

KNOWLEDGE REPRESENTATION AND INTEGRATION FOR PORTFOLIO EVALUATION USING LINEAR BELIEF FUNCTIONS

Liping Liu, *Member, IEEE*, Catherine Shenoy, Prakash P. Shenoy

Abstract—In this paper, we propose a linear belief function approach to evaluating portfolio performance. By drawing on the notion of linear belief functions, we propose an elementary approach to knowledge representation for expert systems using linear belief functions. We show how to use basic matrices to represent market information and financial knowledge, including complete ignorance, statistical observations, subjective speculations, distributional assumptions, linear relations, and empirical asset pricing models. We then appeal to Dempster's rule of combination to integrate the knowledge for assessing the overall belief of portfolio performance, and updating the belief by incorporating additional information. We use an example of three gold stocks to illustrate the approach.

Index Terms — Linear belief functions, Dempster-Shafer belief functions, multivariate normal distribution, knowledge-based systems, portfolio evaluation

I. INTRODUCTION

In portfolio analysis, a financial asset is characterized as a random variable with a probability distribution over its possible returns [1]. A portfolio is a linear combination of asset variables [2] and hence a random variable itself with a return distribution functionally determined by the return distributions of the individual assets. The problem of selecting an optimal portfolio becomes one of ranking probability distributions by their mean-variance positions [1], general moments [3], or coarse approximations [4, 5].

Since the return distribution of a portfolio carries sufficient information to evaluate its performance and risk of various measures, its determination is crucial for the practice of portfolio management and financial decision-making. In the finance literature, there has been ample evidence suggesting that the return distributions are predictable [6]. However, one often finds it difficult to translate this evidence of predictability into practical portfolio advice [7]. Among many questions to be addressed, a most important one is how we could integrate a variety of information concerning asset returns into a predictive return distribution. Two approaches to portfolio selection are commonly used in finance [8]. A “data-based” approach

assumes a functional form for a return distribution and estimates its parameters from the time series of returns. For example, sample estimates of mean and variance may be used to compute the optimal portfolio in a mean-variance framework. This approach ignores potential usefulness of various asset pricing models, including the Capital Asset Pricing Model (CAPM) of Sharpe [9] and Lintner [10], the arbitrage pricing theory of Ross [11], as well as their multifactor extensions [12]. On the other hand, in an approach based on asset pricing models, the optimal portfolio is a combination of benchmark portfolios that expose an investor only to priced sources of risk. For example, the CAPM-based approach prescribes the market portfolio as a single optimal portfolio for every investor. This model-based approach makes no use of historical data on non-benchmark assets. These two typical approaches essentially reflect two extreme positions of selecting information for portfolio evaluation; the first approach bases on historical data while the second one bases on finance theory and its implications. The portfolio literature is silent about what happens in between except for a few recent attempts in combining sample evidence with the CAPM in a Bayesian framework [8].

Besides the selective use of data and models, the finance literature also overlooks a third category of information, which may be called *soft data*, including corporate and government announcements, news reports on industry and political events, as well as subjective judgments. These data are soft in the sense that they often involve ambiguity, leaving room for interpretations and subjective judgments. For example, research has found that economic and monetary factors such as the stance of monetary policy [13] and short-term interest rates [14] predict stock and bond returns. Thus, any news concerning these factors will lead to projections of their movement. The projections along with the empirical finding, of course, provide valuable information for assessing the return distribution of a stock. However, except for event studies investigating how certain events may be evidence to falsify asset-pricing models, the finance literature generally leaves the usage of soft data to financial analysts and individual speculators. Soft data do not enter into any formal modeling. In practice they are holistically combined with empirical data generated by the market, or theoretical models produced by finance theory.

The goal of this study is to propose a linear, also called *normal* (or *Gaussian*), belief function approach to the integration of the information concerning financial assets for modeling return distributions. In particular, we apply the notion of linear belief functions [15, 16], a recent extension of the Dempster-Shafer theory to continuous frames of discernment, to represent knowledge available for determining a portfolio

Manuscript received December 4, 2003. A short version has appeared in C. Meek and U. Kjaerulff (eds.), *Uncertainty in Artificial Intelligence*, 2003, 370–377, Morgan Kaufmann, San Francisco.

L. Liu is with College of Business Administration, University of Akron, Akron, OH 44325-4801, USA (Phone: 330-972-5947; fax: 330-972-6588; e-mail: liu@acm.org).

C. Shenoy is with School of Business, University of Kansas, 1300 Sunnyside Ave., Lawrence, KS 66045-7585 USA (e-mail: cshenoy@ku.edu).

P. P. Shenoy is with School of Business, University of Kansas, 1300 Sunnyside Ave., Lawrence, KS 66045-7585 USA (e-mail: pshenoy@ku.edu).

distribution, including historical data, subjective beliefs, observations, theoretical linear relations, empirical linear regression models, as well as complete ignorance. We then integrate and propagate the knowledge by combining the linear belief functions using Dempster’s rule [17].

A few existing studies have proposed knowledge integration for modeling portfolio distributions. In an earlier study, Black and Litterman [18] suggested using the CAPM as a benchmark toward which an investor can shrink his subjective views about expected returns. The deviation from the CAPM depends on the investor’s confidence in his subjective views. Pástor and Stambaugh [19] carried this idea further and proposed forming an informative prior on the α parameter of the CAPM and updating the prior belief by sample evidence in a Bayesian framework. Using the same approach, Pástor [8] empirically examined the decision problem of forming the optimal portfolio consisting of the market portfolio and one or more non-benchmark assets. Baks *et al.* [20] studied the performance evaluation of portfolios consisting of a risk-less asset, passively managed index funds (benchmark assets), and actively managed mutual funds (non-benchmark assets). The goal of these studies is more on the validation of the CAPM than on the prediction of portfolio returns. Their approach to knowledge integration is restricted to combining sample data with a subjective belief on a theoretical pricing model. Any market evidence such as a fund manager’s skill [20] will have to be translated into such a belief to be combined with sample data. Thus, the approach does not allow a full-fledged combination of market information and financial knowledge. For example, it is not able to combine contradicting views on a same economic factor or multiple empirical pricing models (for different assets or the same asset based on different data).

Shenoy and Shenoy [21] describe an approach based on Bayesian networks and represented market information and financial knowledge as marginal or conditional probabilities. The Bayes net approach is useful in domains where one has a causal model as a source of conditional independence assumptions. However, causal models are not common in portfolio theory. To overcome this limitation, Shenoy and Shenoy [22] propose using belief functions to encode market evidence and factor dependence. Belief function models are based on the semantics of distinct evidence, which may allow easier construction of models in portfolio theory. Shenoy and Shenoy [21, 22] initiated the idea of knowledge integration for portfolio modeling. However, their approaches may not be acceptable to financial analysts and researchers for several reasons. First, their knowledge representation is based on discrete belief functions and Bayesian probabilities, which cannot easily represent linear deterministic relations and theoretical pricing models. For example, they use a large number of ordered pairs to represent simple linear relations. The resulting representation is cumbersome and inefficient. Second, to fit into the finite framework, both approaches discretize continuous variables. This results in a very large state space and makes a problem involving hundreds of assets intractable to solve. Third, instead of focusing on a few moments, such as expected returns and risks as practical portfolio analysis does, their approach focus on computing an approximate posterior distribu-

tion. Unless one uses a fine granularity of discretization, the information on expected returns and risks will be distorted by the computation. But using a fine granularity will make both knowledge representation and problem solving less tractable.

This paper is organized as follows. In Section II, we introduce the concept of linear belief functions and present a matrix approach for knowledge representation and integration. In Section III, we use an example to illustrate our approach to portfolio modeling. This section includes the graphical representation of an expert system, detailed descriptions of how various items of evidence in portfolio modeling are represented as linear belief functions, how inferences are made in the expert system, how new evidence can be added to the model, and what the impact of the new evidence is on the portfolio risk and return distribution. In Section IV, we conclude the paper with a summary and a discussion.

II. THE CONCEPT OF LINEAR BELIEF FUNCTIONS

The Dempster-Shafer theory [17] of belief functions provides a flexible tool for knowledge representation and a rigorous mechanism for knowledge integration. Linear belief functions (LBF) extend the theory to the case when variables of interest are continuous [15, 16]. Examples of such variables include financial asset prices, portfolio performance, and other antecedent and consequent variables. In particular, a LBF can represent both logical and probabilistic knowledge for three types of variables: deterministic such as observables, random whose distribution is normal, and vacuous on which no knowledge bears. Logical knowledge is represented by linear equations. Probabilistic knowledge is represented by normal distributions.

