Communications in Statistics—Theory and Methods, 39: 956–972, 2010

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920902827299

Taylor & Francis
Taylor & Francis Group

Bayesian Updating with Confounded Signals

NEDA FARZINNIA[1] AND KEVIN F. McCARDLE[2]

[1]Internet Marketing Analytic and Customer Insights,
User Analytic and Display Advertising Optimization, eBay, Inc.,
San Jose, California, USA
[2]UCLA Anderson School of Management, University of California,
Los Angeles, California, USA

We develop a Bayesian framework for estimating the means of two random variables when only the sum of those random variables can be observed. Mixture models are proposed for establishing conjugacy between the joint prior distribution and the distribution for observations. Among other desirable features, conjugate distributions allow Bayesian methods to be applied in sequential decision problems.

Keywords Bayes estimates; Bayesian updating; Confounded signal; Conjugacy; Investment strategy; Mixture models.

Mathematics Subject Classification 60G40; 62C10; 62L15; 62L10.

1. Introduction

A firm controls two assets, X and Y, both of which are required as inputs in an investment. For example, a firm opening a specialty restaurant needs both a menu and location. Rather than observing the return on the individual assets (how good or bad is the menu, and how good or bad the location), the firm only observes the aggregate return on the investment (the revenues from the restaurant). The firm is trying to choose between alternative ways of scaling up the investment: open multiple locations using the same menu, expand the menu at the same location, or quit the investment altogether. (See Bernardo and Chowdhry, 2002 for a two-period example, and Farzinnia (2006) for a multi-period example.)

More generally, a firm needs to estimate the means $\mathbf{E}(X)$ and $\mathbf{E}(Y)$ of two random variables X and Y in order to make a decision regarding the random process (X, Y). It can take a series of pairs of draws from the distribution of X and Y, but it does not get to observe the individual draws. Rather, at each period j, it can only observe the sum of the draws $R_j = X_j + Y_j$. Each observation R_j is a confounded signal about the investment random process (X, Y). The firm cannot distinguish the signal from each asset, X and Y, separately and only observes the

Received July 18, 2007; Accepted February 14, 2009
Address correspondence to Neda Farzinnia, Internet Marketing Analytic and Customer Insights, User Analytic and Display Advertising Optimization, eBay, Inc., San Jose, CA 95125, USA; E-mail: nfarzinnia@ebay.com

sum of those signals. Models for uncertain data distributions based on mixtures of standard components, such as mixtures of normal distributions, allow us to use statistical inferences about the relevant model parameters. Escobar and West (1995) and Cao and West (1996) discussed the use of Bayesian inference in models for density estimation using mixture models. We develop a Bayesian approach that allows the evaluation and computation of the posterior distribution and propose two different mixture models. As a natural byproduct we make inferences about the parameters of the population and evaluate these parameter estimates. We start with prior distributions for the parameters of interest and update those distributions as observations are obtained. Given a proper prior, a Bayesian approach to the problem provides estimators which can be written explicitly. The choice of a suitable model is essential for decision making when dealing with complex structures. One way to mitigate complexity is the use of conjugacy; see, for example, DeGroot (2004), McCardle (1985), and Jovanovic et al. (1995).

We are interested in establishing conjugacy between possible joint prior distributions with distributions for the observations, in deriving formulas for the estimates, and in determining the asymptotic properties of those estimates. Formulas are derived for independent Bernoulli draws and a mixture of joint Beta distributions for the prior, and for both independent and correlated normally distributed draws and a mixture of joint normal distributions for the prior. It is worth mentioning that in the Beta-Bernoulli case, although the firm starts with independent prior distributions, once it starts observing the returns, the posterior distributions may no longer be independent. Thus, the assumption of independence in the priors is without loss of generality.

Our interest in the Bernoulli and Normal examples stems from the investment model, where the Bernoulli and Normal specifications make the most sense (for example, Bernardo and Chowdhry, 2002, employ a Bernoulli model). Other conjugate priors could be used in a similar manner and the derivation of their posterior distributions is similar to our examples. Other common hierarchical models in the Bayesian framework are mixtures of gamma distributions and log-normal/normal models. These mixture models have applications in various domains such as engineering, biology, and genetics: see, for example, Wiper et al. (2001), Labbe and Thompson (2005), Rangarajan et al. (2000), Oikonomou (1997), and Lo and Gottardo (2007).

