu = 0.15$, view bias $\delta^v = 0.3$). These are shown in Figure 8 and were computed with $m = 20{,}000$ Gibbs samples divided into $B = 400$ blocks of size $S = 50$ using the procedure described above.

It follows from (30) that an accuracy level of $\tau = 0.015$ ($\delta = 0.75$) is satisfied with probability 0.85 ($\epsilon = 0.15$) with $m = 20{,}000$ samples when $k = 50$. We would require $m = 23{,}500$ samples ($m_L = 19{,}000$, $m_U = 444{,}100$) to achieve an accuracy level of $\tau = 0.015$ with probability 0.90.

For an accuracy level of $\tau = 0.01$ with probability 0.9 ($k = 50$), $m = 53{,}000$ samples are required. With $k = 3$ instead of $k = 50$, $m = 50{,}100$ samples are required, and $|\xi_i(m)| < 0.01$ for all i (not just $i \leq 3$) with probability 0.886. When $k = 5$, $m = 52{,}200$, and $|\xi_i(m)| < 0.01$ for all i with probability 0.897.

Although the probability estimate (30) complements the use of convergence plots to assess whether sufficiently many samples have been generated, $m = 20{,}000$ was sufficient for the standard errors of our Sharpe ratio estimates to be sufficiently small.

5.4. Adaptive Regularization

We derive an alternative expression for the mean and variance of returns that facilitates comparison with the classical Black–Litterman model and reveals intrinsic regularization properties of the GBL model.

Let X_{-r_t} be X with r_t deleted. It follows from Figure 5 (see also Step 3 of the algorithm in Section 4.2) that

$$\mathbb{E}[r_t \mid X_{-r_t}] = \left(P'\Omega^{-1}P + \Sigma^{-1}\right)^{-1}\left[P'\Omega^{-1}\left(V_t - b_t^V\right) + \Sigma^{-1}\mu_t\right],$$

which is random because of its dependence on latent variables b_t^V and μ_t. It follows that

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}[r_t] &= \mathbb{E}_{\mathbb{Q}}\{\mathbb{E}_{\mathbb{Q}}[r_t \mid X_{-r_t}]\} \\ &= \left(P'\Omega^{-1}P + \Sigma^{-1}\right)^{-1}\left[P'\Omega^{-1}\left(V_t - \mathbb{E}_{\mathbb{Q}}\left(b_t^V\right)\right)\right. \\ &\quad \left. + \Sigma^{-1}\mathbb{E}_{\mathbb{Q}}(\mu_t)\right]. \end{aligned} \tag{32}$$

This expression for the expected return is a natural generalization of (4) for the classical model. Indeed, the (unconditional) expected return $\mu = \alpha$ from the CAPM in (4) is replaced by $\mathbb{E}_{\mathbb{Q}}(\mu_t)$ in (32), which includes a correction for equilibrium bias, whereas the other term in (32) is analogous to the view update in (4), which also includes a correction $\mathbb{E}_{\mathbb{Q}}(b^\mu)$ for bias.

To calculate $\mathrm{Var}_{\mathbb{Q}}(r_t)$, we utilize the conditional variance formula

$$\mathrm{Var}_{\mathbb{Q}}(r_t) = \mathbb{E}_{\mathbb{Q}}[\mathrm{Var}_{\mathbb{Q}}(r_t \mid X_{-r_t})] + \mathrm{Var}_{\mathbb{Q}}(\mathbb{E}_{\mathbb{Q}}[r_t \mid X_{-r_t}]).$$

From our specification of the generalized Black–Litterman model (Figure 5),

$$\mathrm{Var}_{\mathbb{Q}}(r_t \mid X_{-r_t}) = \hat{\beta}_r = \left(P'\Omega^{-1}P + \Sigma^{-1}\right)^{-1}$$

is deterministic, so

$$\mathbb{E}_{\mathbb{Q}}[\mathrm{Var}_{\mathbb{Q}}(r_t \mid X_{-r_t})] = \left(P'\Omega^{-1}P + \Sigma^{-1}\right)^{-1}.$$

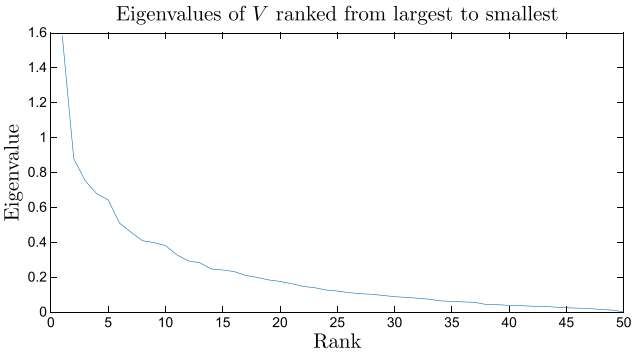
To compute $\mathrm{Var}_{\mathbb{Q}}(\mathbb{E}_{\mathbb{Q}}[r_t \mid X_{-r_t}])$, we use Gibbs sampling to generate

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}\left[r_t \mid X_{-r_t}^{(i)}\right] &= \hat{\alpha}_{r_t}^{(i)} \\ &= \left(P'\Omega^{-1}P + \Sigma^{-1}\right)^{-1} \\ &\quad \cdot \left[P'\Omega^{-1}\left(V_t - b_t^{V(i)}\right) + \Sigma^{-1}\mu_t^{(i)}\right]\end{aligned}$$

(see Step 3 of the algorithm), so

$$\mathrm{Var}(\mathbb{E}[r_t \mid X_{-r_t}]) \approx \frac{1}{m}\sum_{i=1}^{m}\left[\hat{\alpha}_{r_t}^{(i)} - \overline{\alpha}_{r_t}\right]\left[\hat{\alpha}_{r_t}^{(i)} - \overline{\alpha}_{r_t}\right]',$$

Figure 8. (Color online) Eigenvalues of the Matrix V Defined by (26) Ranked from Largest to Smallest for Our Experiment in Section 6 with $N = 50$ Assets and $T = 60$ Months of Historical Data



where

$$\overline{\alpha}_{r_t} = \frac{1}{m}\sum_{i=1}^{m}\hat{\alpha}_{r_t}^{(i)}.$$

It now follows that

$$\begin{aligned}\mathrm{Var}_{\mathbb{Q}}(r_t) &= \left(P'\Omega^{-1}P + \Sigma^{-1}\right)^{-1} + \mathrm{Var}_{\mathbb{Q}}(\mathbb{E}_{\mathbb{Q}}[r_t \mid X_{-r_t}]) \\ &\approx \left(P'\Omega^{-1}P + \Sigma^{-1}\right)^{-1} \\ &\quad + \frac{1}{m}\sum_{i=1}^{m}\left[\hat{\alpha}_{r_t}^{(i)} - \overline{\alpha}_{r_t}\right]\left[\hat{\alpha}_{r_t}^{(i)} - \overline{\alpha}_{r_t}\right]'.\end{aligned} \tag{33}$$

The variance of returns for the generalized model is larger than that of the classical model (4) because of the additional term $\mathrm{Var}_{\mathbb{Q}}(\mathbb{E}_{\mathbb{Q}}[r_t \mid X_{-r_t}])$. The impact of this term on the return distribution can be seen in Figure 7, where the spread of return r_t under the generalized Black–Litterman posterior is higher than that under the classical model.

It follows from (33) that

$$\begin{aligned}&\min_{\pi}\left\{\mathbb{E}_{\mathbb{Q}}[r_t]'\pi - \frac{\gamma}{2}\pi'\mathbb{E}[\mathrm{Var}_{\mathbb{Q}}(r_t)]\pi\right\} \\ &\equiv \mathbb{E}_{\mathbb{Q}}[r_t]'\pi - \frac{\gamma}{2}\pi'\left(P'\Omega^{-1}P + \Sigma^{-1}\right)^{-1}\pi \\ &\quad - \frac{\gamma}{2}\pi'\mathrm{Var}_{\mathbb{Q}}(\mathbb{E}_{\mathbb{Q}}[r_t \mid X_{-r_t}])\pi,\end{aligned}$$

so the mean-variance objective for GBL equals that of the classical model with an additional term, $\pi'\mathrm{Var}_{\mathbb{Q}}(\mathbb{E}_{\mathbb{Q}}[r_t \mid X_{-r_t}])\pi$, which plays the role of an L_2 regularizer. Because the diagonal entries of $\mathrm{Var}_{\mathbb{Q}}(\mathbb{E}_{\mathbb{Q}}[r_t|X_{-r_t}])$ are larger for assets with more estimation uncertainty, the shrinkage for each asset holding is roughly proportional to the level of uncertainty in its expected return estimate. This differs from classical regularization methods in portfolio choice applications where the same penalty is used for all variables (see, e.g., DeMiguel et al. 2009, where it is shown that imposing portfolio constraints is related to regularization).

Observe, too, that $\mathrm{Var}_{\mathbb{Q}}(\mathbb{E}_{\mathbb{Q}}[r_t|X_{-r_t}])$ never completely disappears, because the expected return $\mu_t \sim N(\alpha + b^\mu, \beta)$ and the view bias $b_t^V \sim N(\delta^V, \Theta^V)$ are generated i.i.d. each period, so there will always be some degree of uncertainty if β and Θ^V are nonzero, even if the parameters b^μ, β, δ^V, and Θ^V are known.

Norms of GBL portfolios for training data sets of varying length are reported as part of our experiment in Section 6 and clearly show the impact of regularization.

6. Empirical Tests

We now present empirical tests of the generalized Black–Litterman model with both simulated and real data. Computations in the following section were done using an Intel Core i7-4600U central processing unit at 2.1–2.7 GHz with 16 GB of RAM.

6.1. Controlled Tests

For our controlled tests, we simulate market data from the model in Figure 5 and evaluate the out-of-sample performance of four models:

1. the generalized Black–Litterman model;
2. the classical Black–Litterman model that ignores bias and invests according to the Equations (3) and (4) with knowledge of the parameters $(\alpha, \Sigma, P, \Omega)$;
3. a generalized Black–Litterman model (GBL+) that has learned the value of the bias parameters b^μ and (δ^V, Θ^V) being used in the data-generating model;
4. an investor who does not account for bias and trades the market portfolio (MP).