Linear belief functions have also been studied by Monney [24], who calls these functions “Gaussian hints.” They can also be thought of as an extension of the multivariate normal distribution family as defined by Rao [40], and extended to include complete ignorance by Kenley [41].

Intuitively, a LBF represents knowledge regarding the true value of variables as follows. Assume the truth falls on a hyperplane for sure but we do not know its exact location. We call the hyperplane that contains the truth a *certainty hyperplane*. Also assume that, in some dimensions of the certainty hyperplane, the truth is anywhere from $-\infty$ to $+\infty$ and the probability of being at a particular location is described by a normal distribution. In other dimensions, we have ignorance; the truth is anywhere from $-\infty$ to $+\infty$ but the associated probability is unknown. In the parlance of belief functions, each sub-hyperplane along the dimensions of ignorance is a focal element and the normal distribution across the focal elements defines the basic probability assignment. In symbols, let V be the frame of discernment and C the certainty hyperplane. The focal elements, denoted by B , are the parallel sub-hyperplanes that constitute a partition of C (see Fig. 1).

Example 1: Let X , Y , and E be three variables involved in the normal error model: $Y = 10 - 0.8X + E$, where X is an exogenous variable, Y is an endogenous variable, and E is an independent error term. Let us further assume E is a white noise with standard deviation 5. Then, this model is a perfect

example illustrating the concept of LBFs. This belief function has a 3-dimensional frame of discernment $V = \{(x, y, \epsilon) \mid x, y, \epsilon \in R\}$ and a two-dimensional certainty hyperplane C as determined by the linear equation $Y = 10 - 0.8X + E$ (see Fig. 1). Since E has a known distribution and $E = Y - 10 + 0.8X$, each focal element B is the locus of the points on C that have the same E coordinate, i.e., $\{(x, y, \epsilon) \mid y - 10 + 0.5x = \epsilon \text{ and } \epsilon \text{ is a constant}\}$. This belief function describes our knowledge about X , Y , and E as follows. The true value of X , Y and E , say (x, y, ϵ) , must satisfy $y = 10 - 0.8x + \epsilon$ for sure, i.e., $(x, y, \epsilon) \in C$. The belief that it falls onto a particular straight-line (focal element) is described by the normal distribution $N(0, 25)$. However, we are ignorant regarding where it is located along the line and how likely it is at a particular point.

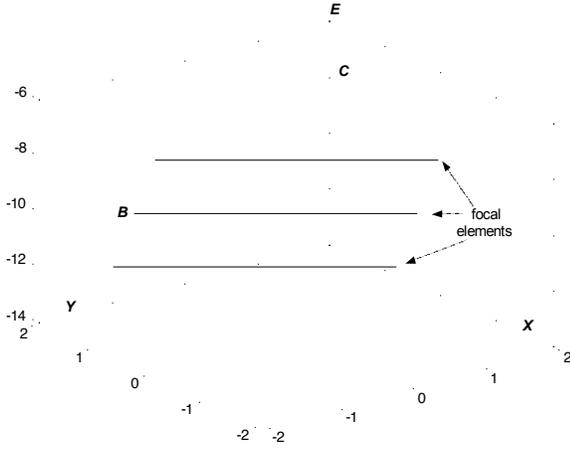


Fig. 1. An illustration of the concept of linear belief functions. The shaded area, which can be imagined to extend infinitely, is the certainty hyperplane C . Parallel lines forming a partition of C are focal elements.

A. Knowledge Representation

Based on the above geometrical description, Shafer [23] and Liu [15] propose two mathematical representations of a LBF: a wide-sense inner product and a linear functional in the variable space, and as their duals over a hyperplane in the sample space. As Liu [15] shows, the two basic operations of LBFs cannot be done using one coherent representation; complex transformations must be performed in order to interplay the two operations. Dempster [16] proposes representing a LBF indirectly by representing each of its degenerate components with a moment matrix. In this sub-section, we re-elaborate on Dempster [16] and present an elementary approach to knowledge representation. To this end, the key is to understand matrix sweeping operations [25] and to represent each of the non-trivial cases [15] using a moment matrix, or its swept form. In general, assume X is a vector of multiple normal variables with mean μ and covariance Σ . Then, the multivariate normal distribution can be equivalently represented as a moment matrix:

$$M(X) = \begin{pmatrix} \mu \\ \Sigma \end{pmatrix}. \quad (1)$$

If the distribution is non-degenerate, i.e., Σ has a full rank and its inverse exists, the moment matrix can be fully swept:

$$M(\bar{X}) = \begin{pmatrix} \mu \Sigma^{-1} \\ -\Sigma^{-1} \end{pmatrix}. \quad (2)$$

Except for normalization constant, Equation (2) completely determines the normal density function for X [see, e.g., 26]. Therefore, $M(\bar{X})$ represents the probability distribution of X in the potential form [27].

These two simple matrices allow us to represent three special cases of LBFs. First, for an ordinary normal probability distribution $M(X)$ in (1) represents it. Second, suppose one makes a direct observation on X and obtains a value μ . In this case, since there is no uncertainty, both variance and covariance vanish, i.e., $\Sigma = 0$. Thus, a direct observation can be represented as:

$$M(X) = \begin{pmatrix} \mu \\ 0 \end{pmatrix}. \quad (3)$$

Third, suppose one is completely ignorant about X . This is a very thorny case in Bayesian statistics since the density function does not exist. As a walk around, one typically has to approximate it by a prior with large-scale parameters [28]. In the Dempster-Shafer theory, one represents it by assigning the whole belief mass to the entire frame, i.e., the bpa function is $m(V) = 1$, meaning the truth is somewhere in V for sure but there is no evidence on the whereabouts exactly or probabilistically. By using the fully swept moment matrix in (2), we represent the vacuous LBF as a zero matrix in the swept form as in (4). One way to understand the representation is to imagine complete ignorance as the limiting case when the variance of X approaches to ∞ , where one can show that $\Sigma^{-1} = 0$ and hence Equation (2) vanishes into (4). Another way to understand the vacuous LBF is as the neutral element for combination:

$$M(\bar{X}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (4)$$

To represent the remaining three special cases, we need the concept of partial sweeping [25, 29]. Unlike a full sweeping, a partial sweeping is a transformation on a subset of variables. Suppose X and Y are two vectors of normal variables with the joint moment matrix:

$$M(X, Y) = \begin{bmatrix} \mu_1 & \mu_2 \\ \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}. \quad (5)$$

Then $M(X, Y)$ may be partially swept. For example, we can define the partial sweeping on X as follows:

$$M(\bar{X}, Y) = \begin{bmatrix} \mu_1(\Sigma_{11})^{-1} & \mu_2 - \mu_1(\Sigma_{11})^{-1}\Sigma_{12} \\ -(\Sigma_{11})^{-1} & (\Sigma_{11})^{-1}\Sigma_{12} \\ \Sigma_{21}(\Sigma_{11})^{-1} & \Sigma_{22} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12} \end{bmatrix}. \quad (6)$$

If X is one-dimensional, a partial sweeping replaces the variance of X by its negative inverse and multiplies the inverse with other elements. If X is multidimensional, the operation involves the inverse of the covariance matrix of X and other multiplications. A swept matrix obtained from a partial sweep-

ing on a subset of variables can be equivalently obtained by a sequence of partial sweepings on each individual variable in the subset and the order of the sequence does not matter [25]. Therefore, Equation (6) is well defined. Similarly, a fully swept matrix is the result of partial sweepings on all variables.

We can make two observations from (6). First, after the partial sweeping on X , the mean vector and covariance matrix of X are respectively $\mu_1(\Sigma_{11})^{-1}$ and $-(\Sigma_{11})^{-1}$, which are the same as that of a full sweeping of the marginal moment matrix of X . I.e., if we restrict the moment matrix $M(X, Y)$ in (5) to X indices and perform a full sweeping on X , the mean vector and covariance matrix will be also transformed into $\mu_1(\Sigma_{11})^{-1}$ and $-(\Sigma_{11})^{-1}$, respectively. Thus, the elements corresponding to X in (6) represent the marginal distribution of X in potential form. Second, according to statistics [see, e.g., 30, p. 163], $\mu_2 - \mu_1(\Sigma_{11})^{-1}\Sigma_{12}$ is the conditional mean of Y given $X = 0$; $\Sigma_{22} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12}$ is the conditional covariance matrix of Y given $X = 0$; and $(\Sigma_{11})^{-1}\Sigma_{12}$ is the slope of the regression model of Y on X . Therefore, the elements corresponding to Y indices and the intersection of X and Y in $M(\bar{X}, Y)$ represents the conditional distribution of Y given $X = 0$.