2. Beta-Bernoulli Model

2.1. *Distribution of the Observations*

At period $j = 1, 2, 3, \ldots$, we denote the observations by $R_j = X_j + Y_j$. We assume that X_j and Y_j are drawn from independent Bernoulli distributions with parameters $p^{x^\star}$ and $p^{y^\star}$, respectively. We assume X_j and Y_j each take on one of two values: H(high) or L(low). Thus, $\mathbf{Pr}(X_j = H) = p^{x^\star} = 1 - \mathbf{Pr}(X_j = L)$ and $\mathbf{Pr}(Y_j = H) = p^{y^\star} = 1 - \mathbf{Pr}(Y_j = L)$. The observed value R_j in period j takes one of the following values:

$$R_j = X_j + Y_j = \begin{cases} 2H & \text{with probability } p^{x^\star} p^{y^\star}, \\ H + L & \text{with probability } p^{x^\star}(1 - p^{y^\star}) + p^{y^\star}(1 - p^{x^\star}), \\ 2L & \text{with probability } (1 - p^{x^\star})(1 - p^{y^\star}). \end{cases}$$

When the observation is $2H$, we know that both X and Y are high. Similarly, when the observation is $2L$, we know that both X and Y are low. However, when the observation is $H+L$, we cannot disentangle whether X is high while Y is low, or vice versa.

2.2. *Prior Distribution*

Our beliefs about the values of $p^{x^\star}$ and $p^{y^\star}$ are contained in the random variables P^x and P^y, for which we have a prior probability distribution. For tractability we assume that the firm's prior distribution and the distribution which generates the returns of the investment form a conjugate pair. We consider the case in which the prior distribution of (P^x, P^y) is a product of independent Beta distributions with parameters (α_1, β_1) and (α_2, β_2), respectively, with $\alpha_i > 0$ and $\beta_i > 0$ for $i = 1, 2$. We denote this by $(P^x, P^y) \sim Beta(\alpha_1, \beta_1) \times Beta(\alpha_2, \beta_2)$. The density $g(p, q)$ of the prior distribution of (P^x, P^y) is

$$g(p^x, p^y) = \left[\frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)}(p^x)^{\alpha_1-1}(1-p^x)^{\beta_1-1}\right]\left[\frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)}(p^y)^{\alpha_2-1}(1-p^y)^{\beta_2-1}\right],$$

where $0 < p^x, p^y < 1$. Here, $\Gamma(\delta)$ denotes the gamma function: $\Gamma(\delta) = \int_0^\infty x^{\delta-1}e^{-x}\,dx$. An intuitive interpretation of the parameters of the Beta in the prior distribution is that of imagined counts for the two outcomes, values of H and values of L, respectively.

2.3. *Posterior Distribution*

After observing the first value, R_1, we update the prior distribution in a Bayesian fashion, resulting in the posterior distribution for P^x and P^y. The posterior density $h(p^x, p^y \mid R_1)$ is obtained by Bayes theorem as follows:

$$\begin{aligned} h\left(p^x, p^y \mid R_1\right) &= \frac{f\left(R_1 \mid p^x, p^y\right) g\left(p^x, p^y\right)}{\int_{p^x}\int_{p^y} f\left(R_1 \mid p^x, p^y\right) g\left(p^x, p^y\right) dp^x\, dp^y} \\ &\propto f\left(R_1 \mid p^x, p^y\right) g\left(p^x, p^y\right), \end{aligned}$$

where $\propto$ denotes "is proportional to"; that is, the posterior distribution is proportional to the prior times the likelihood.

The observation in the first period, R_1, takes on one of the three values: $2H$, $2L$, or $H+L$. We consider the three possible values in turn and derive the posterior distribution accordingly.