GBL+ is the GBL model with a degenerate prior on the values of b^μ and (δ^V, Θ^V) used in the data-generating model. GBL trades identically to GBL+ once it has an infinite amount of historical data, so the out-of-sample performance of GBL+ is an upper bound of that of GBL.

6.1.1. Data-Generating Model.

The parameters α and Σ are generated randomly but consistent with the CAPM, whereas returns in the data-generating model are assumed to follow (6) and (7). We assume that the return of the market portfolio has mean 0.1 and standard deviation 0.2. For the purposes of this experiment, we assume that the variance of the return bias in the data-generating model $\beta = 0.35^2 \times I$, and that the GBL investor adopts the same value of β. A discussion of approaches to selecting β and a study of the sensitivity of GBL performance to misspecification is also provided.

Views in the data-generating model are generated according to (8) and (9). For the purposes of this experiment, we assume there are $K = 15$ views with link matrix

$$P = \left[I_K \ \underline{0}_{K\times(N-K)} \right] \in \mathbb{R}^{K\times N}.$$

We use $\Omega = 0.75^2 \times I$ and $\Theta^V = 0.85^2 \times I$ in the data-generating model throughout. We vary the means of the view and equilibrium biases δ^V and b^μ depending on the experiment.

6.1.2. GBL Investor.

We assume that the parameters α, Σ, β, P, and Ω in the data-generating model are known to the GBL investor (Table 1).

The prior on the view bias (δ^V, Θ^V) is normal inverse Wishart with parameters $(\eta_0^V, \kappa_0^V, \nu_0^V, \Lambda_0^V)$. In this paper, there are $K = 15$ views, and we choose

$$\begin{aligned} \eta_0^V &= \underline{0} \in \mathbb{R}^K, \\ \kappa_0^V &= K + 4 = 19, \\ \nu_0^V &= K + 4 = 19, \\ \Lambda_0^V &= (\nu_0^V - K - 1) \times 0.85^2 \times I = 3 \times 0.85^2 \times I. \end{aligned} \tag{34}$$

With these choices, $\mathbb{E}[\delta^V] = 0$, and

$$\mathrm{Var}(\delta^V) \approx \frac{\mathbb{E}[\Theta^V]}{\kappa_0^V} \times I = 0.195^2 \times I$$

$(\mathbb{E}[\Theta^V] = 0.85^2 I)$. In other words, 95% of the prior distribution for each component of the view bias approximately covers the interval $0 \pm 2 \times 19.5\% \equiv \pm 39\%$.

The prior on the equilibrium bias b^μ is normal, $N(\delta^\mu, \Theta^\mu)$. For all experiments,

$$\begin{aligned} \delta_0^\mu &= \underline{0} \in \mathbb{R}^N, \\ \Theta_0^\mu &= 0.35^2 \times I; \end{aligned} \tag{35}$$

that is, the expected bias (under the prior) $\mathbb{E}(b^\mu) = 0$, and the standard deviation of the bias for each asset is 35%, so 95% of the prior approximately covers the interval $0 \pm 2 \times 35\% = \pm 70\%$.

6.2. Simulation Experiments

Although GBL accounts for potential misspecification in BL by allowing for biases and errors in the equilibrium and the views, the downside is that significantly more parameters need to be estimated. In this section, we compare the benefits of improving the model to the cost of increasing estimation uncertainty.

We begin with experiments where data are generated according to the model described in Section 3 with the parameter values listed in Section 6.1.1 to illustrate the impact of N, T, and K on out-of-sample performance. Following this, we evaluate GBL when modeling assumptions are violated, including misspecification of β and nonnormality of views and returns in the data-generating model.

We evaluate GBL by comparing its out-of-sample Sharpe ratio to the benchmarks GBL+, BL, and MP. The difference between BL and MP can be interpreted as the value of having (biased) expert views, whereas the difference between GBL+ and BL is the potential

improvement over BL if we adopt GBL and successfully learn all the parameters. GBL+ is an upper bound on the performance of GBL that is achieved when the size of the training set $T \to \infty$ and measures the *potential* benefit of using GBL. The difference between GBL and GBL+ can be attributed to estimation uncertainty. Because there is a risk-free asset, the efficient frontier is linear, and the Sharpe ratio is equal to its slope. The Sharpe ratio is also independent of the risk-aversion parameter.

6.2.1. Experiment. We simulate returns and views from the data-generating model described in Section 6.1.1 and report out-of-sample Sharpe ratios of GBL (with randomly generated training data of lengths $T = 0, 12, 60$, and 120 months), GBL+, BL, and MP for different view and equilibrium biases in the data-generating model. Sharpe ratios are computed by generating a history of length T and applying the resulting portfolios to one realization of period $T + 1$ asset returns. This is repeated 200 times, and the collection of 200 (period $T + 1$) investment returns are used to compute the Sharpe ratio for the portfolio. Common histories and period $T + 1$ returns are used across all experiments, and a common random seed is used in all implementations of Gibbs sampling to reduce sampling variability. Standard errors for Sharpe ratios are also reported using Opdyke (2007) (which accounts for dependence in the samples of the investment returns induced by using a common random seed in Gibbs sampling for each of the 200 histories).

Convergence Plots. For each of the $M = 200$ histories, $m = 20{,}000$ samples from the posterior are generated using Gibbs sampling. Figure 9 shows estimates of the portfolio for one such history (history 100 in our experiment) consisting of $T = 60$ months of training data and $N = 15, 50$, and 250 assets, whereas Figure 10 shows estimates of three key parameters, the equilibrium and view biases b^μ and δ^V and the expected return $\mathbb{E}(r_{T+1})$ for each asset, under the posterior. These plots can be used to assess convergence of key parameter estimates and the optimal portfolio holdings. Both plots are typical for all simulated histories and appear stable.

6.2.2. Discussion. Our first experiment is divided into two parts, characterized by the nature of the data-generating model. In the first part (Tables 2–5), the equilibrium bias in the data-generating model has zero mean ($b^\mu = 0$). This is consistent with the assumptions of BL and is a setting where BL should do well. For the same reason, this is a setting where the additional flexibility of GBL is unnecessary, and we are interested to see the impact of this redundancy on the out-of-sample performance. (Tables 2–4 have $K = 15$ views, whereas Table 5 reports the same experiment except with $K = N$ views, one for each asset.) In the second part (Tables 6–8), the mean equilibrium bias in the data-generating model is nonzero ($b^\mu = 0.15$). This is a setting where BL incorrectly assumes that biases are zero. We would like to evaluate the cost of misspecification in BL (i.e., the difference between BL and GBL+) and the degree to which this can be bridged when GBL is calibrated using a limited data set.

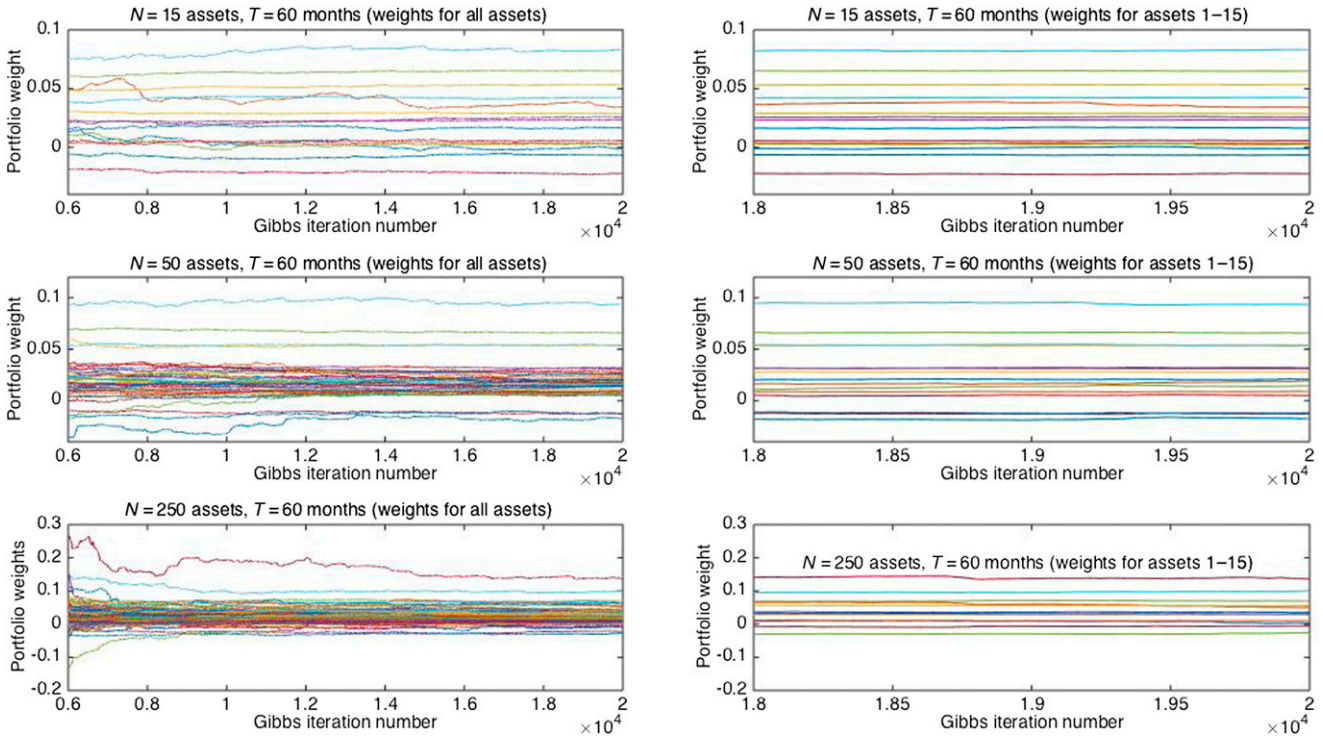
The results reported in Tables 2–5 and 6–8 are consistent with our intuition about the performance of BL in the two settings, respectively. Specifically, BL does well when $b^\mu = 0$ and $\delta^V = 0$ in Tables 2–5, with its out-of-sample Sharpe ratio being almost at the level of GBL+. (There is a small difference because BL ignores the variability in the equilibrium and view biases, and though this difference increases when the view bias δ^V deviates from 0, it remains relatively small.) In contrast, the difference between GBL+ and BL is large in Tables 6–8, meaning that erroneously assuming that biases are zero in BL has a significant impact on its out-of-sample performance in these tests.