These semantics render the partial sweeping operation a useful method for manipulating multivariate normal distributions. They also form the basis of the moment matrix representations for the three remaining important cases of LBFs, including proper belief functions, linear equations, and linear regression models.

Proper Linear Belief Functions: For variables X and Y , assume there exists a piece of evidence justifying a normal distribution for variables Y while bearing no-opinions for variables X . Also, assume that X and Y are not perfectly linearly related, i.e., their correlation is less than 1. This case involves a mix of an ordinary normal distribution for Y and a vacuous belief function for X . Thus, we represent it using a partially swept matrix $M(\bar{X}, Y)$ as in (7). This is how we could understand the representation. Since we are ignorant on X , we use its swept form and set $\mu_1(\Sigma_{11})^{-1} = 0$ and $-(\Sigma_{11})^{-1} = 0$. Since the correlation between X and Y is less than 1, the regression coefficient of X on Y approaches to 0 when the variance of X approaches to ∞ . Therefore, $(\Sigma_{11})^{-1}\Sigma_{12} = 0$. Similarly, one can prove that $\mu_1(\Sigma_{11})^{-1}\Sigma_{12} = 0$ and $\Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12} = 0$. Thus, Equation (6) reduces to (7).

$$M(\bar{X}, Y) = \begin{bmatrix} 0 & \mu_2 \\ 0 & 0 \\ 0 & \Sigma_{22} \end{bmatrix}. \quad (7)$$

Linear Equations: Suppose X and Y are two row vectors, and $Y = XA + b$, where A and b are the coefficient matrices. We represent the equation using a partially swept matrix as in (8). We can understand the representation based on the fact that a linear equation contains two pieces of knowledge: (1) complete ignorance about all variables; and (2) a degenerate

conditional distribution of dependent variables given independent variables. Since X is an independent vector in the equation, we are completely ignorant about it. Thus, $\mu_1(\Sigma_{11})^{-1} = 0$ and $-(\Sigma_{11})^{-1} = 0$. Given $X = 0$, Y is completely determined to be b . Thus, the conditional mean of Y is b and the conditional variance is 0. Also, the regression coefficient matrix is A . Thus, Equation (6) is reduced to (8).

$$M(\bar{X}, Y) = \begin{bmatrix} 0 & b \\ 0 & A \\ A^T & 0 \end{bmatrix}. \quad (8)$$

Note that the knowledge to be represented in linear equations is very close to that in a proper LBF, except that the former assumes a perfect correlation between X and Y while the latter does not. This observation is interesting; it characterizes the difference between partial ignorance and linear equations in one parameter — correlation.

Linear Regression Models: A linear regression model is a more general and interesting case than previous ones. Suppose X and Y are two vectors and $Y = XA + b + E$, where A and b are the appropriate coefficient matrices and E is an independent white noise satisfying $E \sim N(0, \Sigma)$. We represent the model as the following partially swept matrix:

$$M(\bar{X}, Y) = \begin{bmatrix} 0 & b \\ 0 & A \\ A^T & \Sigma \end{bmatrix}. \quad (9)$$

This linear regression model may be considered as the combination of two pieces of knowledge (see next subsection), one is specified by the linear equation involving three variables X , Y , and E , and the other is a simple normal distribution of E , i.e., $E \sim N(0, \Sigma)$. Alternatively, one may consider it similar to a linear equation, except that, given $X = 0$, Y is not completely determined to be b . Instead, the conditional mean of Y is b while the conditional variance is Σ . Note that, in this alternative interpretation, a linear regression model forms a basic building block for knowledge representation and is encoded as one moment matrix. Besides, the noise term E does not appear in the representation. Therefore, it makes the representation more efficient.

From representing the six special cases, we see a clear advantage of the moment matrix representation, i.e., it allows a unified representation for seemingly diverse types of knowledge, including linear equations, joint and conditional distributions, and ignorance. The unification is significant not only for knowledge representation in artificial intelligence but also for statistical analysis and engineering computation. For example, the representation treats the typical logical and probabilistic components in statistics — observations, distributions, improper priors (for Bayesian statistics), and linear equation models — not as separate concepts, but as manifestations of a single concept. It allows one to see the inner connections between these concepts or manifestations and to interplay them for computational purposes.

Example 2. For the linear regression model in Example 1, it can be compactly represented as a partially swept matrix:

$$M(\vec{X}, Y) = \begin{bmatrix} 0 & 10 \\ 0 & -.8 \\ -.8 & 25 \end{bmatrix}. \quad (10)$$

Obviously this representation is more efficient than graphical and functional representations; it captures both the linear regression model and distributional assumption in one matrix. Note that, based on Equation (10), we can operate on the regression model in the same way we operate other types of LBFs. For example, we can sweep $M(\vec{X}, Y)$ further on Y as if it is a regular normal distribution and obtain a fully swept matrix:

$$M(\vec{X}, \vec{Y}) = \begin{bmatrix} 0.32 & 0.40 \\ -0.0256 & -0.032 \\ -0.032 & -0.04 \end{bmatrix}. \quad (11)$$

B. Knowledge Operations

There are two basic operations for making inferences in expert systems using LBFs: combination and marginalization. Combination corresponds to the integration of knowledge whereas marginalization corresponds to the coarsening of knowledge. Making an inference involves combining relevant knowledge into a full body of knowledge and then projecting the full body of knowledge to a partial domain, in which an inference question is to be answered.

Marginalization. Marginalization projects a LBF into one with fewer variables. Expressed as a moment matrix, it is simply the restriction of a non-swept moment matrix to a sub-matrix corresponding to the remaining variables. For example, for the joint distribution $M(X, Y)$ in (5), its marginal to Y is:

$$M^{\downarrow Y}(X, Y) = \begin{bmatrix} \mu_2 \\ \Sigma_{22} \end{bmatrix}. \quad (12)$$

When removing a variable, it is important that the variable has not been swept on in the corresponding moment matrix, i.e., it does not have an arrow sign above the variable. For example, projecting the matrix $M(\vec{X}, Y)$ in (6) to Y produces

$$M^{\downarrow Y}(\vec{X}, Y) = \begin{bmatrix} \mu_2 - \mu_1(\Sigma_{11})^{-1}\Sigma_{12} \\ \Sigma_{22} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12} \end{bmatrix}, \quad (13)$$

which is not the same LBF of Y as represented in (6). However, it is easy to see that removing any or all variables in Y from the partially swept matrix $M(\vec{X}, Y)$ will still produce the correct result—a matrix representing the same LBF for the remaining variables as in (6).

To remove a variable that has been already swept on, we have to reverse the sweeping using partial or full reverse sweepings. Assume $M(\vec{X})$ is a fully swept moment matrix,

$$M(\vec{X}) = \begin{bmatrix} \vec{\mu} \\ \vec{\Sigma} \end{bmatrix}. \quad (14)$$

Then a full reverse sweeping on $M(\vec{X})$ will recover the moment matrix $M(X)$ as follows:

$$M(X) = \begin{bmatrix} -\vec{\mu}\vec{\Sigma}^{-1} \\ -\vec{\Sigma}^{-1} \end{bmatrix}. \quad (15)$$

If a moment matrix is in a partially swept form, say $M(\vec{X}, Y)$:

$$M(\vec{X}, Y) = \begin{bmatrix} \vec{\mu}_1 & \vec{\mu}_2 \\ \vec{\Sigma}_{11} & \vec{\Sigma}_{12} \\ \vec{\Sigma}_{21} & \vec{\Sigma}_{22} \end{bmatrix}, \quad (16)$$

its partially reverse sweeping on X is defined as follows:

$$M(X, Y) = \begin{bmatrix} -\vec{\mu}_1(\vec{\Sigma}_{11})^{-1} & \vec{\mu}_2 - \vec{\mu}_1(\vec{\Sigma}_{11})^{-1}\vec{\Sigma}_{12} \\ -(\vec{\Sigma}_{11})^{-1} & -(\vec{\Sigma}_{11})^{-1}\vec{\Sigma}_{12} \\ -\vec{\Sigma}_{21}(\vec{\Sigma}_{11})^{-1} & \vec{\Sigma}_{22} - \vec{\Sigma}_{21}(\vec{\Sigma}_{11})^{-1}\vec{\Sigma}_{12} \end{bmatrix}. \quad (17)$$

Comparing Equations (15) and (17) with Equations (2) and (6) respectively, we see that reverse sweepings are similar to those of forward ones, except for a sign difference for some multiplications. However, forward and reverse sweepings are opposite operations. It can be easily shown that applying the fully reverse sweeping to $M(\vec{X})$ in (6) will recover the initial moment matrix $M(X)$ in (5). It can also be proved that applying a partial reverse sweeping on X to the matrix $M(\vec{X}, Y)$ in (6) will recover the moment matrix $M(X, Y)$ in (4). As a matter of fact, Liu [25] proves that a moment matrix will be recovered through a reverse sweeping after a forward sweeping on the same set of variables. It can be also recovered through a forward sweeping after a reverse sweeping. Intuitively, a partial forward sweeping factorizes a joint into a marginal and a conditional, whereas a partial reverse sweeping multiplies them into a joint.

