Inverse Optimization: A New Perspective on the Black-Litterman Model

Dimitris Bertsimas
Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139,
dbertsim@mit.edu

Vishal Gupta
Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139,
vgupta1@mit.edu

Ioannis Ch. Paschalidis
Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts 02215,
yannisp@bu.edu

The Black-Litterman (BL) model is a widely used asset allocation model in the financial industry. In this paper, we provide a new perspective. The key insight is to replace the statistical framework in the original approach with ideas from inverse optimization. This insight allows us to significantly expand the scope and applicability of the BL model. We provide a richer formulation that, unlike the original model, is flexible enough to incorporate investor information on volatility and market dynamics. Equally importantly, our approach allows us to move beyond the traditional mean-variance paradigm of the original model and construct "BL"-type estimators for more general notions of risk such as coherent risk measures. Computationally, we introduce and study two new "BL"-type estimators and their corresponding portfolios: a mean variance inverse optimization (MV-IO) portfolio and a robust mean variance inverse optimization (RMV-IO) portfolio. These two approaches are motivated by ideas from arbitrage pricing theory and volatility uncertainty. Using numerical simulation and historical backtesting, we show that both methods often demonstrate a better risk-reward trade-off than their BL counterparts and are more robust to incorrect investor views.

1. Introduction

The Black-Litterman (BL) model is a widely used asset allocation model in the financial industry. Introduced in Black and Litterman (1992), the model uses an equilibrium analysis to estimate the returns of uncertain investments and employs a Bayesian methodology to "blend" these equilibrium estimates with an investor's private information, or *views*, about the investments. Computational experience has shown that the portfolios constructed by this method are more stable and better diversified than those constructed from the conventional mean-variance approach. Consequently, the model has found much favor with practitioners. The U.S. investment bank Goldman Sachs regularly publishes recommendations for investor allocations based on the BL model and has issued reports describing the firm's experience using the model (Bevan and Winkelmann 1998). A host of other firms (Zephyr Analytics, BlackRock, Neuberger Berman, etc.) also use the BL model at the core of many of their investment analytics.

The model, however, does have its shortcomings. First, it is somewhat limited in the way it allows investors to specify private information. Namely, it only allows investors to specify views on asset returns, but not on their volatility

or market dynamics. Secondly, and more restrictively, the model is predicated upon the mean-variance approach to portfolio allocation. A host of theoretical and empirical work suggests variance may not be a suitable proxy for risk. As a consequence, other risk measures such as value at risk (VaR) or conditional value at risk (CVaR) have been explored. (See, for example, Artzner et al. 1999, Bertsimas et al. 2004, Grootveld and Hallerbach 1999, Harlow 1991, Jorion 1997, Rockafellar and Uryasev 2002.)

Subsequent research has tried to address these shortcomings. We mention only a few examples and refer the reader to Walters (2010) and the references therein for a more complete survey. Giacometti et al. (2007), Martellini and Ziemann (2007), and Meucci (2005) extend beyond the mean-variance paradigm, using "fat-tailed" distributions to model asset returns, coupled with VaR or CVaR, and views on the tail behavior of returns. By contrast, Pástor (2000) and Pástor and Stambaugh (2000) build upon the Bayesian interpretation of the original model, using a general pricing model as their "prior." Finally, Meucci (2008) and Meucci et al. (2011) formulate a further generalization within the statistical framework, modeling "risk factors" (in lieu of asset returns) by a general distribution and solving

a minimum entropy optimization to update this distribution to one that incorporates a very general set of views. In most cases, this optimization is solved numerically, and in the case when the distribution is constrained to belong to a parametric family, the optimization is usually nonconvex.

In this paper, we provide a new *optimization-driven* perspective on the BL model that avoids many of the statistical assumptions of previous approaches. The key insight is to characterize the BL estimator as the solution to a particular convex optimization problem. More specifically, we formulate a problem in *inverse optimization*—a setup where one is given an optimal solution to an optimization problem and seeks to characterize the cost function and/or other problem data. Initial research on inverse optimization focused on specific combinatorial optimization problems (see Heuberger 2004 for a survey). Later, Ahuja and Orlin (2001) provided a unified approach for linear programming problems using duality. Iyengar and Kang (2005) extended these ideas to conic optimization with an application to managing portfolio rebalancing costs.

Previous research, starting with Black and Litterman (1992), has drawn some connections between the BL model and what many authors call "reverse optimization." The key idea, formalized in He and Litterman (1999), is to solve the first-order optimality condition of a Lagrangean relaxation of the Markowitz problem to derive the equilibrium returns. It may not be clear how to generalize this approach to other constrained asset allocation problems. Often, as in Herold (2005), which considers a budget constraint and Da Silva et al. (2009), which considers information-ratio portfolio optimization, the authors are forced to rely on ad hoc arguments. Furthermore, "reverse optimization" only provides the equilibrium estimates. Most authors still use a statistical approach to blend in the views.

To the best of our knowledge, our work is the first to offer a rigorous inverse optimization interpretation of the BL model. This has several advantages. First, it allows us to completely characterize the set of "input data" in the Markowitz problem in equilibrium and to incorporate more general views, e.g., on volatility. Second, it enables us to move beyond mean-variance and adopt more general risk metrics by leveraging robust optimization ideas. Further, our approach does not rely on specific distributional assumptions on market returns; we construct the equilibrium estimates and blend them with views using optimization techniques. As such, our approach unifies several threads in the literature (Herold 2005, Krishnan and Mains 2005, Martellini and Ziemann 2007). It can also be potentially combined with the ideas in Meucci (2008) and Meucci et al. (2011) of modelling risk factors instead of asset returns.

We summarize the major contributions of our paper as follows:

1. By linking the BL model with inverse optimization, we provide a novel formulation that constructs equilibrium-based estimators of the mean returns *and* the covariance matrix.

2. Our framework permits greater freedom in the type of private information and investor views expressible in the model. Specifically, our formulation allows investors to incorporate views on volatility and market dynamics. We leverage this freedom to create a new asset allocation procedure we call the "mean-variance inverse optimization" (MV-IO) approach, which uses BL-type estimators in a setting motivated by the arbitrage pricing theory.

3. By drawing on ideas from robust optimization, we generalize the traditional Markowitz portfolio problem to one that can encompass risk metrics such as VaR, CVaR, and generic coherent risk measures.

4. We then use inverse optimization to construct BL-type estimators in this new model. We illustrate the power of this method with what we call the robust mean variance inverse optimization (RMV-IO) approach, which accommodates volatility uncertainty.

5. We provide computational evidence to show that MV-IO and RMV-IO portfolios often provide better risk-reward profiles and are more robust to incorrect or extreme views than their BL counterparts.

The remainder of the paper is organized as follows: Section 2 briefly reviews the BL model. Section 3 derives the BL model from principles in inverse optimization and shows how it can be extended to include volatility estimation and views on volatility. Section 4 introduces a more general portfolio allocation problem and constructs a BL-type estimator in this setting. Section 5 provides numerical results illustrating the strengths and drawbacks of these new estimators compared to the original estimator. Finally, §6 concludes and discusses some avenues for future research.

Throughout this paper we will use bold-faced capital letters ($\boldsymbol{\Sigma}, \mathbf{P}, \ldots$) to indicate matrices, bold-faced lowercase letters to indicate vectors ($\mathbf{x}, \mathbf{r}, \ldots$) and ordinary letters ($L, t, \ldots$) to indicate scalars. The vector $\mathbf{e}$ refers to the vector of all ones, $\mathbf{0}$ is the vector of all zeroes, and $\mathbf{I}$ is an identity matrix. Finally, for a matrix $\mathbf{A}$, we write $\mathbf{A} \succeq \mathbf{0}$ to indicate that $\mathbf{A}$ is positive semidefinite and $\mathbf{A} \succ \mathbf{0}$ to indicate that $\mathbf{A}$ is positive definite.

2. The Black-Litterman Model

This section reviews the original BL model to keep the paper self-contained. For a complete treatment, see Black and Litterman (1992) or He and Litterman (1999).

Consider a market with n risky assets and one riskless asset where investors seek to maximize their expected return subject to a threshold level of risk. We initially define risk to be measured by the standard deviation of returns. In §4, we will generalize the notion of risk considerably. In other words, investors solve the Markowitz portfolio allocation problem

(Markowitz) $\max_{\mathbf{x}} \{ \boldsymbol{\mu}'\mathbf{x} + (1 - \mathbf{e}'\mathbf{x})r_f : \sqrt{\mathbf{x}'\boldsymbol{\Sigma}\mathbf{x}} \leqslant L \}$,   (1)

where $\mathbf{r} \in \mathbb{R}^n$ is the random vector of the risky asset returns, $\boldsymbol{\mu} \equiv \mathbb{E}[\mathbf{r}]$ is the vector of mean asset returns, $r_f \in \mathbb{R}_+$ is

the return on the riskless asset, $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ is the covariance matrix of asset returns, $\mathbf{x} \in \mathbb{R}^n$ is the fraction of wealth invested in each risky asset, and L is an investor-specific threshold level of risk. We assume that an investor can borrow or invest at the riskfree rate.

The primary difficulty in solving (1) is finding a stable estimation procedure for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The BL model suggests such a procedure. At a high level, the procedure blends two sources of information:

• *Market Equilibrium*: If all investors solve (1) for some investor-specific L, then there exists a value $\delta \in \mathbb{R}_+$ such that the expected returns $\boldsymbol{\mu}$ satisfy

$$\boldsymbol{\mu} = r_f \mathbf{e} + 2\delta \boldsymbol{\Sigma} \mathbf{x}^{\mathrm{mkt}}. \tag{2}$$

Here, $\mathbf{x}^{\mathrm{mkt}}$ is the percentage market capitalization of each asset. By multiplying (2) by $\mathbf{x}^{\mathrm{mkt}}$, observe that $\delta = (\boldsymbol{\mu} - r_f \mathbf{e})' \mathbf{x}^{\mathrm{mkt}} / (2\mathbf{x}^{\mathrm{mkt}'} \boldsymbol{\Sigma} \mathbf{x}^{\mathrm{mkt}})$. In this sense, δ measures the risk-reward trade-off for the market portfolio. Finally, there is some aggregate, but unobservable, value of L such that $\mathbf{x}^{\mathrm{mkt}}$ is a solution to (1) and

$$\delta = \sqrt{(\boldsymbol{\mu} - r_f \mathbf{e})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - r_f \mathbf{e})} \Big/ (2L).$$

These conditions collectively form the capital asset pricing model equilibrium (CAPM). (See Sharpe 1964 for classical proofs of these results. Alternatively, we will derive these equations as a special case of Theorem 2 in §4.)

• *Private Information:* Any particular investor may have private information, or views, that are not shared by other investors. The model expresses these views through several different portfolios $\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^m$ for which the investor has some estimate of the corresponding mean returns $q_1, q_2, \ldots, q_m$. More compactly,

$$\mathbf{P}\boldsymbol{\mu} = \mathbf{q}, \tag{3}$$

where $\mathbf{P}' = [\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^m]$ and $\mathbf{q}' = (q_1, q_2, \ldots, q_m)$. Assuming that both of these sources of information are at least approximately true, the BL procedure posits the following statistical model:

$$\hat{\mathbf{r}} \equiv r_f \mathbf{e} + 2\hat{\delta}\hat{\boldsymbol{\Sigma}} \mathbf{x}^{\mathrm{mkt}}. \tag{4}$$

Here $\boldsymbol{\mu} = \hat{\mathbf{r}} + \boldsymbol{\epsilon}^r$, $\mathbf{P}\hat{\mathbf{r}} = \mathbf{q} + \boldsymbol{\epsilon}^v$, $\hat{\delta}$ and $\hat{\boldsymbol{\Sigma}}$ are estimates of δ and $\boldsymbol{\Sigma}$, respectively, and $(\boldsymbol{\epsilon}^r, \boldsymbol{\epsilon}^v)$ is a random noise vector. Although $\hat{\boldsymbol{\Sigma}}$ may be estimated from historical data, δ is, itself, not directly observable. Even though there are various proposals in the literature, to the best of our knowledge there is no consensus on how to fit $\hat{\delta}$. Observe that the choice of $\hat{\delta}$ only affects the estimate of $\hat{\mathbf{r}}$ and the resulting portfolio by a constant of proportionality.