In the case of GBL, observe that it performs at about the level of BL in both experiments when there are no training data ($T = 0$). We also see in Table 8 that GBL improves when T increases, quickly outperforming BL and almost reaching the level of GBL+, illustrating the value of learning the biases in this example. There are some unusual features, however, that we wish to highlight. First, whereas the performance of GBL improves in Tables 5–8 when N increases (T fixed), the opposite occurs in Tables 2–4, where it gets worse. Second, whereas GBL improves monotonically in T in Tables 6–8, this is not the case in Tables 2–5, where performance gets worse between $T = 0$ and 12 months. We provide explanations of both observations below.

Performance of GBL as a Function of N. The out-of-sample performance of GBL is determined by two factors, the *value of the investment opportunities* and *estimation uncertainty*. When N increases, there are more opportunities, but also more parameters to estimate, and the out-of-sample performance of GBL depends on the relative impact of each. To understand the behavior of GBL, it is helpful to disentangle these effects. We measure the value of the investment opportunities using the out-of-sample Sharpe ratio of GBL+, and the impact of estimation uncertainty using the ratio of the out-of-sample Sharpe ratio of GBL to that of GBL+.

In this example, the Sharpe ratio of GBL+ (investment potential) increases in N in all cases, though this increase is small in Table 4. (The only difference

Figure 9. Estimates of Portfolio Weights for a Given $T = 60$ Months of Data as a Function of the Number of Gibbs Samples (200 Such Historical Samples Were Used in Our Simulation Study)



Notes. The plots on the left show estimates of portfolio weights for iterations 6,000 to $M = 20{,}000$ of Gibbs sampling, whereas the plots on the right zero in on iterations 18,000 to 20,000 for the first 15 assets. The N portfolio weights obtained at iteration 20,000 were used in the simulation of out-of-sample performance. There are $K = 15$ views and $b^\mu = 0.15$ and $\delta^V = 0.30$ in the data-generating model.

between Tables 4 and 5 is the number of views, so the improvement in GBL+ in Table 5 occurs because the additional views improve the returns forecast). On the other hand, Figure 11 shows that the performance of GBL relative to GBL+ in Tables 4 (top), 5 (middle), and 8 (bottom) drops as N increases, which is consistent with the intuition that more estimation uncertainty hurts out-of-sample performance. GBL gets worse in Table 4 because GBL+ hardly improves when N increases (the potential associated with these new assets is small), so even a small drop-off in performance relative to GBL+ due to estimation uncertainty will result in the Sharpe ratio getting worse. On the other hand, the increase in GBL+ exceeds the drop-off from estimation uncertainty in Tables 5 and 8, which leads to the improvement in the Sharpe ratio of GBL as N increases.

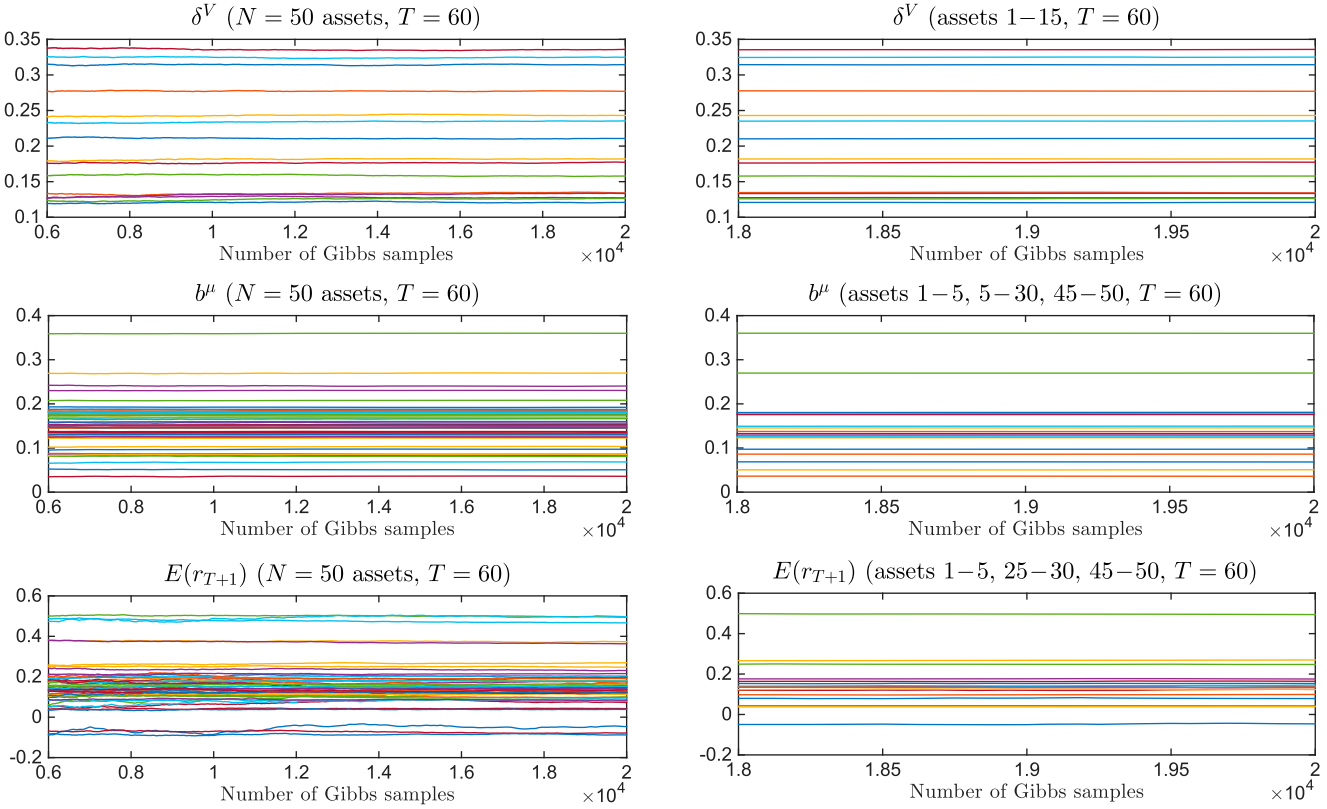
Dependence on T. To explain why the Sharpe ratio of GBL is not monotone in T in Table 4, we take one of the 200 simulated historical data sets from our experiment and examine the posterior distribution for the equilibrium bias parameter b_i^μ after $T = 12$, 60, and 120 months of training data for a collection of assets i.

We begin with the case when $b^\mu = 0.15$ and $\delta^V = 0.30$ in the data-generating model. The prior on the bias parameters has mean zero (see (34) and (35)), and we see from Table 8 that the out-of-sample Sharpe ratio of GBL improves when T increases.

Figure 12 shows estimates of the bias (i.e., posterior means) $\{\mathbb{E}_{\mathbb{Q}_T}(b_i^\mu),\ i = 1, \ldots, N\}$ for all $N = 100$ assets after $T = 0, 12, 60$, and 120 months of data. Although the distance between $\mathbb{E}_{\mathbb{Q}_T}(b_i^\mu)$ and 0.15 is large for some of the assets, the general tendency is to learn the bias, with the estimates being generally positive and becoming more concentrated around 0.15 as T increases. This improvement in the bias estimate leads to a monotonic improvement in the performance of GBL.

As noted, we do not always have monotonicity in the Sharpe ratio of GBL, which decreases when going from $T = 0$ to $T = 12$ in Table 4 before improving again. This occurs because the prior mean for these parameters in GBL is zero (i.e., $\mathbb{E}_{\mathbb{Q}_0}(b_i^\mu) = 0$), which just happens to coincide with the value of the bias parameters b^μ and δ^V in the data-generating model; that is, GBL does well when there are no training data ($T = 0$) because it was lucky in its choice of prior. For the same reason, we see from Figure 13 that data add nothing but noise to the estimate $\mathbb{E}_{\mathbb{Q}_T}(b_i^\mu)$, and the

Figure 10. Estimates of the Bias Parameters (δ^V, b^μ) and Expected Returns for $N = 50$ Assets with $T = 60$ Months of Data as a Function of the Number of Gibbs Samples



Notes. The same historical data set was used in Figure 9 for $N = 50$ assets. The plots on the left show estimates for iterations 6,000 to $M = 20{,}000$ of Gibbs sampling, whereas the plots on the right zero in on iterations 18,000 to 20,000 for the first 15 assets. The estimates obtained at iteration 20,000 were used in the simulation of out-of-sample performance. There are $K = 15$ views and $b^\mu = 0.15$ and $\delta^V = 0.30$ in the data-generating model.

performance of GBL with $T = 12$ months of training data is worse. The performance of GBL improves as T increases beyond 12 because the additional data allow GBL to correct this mistake, though it does not reach the level of $T = 0$ with the quantity of data provided, which is insufficient to eliminate all of the estimation uncertainty (Figure 13).

Shrinkage Effects. Table 9 shows the average l^1 norm of the GBL, GBL+, BL and MP portfolios over the 200 randomly sampled histories generated for the simulation study when $b^\mu = 0.15$ and $\delta^V = 0.3$ in the data-generating model. The risk-aversion parameter is fixed at $\gamma = 50$ for all experiments. The effects of shrinkage on GBL portfolios is clear; they are smallest when there is no historical data and the shrinkage penalty is the largest, and grow as the model learns and the shrinkage penalty decreases.

Computational Times. Table 10 shows the average time to generate 20,000 Gibbs samples for each entry of Table 4 (averaged over 200 histories). Table 11 shows average computational times for Table 5, where there are $K = N$ instead of $K = 15$ views. For example, it takes on average 54.1 seconds to generate 20,000 Gibbs samples when there are $N = 250$ assets, $K = 15$ views, and $T = 120$ months of historical data.