Combination. According to Dempster's rule, the combination of belief functions may be expressed as the intersection of focal elements and the multiplication of probability density functions. Liu [15] applies the rule to LBFs in particular and obtains a formula of combination in terms of density functions. Later he proves a claim by Dempster [16] and re-expresses the formula as the sum of two fully swept matrices [25]. Mathematically, assume $M_1(\vec{X})$ and $M_2(\vec{X})$ are two LBFs for the same vector of variables X :

$$M_1(\vec{X}) = \begin{bmatrix} \vec{\mu}_1 \\ \vec{\Sigma}_1 \end{bmatrix}, \quad M_2(\vec{X}) = \begin{bmatrix} \vec{\mu}_2 \\ \vec{\Sigma}_2 \end{bmatrix}. \quad (18)$$

Then their combination is a fully swept matrix:

$$M(\vec{X}) = \begin{bmatrix} \vec{\mu}_1 + \vec{\mu}_2 \\ \vec{\Sigma}_1 + \vec{\Sigma}_2 \end{bmatrix}. \quad (19)$$

Equation (19) is often used for multiplying two normal distributions [e.g., see 26]. Here we use it to define the combination of two LBFs, which include normal distributions as a special case. Also, note that a vacuous LBF (see Equation (4)) is the neutral element for combination. When applying Equation (19), we need to consider two special cases. First, if two matrices to be combined have different dimensions, then one or both matrices must be vacuously extended, i.e., assuming ignorance on the variables that are not present in each matrix. For example, if $M_1(X, Y)$ and $M_2(X, Z)$ are to be com-

bined, we will first extend them into $M_1(X, Y, \bar{Z})$ and $M_2(X, \bar{Y}, Z)$ respectively such that $M_1(X, Y, \bar{Z})$ is ignorant about Z and $M_2(X, \bar{Y}, Z)$ is ignorant about Y . The vacuous extension operation was initially proposed by Kong [31] for discrete belief functions. Second, if a variable has zero variance, it will not permit a sweeping operation. In this case, we can pretend the variance to be an extremely small number, say ϵ , and perform the desired sweeping and combination. We can then apply a reverse sweeping to the combined matrix on the same variable and let $\epsilon \rightarrow 0$. Since zero variance means complete certainty about a variable, this ϵ -procedure will vanish ϵ terms in the final result.

In general, to combine two LBFs, their moment matrices must be fully swept. However, one may combine a fully swept matrix with a partially swept one directly if the variables of the former matrix have been all swept on in the later. We can use the linear regression model, $Y = XA + b + E$, to illustrate the property. As we mentioned, the regression model may be considered as the combination of two pieces of knowledge: one is specified by the linear equation involving three variables X , Y , and E , and the other is a simple normal distribution of E , i.e., $E \sim N(0, \Sigma)$. Let $M_1(\bar{X}, \bar{E}, Y)$ and $M_2(\bar{E})$ be their moment matrices respectively:

$$M_1(\bar{X}, \bar{E}, Y) = \begin{bmatrix} 0 & 0 & b \\ 0 & 0 & A \\ 0 & 0 & I \\ A^T & I & 0 \end{bmatrix}, \quad (20)$$

$$M_2(\bar{E}) = \begin{bmatrix} 0 \\ -\Sigma^{-1} \end{bmatrix}. \quad (21)$$

Then the two matrices can be combined directly without sweeping $M_1(\bar{X}, \bar{E}, Y)$ on Y first. The result of the combination is a partially swept matrix as follows:

$$M(\bar{X}, \bar{E}, Y) = \begin{bmatrix} 0 & 0 & b \\ 0 & 0 & A \\ 0 & -\Sigma^{-1} & I \\ A^T & I & 0 \end{bmatrix}. \quad (22)$$

If we apply a reverse sweeping on E and then remove E from the matrix in Equation (22), we will obtain the same representation of the regression model as Equation (9).

Example 3. Assume that, besides the linear regression model in Example 1, there is independent evidence saying that $X \sim N(2, 4)$, whose moment matrix is $\begin{pmatrix} 2 & 4 \end{pmatrix}^T$. Through a vacuous extension, we obtain an extended matrix in partially swept form:

$$M(X, \bar{Y}) = \begin{bmatrix} 2 & 0 \\ 4 & 0 \\ 0 & 0 \end{bmatrix}. \quad (23)$$

Fully sweeping the above matrix on X results in:

$$M(\bar{X}, \bar{Y}) = \begin{bmatrix} 0.5 & 0 \\ -0.25 & 0 \\ 0 & 0 \end{bmatrix}. \quad (24)$$

Then we can combine this piece of evidence with the linear regression model in Example 1 by adding the matrices in (11) and (24) and obtain a combined LBF in its potential form:

$$M(\bar{X}, \bar{Y}) = \begin{bmatrix} 0.82 & 0.40 \\ -0.2726 & -0.032 \\ -0.032 & -0.04 \end{bmatrix}. \quad (25)$$

By performing reverse sweepings on both X and Y , we obtain the moment matrix as follows:

$$M(X, Y) = \begin{bmatrix} 2 & 8.4 \\ 4 & -3.2 \\ -3.2 & 27.56 \end{bmatrix}. \quad (26)$$

The result implies that, by integrating the new body of knowledge on X with the linear regression model, we now have more precise knowledge on Y : it is in a negative relationship with X due to covariance -3.2 ; in addition, it is on average 8.4 with standard deviation 5.25 . In contrast, the linear regression model does not tell us anything except that Y is negatively related to X .

Note that $N(2, 4)$ is a marginal distribution of X and a linear regression model represents a conditional distribution of Y given X . Therefore, their combination is actually the familiar operation of multiplying a marginal with a conditional to produce a joint distribution. In fact, the simple formula in Equation (19) unifies many seemingly diverse operations, including Bayesian conditioning, solving linear equations, multiplying probability distributions, and, most importantly, integrating independent belief functions. For example, combining linear equations corresponds to logically solving the equations. For another example, statistical inference on linear models can be reduced to the one of combining LBFs [16, 32]. For yet another example, the Kalman filter results from judging all the components, such as state equations and observation equations, to be independent LBFs, and combining them into a single belief function according to Equation (9) [16].

III. PORTFOLIO EVALUATION

In this section, we use linear belief functions to construct a simple portfolio model consisting of three stocks. Each of the stocks is a gold-mining stock and can be affected by changes in the stock market, the gold price, or other unspecified factors. This example is small enough to illustrate the computations, yet it includes the features of a large class of portfolios. It builds on the traditional framework of finance models such as the CAPM. In a multi-factor model the return on a stock is represented as a regression model:

$$r_i = \alpha + \beta_1 f_1 + \beta_2 f_2 + \dots + \beta_k f_k + \epsilon_i \quad (27)$$

where r_i is the return on stock i , f_k is the return on factor k ,

β_k is the responsiveness of the stock i to factor k , and ϵ_i is a random component of the return due to firm specific effects. A multifactor model makes the following typical assumptions

[12]: (1) the factors f_1 to f_k are uncorrelated; (2) the expected value of the firm specific component, $E(\varepsilon_i)$, is equal to zero; (3) there is no correlation between ε_i and any factor, $f_1 \dots f_k$; and (4) the correlation between the firm specific effects for any firms i and j is equal to zero, i.e., $\rho(\varepsilon_i, \varepsilon_j) = 0$.

The multifactor model in (27) is a conditional probability model, where the distribution of the stock return is conditioned on the factor returns and the firm specific effects. Based on these observation, Shenoy and Shenoy [21] discretize returns as well as factors and represent the model as a Bayesian net model. Later they employ discrete belief functions to represent the Bayesian structure [22]. However, the multifactor model is much better represented as a linear belief function and encoded as a partially swept matrix like Equations (9) and (10).