By assuming that $(\boldsymbol{\epsilon}^r, \boldsymbol{\epsilon}^v)$ is distributed as a multivariate normal distribution, i.e., $(\boldsymbol{\epsilon}^r, \boldsymbol{\epsilon}^v) \sim N(\mathbf{0}, \boldsymbol{\Omega})$ for some covariance matrix $\boldsymbol{\Omega}$, and that several other key quantities are Gaussian and mutually independent, Black and

Litterman (1992) estimates $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ by the Bayesian update[1] $(\boldsymbol{\mu}^{\mathrm{BL}}, \boldsymbol{\Sigma}^{\mathrm{BL}})$ where

$$\boldsymbol{\mu}^{\mathrm{BL}} = \boldsymbol{\Sigma}_2^{-1} \begin{pmatrix} \mathbf{I} \\ \mathbf{P} \end{pmatrix}' \boldsymbol{\Omega}^{-1} \begin{pmatrix} \hat{\mathbf{r}} \\ \mathbf{q} \end{pmatrix}, \tag{5}$$

$$\boldsymbol{\Sigma}_2 = \left[\begin{pmatrix} \mathbf{I} \\ \mathbf{P} \end{pmatrix}' \boldsymbol{\Omega}^{-1} \begin{pmatrix} \mathbf{I} \\ \mathbf{P} \end{pmatrix} \right], \quad \boldsymbol{\Sigma}^{\mathrm{BL}} = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2. \tag{6}$$

The quantities $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}_1$ are unobservable and require exogenous specification. A common approximation used in He and Litterman (1999) is

$$\boldsymbol{\Sigma}_1 = \hat{\boldsymbol{\Sigma}}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \tau_0 \hat{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(\tau_1, \ldots, \tau_m) \end{pmatrix}, \tag{7}$$

where $\tau_i \in \mathbb{R}_+$, $i = 0, \ldots, m$ represent the investor's relative confidence in each view.

As pointed out in He and Litterman (1999), the original paper Black and Litterman (1992) does not mention Equation (6) for $\boldsymbol{\Sigma}^{\mathrm{BL}}$. Indeed, many references refer to Equation (5) alone as the BL model. However, Equation (6) and the subsequent approximations are critically needed to solve the allocation Problem (1). As we will see, our approach avoids such approximations by directly estimating the covariance matrix $\boldsymbol{\Sigma}$.

Despite the number of statistical assumptions and exogenously specified parameters, this procedure yields stable estimates with good practical performance. We illustrate this procedure with an example that we will continue to develop in the sequel.

2.1. An Example: Applying the BL Procedure

In the U.S. equity market, every stock is given a Global Industrial Classification Standard (GICS) sector classification describing the industry to which the underlying company belongs. The list of 10 possible classifications can be found in Table 1 with their average historic returns over the previous 24 months, denoted $\boldsymbol{\mu}^{\mathrm{hist}}$, and market share, denoted $\mathbf{x}^{\mathrm{mkt}}$, as of May 1, 1998.[2] We also present the historical covariance between the sectors in Table EC.1 of the electronic companion which is available as part of the online version at http://dx.doi.org/10.1287/opre.1120.1115. We will use the BL procedure to construct a portfolio of these sectors.

First, consider using the given historical return and covariance information directly in the Markowitz formulation (1). The resulting portfolio, denoted $\mathbf{x}^{\mathrm{hist}}$, is displayed in Table 1. (All portfolios have been scaled to sum to 100 for ease of comparison.) It has large long positions in some assets, large short positions in others, and almost no investment in others—an allocation that does not match our intuition of a "well-diversified" portfolio. This phenomenon is common and highlights one of the difficulties in using the Markowitz formulation (1) in practice.

Table 1.   Differing estimates for mean returns and their corresponding optimal portfolios.

	$\boldsymbol{\mu}^{\text{hist}}$	$\mathbf{x}^{\text{mkt}}$	$\mathbf{x}^{\text{hist}}$	$\hat{\mathbf{r}}$	$\boldsymbol{\mu}^{\text{BL}}$	$\mathbf{x}^{\text{BL}}$	$\boldsymbol{\mu}^{1}$	$\mathbf{x}^{1}$	$\boldsymbol{\mu}^{\text{MV}}$	$\mathbf{x}^{\text{MV}}$	$\boldsymbol{\mu}^{2}$	$\boldsymbol{\mu}^{\text{Rob}}$	$\mathbf{x}^{\text{Rob}}$
Energy	23.47	7.15	120.56	2.25	2.20	5.03	2.19	5.06	2.16	6.03	2.23	2.08	5.57
Materials	13.21	4.54	−863.99	3.40	3.40	8.58	3.39	7.12	3.24	5.75	3.04	2.96	4.07
Industrials	22.92	10.92	251.73	4.10	4.11	10.94	4.10	10.51	4.10	10.93	4.04	4.00	11.03
Consumer discretionary	23.96	10.77	361.07	3.50	3.43	5.56	3.41	3.48	3.37	8.45	3.56	3.28	7.68
Consumer staple	26.22	13.47	−171.69	3.54	3.62	17.36	3.70	19.95	3.97	16.16	3.70	4.20	20.32
Healthcare	29.41	13.57	83.03	4.30	4.22	9.20	4.27	12.81	4.26	11.91	4.37	4.23	12.40
Financials	37.67	15.81	871.92	4.13	4.18	16.93	4.25	20.49	4.52	17.26	4.35	4.73	21.27
Information technology	30.16	15.16	−86.61	6.81	6.80	16.09	6.77	15.30	6.76	15.41	6.65	6.52	14.66
Telecommunications	23.58	5.88	−26.45	3.56	3.48	2.61	3.38	−1.90	3.09	3.47	3.25	2.67	−1.06
Utilities	13.90	2.74	−439.56	1.67	1.73	7.71	1.70	7.18	1.66	4.63	1.43	1.52	4.07

Notes. All returns are annualized (%). The columns $\boldsymbol{\mu}^{\text{hist}}$ and $\mathbf{x}^{\text{mkt}}$ represent historical data taken as of May 1, 1998. The remaining columns are different estimates of the mean return and corresponding portfolios that we develop throughout the paper.

Next, consider using (4) to estimate those mean returns $\hat{\mathbf{r}}$ that are consistent with the CAPM equilibrium, also displayed in Table 1. We have followed He and Litterman (1999) and exogenously specified $\hat{\delta} = 1.25$. This value was chosen by those authors based on their extensive experience with the BL model in the U.S. equity market. By construction the optimal portfolio corresponding to $\hat{\mathbf{r}}$ is the market portfolio $\mathbf{x}^{\text{mkt}}$. If the investor had no other views, she should invest in $\mathbf{x}^{\text{mkt}}$.

Suppose, though, that the investor has the additional view that the portfolio

$$\mathbf{p} = \big[-10\%, 0\%, 0, -20\%, 40\%, -10\%, 30\%, 0\%, \\ -40\%, 10\%\big] \quad (8)$$

should yield, on average, a return of 1%. Under the CAPM equilibrium, the expected return on this portfolio is $\mathbf{p}'\hat{\mathbf{r}} = 0.3862\%$; thus, this view represents "new" information. Using the update Equations (5), we calculate $\boldsymbol{\mu}^{\text{BL}}$ and $\boldsymbol{\Sigma}^{\text{BL}}$ and the resulting new optimal portfolio $\mathbf{x}^{\text{BL}}$. These are shown in Table 1 above and Table EC.2 in the electronic companion. We have taken $\tau_0 = 1/24$ (corresponding to using 24 historical observations) and set $\tau_1 = 0.02\tau_0$, representing a belief 50 times stronger in the portfolio view than in the equilibrium estimate. In contrast to the portfolio generated by historical returns, the Black-Litterman portfolio is considerably more diversified. This stability in the allocations is typical and one of the strengths of this procedure.

Three observations are in order:

1. The BL model distinguishes between the covariance matrix $\boldsymbol{\Sigma}$, for which it assumes good historical estimates are available, and the mean return $\boldsymbol{\mu}$, for which it assumes such estimates are not available. In light of the variety of approximations like (6) and (7), these historical estimates may significantly affect the resulting portfolio.

2. Investor views are notoriously difficult to specify. The form (3) limits the applicability of the model, prohibiting, for example, views on volatility.

3. Finally, the above derivation is firmly based upon the mean-variance approach to portfolio allocation; it essentially assumes that, in equilibrium, a typical investor solves the Markowitz problem. In reality, however, a typical investor may be subject to no-shorting, budget, or margin constraints. She may also measure risk in terms of value at risk (VaR) or conditional value at risk (CVaR). It is not immediately clear how to modify the above methodology in these cases.

3. Reinterpreting the Black-Litterman Model Through Inverse Optimization

In this section, we provide a novel derivation of Equation (5). The key insight is to use inverse optimization to characterize the BL estimate as the solution to a particular convex optimization problem, thereby eliminating the need for a statistical model. We will then show how to use this alternative formulation to address some of the above criticisms.

We begin by defining an inverse optimization problem. Consider the problem defined by

$$\min_{\mathbf{x} \in \mathscr{X}(\boldsymbol{\zeta})} f(\mathbf{x}; \boldsymbol{\zeta}), \quad (9)$$

where $f(\mathbf{x}; \boldsymbol{\zeta})$ is some objective function depending on the decision variables $\mathbf{x} \in \mathbb{R}^n$ and some data $\boldsymbol{\zeta} \in \mathbb{R}^m$. In general, the feasible region $\mathscr{X}(\boldsymbol{\zeta})$ may also depend on the data $\boldsymbol{\zeta}$. Usually, we assume that the data parameters $\boldsymbol{\zeta}$ are known, and we seek an optimal solution $\mathbf{x}^*$ to (9). We call this process the forward problem. In inverse optimization, we are given a solution $\mathbf{x}^*$ and seek a value of $\boldsymbol{\zeta}$ such that $\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathscr{X}(\boldsymbol{\zeta})} f(\mathbf{x}; \boldsymbol{\zeta})$. In the case where there may be multiple such $\boldsymbol{\zeta}$, it is desirable to characterize the full set of $\boldsymbol{\zeta}$, which will make the given solution optimal.

As an example, consider the Markowitz portfolio allocation problem (1). The data $\boldsymbol{\zeta}$ correspond to $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r_f, L)$. In the forward problem, we assume that these data are known, and we seek an optimal portfolio $\mathbf{x}^*$. In the inverse problem, we are given a candidate portfolio $\mathbf{x}^*$ and seek $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r_f, L)$, which make this portfolio optimal. We now characterize the set of such data.