The observation in the first period is $2H$ with probability $p^{x^\star}p^{y^\star}$. Using the density of a beta distribution, it immediately follows that

$$(P^x, P^y \mid R_1 = 2H) \sim Beta(\alpha_1+1, \beta_1) \times Beta(\alpha_2+1, \beta_2).$$

Similarly, if $R_1 = 2L$, which it is with probability $(1-p^{x^\star})(1-p^{y^\star})$, the posterior density of (P^x, P^y) becomes

$$(P^x, P^y \mid R_1 = 2L) \sim Beta(\alpha_1, \beta_1+1) \times Beta(\alpha_2, \beta_2+1).$$

The value of $H+L$ is observed with probability $p^{x^\star}(1-p^y)^\star + p^{y^\star}(1-p^x)^\star$. Consequently, the posterior distribution of (P^x, P^y) for an observation of $H+L$ is

$$(P^x, P^y \mid R_1 = H+L) \sim \left(\frac{\alpha_2\beta_1}{\alpha_1\beta_2+\alpha_2\beta_1}\right) Beta(\alpha_1, \beta_1+1) \times Beta(\alpha_2+1, \beta_2) + \left(\frac{\beta_2\alpha_1}{\alpha_1\beta_2+\alpha_2\beta_1}\right) Beta(\alpha_1+1, \beta_1) \times Beta(\alpha_2, \beta_2+1). \quad (1)$$

If the observation is $H+L$, the posterior distribution is a mixture of products of independent Beta distributions. Equation (1) represents this mixture and it contains two terms. The coefficient of the first term is the prior probability of a high value for X and a low value for Y, while the coefficient of the second term is the prior probability of a low value for X and a high value for Y. When the observation in the first period is $2H$ or $2L$, the posterior distribution is a product of independent beta distributions.

At stage j, the firm has observed j outcomes. Let n_j be the number of times out of j that the observation is $2H$; m_j, the number of times that the observation is $2L$; and l_j, the number of times the observation is $H+L$. Thus, $0 \le n_j, m_j, l_j \le j$ and $n_j + m_j + l_j = j$. Suppose the observed outcome is $2H$ for the first n periods. Then the posterior distribution of P^x and P^y given n observations of $2H$ is

$$(P^x, P^y \mid R_1 = R_2 = \cdots = R_n = 2H) \sim Beta(\alpha_1+n, \beta_1) \times Beta(\alpha_2+n, \beta_2). \quad (2)$$

Similarly, if the observed outcome is $2L$ for the first m periods, the posterior distribution of P^x and P^y is

$$(P^x, P^y \mid R_1 = R_2 = \cdots = R_m = 2L) \sim Beta(\alpha_1, \beta_1+m) \times Beta(\alpha_2, \beta_2+m). \quad (3)$$

Proposition 2.1 describes how the posterior distribution of P^x and P^y given l observed returns of $H+L$ is obtained. We introduce some notation. Let

$$\phi_{l,k} = \binom{l}{k} \frac{\Gamma(\alpha_1+k)}{\Gamma(\alpha_1)} \frac{\Gamma(\beta_1+l-k)}{\Gamma(\beta_1)} \frac{\Gamma(\alpha_2+l-k)}{\Gamma(\alpha_2)} \frac{\Gamma(\beta_2+k)}{\Gamma(\beta_2)} \quad k = 0, 1, 2, \ldots, l \quad (4)$$

and

$$\varphi_{l,k} = \phi_{l,k} / \sum_{i=0}^{l} \phi_{l,i}. \quad (5)$$

The first subscript in $\phi_{l,k}$ represents the total number of times that the observation is $H+L$, of which k times, the second subscript, represents the number of times that X is high. For example, $l=1$ and $k=0$ implies a low value of X and a high value of Y, while $l=1$ and $k=1$ implies a high value of X and a low value of Y.

Proposition 2.1. *If the observed outcome is $H+L$ for the first l periods, then the posterior distribution of P^x and P^y is a mixture of $l+1$ terms and is given by*

$$(P^x, P^y \mid R_1 = \cdots = R_l = H+L) \sim \sum_{k=0}^{l} \varphi_{l,k} Beta(\alpha_1+k, \beta_1+l-k) \times Beta(\alpha_2+l-k, \beta_2+k).$$

Proof. To find the posterior distribution of P^x and P^y after observing $H + L$ for l periods, recall that if the observation after one period is $H + L$, then by Eq. (1) the posterior distribution of P^x and P^y is

$$(P^x, P^y \mid R_1 = H + L) \sim \sum_{k=0}^{1} \varphi_{1,k} Beta(\alpha_1 + k, \beta_1 + 1 - k) \times Beta(\alpha_2 + 1 - k, \beta_2 + k).$$