A. A Valuation Network for Asset Pricing

Let us consider a portfolio consisting of three gold-mining stocks—Stock 1 (S_1), Stock 2 (S_2), and Stock 3 (S_3). Assume each stock is influenced by three factors: the stock market (M), the price of gold (G), and other firm specific factors. We denote the firm specific factors for the three firms as F_1 , F_2 , and F_3 , respectively. The firm specific factors include anything that affects the firm but is not already included in the other factors. Finally, we are interested in the return distribution of the portfolio (P).

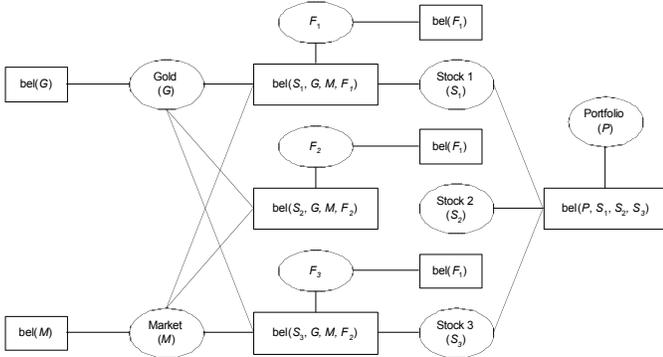


Fig. 2. A valuation network for a portfolio of three gold mining stocks. The elliptical nodes represent variables and the rectangular nodes represent linear belief functions. The edges between a rectangular node and elliptical nodes denote the domain of the belief function.

Figure 2 shows a graphical structure called *valuation network* for modeling the portfolio performance. A valuation network is a bi-partite graph in which edges connect only between variable nodes and belief function nodes [33, 34]. In our example, there are nine variables represented by nine elliptical nodes. There are nine belief functions represented by nine rectangular nodes. Among them, four LBFs capture structural relations among the variables and the other five capture beliefs respectively on gold price, market level, and each of the three firm-specific factors.

Traditional asset pricing models dictate the overall graphical structure and the domain of each structural belief function.

The domain of $bel(S_1, G, M, F_1)$ consists of four variables: S_1 , F_1 , G , and M . It intends to represent how the return of Stock 1 is related to the values of gold, market, and F_1 . It may capture the capital asset pricing model, which describes individual stock return as a conditional expectation of the market return and a residual term, or a multifactor model, which stipulates a linear regression model between S_1 and its ensuing factors G , M and F_1 . Other functions, $bel(S_2, G, M, F_2)$ and $bel(S_3, G, M, F_3)$, have a similar interpretation. The function $bel(P, S_1, S_2, S_3)$ bears on the variables P , S_1 , S_2 , and S_3 . It shows how the portfolio return is related to the individual stock returns. Since a portfolio is a linear blend of individual assets [2], portfolio return is deterministically a weighted average of stock returns. Thus, $bel(P, S_1, S_2, S_3)$ could be a linear equation that links P with its constituent stocks: S_1 , S_2 , and S_3 . In sum, all the four structural belief functions in Figure 2 are special cases of a LBF and can be represented as partially swept moment matrices as we have shown in Section 2.

A LBF on an individual factor, such as $bel(G)$, $bel(M)$, and $bel(F_k)$ ($k = 1, 2, 3$), represents a belief on the true value of the factor. It takes one of many alternative forms. It could represent a distributional assumption as we typically make about residual variables. It could represent a subjective belief that, for example, the gold price is 34 dollars on average with standard deviation 2 dollars. It could represent an observation that the factor takes on a certain value. It could represent an empirical distribution obtained from historical data. It could also be a vacuous belief function, meaning that we are completely ignorant about the factor. Regardless which form it eventually takes, it is a LBF and can be represented as a moment matrix.

It is important to note that the number of belief functions depends on the number of available pieces of knowledge regarding the variables. Therefore, there can be more or less than nine belief functions for our simple portfolio example. First, in the extreme case when there is no knowledge on any variable, there will be no belief function. If necessary, we can represent such a case using a vacuous belief function, whose domain consists all the variables involved. Second, in some reasonable situations, we may have multiple pieces of knowledge regarding one set of variables. For example, we may have two independent sources of information both bearing knowledge on gold price (G). Then, we will represent the knowledge using two rectangular nodes, instead of one as in Figure 2, connecting to the elliptical node G . Third, there may also exist knowledge that bears on a set of variables that are not connected by any belief function node in Figure 2. For example, there may be a piece of evidence justifying a relationship between gold price (G) and market return (M). Note that a multifactor model typically assumes the independence of G and M for validating related statistical techniques. Such assumptions do not have to be held in our linear belief function framework. Instead, what is critical to our approach is that

each belief function entering into the network is based on a distinct piece of evidence or has independent stochastic disturbance.

B. Specifying Linear Belief Functions

To be consistent with traditional asset pricing models such as the arbitrage pricing theory [11], we measure the variables as rates of return or change. Thus, P , S_1 , S_2 , and S_3 are respectively the rate of return of the portfolio and each of the individual gold-mining stocks. M and G are the rates of return of the overall stock market and the gold sector. F_1 , F_2 , and F_3 , are the noise terms associated with individual firm specific factors and are also measured as rates of change.

Let us first specify $bel(G)$, $bel(M)$, $bel(F_1)$, $bel(F_2)$, and $bel(F_3)$. Each of these functions has a single variable in its domain. Their role is similar to the prior distributions in a Bayesian probability model. However, unlike a Bayesian model, if we do not have any information regarding a variable, we can still use a vacuous belief function, instead of a non-informative prior, to represent the situation.

We illustrate the construction of $bel(G)$ with an example.

One kind of information that frequently affects gold prices is news that a central bank is selling a large amount of gold. Based on historical data or personal experience, assume that the volume of the transaction could negatively impact the gold price by 5% on the average. However, the actual rate of change could vary with standard deviation 2%. We represent this piece of information as moment matrix:

$$M(G) = \begin{pmatrix} -0.05 \\ 0.02^2 \end{pmatrix}. \quad (28)$$

In addition, let us assume there is breaking news that, for example, China is joining the WTO. In this case, no body understands exactly the impact of the news because there are no historical data. However, it sparks many speculations. One of them suggests that China's entry to WTO could boost the stock market by 10% on the average with a wide spread of 8%, meaning there is still 11% of probability that the news had negative impact on the stock market. We can represent the speculation as the following moment matrix:

$$M(M) = \begin{pmatrix} 0.10 \\ 0.08^2 \end{pmatrix}. \quad (29)$$

$bel(F_1)$, $bel(S_2)$, and $bel(S_3)$, can be similarly specified.

For now, we will assume that they are all vacuous. When we receive evidence about these firm specific factors, we can update these beliefs accordingly. Therefore, their current moment matrices in fully swept form are as follows:

$$M(\vec{F}_k) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \text{ for } k = 1, 2, 3. \quad (30)$$

Next let us specify $bel(S_1, G, M, F_1)$, $bel(S_2, G, M, F_2)$ and $bel(S_3, G, M, F_3)$, which describe how each of the three gold stocks is related to the gold, market and the corresponding firm specific factor. Suppose we base on historical information and perform linear regressions according to the multi-

factor model in (27). The regression model of Stock k given the return on the market (M) and the return on gold (G) is given by $S_k = \alpha_k + \beta_{1k}G + \beta_{2k}M + F_k$, where F_k is the firm specific residual term, $k = 1, 2, 3$. Assume Table 1 summarizes the regression result. According to the table, the estimated regression model for Stock 1 is $S_1 = 0.03 + 0.6G + 0.4M$ with a standard error 0.08. It means that, given $G = 0$ and $M = 0$, the average return of Stock 1 is 0.03. The standard error term suggests that actual rate of return varies with a standard deviation 8%, i.e., $F_1 \sim N(0, 0.08^2)$. Therefore, $bel(S_1, G, M, F_1)$ can be represented as follows:

$$M(S_1, \vec{G}, \vec{M}, \vec{F}_1) = \begin{bmatrix} 0.03 & 0 & 0 & 0 \\ 0 & 0.60 & 0.40 & 1.0 \\ 0.60 & 0 & 0 & 0 \\ 0.40 & 0 & 0 & 0 \\ 1.0 & 0 & 0 & -0.08^{-2} \end{bmatrix}. \quad (31)$$

TABLE 1.
MULTIPLE REGRESSION RESULT

Stock	α_k	β_{1k}	β_{2k}	Standard Error
1	0.03	0.60	0.40	0.08
2	0.03	0.45	0.25	0.04
3	0.03	0.50	0.30	0.05