PROPOSITION 1. *Consider the Markowitz problem* (1), *and assume we are given a candidate portfolio* $\mathbf{x}^*$. *In addition, suppose that it is known a priori that* $\boldsymbol{\Sigma} \succeq \mathbf{0}$ *and* $L > 0$. *Then,* $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r_f, L)$ *solve the inverse problem for* (1) *if and only if there exists a* δ *such that*:

$$\mathbf{x}^{*'}\boldsymbol{\Sigma}\mathbf{x}^* \leqslant L^2, \quad \delta(L^2 - \mathbf{x}^{*'}\boldsymbol{\Sigma}\mathbf{x}^*) = 0,$$
$$\boldsymbol{\mu} - r_f\mathbf{e} - 2\delta\boldsymbol{\Sigma}\mathbf{x}^* = \mathbf{0}, \quad \delta \geqslant 0, \ L > 0, \ \boldsymbol{\Sigma} \succeq \mathbf{0}. \tag{10}$$

PROOF. For any fixed values of the data with $\boldsymbol{\Sigma} \succeq \mathbf{0}$, $L > 0$, the forward problem is convex and satisfies a Slater condition. Thus, it is necessary and sufficient that any optimal solution $\mathbf{x}$ satisfies the Karush-Kuhn-Tucker (KKT) conditions (Bertsekas 1999). The KKT conditions are precisely the Equations (10), with $\mathbf{x}^*$ replaced by $\mathbf{x}$ and the constraints $\boldsymbol{\Sigma} \succeq \mathbf{0}$ and $L > 0$ omitted. (We have rewritten the constraint $\sqrt{\mathbf{x}'\boldsymbol{\Sigma}\mathbf{x}} \leqslant L$ as $\mathbf{x}'\boldsymbol{\Sigma}\mathbf{x} \leqslant L^2$ for convenience.)

Now consider the inverse problem. We have already established that it is both necessary and sufficient for $\mathbf{x}^*$ to satisfy the system (10) to be optimal. If we now interpret $\mathbf{x}^*$ as given and view the remaining parameters as unknown, it follows that any solution to the inverse problem must satisfy this system of equations with $\mathbf{x}^*$. □

Because $\boldsymbol{\Sigma}$ is a covariance matrix, the assumption $\boldsymbol{\Sigma} \succeq \mathbf{0}$ is natural. On the other hand, it is possible that $L = 0$, but this case is somewhat degenerate. Intuitively, it corresponds to a situation where an investor will tolerate no risk. It is possible to show that the system (10) is still necessary and sufficient in this case. We provide the details in the e-companion.

Using the CAPM and the above proposition, we know that there exists an aggregate value of L such that $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r_f, L)$ satisfies system (10) with $\mathbf{x}^*$ replaced by $\mathbf{x}^{\text{mkt}}$. Observe further that if $\boldsymbol{\Sigma}, \delta$ are replaced by the estimates $\hat{\boldsymbol{\Sigma}}, \hat{\delta}$, respectively, in (10), we retrieve the equilibrium estimates (4). More importantly, as stated, (10) treats $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ on the same footing; it gives conditions that they must both satisfy in equilibrium. We summarize these observations in the following theorem.

THEOREM 1. *Suppose all investors solve* (1) *for some investor-specific value* L. *Then there must exist values of* $\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}$ *such that*

$$\boldsymbol{\mu} - r_f\mathbf{e} - \bar{\boldsymbol{\Sigma}}\mathbf{x}^{\text{mkt}} = \mathbf{0}, \quad \bar{\boldsymbol{\Sigma}} \succeq \mathbf{0}. \tag{11}$$

PROOF. Apply our previous comments and make the change of variables $\bar{\boldsymbol{\Sigma}} = 2\delta\boldsymbol{\Sigma}$ in (10). □

System (11) is a system of linear matrix inequalities (LMIs). LMIs are tractable both theoretically and practically. Optimization problems over LMIs are often representable as semidefinite optimization problems for which commercial and open source software is available.

We now proceed to consider the role of private information. We will first consider private information in the form

of (3) by adding these constraints to the above system. The resulting system is still an LMI, but it may be infeasible. Consequently, we seek the smallest perturbation such that the system is feasible. This yields the following optimization problem.

$$\min_{\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}, t}\left\{t: \left\|\begin{pmatrix}\boldsymbol{\mu} - r_f\mathbf{e} - \bar{\boldsymbol{\Sigma}}\mathbf{x}^* \\ \mathbf{P}\boldsymbol{\mu} - \mathbf{q}\end{pmatrix}\right\| \leqslant t, \ \bar{\boldsymbol{\Sigma}} \succeq \mathbf{0}\right\}, \tag{12}$$

where $\|\cdot\|$ denotes any norm. In contrast to the BL framework, the above optimization problem simultaneously determines $\boldsymbol{\mu}$ and $\bar{\boldsymbol{\Sigma}}$, thereby eliminating the need for approximations like (7).

For many common norms, (12) can be recast as a semidefinite optimization problem. We show this reformulation for a weighted ℓ_2 norm—the case of weighted ℓ_1 and weighted ℓ_∞ norms are similar—and, most importantly, prove that this formulation generalizes the BL result.

PROPOSITION 2. *Consider the problem* (12) *under the weighted* ℓ_2 *norm* $\|\mathbf{z}\|_2^{\boldsymbol{\Omega}} = \sqrt{\mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z}}$, *where* $\boldsymbol{\Omega} \succ 0$.

(a) *Problem* (12) *can be written as the semidefinite optimization problem*

$$\min_{\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}, t, \mathbf{u}}\left\{t: \mathbf{u} = \boldsymbol{\Omega}^{-1/2}\begin{pmatrix}\boldsymbol{\mu} - r_f\mathbf{e} - \bar{\boldsymbol{\Sigma}}\mathbf{x}^* \\ \mathbf{P}\boldsymbol{\mu} - \mathbf{q}\end{pmatrix},\right.$$
$$\left.\begin{pmatrix}t & \mathbf{u}' \\ \mathbf{u} & \mathbf{I}\end{pmatrix} \succeq \mathbf{0}, \ \bar{\boldsymbol{\Sigma}} \succeq \mathbf{0}\right\}. \tag{13}$$

(b) *If we further fix* $\bar{\boldsymbol{\Sigma}} = 2\hat{\delta}\hat{\boldsymbol{\Sigma}}$, *then an optimal solution to* (12) *is given by* (5), *i.e.*,

$$\boldsymbol{\mu} = \left[\begin{pmatrix}\mathbf{I} \\ \mathbf{P}\end{pmatrix}'\boldsymbol{\Omega}^{-1}\begin{pmatrix}\mathbf{I} \\ \mathbf{P}\end{pmatrix}\right]^{-1}\begin{pmatrix}\mathbf{I} \\ \mathbf{P}\end{pmatrix}'\boldsymbol{\Omega}^{-1}\begin{pmatrix}\hat{\mathbf{r}} \\ \mathbf{q}\end{pmatrix}. \tag{14}$$

PROOF. (a) Notice that when $t \geqslant 0$, we can use Schur-Complements (Boyd and Vandenberghe 2004) to write

$$\mathbf{u}'\mathbf{u} \leqslant t^2 \quad \Leftrightarrow \quad \begin{pmatrix}t & \mathbf{u}' \\ \mathbf{u} & \mathbf{I}\end{pmatrix} \succeq \mathbf{0}.$$

The semidefinite programming formulation then follows immediately from the definition of $\mathbf{u}$.

(b) When we fix $\bar{\boldsymbol{\Sigma}}$ at the historical estimates, Problem (12) can be rewritten using Equation (4) as an unconstrained optimization problem:

$$\min_{\boldsymbol{\mu}}\left(\begin{pmatrix}\mathbf{I} \\ \mathbf{P}\end{pmatrix}\boldsymbol{\mu} - \begin{pmatrix}\hat{\mathbf{r}} \\ \mathbf{q}\end{pmatrix}\right)'\boldsymbol{\Omega}^{-1}\left(\begin{pmatrix}\mathbf{I} \\ \mathbf{P}\end{pmatrix}\boldsymbol{\mu} - \begin{pmatrix}\hat{\mathbf{r}} \\ \mathbf{q}\end{pmatrix}\right).$$

Since $\boldsymbol{\Omega} \succ 0$, we can rewrite this as

$$\min \left\|\boldsymbol{\Omega}^{-1/2}\begin{pmatrix}\mathbf{I} \\ \mathbf{P}\end{pmatrix}\boldsymbol{\mu} - \boldsymbol{\Omega}^{-1/2}\begin{pmatrix}\hat{\mathbf{r}} \\ \mathbf{q}\end{pmatrix}\right\|_2^2.$$

This last optimization is of the form $\min_{\mathbf{y}} \|\mathbf{A}\mathbf{y} - \mathbf{b}\|_2^2$. The solution is well known to be $\mathbf{y} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{b}$. Applying this formula above yields the result. □

We conclude this section with some remarks about our approach:

First, because Equations (14) and (5) are identical, a consequence of Proposition 2(b) is that the inverse optimization framework generalizes the BL model.

Second, unlike the original BL approach, it is straightforward to adapt the previous method to incorporate additional constraints on the forward problem. From a modeling point of view, we may believe a typical investor does not solve Problem (1), but rather is additionally constrained by a budget, no-shorting, or margin constraint. Extending our approach to such cases is a simple recipe: First, argue that in equilibrium $\mathbf{x}^{\mathrm{mkt}}$ is an optimal solution to the forward problem. Then, solve the corresponding inverse problem. Finally, replace (11) with the solution to the inverse problem.

For example, in the case when the investor has a budget constraint on the risky assets ($\mathbf{e}'\mathbf{x} = 1$), following the above procedure yields the optimization problem

$$\min_{\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}}, \theta, t} \left\{ t: \left\| \begin{pmatrix} \boldsymbol{\mu} - \theta\mathbf{e} - \bar{\boldsymbol{\Sigma}}\mathbf{x}^* \\ \mathbf{P}\boldsymbol{\mu} - \mathbf{q} \end{pmatrix} \right\| \leqslant t, \; \bar{\boldsymbol{\Sigma}} \succeq \mathbf{0} \right\}, \quad (15)$$

If we fix $\bar{\boldsymbol{\Sigma}}$ at its historical estimate, this approach provides an alternative derivation of the results in Herold (2005). Moreover, the results in §4 can also be seen as an application of this procedure in the case where the forward problem captures a more general notion of risk. (See Proposition 6 and Theorem 2 in §4.3.)

Finally, an important advantage of the inverse optimization perspective is the ability to incorporate a greater variety of investor views. One can introduce any linear or semidefinite constraint on the entries of $\boldsymbol{\mu}$ and $\bar{\boldsymbol{\Sigma}}$ without affecting the tractability of the (13). Practically, this makes it possible to model a rich variety of views that are unavailable in the traditional approach. For example, in liquid options markets, an investor may have information on the volatility σ of a basket of assets $\mathbf{b}$. She may then impose the constraint $\|\mathbf{b}'\bar{\boldsymbol{\Sigma}}\mathbf{b} - \sigma^2\| \leqslant \epsilon$.

As a more interesting example, suppose the investor believes that asset returns follow a factor model. Factor models are common in the financial literature and are related to the arbitrage pricing theory. In factor models, the first few eigenvalues of $\boldsymbol{\Sigma}$ dominate the rest of the spectrum and represent macroeconomic conditions that are slow-changing. (See Connor and Korajczyk 1995 for further discussion.) The other eigenvalues are assumed to be small, random noise. The investor might then estimate $\bar{\boldsymbol{\Sigma}}$ by imposing the constraints

$$\sum_{i=1}^{k} \lambda_i \geqslant \alpha \cdot \mathrm{trace}(\bar{\boldsymbol{\Sigma}}), \quad \|\bar{\boldsymbol{\Sigma}}\mathbf{v}^i - \lambda_i\mathbf{v}^i\| \leqslant \epsilon \quad i = 1, \ldots, k, \quad (16)$$

where $\alpha \in (0, 1)$, λ_i is the ith largest eigenvalue of $2\hat{\delta}\hat{\boldsymbol{\Sigma}}$, $\mathbf{v}^i$ is its corresponding eigenvector, and k is a small number like 2 or 3. The first of these constraints ensure that the macroeconomic factors constitute a large part of the spectrum and the second ensures that they are slow-changing. We will refer to Problem (13) with the additional constraints (16) as the *mean-variance inverse optimization* (MV-IO) approach.