Whereas run times increase along the rows (N increasing, T and K fixed) and down the columns (T increasing, N and K fixed), there is a bigger jump when going from a given (N, T) entry in Table 10 to the same entry in Table 11. For instance, although moving from (N, T) = (15, 0) to (N, T) = (250, 0) in Table 10 corresponds to an increase in N by 235 assets ($T = 0$ and $K = 15$ is fixed), and moving from (N, T) = (250, 0) in Table 10 to the same entry in Table 11 is an increase in K of 235 views ($T = 0$ and $N = 250$ fixed), the jump in computation time from the change in K is significantly larger. (Likewise, the increase in run times when going downs the columns of Table 11, so T is increasing while N and $K = N$ are fixed, is modest.) This is consistent with our discussion in Section 4.2, where the number of samples required in each iteration of Gibbs sampling increases linearly in T and N, and quadratically in K (other variables remaining fixed).

Table 2. $b^{\mu} = 0$ and $N = 20$ in the Data-Generating Model

	Mean view bias δ^V		
Portfolio ($N = 20$ assets)	–0.3	0	0.3
GBL+	1.30 (±0.08)	1.30 (±0.08)	1.30 (±0.08)
BL	1.25 (±0.08)	1.30 (±0.08)	1.27 (±0.09)
GBL (history $T = 0$ months)	1.32 (±0.09)	1.34 (±0.08)	1.29 (±0.08)
GBL (history $T = 12$ months)	1.13 (±0.08)	1.13 (±0.08)	1.08 (±0.07)
GBL (history $T = 60$ months)	1.24 (±0.08)	1.24 (±0.08)	1.27 (±0.08)
GBL (history $T = 120$ months)	1.27 (±0.08)	1.26 (±0.08)	1.27 (±0.08)
MP	0.23 (±0.07)	0.23 (±0.07)	0.23 (±0.07)

Table 3. $b^{\mu} = 0$ and $N = 100$ in the Data-Generating Model

	Mean view bias δ^V		
Portfolio ($N = 100$ assets)	–0.3	0	0.3
GBL+	1.33 (±0.08)	1.33 (±0.08)	1.33 (±0.08)
BL	1.27 (±0.08)	1.31 (±0.08)	1.27 (±0.09)
GBL (history $T = 0$ months)	1.26 (±0.08)	1.28 (±0.08)	1.22 (±0.08)
GBL (history $T = 12$ months)	0.80 (±0.08)	0.80 (±0.07)	0.77 (±0.07)
GBL (history $T = 60$ months)	1.06 (±0.08)	1.05 (±0.08)	1.05 (±0.08)
GBL (history $T = 120$ months)	1.14 (±0.10)	1.15 (±0.10)	1.14 (±0.10)
MP	0.33 (±0.07)	0.33 (±0.07)	0.33 (±0.07)

Table 4. $b^{\mu} = 0$ and $\delta^V = 0$ in the Data-Generating Model

	Number of assets (N)				
Portfolio	15	20	50	100	250
GBL+	1.31 (±0.08)	1.30 (±0.08)	1.33 (±0.08)	1.33 (±0.08)	1.34 (±0.08)
BL	1.30 (±0.08)	1.30 (±0.08)	1.33 (±0.08)	1.31 (±0.08)	1.33 (±0.09)
GBL (history $T = 0$ months)	1.35 (±0.09)	1.34 (±0.08)	1.33 (±0.09)	1.28 (±0.08)	1.19 (±0.08)
GBL (history $T = 12$ months)	1.21 (±0.08)	1.13 (±0.08)	0.93 (±0.08)	0.80 (±0.07)	0.63 (±0.07)
GBL (history $T = 60$ months)	1.29 (±0.09)	1.24 (±0.08)	1.19 (±0.08)	1.05 (±0.08)	0.96 (±0.09)
GBL (history $T = 120$ months)	1.25 (±0.09)	1.26 (±0.08)	1.27 (±0.08)	1.15 (±0.10)	1.13 (±0.09)
MP	0.22 (±0.07)	0.23 (±0.07)	0.49 (±0.07)	0.33 (±0.07)	0.47 (±0.07)

Table 5. $b^{\mu} = 0$ and $\delta^V = 0$ in the Data-Generating Model with $K = N$ Views

	Number of assets (N)				
Portfolio	15	20	50	100	250
GBL+	1.31 (±0.08)	1.45 (±0.09)	2.11 (±0.10)	2.61 (±0.15)	3.73 (±0.26)
BL	1.30 (±0.08)	1.46 (±0.09)	2.13 (±0.11)	2.64 (±0.15)	3.76 (±0.26)
GBL (history $T = 0$ months)	1.35 (±0.09)	1.54 (±0.10)	2.23 (±0.14)	2.52 (±0.21)	3.26 (±0.31)
GBL (history $T = 12$ months)	1.21 (±0.08)	1.31 (±0.09)	2.08 (±0.12)	2.38 (±0.15)	3.03 (±0.28)
GBL (history $T = 60$ months)	1.29 (±0.09)	1.44 (±0.10)	2.08 (±0.12)	2.50 (±0.17)	3.47 (±0.44)
GBL (history $T = 120$ months)	1.25 (±0.09)	1.43 (±0.09)	2.11 (±0.11)	2.46 (±0.15)	3.63 (±0.25)
MP	0.22 (±0.07)	0.23 (±0.07)	0.49 (±0.07)	0.33 (±0.07)	0.47 (±0.07)

6.3. Robustness Tests

6.3.1. Misspecification of the Model of Returns and Views. GBL assumes that the distribution of views and returns r_t, μ_t, V_t, and b_t^V are (conditionally) multivariate normal. To test the robustness of GBL to violation of this assumption, we apply it to a data-generating model where returns and views are multivariate t-distributions. (The correlation structure used in the tests in Section 6.2 is retained, however). GBL, BL, GBL+, and MP continue to make the same assumptions as before. Sharpe ratios are reported in Table 12.

Table 6. $b^{\mu} = 0.15$ and $N = 20$ in the Data-Generating Model

	Mean view bias δ^V		
Portfolio ($N = 20$ assets)	−0.3	0	0.3
GBL+	1.95 (±0.11)	1.95 (±0.11)	1.95 (±0.11)
BL	0.99 (±0.08)	1.28 (±0.09)	1.49 (±0.09)
GBL (history $T = 0$ months)	1.10 (±0.08)	1.41 (±0.09)	1.64 (±0.10)
GBL (history $T = 12$ months)	1.67 (±0.09)	1.71 (±0.09)	1.73 (±0.09)
GBL (history $T = 60$ months)	1.86 (±0.10)	1.85 (±0.10)	1.90 (±0.11)
GBL (history $T = 120$ months)	1.87 (±0.11)	1.85 (±0.11)	1.88 (±0.11)
MP	−0.21 (±0.07)	−0.21 (±0.07)	−0.21 (±0.07)

Table 7. $b^{\mu} = 0.15$ and $N = 100$ in the Data-Generating Model

	Mean view bias δ^V		
Portfolio ($N = 100$ assets)	−0.3	0	0.3
GBL+	4.05 (±0.22)	4.05 (±0.22)	4.05 (±0.22)
BL	1.17 (±0.09)	1.45 (±0.09)	1.64 (±0.10)
GBL (history $T = 0$ months)	1.36 (±0.09)	1.67 (±0.10)	1.92 (±0.11)
GBL (history $T = 12$ months)	3.03 (±0.16)	3.07 (±0.16)	3.09 (±0.17)
GBL (history $T = 60$ months)	3.50 (±0.19)	3.51 (±0.19)	3.55 (±0.19)
GBL (history $T = 120$ months)	3.71 (±0.22)	3.70 (±0.22)	3.73 (±0.22)
MP	0.63 (±0.07)	0.63 (±0.07)	0.63 (±0.07)

Table 8. $b^{\mu} = 0.15$ and $\delta^V = 0.3$ in the Data-Generating Model

	Number of assets N				
Portfolio	15	20	50	100	250
GBL+	1.71 (±0.10)	1.95 (±0.11)	2.79 (±0.14)	4.05 (±0.22)	6.30 (±0.33)
BL	1.47 (±0.09)	1.49 (±0.09)	1.59 (±0.09)	1.64 (±0.10)	1.63 (±0.10)
GBL (history $T = 0$ months)	1.60 (±0.10)	1.64 (±0.10)	1.61 (±0.11)	1.92 (±0.11)	1.43 (±0.09)
GBL (history $T = 12$ months)	1.63 (±0.09)	1.73 (±0.09)	2.29 (±0.14)	3.09 (±0.17)	4.64 (±0.22)
GBL (history $T = 60$ months)	1.70 (±0.09)	1.90 (±0.11)	2.70 (±0.15)	3.55 (±0.19)	5.87 (±0.32)
GBL (history $T = 120$ months)	1.60 (±0.09)	1.88 (±0.11)	2.62 (±0.14)	3.73 (±0.22)	5.74 (±0.30)
MP	−0.28 (±0.07)	−0.21 (±0.07)	0.47 (±0.07)	0.63 (±0.07)	0.70 (±0.07)

Although the Sharpe ratios of all portfolios are lower, the qualitative performance of all portfolios mirrors that in Tables 7 and 8 (column $N = 100$), with the performance of GBL being around or below the level of BL when there is no history, but substantially exceeding the performance of BL with 12 months of data.