Note that here we employed Equation (22) to represent a linear regression. There are two cases in which we may want to make residual terms explicit: (1) when one currently has non-vacuous knowledge about them; and (2) when one wants to have the flexibility to enter additional information on the residuals later. In our example, the valuation network (see Figure 2) makes all firm specific factors explicit and thus our matrix representation corresponds to it. The other two regression models can be similarly represented as follows:

$$M(S_2, \vec{G}, \vec{M}, \vec{F}_2) = \begin{bmatrix} 0.03 & 0 & 0 & 0 \\ 0 & 0.45 & 0.25 & 1.0 \\ 0.45 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 \\ 1.0 & 0 & 0 & -0.04^{-2} \end{bmatrix} \quad (32)$$

$$M(S_3, \vec{G}, \vec{M}, \vec{F}_3) = \begin{bmatrix} 0.03 & 0 & 0 & 0 \\ 0 & 0.50 & 0.30 & 1.0 \\ 0.50 & 0 & 0 & 0 \\ 0.30 & 0 & 0 & 0 \\ 1.0 & 0 & 0 & -0.05^{-2} \end{bmatrix} \quad (33)$$

Finally, we describe the function $bel(P, S_1, S_2, S_3)$ that describes the relationship between the three stocks and the portfolio. This relationship is a deterministic one depending on the number of shares of each stock that constitutes the portfolio. Assume the portfolio consists of 20% of Stock 1, 70% of Stock 2, and 10% of Stock 3. Then, P is simply the weighted average of the returns of the three stocks, i.e.,

$P = 0.2S_1 + 0.7S_2 + 0.1S_3$. This equation can be represented as the following partially swept matrix:

$$M(P, \bar{S}_1, \bar{S}_2, \bar{S}_3) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.2 & 0.7 & 0.1 \\ 0.2 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 \\ 0.1 & 0 & 0 & 0 \end{bmatrix}. \quad (34)$$

C. Making Inferences

Once we have represented all the knowledge related to the stocks, we can make inferences about each one and the overall portfolio as well by first combining the knowledge and then projecting the full body of knowledge to a particular variable of interest. This process translates into combining LBFs and marginalizing the joint LBF into a subset of variables. In our portfolio model, we are interested in computing the marginal of the joint belief function for the portfolio variable P and for each individual stock variables, S_1 , S_2 , and S_3 .

When there are a large number of variables, it may be more efficient to compute the marginal of the joint without explicitly computing the joint using local computation [35]. Liu [25] shows that the combination and marginalization of LBFs satisfies the three axioms of Shenoy and Shafer [36] and therefore justifies the feasibility of doing local computations over LBFs. By following the Shafer-Shenoy architecture [37], he also shows a detailed implementation including the construction of a join-tree, the propagation of messages across the join-tree, and computation of marginals by gathering the messages. In this subsection we follow the same architecture to compute the marginals of interest.

Step 1. Combine $M(S_k, \bar{G}, \bar{M}, \bar{F}_k)$ with the corresponding $M(\bar{F}_k)$ in Equation (30), where $k = 1, 2, 3$. Then apply a reverse sweeping on \bar{F}_k and remove the residual variable. The resulting matrices are shown in Table 2.

TABLE 2
PARTIAL SWEEP MATRICES AFTER REMOVING \bar{F}_k

$M(S_1, \bar{G}, \bar{M})$			$M(S_2, \bar{G}, \bar{M})$			$M(S_3, \bar{G}, \bar{M})$		
.03	0	0	.03	0	0	.03	0	0
.0064	.6	.4	.0016	.45	.25	.0025	.5	.3
.6	0	0	.45	0	0	.5	0	0
.4	0	0	.25	0	0	.3	0	0

Step 2. In order to combine the matrices in Table 2 as well as the belief functions in Equations (28–29), we apply a forward sweeping to $M(S_k, \bar{G}, \bar{M})$ on S_k to make it fully swept, where $k = 1, 2, 3$. The results are shown in Table 3.

TABLE 3
SWEEPING OF $M(S_k, \bar{G}, \bar{M})$ ON S_k

$M(\bar{S}_1, \bar{G}, \bar{M})$			$M(\bar{S}_2, \bar{G}, \bar{M})$		
4.6875	-2.8125	-1.875	18.75	-8.4375	-4.6875
-156.25	93.75	62.5	-625	281.25	156.25
93.75	-56.25	-37.5	281.25	-126.563	-70.3125
62.5	-37.5	-25	156.25	-70.3125	-39.0625
$M(\bar{S}_3, \bar{G}, \bar{M})$					
12	-6	-3.6			
-400	200	120			
200	-100	-60			

Step 3. Sweep the moment matrices in Equations (28–29) and then combine them with all the fully swept matrices in Table 3 using the formula in Equation (19). The result is a fully swept matrix of S_1 , S_2 , S_3 , G , and M as shown in Table 4.

TABLE 4
FULLY SWEEP MOMENT MATRIX $M(\bar{S}_1, \bar{S}_2, \bar{S}_3, \bar{G}, \bar{M})$

4.6875	18.7500	12.0000	-142.2500	5.4625
-156.2500	0.0000	0.0000	93.7500	62.5000
0.0000	-625.0000	0.0000	281.2500	156.2500
0.000	0.000	-400.000	200.000	120.0000
93.7500	281.2500	200.000	-2782.8125	-167.813
62.5000	156.2500	120.0000	-167.8125	-256.313

Step 4. To combine the result with the belief function $bel(P, S_1, S_2, S_3)$ in Equation (34), we just need a marginal for S_1 , S_2 , and S_3 from the combined belief function $M(\bar{S}_1, \bar{S}_2, \bar{S}_3, \bar{G}, \bar{M})$. Therefore, instead of combining them directly, we can reduce $M(\bar{S}_1, \bar{S}_2, \bar{S}_3, \bar{G}, \bar{M})$ by removing the variables G and M first. To this end, we apply reverse sweepings on G and M to the matrix $M(\bar{S}_1, \bar{S}_2, \bar{S}_3, \bar{G}, \bar{M})$ in Table 4. The result is a partially swept matrix $M(\bar{S}_1, \bar{S}_2, \bar{S}_3, G, M)$ as shown in Table 5. Then we marginalize the result to the variables S_1 , S_2 , and S_3 by removing G and M . The marginal in the swept form is shown in Table 6.

TABLE 5
PARTIALLY SWEEP MOMENT MATRIX $M(\bar{S}_1, \bar{S}_2, \bar{S}_3, G, M)$

3.1373	12.3171	7.9324	-0.0546	0.0570
-139.9656	41.6374	31.6613	0.0198	0.2309
41.6374	-517.7683	81.2827	0.0669	0.5658
31.6613	81.2827	-338.3018	0.0454	0.4384
0.0198	0.0669	0.0454	0.0004	-0.0002
0.2309	0.5658	0.4384	-0.0002	0.0041

TABLE 6
FULLY SWEEP MOMENT MATRIX $M(\bar{S}_1, \bar{S}_2, \bar{S}_3)$

3.1373	12.3171	7.9324
-139.9656	41.6374	31.6613
41.6374	-517.7683	81.2827
31.6613	81.2827	-338.3018

Step 5. Since the variables S_1 , S_2 , and S_3 have all been swept on in the matrix in Equation (34), we can directly add the matrix in Table 6 with the matrix in Equation (34). The result is a partially swept matrix for P , S_1 , S_2 , and S_3 as shown in Table 7.

TABLE 7

PARTIALLY SWEEPED MOMENT MATRIX $M(P, \bar{S}_1, \bar{S}_2, \bar{S}_3)$

0.0000	3.1373	12.3171	7.9324
0.0000	0.2000	0.7000	0.1000
0.2000	-139.9656	41.6374	31.6613
0.7000	41.6374	-517.7683	81.2827
0.1000	31.6613	81.2827	-338.3018

Step 6. Finally, to obtain the marginals for each individual stock and the portfolio as a whole, we need to apply reverse sweepings on S_1 , S_2 , and S_3 to the matrix $M(P, \bar{S}_1, \bar{S}_2, \bar{S}_3)$. The result is shown in Table 8.