Although the constraints (16) do not affect the computational tractability of (13), they do represent a sophisticated view on market dynamics; future returns are well approximated through a small number of macroeconomic factors. In what follows, we determine the factors numerically via the historical spectrum. One could instead combine our approach with the results from Fama and French (1993) where the authors seek to identify the factors via economic principles.

3.1. An Example: Applying Inverse Optimization to the Markowitz Framework

Using the same data and investor's view from our previous example, we illustrate the inverse optimization approach. First, suppose we solve problem (13) under the additional constraint that $\bar{\boldsymbol{\Sigma}} = 2\hat{\delta}\hat{\boldsymbol{\Sigma}}$. The corresponding estimate of the returns, denoted by $\boldsymbol{\mu}^1$, and the optimal portfolio, denoted by $\mathbf{x}^1$, are shown in Table 1.[3] As proved in Proposition 2, $\boldsymbol{\mu}^1$ equals the BL estimate $\boldsymbol{\mu}^{\mathrm{BL}}$. However, there is a slight difference between the BL portfolio $\mathbf{x}^{\mathrm{BL}}$ and $\mathbf{x}^1$. This can be attributed to the approximation (6) used in the BL procedure. Under the inverse optimization procedure we present in this example, the updated covariance estimate is identical to the historical estimate, not $\boldsymbol{\Sigma}^{\mathrm{BL}}$.

A second, more interesting example is the MV-IO approach. The first three eigenvalues of the historical covariance matrix (see e-companion) account for almost 87% of its spectrum. This suggests that a factor model may be an appropriate approximation. Solving (13) under the additional constraints (16) yields the estimate $\boldsymbol{\mu}^{\mathrm{MV}}$ and portfolio $\mathbf{x}^{\mathrm{MV}}$ in Table 1 and covariance estimate $(2\hat{\delta})^{-1}\bar{\boldsymbol{\Sigma}}$ in Table EC.3 of the e-companion.[4] For an investor who only believes that the major market factors are likely to remain constant in the future and that returns will be close to a CAPM equilibrium, this portfolio more closely accords with her views.

4. Beyond Mean-Variance Portfolio Allocation

The models of the previous section assume that investors measure risk in terms of standard deviation. Alternative measures of risk have been suggested, including value at risk (VaR), conditional value at risk (CVaR), and more generic coherent risk measures. Many practitioners believe these measures better represent investor behavior. This section is concerned with generalizing the previous results to a model capable of capturing some of these measures.

4.1. Definitions

Given a random variable Z, its value at risk is defined by $\mathrm{VaR}_\alpha(Z) = \inf\{z \in \mathbb{R} : \mathbb{P}(z + Z \geqslant 0) \leqslant 1 - \alpha\}$ for any $\alpha \in (0, 1)$. In other words, $\mathrm{VaR}_\alpha(Z)$ is the negative of the α-quantile. One way of measuring portfolio risk would be to posit a distribution for $\mathbf{r}$ and place a limit on the $\alpha\%$ worst-case losses with respect to some benchmark, such as the risk-free rate r_f. The corresponding optimization problem has the form

$$\max_{\mathbf{x}}\big\{\boldsymbol{\mu}'\mathbf{x} + (1 - \mathbf{e}'\mathbf{x})r_f : \mathrm{VaR}_\alpha((\mathbf{r} - r_f\mathbf{e})'\mathbf{x}) \leqslant L\big\}, \qquad (17)$$

for some value L.

Unfortunately, for an arbitrary distribution of $\mathbf{r}$, the feasible region of (17) may be nonconvex. This poses computational difficulties. A popular alternative that maintains the convexity of the problem is to use a *coherent* risk measure (Artzner et al. 1999).

A function ρ of a bounded random variable Z is a coherent risk measure if it satisfies the following conditions:

(a) *Monotonicity*: If $Z \geqslant Y a.s.$, then $\rho(Z) \leqslant \rho(Y)$.

(b) *Translation Invariance*: If $c \in R$, then $\rho(Z + c) = \rho(Z) - c$.

(c) *Convexity*: If $\lambda \in [0, 1]$, then $\rho(\lambda Z + (1 - \lambda)Y) \leqslant \lambda\rho(Z) + (1 - \lambda)\rho(Y)$.

(d) *Positive Homogeneity*: If $\lambda \geqslant 0$, then $\rho(\lambda Z) = \lambda\rho(Z)$.

A canonical example of a coherent risk measure is CVaR defined by $\mathrm{CVaR}_\alpha(Z) = -\mathbb{E}[Z \mid Z \leqslant \mathrm{VaR}_\alpha(Z)]$. For a generic coherent risk measure ρ, we formulate the portfolio optimization problem

$$\max_{\mathbf{x}}\big\{\boldsymbol{\mu}'\mathbf{x} + (1 - \mathbf{e}'\mathbf{x})r_f : \rho\big((\mathbf{r} - r_f\mathbf{e})'\mathbf{x}\big) \leqslant L\big\}. \qquad (18)$$

Notice that this problem is always convex by Property (c) regardless of the choice of distribution of $\mathbf{r}$ or the coherent risk measure ρ.

We next unify these frameworks by introducing a general portfolio allocation problem.

4.2. A General Portfolio Allocation Problem

Consider the following optimization problem:

$$P(L): \quad \max_{\mathbf{x}}\big\{\boldsymbol{\mu}'\mathbf{x} + (1 - \mathbf{e}'\mathbf{x})r_f : (\mathbf{r} - r_f\mathbf{e})'\mathbf{x}$$
$$\geqslant -L \; \forall \mathbf{r} \in \mathcal{U}\big\}, \qquad (19)$$

for some set $\mathcal{U}$. Problem (19) is an example of a robust optimization problem. It has a natural interpretation; for every possible scenario $\mathbf{r} \in \mathcal{U}$, we require that the return on our portfolio be at least $-L$. The tractability of a robust optimization problem depends both on the structure of the problem and the form of the uncertainty set $\mathcal{U}$. (See Ben-Tal and Nemirovski 2008, Bertsimas et al. 2011 for a general survey.) We will use the notation $P(L)$ when we would like to make the dependence on L explicit in (19).

We next show that the problem (19) has many commonly studied problems in asset allocation as special cases.

PROPOSITION 3. *Consider the following uncertainty sets*:

$$\mathcal{U}_1 = \big\{\mathbf{r} : (\mathbf{r} - r_f\mathbf{e})'\boldsymbol{\Sigma}^{-1}(\mathbf{r} - r_f\mathbf{e}) \leqslant 1\big\},$$

$$\mathcal{U}_2 = \big\{\mathbf{r} : (\mathbf{r} - \boldsymbol{\mu} + r_f\mathbf{e})'\boldsymbol{\Sigma}^{-1}(\mathbf{r} - \boldsymbol{\mu} + r_f\mathbf{e}) \leqslant z_\alpha^2\big\},$$

$$\mathcal{U}_3 = \big\{\mathbf{r} : (\mathbf{r} - \boldsymbol{\mu} + r_f\mathbf{e})'\boldsymbol{\Sigma}^{-1}(\mathbf{r} - \boldsymbol{\mu} + r_f\mathbf{e}) \leqslant 2\pi\alpha^2 e^{-z_\alpha^2/2}\big\}.$$

(a) *Problem* (19) *with* $\mathcal{U} = \mathcal{U}_1$ *is equivalent to the Markowitz problem* (1).

In the special case when $\mathbf{r}$ *is distributed as a multivariate Gaussian*, $\mathbf{r} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

(b) *Problem* (19) *with* $\mathcal{U} = \mathcal{U}_2$ *and* $\alpha \leqslant 1/2$ *is equivalent to the Value at Risk formulation* (17).

(c) *Problem* (19) *with* $\mathcal{U} = \mathcal{U}_3$ *is equivalent to the coherent risk measure formulation* (18) *for CVaR*.

PROOF. The proof of (a) can be found in Bertsimas and Brown (2009) and Natarajan et al. (2009). All three claims, though, follow directly from the observation that the minimum of a linear function over an ellipsoid admits a closed-form solution, and the expectations defining VaR and CVaR can be computed explicitly when $\mathbf{r}$ is Gaussian. We omit the details. $\square$

PROPOSITION 4 (BERTSIMAS AND BROWN 2009, NATARAJAN ET AL. 2009). *For a given probability measure* $\mathbb{P}$ *of* $\mathbf{r}$ *and a given coherent risk measure* ρ, *there exists a convex set* $\mathcal{U}(\rho, \mathbb{P})$ *such that the Problem* (19) *with* $\mathcal{U} = \mathcal{U}(\rho, \mathbb{P})$ *is equivalent to* (18).

The proof of the above proposition is beyond the scope of this paper. We only mention that for many common risk measures and probability measures, like CVaR_α over a discrete distribution, the corresponding set $\mathcal{U}(\rho, \mathbb{P})$ admits a simple, tractable description.

4.3. Characterizing Mean Returns in Equilibrium

In this section we generalize the BL framework by constructing BL-type estimates for problem (19). To this end, we must first argue that $\mathbf{x}^{\mathrm{mkt}}$ is an observable, optimal solution. The following proposition follows directly by scaling any given solution.

PROPOSITION 5. *Let* $\mathbf{x}^*(L)$ *denote an optimal solution to* $P(L)$ *when it exists.*

(a) *For any* $\delta \geqslant 0$, $\delta\mathbf{x}^*(L)$ *is an optimal solution to* $P(\delta L)$.

(b) *If* $\mathbf{x}^*(L)$ *exists and is unique for some* $L > 0$, *then* $\mathbf{x}^*(L)$ *exists and is unique for all* $L > 0$.

We now assert that in equilibrium, $\mathbf{x}^{\mathrm{mkt}}$ is an optimal portfolio.

PROPOSITION 6. *Assume that all investors solve* $P(L)$ *for some investor-specific value of L and that for some value of* $L > 0$ *the solution* $\mathbf{x}^*(L)$ *is unique. Then, there exists an* L^* *such that* $\mathbf{x}^{\mathrm{mkt}}$ *is an optimal solution to* $P(L^*)$.

PROOF. Recall that $\mathbf{x}^{\text{mkt}}$ is the percentage of total wealth invested in each asset. Let $p(l)$ be the fraction of wealth held by investors who have risk preference $L = l$. By Proposition 5(a) we have for each asset $i = 1, \dots, n$

$$x_i^{\text{mkt}} = \int_{l=0}^{\infty} x_i^*(l) p(l)\, dl = x_i^*(1) \int_{l=0}^{\infty} l p(l)\, dl.$$

Letting $L^* = \int_{l=0}^{\infty} l p(l)\, dl$ and applying Proposition 5(a) again yields $x_i^{\text{mkt}} = x_i^*(L^*)$, as desired. $\square$

The assumption that $\mathbf{x}^*(L)$ is unique for some $L > 0$ can be relaxed in Proposition 6 at the expense of more notation. We omit the details.

We now proceed to use inverse optimization to characterize the set of data for which $\mathbf{x}^{\text{mkt}}$ is optimal. For the remainder of this section we will assume $\mathcal{U}$ has the form

$$\mathcal{U} = \left\{ \mathbf{r} \colon \exists\, \mathbf{v} \text{ s.t. } \mathbf{Fr} + \mathbf{Gv} - \mathbf{g} \in K \right\}. \tag{20}$$

Here K is a proper cone (i.e., convex, pointed, closed, and with nonempty interior) such as the nonnegative orthant, the second-order cone, or the semidefinite cone. Examples of sets $\mathcal{U}$ describable by (20) are polyhedra and intersections of ellipsoids, which subsume our previous examples. Finally, in Natarajan et al. (2009), it is shown that under some additional technical conditions on $\mathbf{r}$ and ρ, the set $\mathcal{U}(\rho, \mathbb{P})$ is describable by (20). In this sense, (20) is not a very restrictive assumption.