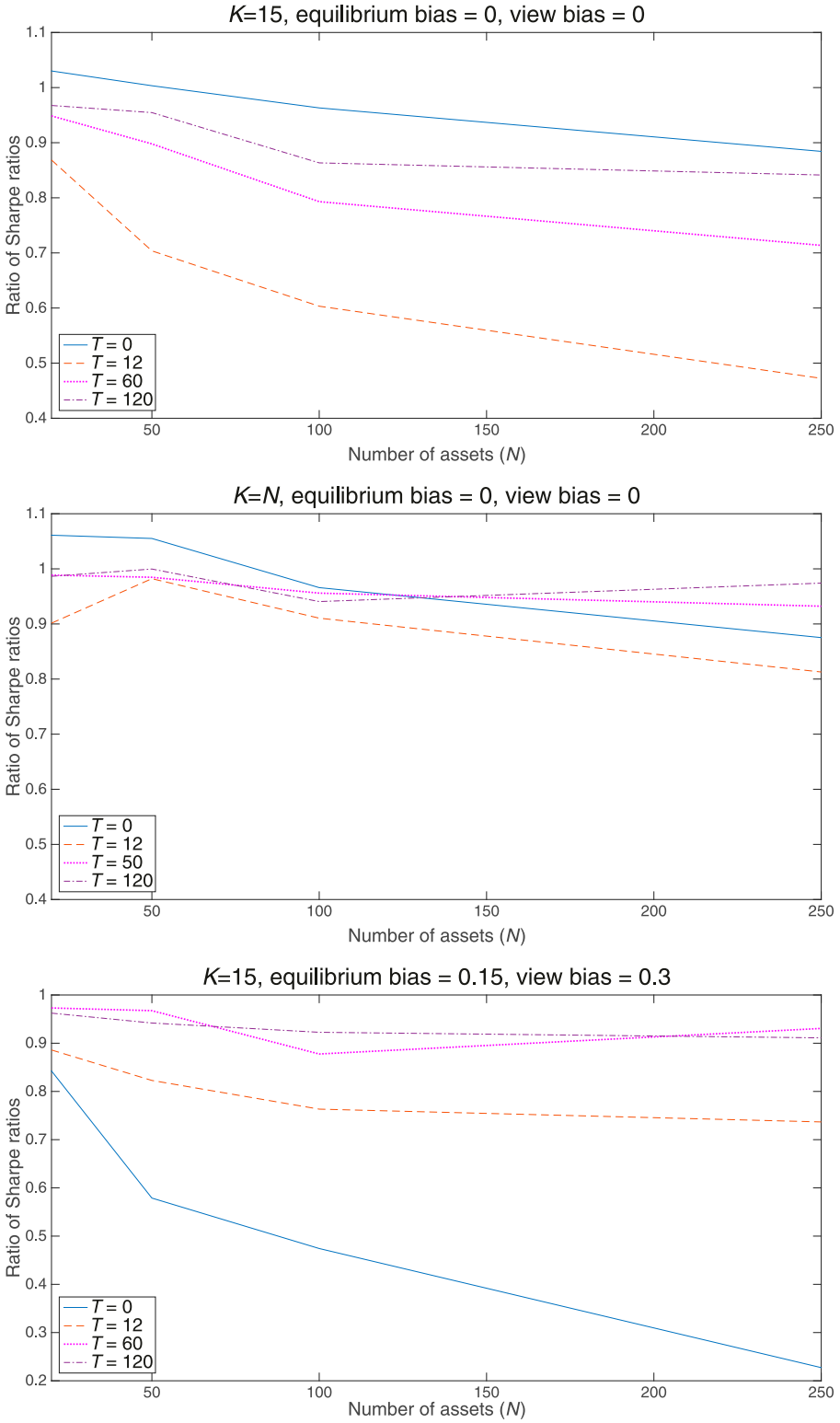
6.3.2. Misspecification of β. For this particular subsection, let β_{GBL} denote the value of β that is used in GBL, and let β_{DGM} be the one used in the data-generating model.

In applications, there are several ways of choosing β_{GBL}. One approach is to choose a matrix such that the spread of μ_t around $\alpha + b^{\mu}$ is reasonable. In our experiment, $\beta_{GBL} = 0.35^2 \times I$, so 95% of the deviations under this model are within ±70% of the mean (quite a conservative/large choice).

Although specifics of the application can lead to reasonable specifications of this parameter, this approach is admittedly quite ad hoc, and there are natural concerns about misspecification. To address this, we conduct an out-of-sample test with $N = 100$ assets and our existing data-generating model (which has $\beta_{DGM} = 0.35^2 \times I$) and evaluate the impact when GBL erroneously assumes that $\beta_{GBL} = (1 + \delta) \times \beta_{DGM}$ over a range of values for δ. As in our other tests, we include GBL+, BL, and MP. Note, however, that BL and MP do not depend on β, whereas GBL+ uses the same value of β as the data-generating model, so the performance of GBL+, BL, and MP are unchanged from the test where β_{GBL} is correctly specified (see column $N = 100$ in Table 8).

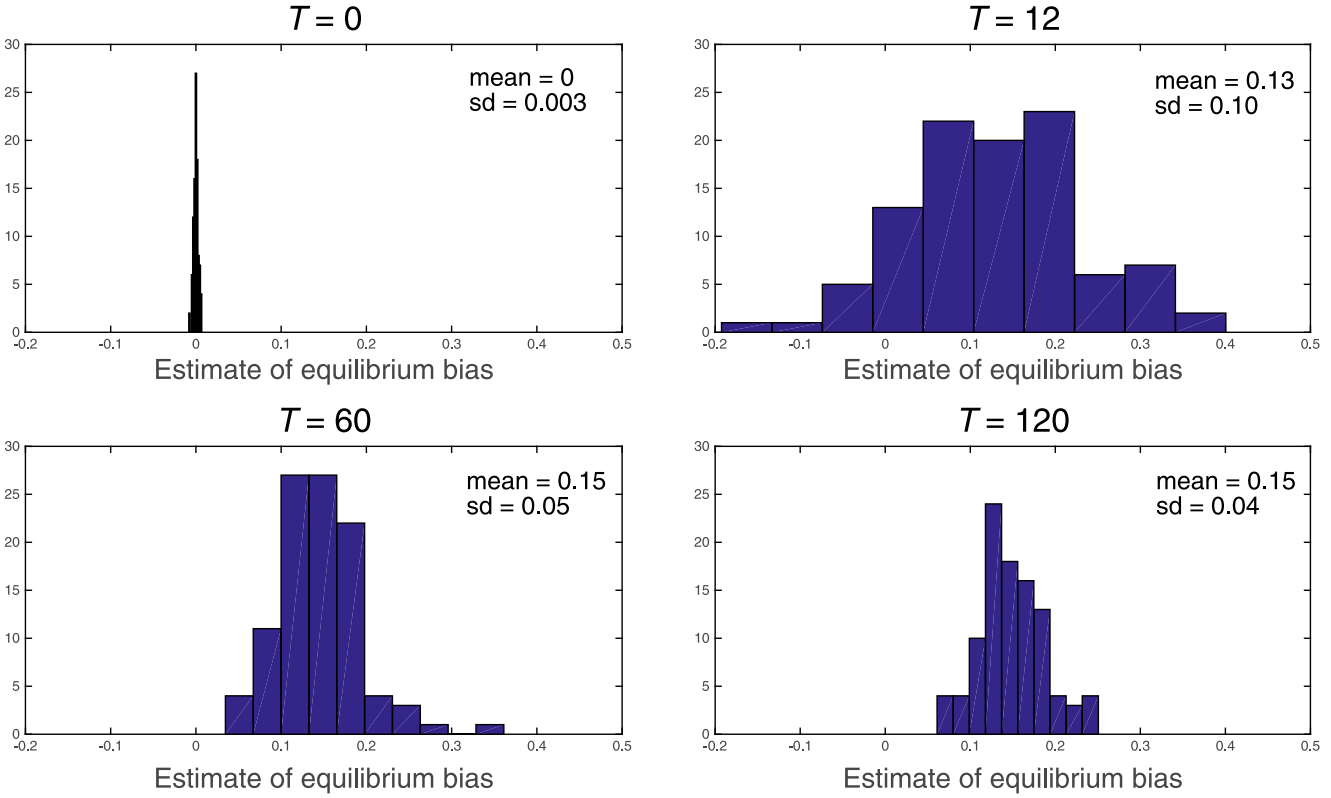
We see from Table 13 that although there is some drop in performance as δ increases from 0 (though there is also improvement, in this particular example, when δ decreases), the same qualitative properties for GBL as the experiments where it correctly specifies β_{GBL} are apparent: GBL learns quickly, and its

Figure 11. (Color online) Plot of the Ratio of the Sharpe Ratio of GBL to the Sharpe Ratio of GBL+ for Tables 4 (Top), 5 (Middle), and 8 (Bottom)



Note. There are $K = 15$ views in the top and bottom plots, and $K = N$ in the middle plot.

Figure 12. (Color online) Distribution of Posterior Estimates $\{\mathbb{E}_{\mathbb{Q}_T}(b_i^\mu),\ i = 1, \ldots, N\}$ When $b^\mu = 0.15$ and $\delta^V = 0.3$ in the Data-Generating Model with $N = 100$ Assets and $K = 15$ Views



Notes. As seen in the first plot, the mean of the prior on b_i^μ is 0 (for all i), which differs from its actual value (0.15). The model starts to correct for this as T increases, with the distribution of estimates of b^μ becoming more centered around 0.15. sd, Standard deviation.

performance exceeds that of BL with 12 months of data in all cases except $\delta = 0.4$, where 60 months of data are needed.

A second approach is based on the observation that the covariance matrix of returns r_t under the data-generating model is given by

$$\mathrm{Var}(r_t) = \mathbb{E}\big(\mathrm{Var}(r_t \mid \mu_t)\big) + \mathrm{Var}\big(\mathbb{E}(r_t \mid \mu_t)\big) = \Sigma + \beta$$

(conditional variance formula); that is, the covariance of returns r_t under the data-generating model comes from the variability β of the expected return μ_t around $\alpha + b^\mu$ and variability Σ of returns r_t around the realized expected return μ_t.