If the observed value is $H + L$ for l periods, then there are $l + 1$ possible combinations; $k = 0$ implies l low values of X (and l high values of Y), $k = 1$ implies one high value of X (and $l - 1$ high values of Y), ..., and $k = l$ implies l high values of X (and l low values of Y). Thus, the posterior density of P^x and P^y is obtained as follows:

$$h(p^x, p^y \mid R_1, \ldots, R_l) = \frac{\sum_{k=0}^{l} \binom{l}{k} (p^x)^{\alpha_1+k-1}(1-p^x)^{\beta_1+l-k-1}(p^y)^{\alpha_2+l-k-1}(1-p^y)^{\beta_2+k-1}}{\sum_{k=0}^{l} \binom{l}{k} \Gamma(\alpha_1+k)\Gamma(\beta_1+l-k)\Gamma(\alpha_2+l-k)\Gamma(\beta_2+k)}.$$

Using $\phi_{l,k}$ defined in (4), the posterior distribution of P^x and P^y after observing an outcome of $H + L$ for l periods is

$$\begin{aligned}(P^x, P^y \mid R_1, \ldots, R_l) &\sim \sum_{k=0}^{l} \frac{\phi_{l,k} Beta(\alpha_1 + k, \beta_1 + l - k) \times Beta(\alpha_2 + l - k, \beta_2 + k)}{\sum_{i=0}^{l} \phi_{l,i}} \\ &= \sum_{k=0}^{l} \varphi_{l,k} Beta(\alpha_1 + k, \beta_1 + l - k) \times Beta(\alpha_2 + l - k, \beta_2 + k).\end{aligned}$$

The posterior distribution after l observations of $H + L$ is a mixture of products of independent Betas with parameters that correspond to the possible outcomes of X and Y over l periods.

The next theorem pulls Eqs. (2) and (3) and Proposition (2.1) together.

Theorem 2.1. *At stage j the posterior distribution of P^x and P^y given n_j observations of $2H$, m_j observations of $2L$, and l_j observations of $H + L$, where $n_j + m_j + l_j = j$, is a mixture of products of independent Betas*:

$$(P^x, P^y \mid R_1, \ldots, R_j) \sim \sum_{k=0}^{l_j} \varphi_{l_j,k} Beta(\alpha_{1j} + k, \beta_{1j} + l_j - k) \times Beta(\alpha_{2j} + l_j - k, \beta_{2j} + k),$$

where $\alpha_{1j} = \alpha_1 + n_j$, $\alpha_{2j} = \alpha_2 + n_j$, and $\beta_{1j} = \beta_1 + m_j$, $\beta_{2j} = \beta_2 + m_j$.

Proof. We observe the outcomes of the draws for j periods. Each observation is independent of the previous ones, conditioned on P^x and P^y. Because of the conditional independence, the order of the observations is not important. We can update the observations of $2H$ first, the observations of $2L$ second, and then the observations of $H + L$. First, using Eq. (2), we obtain the updated posterior distribution for the n_j observed values of $2H$

$$\begin{aligned}(P^x, P^y \mid R_1, \ldots, R_{n_j}) &\sim Beta(\alpha_1 + n_j, \beta_1) \times Beta(\alpha_2 + n_j, \beta_2) \\ &= Beta(\alpha_{1j}, \beta_1) \times Beta(\alpha_{2j}, \beta_2).\end{aligned}$$

Then using Eq. (3), we update the posterior distribution for the m_j observed values of $2L$. This updated posterior distribution is

$$\begin{aligned}(P^x, P^y \mid R_1, \ldots, R_{n_j+m_j}) &\sim Beta(\alpha_{1j}, \beta_1 + m_j) \times Beta(\alpha_{2j}, \beta_2 + m_j)\\ &= Beta(\alpha_{1j}, \beta_{1j}) \times Beta(\alpha_{2j}, \beta_{2j}).\end{aligned}$$

Finally, we update the posterior distribution for the l_j values of $H + L$. Using Proposition 2.1 the updated posterior distribution is

$$(P^x, P^y \mid R_1, \ldots, R_j) \sim \sum_{k=0}^{l_j} \varphi_{l_j,k} Beta(\alpha_{1j} + k, \beta_{1j} + l_j - k) \times Beta(\alpha_{2j} + l_j - k, \beta_{2j} + k).$$

This completes the proof. □

The posterior distribution after observing the outcomes for j periods, when there are outcomes of $2H$, $2L$, and $H + L$, is a mixture of products of independent Beta distributions.