We now prove the main result of this section. We will write $\mathbf{x} \geqslant_K \mathbf{y}$ whenever $\mathbf{x} - \mathbf{y} \in K$.

THEOREM 2. *Suppose* (19) *is feasible and has a unique optimum for some $L > 0$. Assume further that $\mathcal{U}$ in (20) has a strictly interior point. Then, if all investors solve* (19) *for some investor-specific value of L, there must exist values* $\mathbf{p}, \Gamma, \boldsymbol{\beta}$ *such that* $(\boldsymbol{\mu}, r_f, L)$ *satisfy*

$$(\boldsymbol{\mu} - r_f \mathbf{e})' \mathbf{x}^{\text{mkt}} = \Gamma L,$$
$$\mathbf{F}' \mathbf{p} = \mathbf{x}^{\text{mkt}}, \quad \mathbf{G}' \mathbf{p} = \mathbf{0}, \quad \mathbf{g}' \mathbf{p} \geqslant -L,$$
$$\Gamma r_f \mathbf{e} - \boldsymbol{\beta} = \boldsymbol{\mu} - r_f \mathbf{e}, \tag{21}$$
$$\mathbf{F} \boldsymbol{\beta} + \mathbf{G} \mathbf{w} \geqslant_K \Gamma \mathbf{g},$$
$$\mathbf{p} \in K^*, \quad \Gamma \geqslant 0,$$

where K^* is the dual cone to K, i.e., $K^* = \{\mathbf{y} \colon \mathbf{y}' \mathbf{x} \geqslant 0\ \forall\, \mathbf{x} \in K\}$.

PROOF. We first rewrite (19) as a conic optimization problem. Note

$$\left\{ \mathbf{x} \in \mathbb{R}^n \colon (\mathbf{r} - r_f \mathbf{e})' \mathbf{x} \geqslant -L\ \forall\, \mathbf{r} \in \mathcal{U} \right\}$$
$$= \left\{ \mathbf{x} \in \mathbb{R}^n \colon \min_{\mathbf{r} \in \mathcal{U}} \mathbf{r}' \mathbf{x} \geqslant -L + r_f \mathbf{e}' \mathbf{x} \right\}.$$

For any $\mathbf{x}$ feasible in (19), the optimization problem $\min_{\mathbf{r} \in \mathcal{U}} \mathbf{r}' \mathbf{x}$ must be bounded for that $\mathbf{x}$. Because the set

$\mathcal{U}$ has a strictly feasible point, it follows that this optimization further satisfies a Slater condition, and, consequently, satisfies strong conic duality. Thus, its optimal value is equal to the optimal value of the following program: $\max_{\mathbf{p}} \{\mathbf{g}' \mathbf{p} \colon \mathbf{F}' \mathbf{p} = \mathbf{x}, \mathbf{G}' \mathbf{p} = \mathbf{0}, \mathbf{p} \in K^*\}$. Consequently, we can rewrite the original optimization problem (19) as

$$\max_{\mathbf{x}, \mathbf{p}, \mathbf{v}} \left\{ r_f + (\boldsymbol{\mu} - r_f \mathbf{e})' \mathbf{x} \colon \mathbf{g}' \mathbf{p} \geqslant -L + r_f \mathbf{e}' \mathbf{x}, \mathbf{F}' \mathbf{p} = \mathbf{x}, \right.$$
$$\left. \mathbf{G}' \mathbf{p} = \mathbf{0}, \mathbf{p} \in K^* \right\}. \tag{22}$$

By assumption, this problem has a unique optimal solution for some $L > 0$, and so by Proposition 5(b), it has a unique optimal solution for all $L > 0$. This implies that this problem is bounded above and we can apply strong duality again. Its dual problem is

$$\min_{\Gamma, \boldsymbol{\beta}, \mathbf{w}} \left\{ \Gamma L \mid \Gamma r_f \mathbf{e} - \boldsymbol{\beta} = \boldsymbol{\mu} - r_f \mathbf{e}, \mathbf{F} \boldsymbol{\beta} \right.$$
$$\left. + \mathbf{G} \mathbf{w} \geqslant_K \Gamma \mathbf{g}, \Gamma \geqslant 0 \right\}. \tag{23}$$

Finally, by Proposition (6) $\mathbf{x}^{\text{mkt}}$ is an optimal solution for some value of L. The system (21) is a restatement of this optimality in terms of the conditions for strong conic duality. Namely, $\mathbf{x}^{\text{mkt}}$ is primal feasible (Problem (22)), there exist dual feasible variables (Problem (23)), and the objective value of the primal and dual problems are equal. Note we can use the fact that $\mathbf{e}' \mathbf{x}^{\text{mkt}} = 1$ to simplify these systems of equations slightly. $\square$

As in §3 we can also integrate investor views by adding appropriate constraints and solving for the minimal perturbation such that the resulting system is feasible. We illustrate this idea in the next section.

4.4. The Robust Mean Variance Inverse Optimization (RMV-IO) Approach

In the Markowitz formulation, an investor assumes that future volatility is determined by the covariance matrix $\boldsymbol{\Sigma}$. An investor without good volatility information may be uncomfortable specifying a single matrix $\boldsymbol{\Sigma}$. Rather, she might specify a small collection of volatility scenarios $\{\boldsymbol{\Sigma}^1, \boldsymbol{\Sigma}^2, \dots\}$ and insist that her portfolio not incur too much risk under any scenario. In this way, she forms a portfolio robust to her own volatility uncertainty.

As an example, suppose an investor believes returns are driven by a factor model and can estimate these market factors. She might then solve the optimization problem:

$$\max_{\mathbf{x}} \left\{ r_f + (\boldsymbol{\mu} - r_f \mathbf{e})' \mathbf{x} \colon \sqrt{\mathbf{x}' \boldsymbol{\Sigma} \mathbf{x}} \leqslant L, \forall\, \boldsymbol{\Sigma} \in \hat{\mathcal{U}} \right\}, \tag{24}$$

where

$$\hat{\mathcal{U}} \equiv \left\{ \boldsymbol{\Sigma} \succeq \mathbf{0} \colon \| \boldsymbol{\Sigma} \mathbf{v}^i - \lambda_i \mathbf{v}^i \| \leqslant \epsilon,\ i = 1, \dots, k, \right.$$
$$\left. \operatorname{tr}(\boldsymbol{\Sigma}) \leqslant (1/\alpha) \sum_{i=1}^{k} \lambda_i \right\}.$$

Note in the limit when $k = n$ and $\epsilon = 0$, problem (24) is equivalent to problem (1). The parameters ϵ and α determine the "size" of the uncertainty set. They are often referred to as the "budget" of uncertainty.

To form a BL-type estimator for (24), we first rewrite problem (24) as the equivalent optimization problem (19) with the uncertainty set

$$\mathcal{U} = \left\{ \mathbf{r}: \exists \boldsymbol{\Sigma} \text{ s.t. } \boldsymbol{\Sigma} \succeq \mathbf{0}, \ (\mathbf{r} - r_f \mathbf{e})' \boldsymbol{\Sigma}^{-1} (\mathbf{r} - r_f \mathbf{e}) \leqslant 1, \right.$$
$$\left. \|\boldsymbol{\Sigma} \mathbf{v}^i - \lambda_i \mathbf{v}^i\| \leqslant \epsilon \ \forall i = 1, \ldots, k, \ \mathrm{tr}(\boldsymbol{\Sigma}) \leqslant \frac{1}{\alpha} \sum_{i=1}^{k} \lambda_i \right\}$$

by using the explicit solution to minimizing a linear function over an ellipsoid. We then use Schur complements to rewrite $\mathcal{U}$ in the form (20). Applying Theorem 2, we conclude that there must exist parameters $(\boldsymbol{\mu}, L, z, \boldsymbol{\Sigma}, \theta, \gamma, \mathbf{B}^i, \mathbf{p}^i, \Gamma^i)$ for $i = 1, \ldots, k$ such that

$$(\boldsymbol{\mu} - r_f \mathbf{e})' \mathbf{x}^{\mathrm{mkt}} = Lz, \quad \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\mu} - r_f \mathbf{e} \\ (\boldsymbol{\mu} - r_f \mathbf{e})' & z \end{pmatrix} \succeq \mathbf{0},$$

$$\frac{1}{\alpha} z \sum_{i=1}^{k} \lambda_i \geqslant \mathrm{tr}(\boldsymbol{\Sigma}), \quad \|\boldsymbol{\Sigma} \mathbf{v}^i - \lambda_i \mathbf{v}^i\| \leqslant \epsilon, \quad i = 1, \ldots, k,$$

$$\begin{pmatrix} \theta \mathbf{I} & \frac{1}{2} \mathbf{x}^{\mathrm{mkt}} \\ \frac{1}{2} \mathbf{x}^{\mathrm{mkt}'} & \gamma \end{pmatrix} + \sum_{i=1}^{k} \sum_{j=1}^{n} p_j^i \overline{\mathbf{A}(i, j)} \succeq \mathbf{0},$$

$$\gamma + \epsilon^2 \sum_{i=1}^{k} \Gamma^i + \sum_{i=1}^{k} \mathrm{tr}(\mathbf{B}^i) + \frac{1}{\alpha} \theta \sum_{i=1}^{k} \lambda_i + \sum_{i=1}^{k} \lambda_i \mathbf{v}^i \mathbf{p}^i \leqslant L,$$

$$z, \theta \geqslant 0, \quad \begin{pmatrix} \mathbf{B}^i & \frac{1}{2} \mathbf{p}^i \\ \frac{1}{2} \mathbf{p}^{i'} & \Gamma^i \end{pmatrix} \succeq \mathbf{0} \quad i = 1, \ldots, k.$$

Here we define

$$\mathbf{A}(i, j) \equiv \begin{pmatrix} \mathbf{0}_{n, j-1} & \mathbf{v}^i & \mathbf{0}_{n, n+1-j} \\ 0 & 0 & 0 \end{pmatrix},$$
$$i = 1, \ldots, k, \ j = 1, \ldots, n, \quad \text{and}$$
$$\overline{\mathbf{A}(i, j)} = \tfrac{1}{2} (\mathbf{A}(i, j) + \mathbf{A}(i, j)').$$

It is not clear how to solve this system. In practice, an investor would likely specify some parameters a priori. We will specify that $k = 3$, that $\mathbf{v}^i$ and λ^i have been determined from historical analysis, $\epsilon = 10^{-5}$ and that α is the proportion of the spectrum taken by the first three eigenvalues historically. Then, this system reduces to a tractable LMI, as we now show.

First note that z is similar to the parameter δ in the original BL framework. The case where $z = 0$ is degenerate and corresponds to an instance where an average investor tolerates no risk on her investment and all assets yield the risk-free rate. The more realistic case is when $z > 0$. In this case, the procedure only determines $\boldsymbol{\mu}$ up to proportionality because we may always scale $\boldsymbol{\mu}, \boldsymbol{\Sigma}, z$ by a positive constant. As in the BL framework, then, we must exogenously

specify a value for z. Some intuition for this choice stems from the fact that $z = (\boldsymbol{\mu} - r_f \mathbf{e})' \mathbf{x}^{\mathrm{mkt}} / L$, so that, like δ, z resembles a kind of Sharpe ratio measuring the risk-reward trade-off of the market portfolio.