It follows that one (albeit heuristic) method of specifying β_{GBL} is to estimate the covariance matrix $\mathrm{Var}(r_t)$ of r_t from data, and to attribute some proportion δ of this to β_{GBL} with the remainder to Σ_{GBL}:

$$\beta_{GBL} = \delta \mathrm{Var}(r_t),\ \Sigma_{GBL} = (1-\delta)\mathrm{Var}(r_t). \qquad (36)$$

One advantage is that the resulting covariance structure of r_t under GBL (conditional on the bias parameters) is consistent with the history of returns generated by the data-generating model, though it is unclear how δ should be chosen and whether out-of-sample performance is sensitive to this choice. To address this concern, we adopt the same data-generating model as in our earlier examples and choose

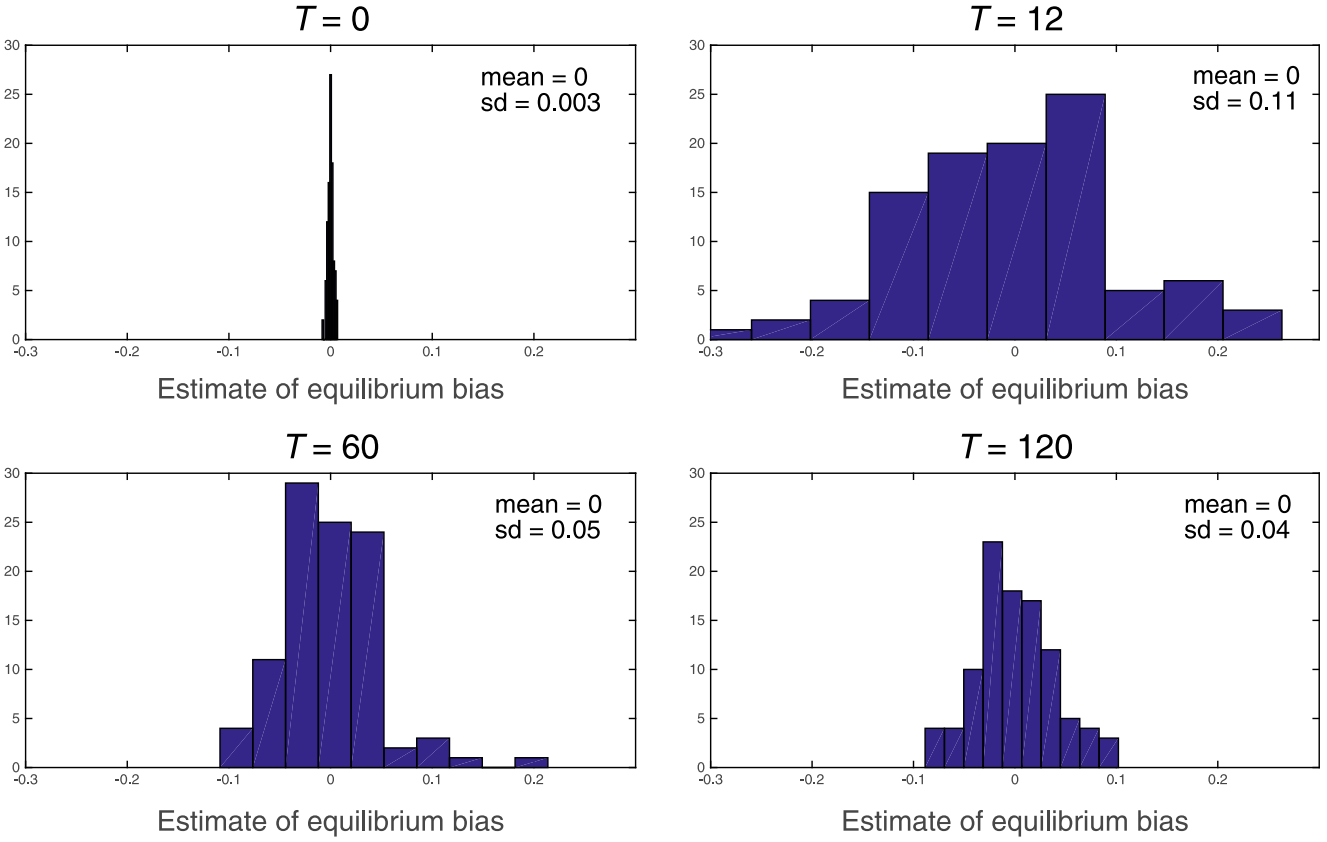
$$\beta_{GBL} = \delta \times (\Sigma_{DGM} + \beta_{DGM}),$$
$$\Sigma_{GBL} = (1-\delta) \times (\Sigma_{DGM} + \beta_{DGM})$$

(recall, $\beta_{DGM} = 0.35^2 \times I$) for a range of δ's. We also include GBL+, BL, and MP, which use correctly specified $\beta = \beta_{DGM}$ and $\Sigma = \Sigma_{DGM}$. Sharpe ratios reported in Table 14 show good performance over a reasonably large range of δ (for this example). This approach is used in the real data experiments in the next section, where $\delta = 1/5$.

6.4. Backtests

In this experiment, we compare the performance of classical Black–Litterman and generalized Black–Litterman models over the 20-year period 1993–2012 on the ten industry portfolios from Kenneth French's data library.[1] The 13-week treasury bill (IRX) is the risk-free asset. To assess the impact of view bias, we

Figure 13. (Color online) Distribution of Posterior Estimates $\{\mathbb{E}_{\mathbb{Q}_T}(b_i^\mu), i = 1, \ldots, N\}$ When $b^\mu = 0$ and $\delta^V = 0$ in the Data-Generating Model with $N = 100$ Assets and $K = 15$ views



Notes. The mean of the prior on b^μ is 0 (for all i), which just happens to be its value in the data-generating model, so in this particular example, data just adds noise and makes the estimate worse (relative to $T = 0$). GBL gradually corrects this error as the amount of data increases. sd, Standard deviation.

Table 9. Average of $\|\pi\|_1$ When $b^\mu = 0.15$ and $\delta^V = 0.3$ in the Data-Generating Model with $K = 15$ Views and Risk Aversion $\gamma = 50$

	Number of assets (N)				
Portfolio	15	20	50	100	250
GBL+	0.3	0.4	1.0	1.9	4.7
BL	0.6	0.6	0.6	0.6	0.6
GBL (history $T = 0$ months)	0.2	0.2	0.3	0.4	0.6
GBL (history $T = 12$ months)	0.4	0.5	1.0	1.8	4.7
GBL (history $T = 60$ months)	0.4	0.5	1.0	2.2	5.0
GBL (history $T = 120$ months)	0.4	0.5	1.0	2.0	5.0

Table 10. Average Time (in Seconds) to Simulate $m = 20{,}000$ Gibbs Samples for the Histories Used to Generate Table 4 ($K = 15$ Views)

	Number of assets (N)				
Portfolio	15	20	50	100	250
GBL (history $T = 0$ months)	3.0	3.2	4.7	8.9	27.0
GBL (history $T = 12$ months)	3.5	3.7	5.9	10.1	31.6
GBL (history $T = 60$ months)	5.6	6.0	8.6	13.7	42.3
GBL (history $T = 120$ months)	6.0	6.9	9.8	19.2	54.1

Table 11. Average Time (in Seconds) to Simulate $m = 20{,}000$ Gibbs Samples for the Histories Used to Generate Table 5 ($K = N$ Views)

	Number of assets (N)				
Portfolio	15	20	50	100	250
GBL (history $T = 0$ months)	3.0	3.7	10.8	38.1	152.9
GBL (history $T = 12$ months)	3.5	4.2	12.4	40.5	161.3
GBL (history $T = 60$ months)	5.6	6.8	15.5	46.0	182.2
GBL (history $T = 120$ months)	6.0	7.6	16.9	58.1	196.6

continue to simulate views. A plot of the asset prices and the market portfolio is shown in Figure 14.

We adopt the following data-generating model for the views. There is one independent view per asset ($K = 10$ in total), the link matrix P is the $K \times K$ identity matrix, $\Omega = 0.5^2 \times I$, and $\Theta^V = 0.2^2 \times I$. We compare BL and GBL for a range of view biases $\delta^V = \delta \times \mathbf{1}$ ($\mathbf{1}$ is a column vector of 1's of length K) with $\delta = 0$, ± 0.2, ± 0.5. Views are simulated using the history of returns from the French data set and this model of views.

Table 12. DGM: $b^\mu = 0.15$, $\delta^V = 0.3$, $N = 100$, and a Multivariate t-Distribution for Returns

	Degrees of freedom			
Portfolio ($N = 100$ assets)	5	8	10	15
GBL+	3.40 (±0.51)	3.78 (±0.22)	3.59 (±0.18)	3.63 (±0.20)
BL	0.92 (±0.12)	1.23 (±0.14)	1.53 (±0.09)	1.22 (±0.09)
GBL (history $T = 0$ months)	1.48 (±0.11)	1.60 (±0.09)	1.84 (±0.11)	1.71 (±0.11)
GBL (history $T = 12$ months)	1.79 (±0.19)	2.91 (±0.17)	3.01 (±0.16)	2.96 (±0.16)
GBL (history $T = 60$ months)	2.15 (±0.17)	3.46 (±0.23)	3.55 (±0.17)	3.42 (±0.20)
GBL (history $T = 120$ months)	2.11 (±0.17)	3.43 (±0.21)	3.42 (±0.19)	3.21 (±0.18)
MP	0.29 (±0.07)	0.60 (±0.08)	0.59 (±0.07)	0.65 (±0.07)

Notes. Returns and views are multivariate t-Distributions in the DGM whereas BL and GBL erroneously assume normality.

Table 13. DGM: $b^\mu = 0.15$, $\delta^V = 0.3$, and $\beta_{DGM} = 0.35^2 \times I$

	Inflation factor δ				
Portfolio	–0.6	–0.3	0	0.3	0.6
GBL+	4.05 (±0.22)	4.05 (±0.22)	4.05 (±0.22)	4.05 (±0.22)	4.05 (±0.22)
BL	1.64 (±0.10)	1.64 (±0.10)	1.64 (±0.10)	1.64 (±0.10)	1.64 (±0.10)
GBL ($T = 0$)	1.86 (±0.10)	1.91 (±0.11)	1.92 (±0.11)	1.92 (±0.11)	1.91 (±0.11)
GBL ($T = 12$)	3.05 (±0.16)	3.11 (±0.17)	3.09 (±0.17)	3.02 (±0.16)	2.93 (±0.16)
GBL ($T = 60$)	3.70 (±0.20)	3.65 (±0.20)	3.55 (±0.19)	3.43 (±0.19)	3.31 (±0.18)
GBL ($T = 120$)	3.84 (±0.22)	3.82 (±0.22)	3.73 (±0.22)	3.61 (±0.21)	3.48 (±0.20)
MP	0.63 (±0.07)	0.63 (±0.07)	0.63 (±0.07)	0.63 (±0.07)	0.63 (±0.07)

Notes. GBL inflates β_{GBL} by a factor of $1 + \delta$, where $\delta = 0, \pm 0.3, \pm 0.6$. BL and MP construct portfolios as in Table 8, which are independent of β. GBL+ trades according to the correctly specified DGM and corresponds to $\delta = 0$.