2.4. *Posterior Means*

We start with prior distributions of P^x and P^y with means $\widehat{p^x}_0 = \frac{\alpha_1}{\alpha_1+\beta_1}$ and $\widehat{p^y}_0 = \frac{\alpha_2}{\alpha_2+\beta_2}$, respectively. Then we observe the draws from the distribution of P^x and P^y and update our beliefs upon observing the values, $R_1, \ldots, R_j$. Note that n_j, m_j, and l_j are sufficient statistics for $R_1, \ldots, R_j$: all information from the observed values, $R_1, \ldots, R_j$ is incorporated in the state variable (n_j, m_j, l_j). Because n_j, m_j, and l_j must sum to j, for a fixed j there are only two degrees of freedom. We denote the current point estimates of $p^{x\star}$ and $p^{y\star}$ by $\widehat{p^x}_j$ and $\widehat{p^y}_j$, respectively, that is, $\widehat{p^x}_j = \mathbf{E}(P^x \mid n_j, m_j, l_j)$ and $\widehat{p^y}_j = \mathbf{E}(P^y \mid n_j, m_j, l_j)$. These posterior means are obtained as follows:

$$\begin{aligned}\widehat{p^x}_j &= \mathbf{E}(P^x \mid n_j, m_j, l_j)\\ &= \int_{p^x}\int_{p^y} p^x h(p^x, p^y \mid n_j, m_j, l_j) dp^x\, dp^y\\ &= \sum_{k=0}^{l_j} \varphi_{l_j,k} \int_{p^x} p^x \frac{1}{B(\alpha_{1j} + k, \beta_{1j} + l_j - k)} (p^x)^{\alpha_{1j}+k-1} (1 - p^x)^{\beta_{1j}+l_j-k-1}\\ &= \sum_{k=0}^{l_j} \varphi_{l_j,k} \frac{\alpha_{1j} + k}{\alpha_{1j} + \beta_{1j} + l_j}.\end{aligned} \tag{6}$$

Similarly,

$$\widehat{p^y}_j = \mathbf{E}(P^y \mid n_j, m_j, l_j) = \sum_{k=0}^{l_j} \varphi_{l_j,k} \frac{\alpha_{2j} + l_j - k}{\alpha_{2j} + \beta_{2j} + l_j}. \tag{7}$$

The posterior means, $\widehat{p^x}_j$ and $\widehat{p^y}_j$, are each bounded between 0 and 1. Therefore, the state space for the posterior means is the unit square. The main diagonal of this square represents equal posterior means for P^x and P^y. The area below this diagonal implies a larger posterior mean at stage j for P^x, and the area above this diagonal implies a larger posterior mean for three possibilities for the prior means of P^x and P^y: the prior mean of P^x is less than the prior mean of P^y, $\frac{\alpha_1}{\alpha_1+\beta_1} < \frac{\alpha_2}{\alpha_2+\beta_2}$ and we start in the area above the main diagonal; the prior mean of P^x is larger than the prior mean of P^y, $\frac{\alpha_1}{\alpha_1+\beta_1} > \frac{\alpha_2}{\alpha_2+\beta_2}$ and we start in the area below the diagonal; or we start on the diagonal where the prior mean of P^x equals the prior mean of P^y. If the effective prior sample size for P^x is the same as the effective prior sample size for P^y, $\alpha_1 + \beta_1 = \alpha_2 + \beta_2$, then the posterior means remain on the same side of the diagonal as the prior means.

In short, the firm starts with a prior belief regarding the return of each one of its assets; that is, $\widehat{p^x}_0$ and $\widehat{p^y}_0$. It then observes the sum of the returns from the investment and updates its beliefs. The posterior estimates depend on the prior information as well as the returns from the investment. The closer the prior means are to each other the less sensitive the estimates will be to the observations; and the further the prior means are from each other, the more sensitive the estimates.