Note that by assumption the set $\mathcal{U}$ is known a priori. Consequently, we can always take $L = L_{\mathrm{mkt}} \equiv \max_{\mathbf{r} \in \mathcal{U}} -\mathbf{r}' \mathbf{x}^{\mathrm{mkt}}$, guaranteeing the existence of the necessary parameters $\gamma, \Gamma^i, \mathbf{B}^i, \theta, \mathbf{p}^i$. This suggests a method for choosing z; choose z such that $L_{\mathrm{mkt}} z$ achieves a target return for the market portfolio. In what follows, we adopt this approach, specifying the target return as the equilibrium return on the market portfolio in the Black-Litterman model. This is done primarily for ease of comparison between the two models.

Combining these observations yields the simpler condition that $\boldsymbol{\mu} \in \mathcal{F}$ where

$$\mathcal{F} = \left\{ \boldsymbol{\mu}: \exists \boldsymbol{\Sigma} \text{ s.t. } (\boldsymbol{\mu} - r_f \mathbf{e})' \mathbf{x}^{\mathrm{mkt}} = z L_{\mathrm{mkt}}, \ \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\mu} \\ \boldsymbol{\mu}' & z \end{pmatrix} \succeq \mathbf{0}, \right.$$
$$\left. \|\boldsymbol{\Sigma} \mathbf{v}^i - z \lambda_i \mathbf{v}^i\| \leqslant \epsilon \ \forall i = 1, \ldots, k, \ \mathrm{tr}(\boldsymbol{\Sigma}) \leqslant z \frac{1}{\alpha} \sum_{i=1}^{k} \lambda_i \right\}.$$

Incorporating investor views of the form (3), we specify a "BL"-type estimate by solving the optimization problem

$$\min_{t, \boldsymbol{\mu}, \boldsymbol{\mu}^{eq}, \boldsymbol{\Sigma}} \left\{ t: \boldsymbol{\mu}^{eq} \in \mathcal{F}, \ \left\| \begin{pmatrix} \boldsymbol{\mu} - r_f \mathbf{e} - \boldsymbol{\mu}^{eq} \\ \mathbf{P} \boldsymbol{\mu} - \mathbf{q} \end{pmatrix} \right\|_{\Omega} \leqslant t \right\}.$$

We will refer to the resulting estimator as the *robust mean-variance inverse optimization* (RMV-IO) *approach*. We continue our running example by applying the RMV-IO approach. The results are shown as $\boldsymbol{\mu}^{\mathrm{Rob}}$ and $\mathbf{x}^{\mathrm{Rob}}$ in Table 1.

5. Numerical Testing

As previously noted, the IO approach is more flexible than the traditional BL approach. The role of numerical testing, then, is not to decide which procedure "outperforms" the other, but rather to help identify *when* and *how* to use this additional flexibility. Our goal is not to be exhaustive, but rather to identify situations in which these models may be useful alternatives to the traditional BL model.

Our study proceeds in two steps: First, we contrast the three portfolios through simulation using stylized examples to isolate the conflating effects of the accuracy of the portfolio views and the equilibrium assumptions. Second, we backtest these portfolios against a set of historical returns.

In order to present the best "apples-to-apples" comparison we have limited ourselves to views on returns—the BL model does not support views on volatility—and have imposed factor structure on our estimate of the historical covariance matrix for the BL model.[5] Imposing factor structure on the historical estimate in the BL model helps ensure that observed differences can be attributed to the IO approach, not simply the use of factor structure.[6]

Before presenting the experiments, we summarize our key insights:

• Both IO portfolios demonstrate out-of-sample variance that is typically better than their BL counterparts. This is especially true of the RMV-IO portfolio.

• BL portfolios are more sensitive to the accuracy of the portfolio views than their IO counterparts. Consequently, when the views are correct, BL portfolios provide slightly higher returns than the IO portfolios. When views are incorrect, however, the BL portfolios' performance varies substantially depending on the realization of the market. By contrast, the performance of the IO portfolios is more consistent.

• As a result of the first two effects, depending on the risk appetite of the investor, IO portfolios may provide a better risk-reward trade-off than their BL counterparts.

• Finally, the magnitude of each of these effects is more evident when the view represents information that is distinct from the principal market factors. When the view represents information about the principal market factors, the methods perform very similarly.

5.1. Simulations

Extending our previous example, we construct portfolios using the three approaches for $\mathbf{p}$ (c.f. (8)) and with $L = 12.87\%$ (annualized), which is the expected standard deviation of the market portfolio based upon historical estimates. For convenience, we have set $r_f = 0$. All other parameters are as previously described.

5.1.1. Sensitivity to Accuracy of View. For simplicity, we first consider the scenario where future returns realize according to a distribution whose covariance matrix is identical to the historical covariance and whose mean value is in accordance with the CAPM equilibrium. In other words, the imposed view is entirely incorrect and the CAPM equilibrium holds entirely. The results for various values of q are shown in Table 2.

When $q = 0$, the additional view is nearly consistent with what is expected in equilibrium ($\mathbf{p}'\hat{\mathbf{r}} = 0.3862\%$). Thus, the

Table 3. Increasing confidence in an incorrect view.

	τ_1/τ_0	0.2	0.02	0.002	0.0002
Return	BL	4.02	4.02	3.45	1.14
	MV-IO	4.14	4.11	3.95	2.28
	RMV-IO	4.08	4.07	3.84	2.13
Std dev	BL	12.49	12.55	14.50	17.88
	MV-IO	12.86	12.78	12.36	8.78
	RMV-IO	12.72	12.68	12.14	8.69
Sharpe ratio	BL	32.17	32.02	23.82	6.38
	MV-IO	32.16	32.16	31.94	26.03
	RMV-IO	32.11	32.09	31.66	24.55

three portfolios are all close to the market portfolio. Differing estimates of the covariance matrix cause differences in the amount invested in the risky assets, but the Sharpe ratio is similar for the three methods.

As $|q|$ increases, all three portfolios underperform relative to the market because the view is increasingly incorrect. For even moderately incorrect views, the BL portfolio has noticeably worse returns and Sharpe ratio than either IO approach. These results suggest that the two IO approaches are more robust to inaccuracy in the views.

This robustness becomes more apparent as the confidence in that view increases. We consider fixing $q = 1$ and $\tau_0 = 1/24$, but vary our confidence in the incorrect view τ_1. The results are displayed in Table 3. The IO portfolios again retain better Sharpe ratios.

5.1.2. Market Perturbation. We next consider a scenario where future asset returns are drawn from a distribution whose covariance matrix differs from the historical covariance matrix by a small, random additive perturbation.[7] Mean returns are then generated according to the CAPM. This scenario is consistent with a factor model. Viewed as functions of this random perturbation, the portfolio's risk and return are random variables. Approximate densities for the portfolio return and portfolio standard deviation are shown in Figure 1. We have taken $q = 5$.

The returns of the two IO portfolios have similar densities. They are better on average and more concentrated

Table 2. Sensitivity to accuracy of the view under CAPM assumptions.

|	q	−10	−5	−2	−1	0	1	2	5	10
Return	Mkt	4.14	4.14	4.14	4.14	4.14	4.14	4.14	4.14	4.14
	BL	3.22	3.75	3.96	4.00	4.02	4.02	4.00	3.84	3.38
	MV-IO	4.08	4.11	4.12	4.13	4.14	4.11	4.10	4.05	3.92
	RMV-IO	3.99	4.09	4.13	4.13	4.15	4.07	4.05	3.97	3.80
Std dev	Mkt	12.87	12.87	12.87	12.87	12.87	12.87	12.87	12.87	12.87
	BL	14.26	13.03	12.54	12.48	12.48	12.55	12.67	13.31	14.67
	MV-IO	12.73	12.80	12.84	12.85	12.87	12.78	12.74	12.61	12.30
	RMV-IO	12.60	12.80	12.87	12.88	12.91	12.68	12.62	12.44	12.04
Sharpe ratio	Mkt	32.17	32.17	32.17	32.17	32.17	32.17	32.17	32.17	32.17
	BL	22.62	28.77	31.55	32.01	32.17	32.02	31.56	28.88	23.07
	MV-IO	32.07	32.10	32.11	32.11	32.17	32.16	32.15	32.09	31.90
	RMV-IO	31.67	31.96	32.08	32.11	32.15	32.09	32.05	31.91	31.56

Figure 1. Comparative performance of BL and MV-IO portfolios after a small perturbation in the covariance matrix.
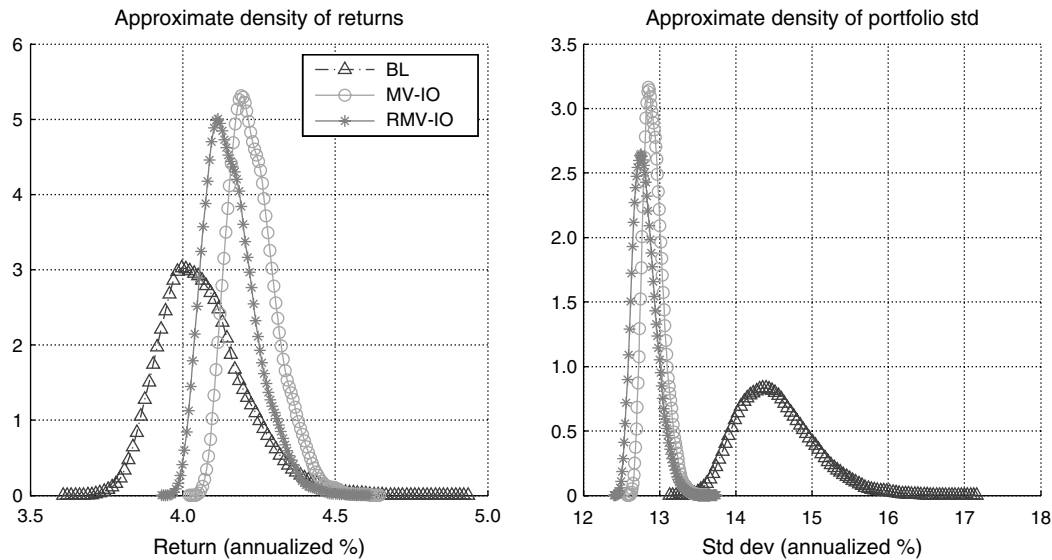
Note. The view $q = 5$.

around their means than the BL portfolio. The standard deviations of the IO portfolios are also smaller on average and have thinner tails than the BL portfolio. Interestingly, the returns for all three portfolios are highly correlated with the realization of the view portfolio. In Figure 2 we have plotted the excess returns over the market portfolio for each approach as a function of the accuracy of the view $\mathbf{p}'\mathbf{r} - q$. The BL portfolio demonstrates a stronger dependence. In scenarios where the view is more correct (i.e., $\mathbf{p}'\mathbf{r} - q \approx 0$), the BL approach outperforms the IO approaches. For these data, however, such scenarios are very rare.

Figure 2. Dependence of portfolio return on accuracy of the view.
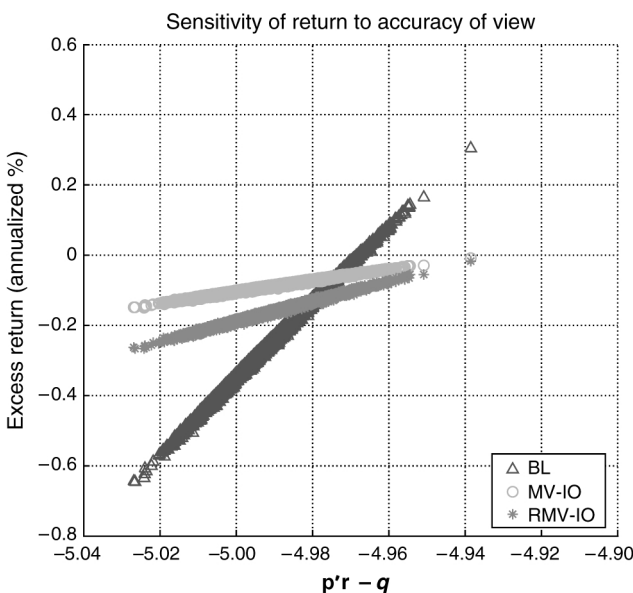
To better understand this linear dependence, we compute the slope of a regression line for each set of returns as a function of the view realization. We provide some results in Table EC.4 of the e-companion. Overall, the slope is generally positive when $q > \mathbf{p}'\hat{\mathbf{r}}$, i.e., the view predicts $\mathbf{p}$ to perform better than it would in equilibrium, and increasing in magnitude in both the confidence τ_1/τ_0 and the extremity $|q - \mathbf{p}'\hat{\mathbf{r}}|$. In He and Litterman (1999), the authors prove these assertions formally for the BL model with a single view under approximation (6). Our numerical study suggests that similar statements hold, at least approximately, for the MV-IO and RMV-IO approaches, but that the dependence is more mild. This insight will provide strong intuition when backtesting against a historical data set.