Table 14. DGM: $b^\mu = 0.15$, $\delta^V = 0.3$, and $N = 100$

	Proportion of variance δ allocated to β_{GBL}		
Portfolio ($N = 100$ assets)	0.75	0.5	0.25
GBL+	4.05 (±0.22)	4.05 (±0.22)	4.05 (±0.22)
BL	1.64 (±0.10)	1.64 (±0.10)	1.64 (±0.10)
GBL (history $T = 0$ months)	1.78 (±0.11)	1.74 (±0.10)	1.70 (±0.10)
GBL (history $T = 12$ months)	3.15 (±0.17)	3.16 (±0.17)	3.18 (±0.17)
GBL (history $T = 60$ months)	3.83 (±0.22)	3.83 (±0.22)	3.84 (±0.22)
GBL (history $T = 120$ months)	3.81 (±0.21)	3.83 (±0.22)	3.84 (±0.22)
MP	0.63 (±0.07)	0.63 (±0.07)	0.63 (±0.07)

Notes. The variance of the equilibrium bias, β_{GBL}, assumed in GBL is the proportion δ of total variance of returns, $\Sigma + \beta$, in the data-generating model.

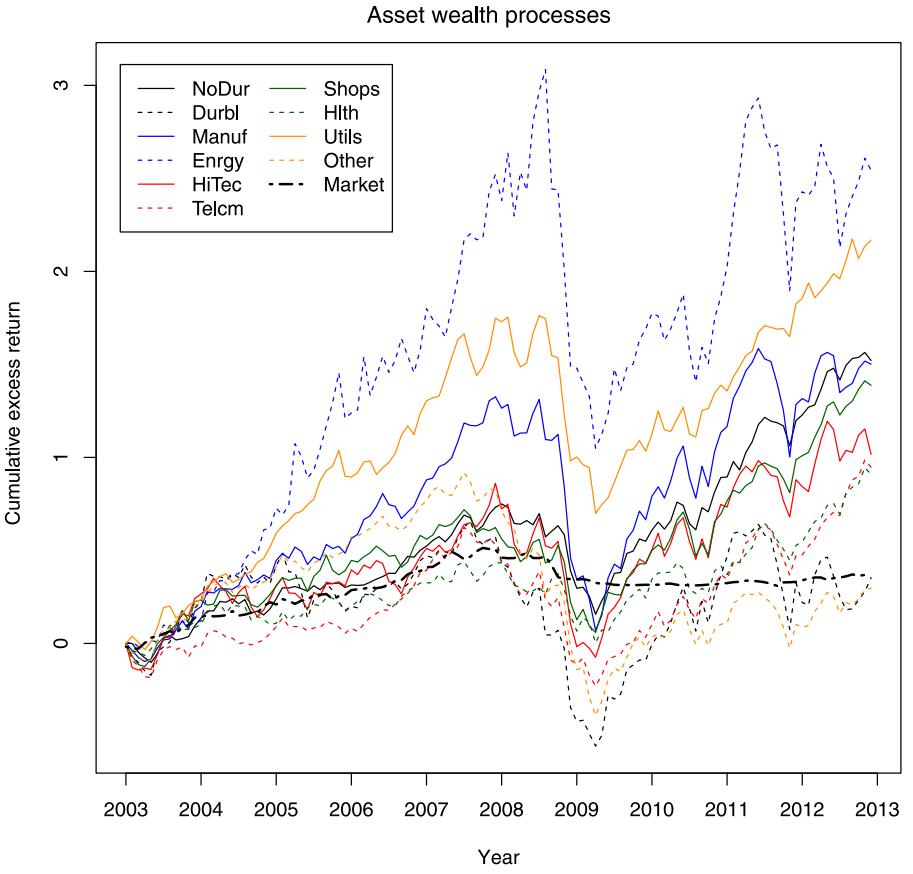
GBL adopts prior distributions on the return and view biases according to the model described in Section 3 with the following parameter values:

- prior on return bias μ: $\delta_0^\mu = \underline{0}$, $\Theta_0^\mu = 0.1^2 \times I$;
- prior on the view bias (δ^V, Θ^V):

$\eta_0^V = \underline{0}$, $\kappa_0^V = 12$, $\nu_0^V = 22$ $(= K + \kappa_0^V)$, and $\Lambda_0^V = 0.5^2 \times (\nu_0^V - K - 1) \times I$.

6.4.1. Experiment. At the start of each month between January 2003 and December 2012, α is calibrated using the preceding 120 months of data and the CAPM. We calibrate Σ and β by constructing the sample covariance matrix using the same data and attributing 1/5 of this to β and the rest to Σ. The remaining parameters in GBL are estimated using Gibbs sampling with the most recent T months of returns and views (we try $T = 0, 30, 60, 90$, and 120 months). GBL portfolios associated with each of these length T histories together with the classical Black–Litterman model are applied to the next month, and the returns are recorded. Sharpe ratios for all portfolios are computed using its realized monthly

Figure 14. Value of the Market Portfolio and Risky Assets in Excess of the Risk-Free Asset



investment returns and recorded in Table 15. This experiment was repeated for views generated with biases $\delta^V = \delta \times I$ with $\delta = 0, \pm 0.2, \pm 0.5$.

As an aside, both the classical and generalized Black–Litterman models are single-period models, and we are comparing the profit and loss generated by resolving each model at the start of each period and implementing the recommended holdings. These holdings differ from those generated by multiperiod generalizations of the Black–Litterman model, which include an additional hedging demand to account for the future evolution of the posterior and the associated resolution of uncertainty (and are also significantly more challenging to compute). We refer to Lim and Wimonkittiwat (2016) for work in this direction.

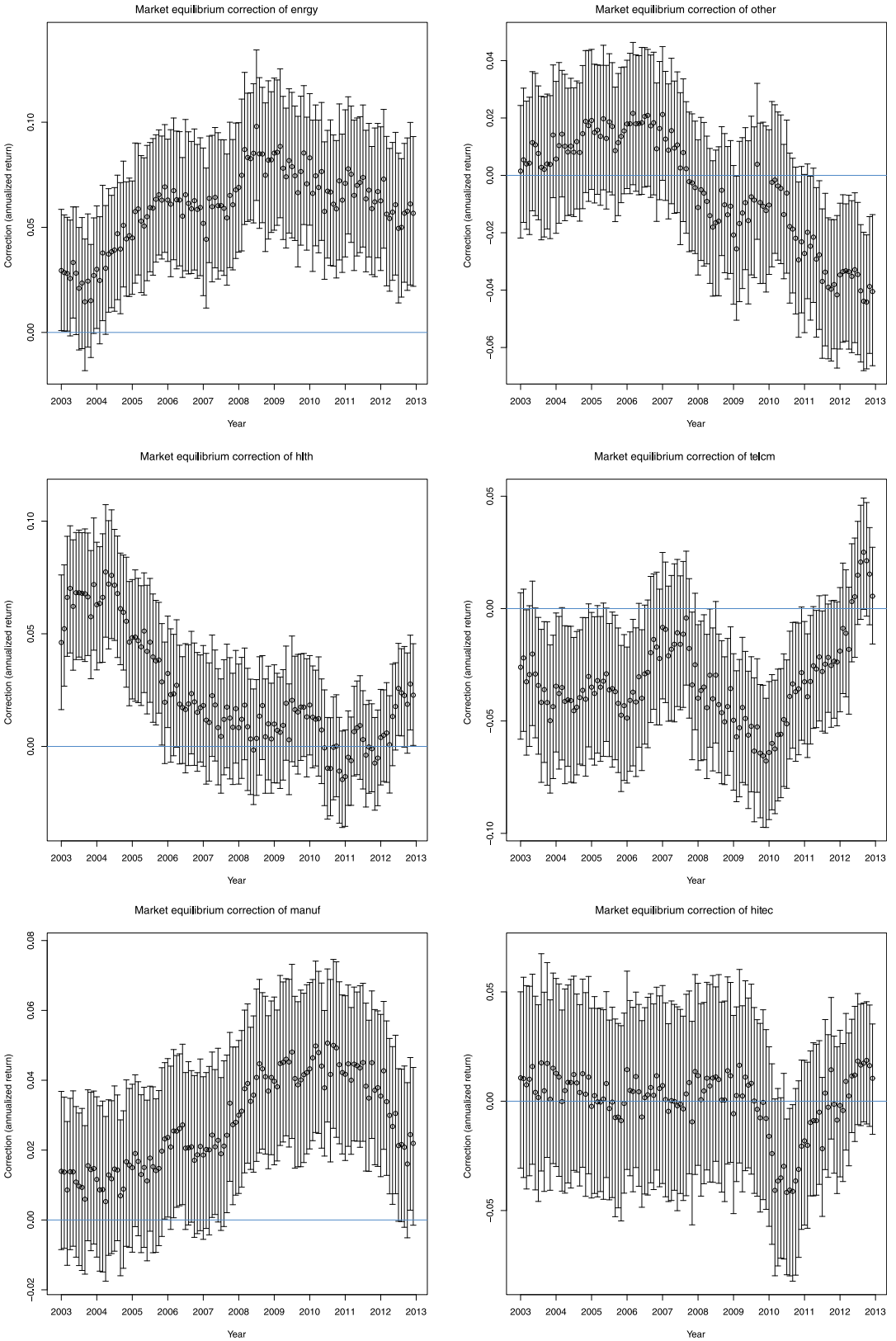
6.4.2. Discussion. The beginnings of the financial crisis can be seen in mid-2007 in both the plot of asset prices in Figure 14 and the changes in equilibrium bias estimates in Figure 15. (August 2007 coincided with the so-called quant meltdown (Khandani and Lo 2011), and the financial crisis occurred in 2008–2009.)

Table 15. Sharpe Ratios for the BL and GBL Portfolios with Histories of Lengths $T = 12, 30, 60, 90$, and 120 Months and Views Generated with Biases $\delta^V = 0, \pm 0.2, \pm 0.5$

	View bias δ^V				
Portfolio	−0.5	−0.2	0	0.2	0.5
BL	0.51 (±0.05)	0.58 (±0.06)	0.62 (±0.07)	0.65 (±0.08)	0.66 (±0.08)
GBL ($T = 12$)	0.62 (±0.08)	0.62 (±0.07)	0.64 (±0.07)	0.69 (±0.08)	0.79 (±0.09)
GBL ($T = 30$)	0.66 (±0.10)	0.62 (±0.07)	0.63 (±0.07)	0.66 (±0.08)	0.75 (±0.11)
GBL ($T = 60$)	0.66 (±0.10)	0.62 (±0.07)	0.63 (±0.07)	0.65 (±0.08)	0.72 (±0.10)
GBL ($T = 120$)	0.64 (±0.09)	0.61 (±0.07)	0.62 (±0.07)	0.64 (±0.08)	0.68 (±0.09)
Market portfolio	0.217	0.217	0.217	0.217	0.217

Notes. BL is sensitive to view bias in the DGM. In the case of GBL, the sensitivity is less and performance improves, relative to BL, as T increases.