TABLE 8

MOMENT MATRIX $M(P, S_1, S_2, S_3)$

0.0343	0.0400	0.0325	0.0350
0.0017	0.0021	0.0017	0.0009
0.0021	0.0076	0.0007	0.0009
0.0017	0.0007	0.0021	0.0006
0.0009	0.0009	0.0006	0.0032

As shown in Table 8, the marginal distributions for Stocks 1, 2, and 3 are respectively $N(0.04, 0.087^2)$, $N(0.033, 0.046^2)$, and $N(0.035, 0.057^2)$. Similarly, the portfolio has a return distribution $N(0.034, 0.041^2)$. The marginals are similar to Bayesian posterior distributions. According to the marginals, the prediction is as follows: the rate of return for Stock 1 is 4% with standard deviation 8.7%, for Stock 2 is 3.3% with a standard deviation 4.6%, for Stock 3 is 3.5% with standard deviation 5.7%, and for the overall portfolio is 3.4% with standard deviation 4.1%. This prediction is made based on all the knowledge available, including the news on gold sale, China's entry to WTO, ignorance about firm-specific factors, empirical multifactor models, and the structure of the portfolio.

Note that, in our example, we assumed the weights of the portfolio to be fixed. Actually, based on the mean and covariance matrix of the three stocks shown in Table 8, one may also compute an optimal portfolio, i.e., the optimal allocation of wealth on the three stocks. Assume a risk-averse mean-variance investor attempts to maximize Sharpe ratio [38]. Then the optimal weights of the three stocks will be 0.13, 0.53, and 0.34 respectively.

D. Dynamic Belief Updating

Besides being a device for integrating evidence from independent sources, the linear belief function approach may be also used to dynamically update predictions on asset performance when additional information becomes available. For example, consider a report that Asian jewelry makers are entering the gold market to stockpile inventory. Considering the size of these jewelry makers and the amount of gold they will stockpile, one estimates that the gold price will rise at a rate between 3~5%. For simplicity, let us assume 3~5% is the 95% confidence interval for the actual gold price. Then we can present the information as a LBF as

$$M(G) = \begin{pmatrix} 0.04 \\ 0.005^2 \end{pmatrix}. \quad (35)$$

To integrate this piece of evidence, we can simply sweep the matrix in Equation (35) and add the swept mean (1600) to

cell (1, 4) and the swept variance (-40000) to cell (5, 4) in Table 4. Then we continue Steps 4–6 to compute the impact of the new evidence on other variables of interest. The final result, after integrating the new evidence into other knowledge as represented in Equations (28–34), is an updated LBF for S_1 , S_2 , S_3 , and P as shown in Table 9. As the table shows, the updated marginal distributions for Stocks 1, 2, and 3 are respectively $N(0.091, 0.086^2)$, $N(0.071, 0.045^2)$, and $N(0.077, 0.056^2)$, and for the overall portfolio is $N(0.075, 0.04^2)$. As expected, comparing with the result in Table 8, we see the new evidence significantly boost the predicted rates of return of all stocks and the portfolio. It also slightly reduced the standard deviation of their distributions, meaning the prediction becomes less uncertain. By using the new mean and covariance matrix, we compute the weights in an optimal portfolio of the three gold stocks to be 0.14, 0.52, and 0.34 respectively.

TABLE 9

UPDATED MOMENT MATRIX $M(P, S_1, S_2, S_3)$

0.0753	0.0908	0.0706	0.0774
0.0016	0.0020	0.0016	0.0008
0.0020	0.0074	0.0006	0.0008
0.0016	0.0006	0.0020	0.0005
0.0008	0.0008	0.0005	0.0031

IV. CONCLUSION

In this paper, we propose a new knowledge-based approach to evaluating portfolio performance in a framework of linear belief functions. We revisited the concept of linear belief function—a new extension to the Dempster-Shafer theory of belief functions. Unlike existing studies that build the concept on the foundation of linear functional spaces [15, 23], this paper attempts to simplify the concept and make it more accessible to financial analysts and researchers. Based on the simple idea of using a moment matrix to represent a multivariate distribution, we defined full and partial sweeping operations over matrices. We explained the semantics of swept matrices and then, based on the semantics, we re-elaborated the Dempster's approach to knowledge representation using moment matrices. As we have shown, using a unified representation device—fully or partially swept moment matrices, we can intuitively represent many types of knowledge, including normal distributions, statistical observations, ignorance, linear equations, and linear regression models, in knowledge-based systems. This new tool bridges the two abstract representations respectively in vector and sample spaces [15, 23]. It allows a direct representation of knowledge in its natural form.

The new representation also simplifies the rules of marginalization and combination of linear belief functions, which are required for making inferences in a knowledge-based system. As we have shown, the marginalization of a linear belief function is simply the project of a partially swept matrix by removing a variable and its corresponding elements from the matrix, if the variable has not been swept on. The combination of two linear belief functions is simply the addition of their fully swept moment matrices, or a partially swept matrix with a fully swept one, if the variables of the later have all been swept on in the former. These rules are consistent with the

Dempster's rule of combination for discrete belief functions [17] and Bayesian rule of conditioning for probabilities.

We then showed how to encode market information and financial knowledge using linear belief functions. We used a portfolio of three gold stocks to illustrate the approach. Each stock is affected by three factors—the overall stock market, the gold market, and firm specific effects. Changes in any of these factors may affect the return of the stock. We divided the task into two parts. The first part focused on the structural relationships between economic variables. In finance theory, they are well described by the capital asset pricing model, the arbitrage pricing theory, and other asset pricing models that have been built on these foundations. It also includes any deterministic relationships among variables of interest. For example, a portfolio is linear blend of constituent stocks, dictating a linear equation between portfolio return and the returns of individual stocks. We used the structured relationships to develop a valuation network, which provides control architecture on how existing evidence may be integrated and additional evidence may be propagated throughout the network. The second part of knowledge representation is to encode factor specific knowledge based on current economic events and their resulting speculations. It is similar to specifying Bayesian priors or the Dempster-Shafer belief functions.

We showed how to make inferences using a valuation-based system of linear belief functions. We illustrated the steps of the local computation algorithm [25]. We also showed how to update the distribution by entering additional information into the system. The output of the computation is a marginal belief function for the portfolio and stock variables and, if desirable, investment decisions based on the marginal. For example, we calculated the weights of an optimal portfolio of the three stocks by maximizing the Sharpe ratio.

Our approach to portfolio evaluation has many advantages over a Bayesian approach. First, it represents all knowledge directly and naturally in a unified representation — moment matrices. For example, it represents a multifactor model naturally as a partially swept matrix, which contains the primitive model information including distributional assumptions. In contrast, a Bayesian approach has to translate each piece of market evidence into a prior on a pricing model [8, 19, 20]. It also represents different information using different devices such as priors, posteriors, observations, and equations. For example, a Bayesian net model requires conceiving and imposing directional structure, in which earlier components of uncertain knowledge are judged asymmetrically prior to later components that are described as conditional probabilities. Second, our approach allows multiple views, coherent or contradicting, expressed on the same variable or same set of variables. It allows multiple asset pricing models for different assets or a same asset to be integrated with other knowledge. Third, our approach can properly represent ignorance without using improper or non-informative priors, which Fisher [39] regarded as “completely bogus” and Shafer [17] considered confusing lack of belief with disbelief.

Our new approach is also superior to the ones proposed in Shenoy and Shenoy [21, 22]. Among many reasons, the most important is that it does not require discretizing continuous

variables and representing linear equation or regression models using subsets of ordered pairs of values. Note that the discretization procedure not only introduces approximation errors to knowledge representation but also makes computation complex and the modeling of a large-size portfolio intractable. For example, assume a portfolio consists of 500 individual stocks and each is stochastically predicted by 5 factors. Suppose one discretizes stock returns and factors into 10 discrete values. Then, to model the performance of this portfolio, one has to cope with a sample space of 10^{1005} values. By using the techniques of local computation, the largest sample space required for the computation of averaging stock returns still have 10^{500} values, which is still too large for current powerful workstations. In contrast, our new approach does not have this problem. By using moment matrices, a computation involving 500 variables has to deal with only 125,750 distinct values, including 500 means and 125,250 variance and covariance values. The complexity becomes manageable.