Finally, a more complete picture of the dependence on q as can be seen in Table EC.5 in the e-companion, where we have summarized portfolio characteristics after 10,000 simulations.

5.1.3. Sensitivity to Choice of Portfolio p. We next consider a different view portfolio:

$$\tilde{\mathbf{p}}' = \big[20\%, 0\%, 10\%, 10\%, 10\%, 10\%, 10\%, -10\%, 20\%, 20\% \big].$$

Intuitively, the difference between $\mathbf{p}$ and $\tilde{\mathbf{p}}$ is that $\mathbf{p}$ describes information distinct from the information in the principle market factors, whereas $\tilde{\mathbf{p}}$ describes information similar to the information in the principle market factors. More formally, the projection of $\mathbf{p}$ onto the space spanned by the first three eigenvectors of $\hat{\boldsymbol{\Sigma}}$ is small, whereas the projection of $\tilde{\mathbf{p}}$ onto the same space is large. Specifically,

$$\frac{\| \operatorname{proj}(\mathbf{p}, < \mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3 >) \|}{\| \mathbf{p} \|} \approx 2\%,$$

$$\frac{\| \operatorname{proj}(\tilde{\mathbf{p}}, < \mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3 >) \|}{\| \tilde{\mathbf{p}} \|} \approx 84\%,$$

where $\mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3$ are the three largest eigenvalues of $\hat{\boldsymbol{\Sigma}}$. Because both IO approaches constrain the eigenvectors of the covariance matrix to be close to their historical values, we expect that the extra degrees of freedom afforded to them in fitting the covariance matrix are not helpful in realizing this view. All three approaches must attempt to realize the view by adjusting the estimate of the mean return away from the CAPM equilibrium. Consequently, the portfolios perform similarly. We display some illustrative results in Table EC.6 of the e-companion. Note that the ratio of portfolio return to standard deviation is nearly identical for all methods.

5.2. Backtesting

We consider a data set of Morgan Stanley Composite Index (MSCI) sector indices from June '92 to Dec '09 with the three-month U.S. Treasury bill as a proxy for the risk-free rate. For each month beginning with June '98, we use the previous 60 months of historical data to fit an empirical covariance matrix and use that month's market capitalization across the sectors to compute the BL, MV-IO, and RMV-IO portfolios. Portfolios are constructed to have a target annualized standard deviation $L = 15.73\%$. We include the Markowitz portfolio with this target standard deviation in our results for comparison.

To emphasize the differences between the methods, we choose a single portfolio view over the entire time horizon, namely, $\mathbf{p}$ defined by (8). Recall that for a portfolio that lives in the space spanned by the first three eigenvectors of the covariance matrix, the performance of all three methods is quite similar. The percentage of $\mathbf{p}$ that lives in this space ranges from 5%–40%. This percentage and the realized returns for this portfolio are plotted in Figure EC.1 of the e-companion. The average realized monthly return for $\mathbf{p}$ is 0.03 %.

5.2.1. Static Views. We first consider the performance of the portfolios when q is the same fixed constant each period. This is a particularly crude belief, but in some ways

it is the best way to disassociate the performance of the methods from the system generating the views. We calculate the mean return and standard deviation experienced over the data set. We plot these metrics in the left two panels of Figures 3. To better understand the risk experienced *within* the period, we also compute the maximal standard deviation achieved over a rolling window of 24 months, plotted in the rightmost panel of Figure 3.

The results for standard deviation and risk are quite striking. For almost all values of q, the IO portfolios adopt less risk than the BL portfolio. Indeed, the difference between the BL approach and the RMV-IO approach can be as great as 6% for overall standard deviation and 11% for maximal standard deviation. These results are intuitive because both IO approaches allow a more flexible volatility estimation, and the RMV-IO approach, in particular, protects against worst-case realizations of volatility. Recall that the target standard deviation for these portfolios was 15.73%, so that the IO portfolios more accurately match the target.

The results for the average return, however, are more subtle. When $q = 0$, the view is almost in accordance with equilibrium, and all three portfolios perform similarly to the Markowitz portfolio. One might be tempted to infer that for $q < 0$ the BL significantly outperforms the IO approaches, whereas for $q > 0$, it underperforms. This inference, though, heavily depends on the choice of out-of-sample window. Compare Figure 3 with Figure 4, where we have performed the same experiment, but only averaged over the period Dec. '99–Dec. '09 (a difference of 18 months). The average return of the BL portfolio demonstrates a very different dependence on q. By contrast, the average return of the IO portfolios and the standard deviation of all three portfolios is still quite similar.

We feel that this difference can at least partially be explained by our insights from §5.1.2. Because the BL portfolio is very sensitive to the realization of the view, the portfolio may experience very large gains and losses. For example, suppose the view predicts $\mathbf{p}$ to perform much better than it would in equilibrium, i.e., $q \gg \mathbf{p}^T \hat{\mathbf{r}}$. If the

Figure 3. Portfolio characteristics for static views (June '98–December '09).
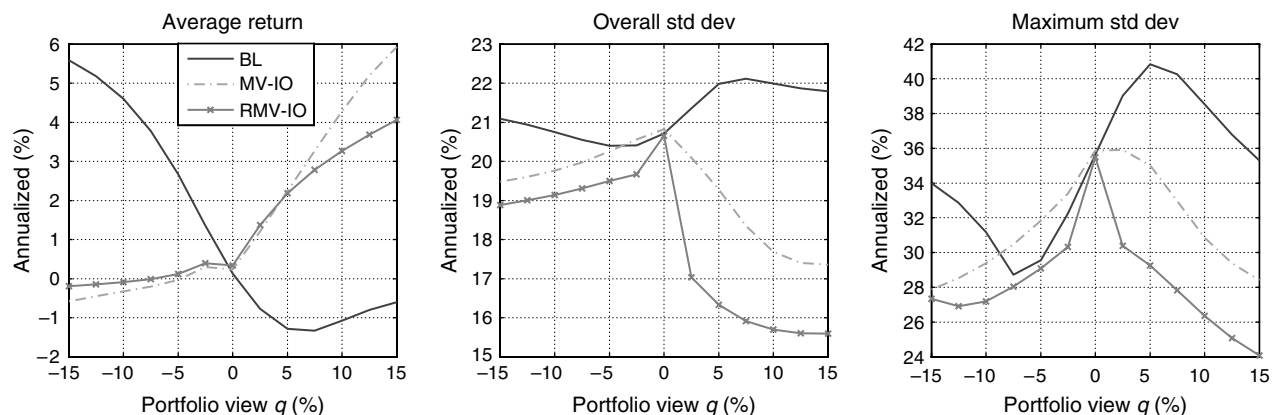
Notes. For comparison, the Markowitz portfolio has an average return of 0.00%, an overall standard deviation of 21.07%, and a maximal standard deviation of 36.17%.

Figure 4. Portfolio characteristics for static views (December '99–December '09).
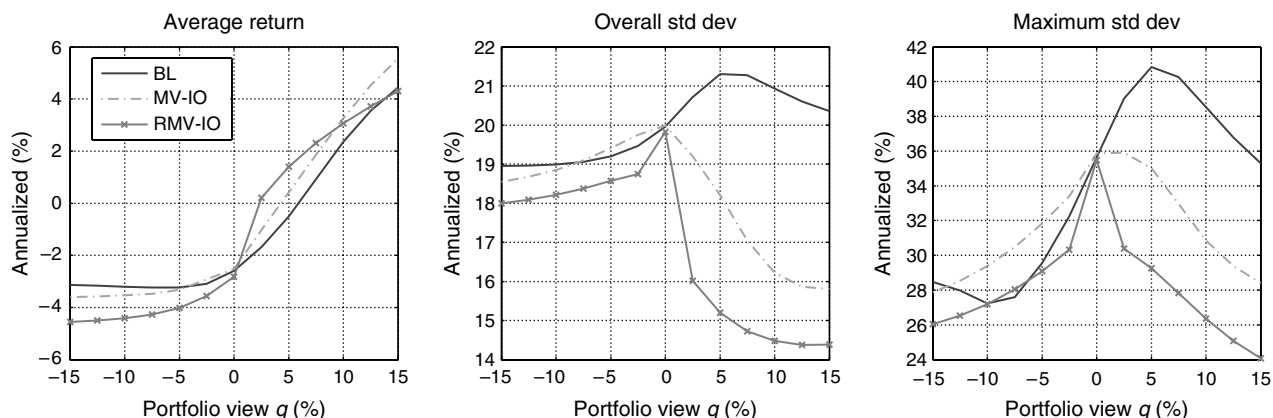
Notes. For comparison, the Markowitz portfolio has an average return of -2.97%, an overall standard deviation of 20.29%, and a maximal standard deviation of 37.17% over the period. The target standard deviation of these portfolios is 15.73%.

realization of the view is also above the market, the BL portfolio will perform well. If the realization of the view is below the market, the BL portfolio will perform badly. This high sensitivity can cause large swings in the portfolio value. Indeed, the returns on the BL portfolio from June '97 to Dec '99 change so dramatically as to cause the difference in Figures 3 and 4. By contrast, because the IO portfolios are less sensitive to the view realization, they are more stable to the choice of window. Depending on the investment objectives, investors may prefer this stability.

5.2.2. Systematic Bias. The previous experiment compared the performance of the methodologies when views were static and (generally) incorrect. Next we consider the case when views are (generally) correct, while still using a realistic view generation mechanism.

To this end, we consider a scenario where the view generation mechanism has a systematic bias. Specifically, at each point in time, the mechanism returns the true return of the portfolio **p** in the next period, plus an offset Δq. When $\Delta q < 0$, we underpredict. When $\Delta q > 0$, we over-

predict. The case $\Delta q = 0$ corresponds to perfect foresight. We plot the average return and overall standard deviation for various values of Δq in the left two panels of Figure 5. We also plot the maximal standard deviation achieved in a rolling 24-month window in the rightmost panel. Recall, the target standard deviation for these portfolios is 15.73%.

Not surprisingly, all methods significantly outperform the market portfolio when Δq is small. This outperformance is precisely the reason that "BL"-type estimators are very popular in pratice. The outperformance can be as much as 20% in annualized return. Due to its higher sensitivity, the BL portfolio has a higher return than either IO approach when the views are very accurate ($\Delta q = 0$), but this difference degrades as the view becomes less accurate. Moreover, as one might suspect, this extra return comes at the cost of increased risk. The BL model generally adopts more risk than either IO approach. The difference can be quite substantial, as much as 6% in overall standard deviation and 11% in maximal standard deviation.