Figure 15. (Color online) Equilibrium Bias Correction for a Selection of Assets



Notes. "enrgy," "hlth," "telcm," "manuf," "hitec" are industry portfolios for the energy, health, telecom, manufacturing, and high tech industries, respectively. 'other' is the industry portfolio for businesses that do not fall into other categories, such as mining, construction, building management, transportation, hotels, entertainment, and so forth. The error bars refer to the interquartile range of the posterior distribution for the equilibrium bias.

Figure 16. Value of Portfolios Net the Value of Investing in the Risk-Free Asset When View Bias $\delta^V = -0.5$ with (Left) $T = 12$ Months and (Right) $T = 60$ Months

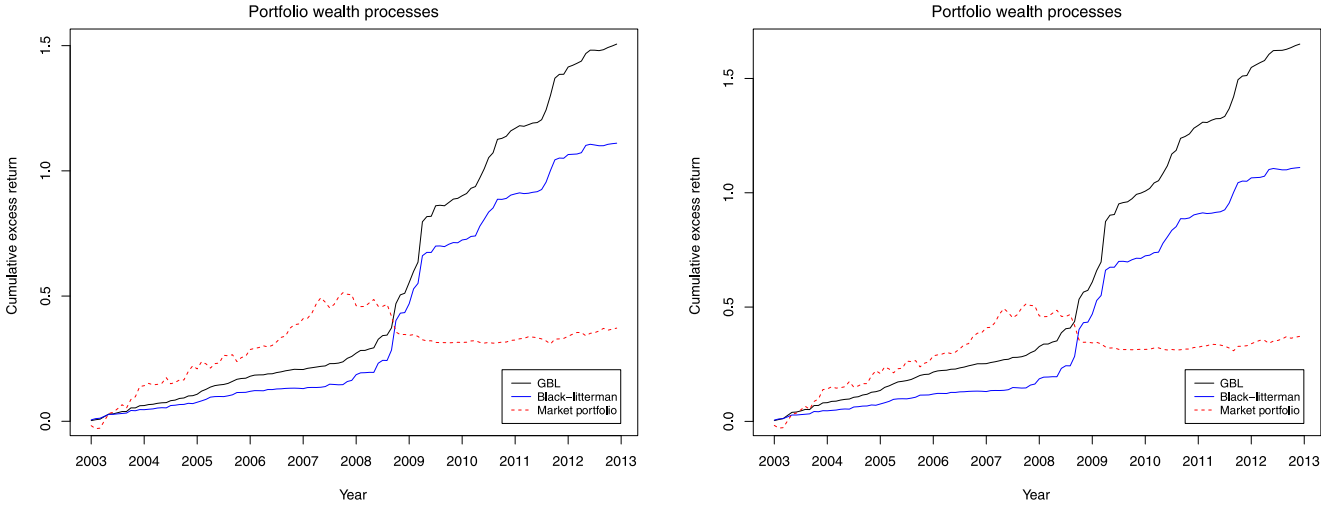


Table 16. Average l^1 Norm of Portfolios with $\gamma = 6.67$

	View bias δ^V				
Portfolio	−0.5	−0.2	0	0.2	0.5
BL	4.3	3.9	3.9	4.2	5.0
GBL ($T = 12$)	2.3	2.6	2.7	2.7	2.4
GBL ($T = 30$)	2.6	3.0	3.1	3.1	2.6
GBL ($T = 60$)	2.8	3.3	3.4	3.4	2.9
GBL ($T = 90$)	3.0	3.4	3.5	3.6	3.1
GBL ($T = 120$)	3.1	3.5	3.6	3.6	3.2

We see from Table 15 that the performance of classical Black–Litterman model is sensitive to the view bias, improving when δ^V increases and diminishing when it becomes more negative. Although a similar pattern is observed in GBL, the drop-off in performance is much less and it improves as T increases. This is consistent with the observations from our simulations.

Plots of the *excess* wealth of BL, GBL, and MP investors over risk-free investment are shown in Figure 16 when the view bias is $\delta^V = -0.5$. (These plots do not show the total wealth of each investor, but the cumulative profit he or she would earn over investing in the risk-free asset every period.) The GBL investor in the plot on the left uses $T = 12$ months of data to calibrate the bias parameters, whereas the one on the right uses $T = 60$ months. The performance of the GBL relative to the BL and MP is consistent with the Sharpe ratios reported in Table 15.

Finally, Table 16 shows the average l^1-norms of BL and GBL portfolios when the risk aversion parameter $\gamma = 6.67$. All GBL portfolios are shrunk closer to zero compared with the BL portfolio, whereas shrinkage for GBL portfolios is greater when the training set horizon is smaller.

7. Conclusion

In this paper we extend the Black–Litterman model to take into account errors in the market equilibrium estimate and the possibility for expert view bias. Gibbs sampling is used to generate samples from the posterior and the resulting estimate the optimal portfolio is shown to be consistent and asymptotically normal, converging to the optimal at rate $1/\sqrt{m}$ (where m is the number of Gibbs samples).

Experiments indicate that generalized Black–Litterman performs well in comparison with its classical counterpart under a variety of conditions and is able to detect monthly discrepancies between the CAPM estimate and the actual returns for historical data, learn the behavior of biased experts, and balance the portfolio accordingly. More generally, this paper shows how the views of multiple experts can be modeled as a Bayesian graphical model and estimated using historical data, which may be of interest in applications that involve the aggregation of expert opinions for the purpose of decision making.

Endnote

[1] See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_10_ind_port.html.

References

Beach SL, Orlov AG (2007) An application of the Black–Litterman model with EGARCH-M-derived views for international portfolio management. *Financial Markets Portfolio Management* 21(2): 147–166.

Bertsimas D, Gupta V, Paschalidis IC (2012) Inverse optimization: A new perspective on the Black–Litterman model. *Oper. Res.* 60(6):1389–1403.

Black F, Litterman R (1991) Asset allocation: Combining investor views with market equilibrium. *J. Fixed Income* 1(2):7–18.

Black F, Litterman R (1992) Global portfolio optimization. *Financial Analysts J.* 48(5):28–43.

DeMiguel V, Garlappi L, Nogales FJ, Uppal R (2009) A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Sci.* 55(5):798–812.

Fama E, French K (1993) Common risk factors in the returns on stocks and bonds. *J. Financial Econom.* 33(1):3–56.

Fama E, French K (1996) Multifactor explanations of asset pricing anomalies. *J. Finance* 51(1):55–84.

Fama E, French K (2004) The capital asset pricing model: Theory and evidence. *J. Econom. Perspect.* 18(3):25–46.

Gelman A, Carlin JB, Stern HS, Rubin DB (2004) *Bayesian Data Analysis* (Chapman & Hall/CRC, Boca Raton, FL).

Gilks WR, Richardson S, Spiegelhalter DJ (1996) Introducing Markov chain Monte Carlo. Gilks WR, Richardson S, Spiegelhalter DJ, eds. *Markov Chain Monte Carlo in Practice* (Chapman and Hall/CRC, Boca Raton, FL), 1–19.

He G (2002) The intuition behind Black–Litterman model portfolios. Preprint, submitted October 28, http://dx.doi.org/10.2139/ssrn.334304.

Idzorek TM (2002) A step-by-step guide to the Black–Litterman model. Satchell S, ed. *Forecasting Expected Returns in the Financial Markets* (Academic Press, London), 17–37.

Khandani AE, Lo AW (2011) What happened to the quants in August 2007? Evidence from factors and transactions data. *J. Financial Markets* 14(1):1–46.

Krishnan H, Mains N (2005) The two-factor Black-Litterman model. *Risk Magazine* 18(7):69–73.

Lim AEB, Wimonkittiwat P (2016) Dynamic portfolio choice under a hidden Markov model. Working paper, National University of Singapore, Singapore.

Meucci A (2006) Beyond Black–Litterman: Views on non-normal markets. *Risk Magazine* 19(2):87–92.

Meyn SP, Tweedie R (1993) *Markov Chains and Stochastic Stability* (Springer-Verlag, Berlin).

Opdyke JD (2007) Comparing Sharpe ratios: So where are the p-values? *J. Asset Management* 8(5):308–336.

Qian E, Gorman S (2001) Conditional distribution in portfolio theory. *Financial Analysts J.* 57(2):44–51.

Roberts GO, Rosenthal JS (2004) General state space Markov chains and MCMC algorithms. *Probab. Surveys* 1:20–71.

Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. *J. Risk* 2(3):21–42.

Satchell S, Scowcroft A (2000) A demystification of the Black–Litterman model: Managing quantitative and traditional portfolio construction. *J. Asset Management* 1(2):138–150.

Shapiro A, Dentcheva D, Ruszczynski A (2014) *Lectures on Stochastic Programming: Modeling and Theory*, 2nd ed., MOS-SIAM Series on Optimization (SIAM, Philadelphia).

Theil H (1971) *Principles of Econometrics* (Wiley, New York).

Walters J (2011) The Black–Litterman model in detail. Working paper, Boston University, Boston.

Walters J (2013) The factor tau in the Black–Litterman model. Working paper, Boston University, Boston.

Shea D. Chen is a special assistant at Eyon Holding Group in Taipei, Taiwan. He received his PhD in industrial engineering and operations research from the University of California, Berkeley.

Andrew E. B. Lim is a professor in the Department of Analytics and Operations and the Department of Finance, and a member of the Institute for Operations Research and Analytics, at the National University of Singapore. He is a past recipient of a National Science Foundation CAREER Award. His research interests are in the areas of stochastic control and optimization, decision making under uncertainty, robust optimization, and financial engineering.