2.5. *Consistency of the Bayes Estimates Despite the Inconsistency of the Posterior Means*

A weakly *consistent estimator* is an estimator that converges in probability to the quantity being estimated as the sample size grows; that is, an estimator $\widehat{\theta}_j$ (where j is the sample size) is consistent for parameter θ if and only if, $\forall \epsilon > 0$, we have $\lim_{j\to\infty} \mathbf{Pr}_\theta \left\{ |\hat{\theta}_j - \theta| < \epsilon \right\} = 1$. An estimator is strongly consistent if $\hat{\theta}_j \xrightarrow{a.s.} \theta$.

Recall that R_j is $2H$ with probability $p^{x^\star} p^{y^\star}$; R_j is $H + L$ with probability $p^{x^\star} + p^{y^\star} - 2p^{x^\star} p^{y^\star}$; and R_j is $2L$ with probability $1 - (p^{x^\star} + p^{y^\star}) + p^{x^\star} p^{y^\star}$. Note that $p^{x^\star}$ and $p^{y^\star}$ are interchangeable in these formulas. The joint distribution of the observations, $R_1, R_2, \ldots, R_j$, given $p^{x^\star}$ and $p^{y^\star}$ is the same as the joint distribution of the observations given $p^{y^\star}$ and $p^{x^\star}$: $f(R_1, \ldots, R_j \mid p^{x^\star}, p^{y^\star}) = f(R_1, \ldots, R_j \mid p^{y^\star}, p^{x^\star})$. We have no way of knowing whether the true value of the parameters are $(p^{x^\star}, p^{y^\star})$ or $(p^{y^\star}, p^{x^\star})$. Therefore, the parameters $p^{x^\star}$ and $p^{y^\star}$ are not identifiable in the model. The non-identifiability of $p^{x^\star}$ and $p^{y^\star}$ is due to the assumption that we can only observe the sum of the values $(X + Y)$ not the individual values X and Y. No method of estimation gives consistent estimates under this structure.

While $\widehat{p^x}_j$ and $\widehat{p^y}_j$ are not consistent estimators of $p^{x^\star}$ and $p^{y^\star}$ individually, the Bayes estimates of the sum and product of $p^{x^\star}$ and $p^{y^\star}$ are strongly consistent; that is,

$$\widehat{p^x}_j + \widehat{p^y}_j \xrightarrow{a.s.} p^{x^\star} + p^{y^\star} \quad \text{and} \quad \widehat{p^x p^y}_j \xrightarrow{a.s.} p^{x^\star} p^{y^\star}$$

as $j \longrightarrow \infty$. Here, $\widehat{p^x p^y}_j = \mathbf{E}(P^x P^y \mid n_j, m_j, l_j)$ and is computed as follows:

$$\widehat{p^x p^y}_j = \mathbf{E}(P^x P^y \mid n_j, m_j, l_j) = \sum_{k=0}^{l_j} \varphi_{l_j,k} \left(\frac{\alpha_{1j} + k}{\alpha_{1j} + \beta_{1j} + l_j} \cdot \frac{\alpha_{2j} + l_j - k}{\alpha_{2j} + \beta_{2j} + l_j} \right). \tag{8}$$

The consistency of these Bayes estimates follows from Doob (1949), who gave general conditions, satisfied here, under which Bayes estimates are strongly consistent.

3. The Impact of Changes in $p^{x\star}$ and $p^{y\star}$

The firm's returns from the investment each period, that is, its observations, depend on $p^{x^\star}$ and $p^{y^\star}$, the true rates of return of the individual assets. The values of $p^{x^\star}$ and $p^{y^\star}$ are not known to the firm. It observes the returns and forms estimates accordingly. Because the firm's posterior estimates depend in part on its prior beliefs, it is not clear how sensitive the posterior estimates are to changes in $p^{x^\star}$ and $p^{y^\star}$. Because the Bayesian updating depends upon $(p^{x^\star}, p^{y^\star})$ in a probabilistic manner, to investigate the impact of changes in $p^{x^\star}$ and $p^{y^\star}$, we simulate the returns from the investment. Then depending on these simulated outcomes, we can find the point estimates, $\widehat{p^x}$ and $\widehat{p^y}$, using Eqs. (6) and (7).

3.1. *Simulations*

We simulate the returns from each one of the assets, X and Y, according to the assumed rates, i.e.,

$$\mathbf{Pr}(R^X = H) = p^{x^\star} \quad \text{and} \quad \mathbf{Pr}(R^Y = H) = p^{y^\star}.$$

We run 1,000 simulations; each simulation contains 50 observations (returns) from the investment. We compute the posterior estimates using the 50 observations. By so doing, we generate a set of 1,000 posterior estimates. We then compute the sample average of these estimates, which we then use as the point estimate of the simulation given $p^{x^\star}$ and $p^{y^\star}$.