Figure 5. Portfolio characteristics for increasingly inaccurate views (June '98–December '09).
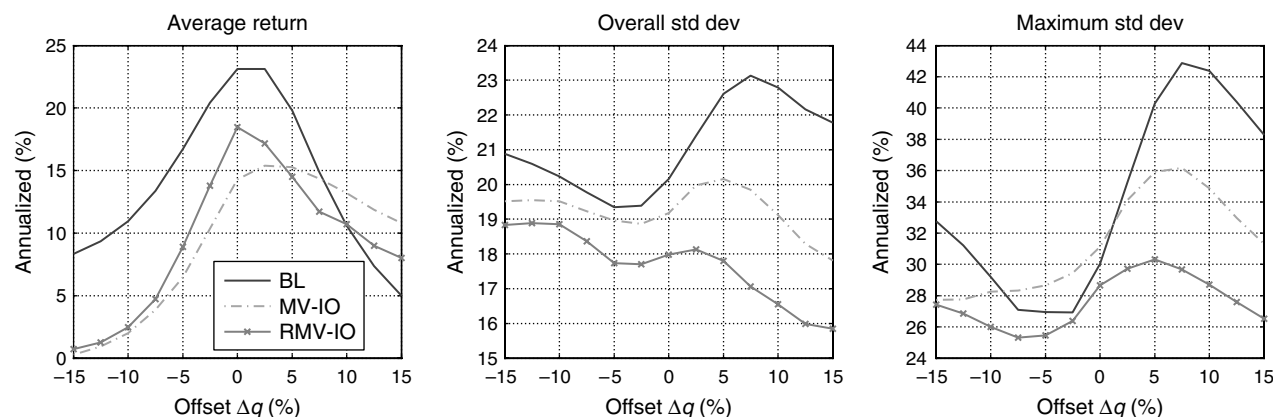
Notes. For comparison, the Markowitz portfolio has an average return of 0.00%, an overall standard deviation of 21.07%, and a maximal standard deviation of 36.17% over the period.

As in the previous example, one might be tempted to draw inferences about the behavior of the portfolios when $\Delta q < 0$ versus when $\Delta q > 0$, but again, these inferences are highly dependent on the choice of out-of-sample window and portfolio $\mathbf{p}$.

5.2.3. Length of Historical Window. To test the robustness of our insights, we reran the previous two experiments for various lengths of historical window. Generally speaking, the precise choice of window does not make a material difference. To illustrate, we have plotted the results of these experiments for a window of length of 72 months in Figures EC.2 and EC.3 of the e-companion.

6. Conclusion and Further Research

In this paper we have used techniques from inverse optimization to create a novel, richer, reformulation of the Black-Litterman (BL) framework. The major advantage of this approach is the increased flexibility for specifying views and the ability to consider more general notions of risk than the traditional mean-variance approach. We have exploited this flexibility to introduce two new BL-type estimators and their corresponding portfolios, a mean-variance inverse optimization (MV-IO) approach and a robust mean-variance inverse optimization (RMV-IO) approach. The major distinction between the approaches is that the first allows investors to capitalize upon any private information they may have on volatility, whereas the second seeks to insulate investors from volatility uncertainty when they have no such information. Computational evidence suggests that these approaches provide certain benefits over the traditional BL model, especially in scenarios where views are not known precisely.

These two are not the only estimators that can be constructed from the inverse optimization methodology. Indeed, it is possible to construct a number of different "BL"-type estimators under different assumptions about investor behavior and market dynamics. For example, one might frame an IO estimator based upon shrinkage techniques for the covariance matrix (following Ledoit and Wolf 2003). Shrinkage estimators have demonstrated significant benefits in other areas of finance. On the other hand, leveraging the explicit formulation of the set $\mathcal{U}(\rho, \mathbb{P})$ for CVaR_α from Natarajan et al. (2009) or Bertsimas and Brown (2009), one might consider a CVaR_α-based estimator where the distribution of $\mathbf{r}$ is determined in a purely data-driven way. Each of these estimators will enjoy its own particular numerical properties.

We believe that computationally and analytically studying the properties of the inverse-optimization based estimators, especially in conjunction with specific view generation mechanisms, is a fruitful and practical area for future research. Given the flexibility and simplicity of the method, it should be possible to leverage the inverse optimization approach to build robust, high-performing portfolio investment strategies.

Electronic Companion

An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/opre.1120.1115.

Endnotes

1. To be precise, Black and Litterman (1992) does not provide an explicit Bayesian interpretation of their result, but instead relies on Thiel's Mixed Estimation Theorem. We do not know who was the first to recognize the result as a Bayesian update. Subsequent research, however, including Litterman's own (He and Litterman 1999), usually describes the initial result in a Bayesian framework.
2. Historical return data in this paper was drawn from the DataStream database using Morgan Stanley Composite Indices (MSCI) sector indices as proxies. We chose to use MSCI sector indices because they more closely resemble the actual investment opportunities for investors. For each sector, the market share of the sector was proxied using the percentage of the S&P 500 composed of stocks from that sector. Constituent lists for the S&P and the price and market share for each constituent stock were taken from the WRDS database. Monthly returns for the previous 24 months were used to construct volatility and correlation estimates.
3. To solve problem (13) and all other optimization problems in this paper, we use CVX, a package for specifying and solving convex optimization problems (Grant and Boyd 2010).
4. We have set $k = 3$, $\epsilon = 10^{-8}$, and α equal to 87% in line with the historical value. Furthermore, when the solution $\bar{\mathbf{\Sigma}}$ is nearly singular, here and in the remainder, we calculate $\mathbf{x}^{\text{MV}}$ using the pseudo-inverse of $\bar{\mathbf{\Sigma}}$ with a tolerance of 5×10^{-4}.
5. Specifically, we have used the estimation method proposed in §9.5.1 of Tsay (2005). Namely, we truncate the spectral decomposition of the traditional sample covariance matrix at the third-largest eigenvalue and add a diagonal matrix to this truncation so that the diagonal elements of our estimate and of the sample covariance matrix match. We use this sum of the truncation and diagonal matrix as the estimate $\hat{\mathbf{\Sigma}}$ in the BL model.
6. In fact, we have repeated all of our experiments using the full sample covariance matrix and a rank 3 estimator for the BL model and obtained similar results in both cases.
7. In detail, for each scenario we generate 10 random vectors uniformly from the surface of the unit sphere, say $\mathbf{v}^1, \dots, \mathbf{v}^{10}$. We then form the matrix $\mathbf{E} = \sum_{i=1}^{10} \mathbf{v}^i \mathbf{v}^{i\prime}$ and apply the scaled perturbation $((5 \times 10^{-3})/\text{trace}(\mathbf{E}))\mathbf{E}$. The choice of the constant 5×10^{-3} was based on the average difference in trace observed between covariance matrices in the data set.

Acknowledgments

References

Ahuja RK, Orlin JB (2001) Inverse optimization. *Oper. Res.* 49(5):771–783.

Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. *Math. Finance* 9(3):203–228.

Ben-Tal A, Nemirovski A (2008) Selected topics in robust convex optimization. *Math. Programming* 112(1):125–158.

Bertsekas DP (1999) *Nonlinear Programming* (Athena Scientific, Belmont, MA).

Bertsimas D, Brown DB (2009) Constructing uncertainty sets for robust linear optimization. *Oper. Res.* 57(6):1483–1495.

Bertsimas D, Brown DB, Caramanis C (2011) Theory and applications of robust optimization. *SIAM Rev.* 53:464–501.

Bertsimas D, Lauprete GJ, Samarov A (2004) Shortfall as a risk measure: Properties, optimization and applications. *J. Econom. Dynam. Control* 28(7):1353–1381.

Bevan A, Winkelmann K (1998) Using the Black-Litterman Global Asset Allocation Model: Three years of practical experience. Fixed Income Research, Goldman Sachs & Company, New York.

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

Boyd SP, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, New York).

Connor G, Korajczyk RA (1995) The arbitrage pricing theory and multifactor models of asset returns. Jarrow RA, Maksimovic V, Ziemba WT, eds. *Finance*. Handbooks in Operations Research and Management Science, Vol. 9 (Elsevier, Amsterdam), 87–144.

Da Silva AS, Lee W, Pornrojnangkool B (2009) The Black-Litterman model for active portfolio management. *J. Portfolio Management* 35(2):61–70.

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

Giacometti R, Bertocchi M, Rachev ST, Fabozzi FJ (2007) Stable distributions in the Black–Litterman approach to asset allocation. *Quant. Finance* 7(4):423–433.

Grant M, Boyd S (2010) CVX: Matlab software for disciplined convex programming, version 1.21. Accessed April 1, 2011, http://cvxr.com/cvx.

Grootveld H, Hallerbach W (1999) Variance vs. downside risk. *Eur. J. Oper. Res.* 114(2):304–319.

Harlow WV (1991) Asset allocation in a downside-risk framework. *Financial Analysts J.* 47(5):28–40.

He G, Litterman R (1999) The intuition behind Black-Litterman model portfolios. Investment Management Research, Goldman Sachs & Company, New York.

Herold U (2005) Computing implied returns in a meaningful way. *J. Asset Management* 6(1):53–64.

Heuberger C (2004) Inverse combinatorial optimization: A survey on problems, methods, and results. *J. Combin. Optim.* 8(3):329–361.

Iyengar G, Kang W (2005) Inverse conic programming with applications. *Oper. Res. Lett.* 33(3):319–330.

Jorion P (1997) *Value at Risk: The New Benchmark for Controlling Market Risk* (McGraw-Hill, New York).

Krishnan HP, Mains NE (2005) The two-factor Black-Litterman model. *Risk* (July).

Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empirical Finance* 10(5):603–621.

Martellini L, Ziemann V (2007) Extending Black-Litterman analysis beyond the mean-variance framework. *J. Portfolio Management* 33(4):33–44.

Meucci A (2005) Beyond Black-Litterman: Views on non-normal markets. http://ssrn.com/abstract=848407.

Meucci A (2008) Fully flexible views: Theory and practice. *Risk* 21(10):97–102.

Meucci A, Ardia D, Keel S (2011) Fully flexible views in multivariate normal markets. http://ssrn.com/abstract=1742559.

Natarajan K, Pachamanova D, Sim M (2009) Constructing risk measures from uncertainty sets. *Oper. Res.* 57(5):1129–1141.

Pástor L (2000) Portfolio selection and asset pricing models. *J. Finance* 55(1):179–223.

Pástor L, Stambaugh RF (2000) Comparing asset pricing models: An investment perspective. *J. Financial Econom.* 56(3):335–381.

Rockafellar RT, Uryasev S (2002) Conditional value-at-risk for general loss distributions. *J. Banking Finance* 26(7):1443–1471.

Sharpe WF (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Finance* 19(3):425–442.

Tsay RS (2005) *Analysis of Financial Time Series*, 2nd ed. (Wiley-Interscience, Hoboken, NJ).

Walters J (2010) The Black-Litterman model: A detailed exploration. Accessed January 1, 2011, http://www.blacklitterman.org.

Dimitris Bertsimas is the Boeing Professor of Operations Research, co-director of the Operations Research Center at Massachusetts Institute of Technology, and a member of the National Academy of Engineering. He works on optimization, applied probability, and data mining and their applications. The paper in this issue is part of the author's research program in understanding the implications of inverse optimization in a variety of fields.

Vishal Gupta is a Ph.D. candidate in operations research at Massachusetts Institute of Technology under the supervision of Dimitris Bertsimas. Prior to matriculating at MIT, he spent several years working in the financial industry. Broadly speaking, he is interested in developing tractable paradigms for modeling uncertainty.

Ioannis Ch. Paschalidis is a professor and Distinguished Faculty Fellow in the College of Engineering at Boston University with a joint appointment in the Department of Electrical and Computer Engineering and the Division of Systems Engineering. He is co-director of the Center for Information and Systems Engineering (CISE). His current research interests lie in the fields of systems and control, networking, applied probability, optimization, operations research, computational biology, and bioinformatics.