REFERENCES

- [1] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. New York, NY: John Wiley and Sons, 1959.
- [2] J. Tobin, "Liquidity preference as behavior towards risk," *Review of Economic Studies*, vol. 25, pp. 65-85, 1959.
- [3] P. A. Samuelson, "The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments," *Review of Economic Studies*, vol. 37, pp. 537-542, 1970.
- [4] L. Liu, "Approximate portfolio analysis," *European Journal of Operational Research*, vol. 119, pp. 35-49, 1999.
- [5] L. Liu, "A new foundation for the mean-variance analysis," *European Journal of Operational Research*, in press, 2003.
- [6] J. Y. Campbell, "Asset pricing at the millennium," *Journal of Finance*, vol. 55, pp. 1515-1567, 2000.
- [7] Y. Ait-Sahalia and M. W. Brandt, "Variable selection for portfolio choice," *Journal of Finance*, vol. 56, pp. 1297-1351, 2001.
- [8] L. Pástor, "Portfolio selection and asset pricing models," *Journal of Finance*, vol. 55, pp. 179-223, 2000.
- [9] W. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *Journal of Finance*, vol. 19, pp. 425-442, 1964.
- [10] J. Lintner, "The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets," *Review of Economics and Statistics*, vol. 47, pp. 13-37, 1965.
- [11] S. A. Ross, "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, vol. 13, pp. 453-475, 1976.
- [12] E. F. Fama, "Multifactor portfolio efficiency and multifactor asset pricing," *Journal of Financial and Quantitative Analysis*, vol. 31, pp. 441-465, 1996.
- [13] G. R. Jensen, J. M. Mercer, and R. R. Johnson, "Business conditions, monetary policy, and expected security returns," *Journal of Financial Economics*, vol. 40, pp. 213-237, 1996.
- [14] J. Y. Campbell, "Stock returns and the term structure," *Journal of Financial Economics*, vol. 18, pp. 373-399, 1987.
- [15] L. Liu, "A theory of Gaussian belief functions," *International Journal of Approximate Reasoning*, vol. 14, pp. 95-126, 1996.
- [16] A. P. Dempster, "Normal belief functions and the Kalman filter," in *Data Analysis from Statistical Foundations*, A. K. M. E. Saleh, ed.: Nova Science Publishers, 2001, pp. 65-84.
- [17] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press, 1976.
- [18] F. Black and R. Litterman, "Global portfolio optimization," *Financial Analysts Journal*, vol. 48, pp. 28-43, 1992.
- [19] L. Pástor and R. F. Stambaugh, "Costs of equity capital and model mispricing," *Journal of Finance*, vol. 54, pp. 67-121, 1999.
- [20] K. P. Baks, A. Metrick, and J. Wachter, "Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation," *Journal of Finance*, vol. 56, pp. 45-85, 2001.
- [21] C. Shenoy and P. P. Shenoy, "Bayesian Network Models of Portfolio Risk and Return," in *Computational Finance*, Y. S. Abu-Mostafa, B.

- LeBaron, A. W. Lo, and A. S. Weigend, Eds. Cambridge, MA: MIT Press, 1999, pp. 87-106.
- [22] C. Shenoy and P. P. Shenoy, "Modeling Financial Portfolios Using Belief Functions," in *Belief Functions in Business Decisions*, R. Srivastava and T. Mock, Eds. Physica-Verlag, Heidelberg: Springer-Verlag Company, 2002, pp. 316-332.
- [23] G. Shafer, "A note on Dempster's Gaussian belief functions," School of Business, University of Kansas, Lawrence, KS, Technical Report 1992.
- [24] P. A. Monney, *A Mathematical Theory of Arguments for Statistical Evidence*. New York, NY: Springer, 2003.
- [25] L. Liu, "Local Computation of Gaussian Belief Functions," *International Journal of Approximate Reasoning*, vol. 22, pp. 217-248, 1999.
- [26] S. L. Lauritzen, "Propagation of probabilities, means and variances in mixed graphical association models," *Journal of American Statistical Association*, vol. 87, pp. 1098--1108, 1992.
- [27] S. L. Lauritzen and D. J. Spiegelhalter, "Local computation with probabilities on graphical structures and their application to expert systems (with discussion)," *Journal of the Royal Statistical Society Series B*, vol. 50, pp. 157-224, 1988.
- [28] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley, 1973.
- [29] A. P. Dempster, *Elements of Continuous Multivariate Analysis*. Reading, MA: Addison-Wesley, 1969.
- [30] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*. New York, NY: Wiley, 1990.
- [31] A. Kong, "Multivariate belief functions and graphical models," in *Department of Statistics*. Cambridge, MA: Harvard University, 1986.
- [32] P. A. Monney, "Analyzing Linear Regression Models with Hints and Dempster-Shafer Theory," *International Journal of Intelligent Systems*, vol. 18, pp. 5-29, 2003.
- [33] P. P. Shenoy, "Valuation-Based Systems for Bayesian Decision Analysis," *Operations Research*, vol. 40, pp. 463-484, 1992.
- [34] P. P. Shenoy, "Representing conditional independence relations by valuation networks," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 2, pp. 143-165, 1994.
- [35] P. P. Shenoy and G. Shafer, "Propagating belief functions using local computations," *IEEE Expert*, vol. 1, pp. 43-52, 1986.
- [36] P. P. Shenoy and G. Shafer, "Axioms for Probability and Belief-Function Propagation," in *Uncertainty in Artificial Intelligence 4*, R. D. Shachter, T. S. Levitt, J. F. Lemmer, and L. N. Kanal, Eds. North-Holland, 1990, pp. 169-198.
- [37] G. Shafer, *Probabilistic Expert Systems*. Philadelphia, PA: SIAM, 1996.
- [38] M. R. Gibbons, S. A. Ross, and J. Shanken, "A test of efficiency of a given portfolio," *Econometrica*, vol. 57, pp. 1221-1152, 1989.
- [39] R. A. Fisher, "Mathematical probability in the natural sciences," *Technometrics*, vol. 1, pp. 21-29, 1959.
- [40] C. R. Rao, *Linear Statistical Inference and Its Applications*. New York: Wiley, second edition, 1973.
- [41] Kenley, C. R. "Influence diagram models with continuous variables." PhD dissertation, EES Department, Stanford University, 1986.



Liping Liu was born in Auhui, China in 1962. He received his BS in Applied Mathematics from Huazhong University of Science and Technology, China, in 1986, BE in River Dynamics from Wuhan University, China, 1987, Master of Engineering in Systems Engineering from Huazhong University of Science and Technology, in 1991, and his Ph.D. in Business from University of Kansas, USA, in 1995.

He is currently an Associate Professor of Management and Information Systems at the University of Akron. His research interests have been in the areas of Uncertainty Reasoning and Decision Making in Artificial Intelligence, Electronic Business, and Systems Design. His articles have appeared in *Decision Support Systems*, *European Journal of Operational Research*, *Journal of Risk and Uncertainty*, and others. He has served as a guest editor for *International Journal of Intelligent Systems*, and a co-editor for *Classic Works on the Dempster-Shafer Theory of Belief Functions*. Dr. Liu is a member of American Association for the Advancement of Science, *Association for Computing Machinery*, *Association for Information Systems*, and *Institute for Operations Research and Management Sciences*. He has won numerous teaching and research awards.

Catherine Shenoy was born in Slaton, Texas. She received a Ph.D in finance from the University of Kansas in 1992, an M.B.A. from the University of Kansas in 1984, and a B.S. in accounting from the University of Kansas in 1980. She is an Assistant Professor and Director of the Applied Portfolio Management Program at the University of Kansas since 1999. Prior to joining the University of Kansas she served the U.S. Department of State as a Foreign Service Officer in Washington D.C. and Tunisia.



Her research areas include corporate governance, corporate financial policies, and applications of artificial intelligence to portfolio analysis. She has published "Modeling Financial Portfolios using Belief Functions" in R. P. Srivastava and T. J. Mock (eds.) *Belief Functions in Business Decisions*. Physica-Verlag, 2002; "Stock Price Decreases Prior to Executive Stock Option Grants" *Journal of Corporate Finance*, 2001; and "Bayesian Network Models of Portfolio Risk and Return", in Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo, and A. S. Weigend (eds.), *Computational Finance 1999*, MIT Press, 2000. She is the recipient of the Guy O. Mabry award for the best publication in the School of Business for the paper "The Firm's Leverage/Cash-Flow Relationship" published in the *Journal of Empirical Finance*.



Prakash P. Shenoy is the Ronald G. Harper Distinguished Professor of Artificial Intelligence in Business, University of Kansas at Lawrence. He received a B.Tech. in Mechanical Engineering from the Indian Institute of Technology, Bombay, India, in 1973, and an M.S. and a Ph.D. in Operations Research from Cornell University in 1975 and 1977, respectively.

His research interests are in the areas of uncertain reasoning and decision analysis. He is the inventor of valuation-based systems, an abstract framework for knowledge representation and inference that includes Bayesian probabilities, Dempster-Shafer belief functions, Spohn's kappa calculus, Zadeh's possibility theory, propositional logic, optimization, solving systems of equations, database retrieval, and other domains.

He serves as the North-American editor of *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, and as an Associate Editor of *Management Science*, and the *International Journal of Approximate Reasoning*, and as an ad-hoc referee for over 30 journals and conferences in Artificial Intelligence and Management Science/Operations Research.