To consider the impact of changes of $p^{x^\star}$, we increase $p^{x^\star}$ keeping everything else constant. We consider three alternatives:

- $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$,
- $\alpha_1 + \beta_1 = \alpha_2 + \beta_2$, but $\alpha_1 \neq \alpha_2$,
- $\alpha_1 + \beta_1 \neq \alpha_2 + \beta_2$.

3.1.1. *Simulation when $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$.* When $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, the prior beliefs regarding each of the assets are identical, hence the posterior estimates will also be identical. The results are summarized in Table 1.

The third column shows that the (equal) posterior estimates approach the average of the true values $p^{x^\star}$ and $p^{y^\star}$. The last column in the table represents the standard error of the estimates. These standard errors are all relatively small, and as

Table 1
The impacts of increasing $p^{x^\star}$ on the posterior estimates when $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, $j = 50$

$(\alpha_1, \beta_1) = (1, 2)$ $p^{x^\star}$	$(\alpha_2, \beta_2) = (1, 2)$ $p^{y^\star}$	$(\widehat{p^x}_0, \widehat{p^y}_0) = (0.3333, 0.3333)$ $\widehat{p^x}_j = \widehat{p^y}_j$	$\hat{\sigma}_{p^x{}_0} = \hat{\sigma}_{p^y{}_0} = 0.2357$ $\widehat{S}_{p^x{}_j} = \widehat{S}_{p^y{}_j}$
0.3	0.2	0.2566	0.00128
0.4	0.2	0.3010	0.00134
0.5	0.2	0.3507	0.00131
0.6	0.2	0.3950	0.00132
0.7	0.2	0.4439	0.00131

Table 2
The impacts of increasing $p^{x\star}$ on the posterior estimates when $\alpha_1 + \beta_1 = \alpha_2 + \beta_2$ and $\alpha_1 < \alpha_2$, $j = 50$

$(\alpha_1, \beta_1) = (1, 2)$	$(\alpha_2, \beta_2) = (2, 1)$	$(\widehat{p^x}_0, \widehat{p^y}_0) = (0.3333, 0.6667)$		$\hat{\sigma}_{p^x_0} = \hat{\sigma}_{p^y_0} = .2357$	
$p^{x\star}$	$p^{y\star}$	$\widehat{p^x}_j$	$\widehat{p^y}_j$	$\widehat{\mathbf{S}}_{p^x_j}$	$\widehat{\mathbf{S}}_{p^y_j}$
0.3	0.2	0.1993	0.3346	0.00163	0.00179
0.4	0.2	0.2336	0.3935	0.00178	0.00196
0.5	0.2	0.2566	0.4590	0.00202	0.00223
0.6	0.2	0.2759	0.5354	0.00234	0.00256
0.7	0.2	0.2798	0.6264	0.00254	0.00268

$p^{x^\star}$ increases the standard errors remain almost unchanged. The results represented in the table are unchanged if the role of $p^{x^\star}$ and $p^{y^\star}$ are interchanged.

3.1.2. *Simulation when* $\alpha_1 + \beta_1 = \alpha_2 + \beta_2$ *but* $\alpha_1 \neq \alpha_2$. If the firm's prior sample sizes for each of the asset returns are the same, but the means are different, then either the mean of P^x is less than mean of P^y or vice versa. Consider the case where $\alpha_1 + \beta_1 = \alpha_2 + \beta_2$ and $\alpha_1 < \alpha_2$ (the opposite case will clearly yield symmetric results). To determine the impact of changes of $p^{x^\star}$, we increase $p^{x^\star}$ keeping everything else constant. The results of the simulation as summarized in Table 2 suggest that as $p^{x^\star}$ increases the posterior estimates also increase. Because the prior mean of P^x is less than the prior mean of P^y, however, $\widehat{p^y}_j$ increases at a faster rate than $\widehat{p^x}_j$.

Table 2 also indicates the significant influence of the prior distributions on $\widehat{p^x}_j$ and $\widehat{p^y}_j$ even for large sample sizes ($j = 50$). Observe that the stochastic order implied by the prior distributions (i.e., P^x stochastically smaller than P^y) is opposite of the "true" values used in the simulation (P^x stochastically larger than P^y). These mistaken prior beliefs persist in the ordering of the posterior distributions. These simulations exemplify the inconsistency of the estimators discussed in the previous section. On the other hand, observe that $\widehat{p^x}_j + \widehat{p^y}_j$ approaches $p^{x^*} + p^{y^*}$.

The last two columns of Table 2 represent the standard errors of the estimates. Note that, unlike in the previous set of simulations, the standard errors increase with p^{x^*}.

3.1.3. *Simulation when* $\alpha_1 + \beta_1 \neq \alpha_2 + \beta_2$. Now, we simulate the returns in a case where $\alpha_1 + \beta_1 \neq \alpha_2 + \beta_2$ and investigate the impact of changes in $p^{x^\star}$.

Table 3 provides the output of simulations in the case where $\alpha_1 + \beta_1 = 2 < 6 = \alpha_2 + \beta_2$, but $\alpha_1/(\alpha_1 + \beta_1) = 0.5 = \alpha_2/(\alpha_2 + \beta_2)$; that is, the prior means are the same, but the prior sample size is smaller for P^x than for P^y. The results in Table 3 suggest that as $p^{x^\star}$ increases both $\widehat{p}^x_j$ and $\widehat{p}^y_j$ increase. Because the prior sample size for P^x is less than for P^y, $\widehat{p}^x_j$ changes more relative to the prior than does $\widehat{p}^y_j$. The standard errors reported in the last two columns also show that the estimate of P^x which starts with a greater variance than that of P^y, maintains that greater variance through the updating process.

Overall, the results from the simulations confirm the sensitivity of the posterior estimates to the prior beliefs. Because the firm only observes the sum of the returns from the assets, mistaken prior beliefs can persist. On the other hand, if the prior

Table 3
The impacts of increasing $p^{x^\star}$ on the posterior estimates when $\alpha_1+\beta_1 \neq \alpha_2+\beta_2$, $j=50$

$(\alpha_1,\beta_1)=(1,1)$	$(\alpha_2,\beta_2)=(3,3)$	$(\widehat{p^x}_0,\widehat{p^y}_0)=(0.5,0.5)$		$\hat{\sigma}_{p^x{}_0}=0.0833333$	$\hat{\sigma}_{p^y{}_0}=0.0357142$
$p^{x^\star}$	$p^{y^\star}$	$\widehat{p^x}_j$	$\widehat{p^y}_j$	$\widehat{\mathbf{S}}_{p^x_j}$	$\widehat{\mathbf{S}}_{p^y_j}$
0.3	0.2	0.2026	0.3286	0.00178	0.00156
0.4	0.2	0.2498	0.3713	0.00199	0.00156
0.5	0.2	0.2976	0.4193	0.00218	0.00150
0.6	0.2	0.3483	0.4641	0.00236	0.00134
0.7	0.2	0.4068	0.4972	0.00262	0.00113

beliefs have the correct order, then the posterior estimates converge to the true values.

4. A Normal Model

In this section, we assume that the firm's prior distribution and the distribution which generates the observations follow normal distributions.

4.1. *Distribution of the Observed Values*

We now assume that the values X_j and Y_j are each drawn from independent normal distributions with means $\theta_X^\star$ and $\theta_Y^\star$ and variances σ_X^2 and σ_Y^2, respectively. Thus, $X_j \mid \theta_X^\star \sim \mathbf{N}(\theta_X^\star, \sigma_X^2)$ and $Y_j \mid \theta_Y^\star \sim \mathbf{N}(\theta_Y^\star, \sigma_Y^2)$. For simplicity we assume that the variances of the observations, σ_X^2 and σ_Y^2, are fixed and known. As before, we observe the sum of X and Y, but we do not know the relative magnitude of X and Y separately.

At each stage the observed value equals the sum of the separate values, $R_j = X_j + Y_j$. The observed value in period j given $\theta_X^\star$ and $\theta_Y^\star$ follows a normal distribution with mean $\theta_X^\star + \theta_Y^\star$ and variance $\sigma_X^2 + \sigma_Y^2$. For ease of notation we define $\sigma^2 = \sigma_X^2 + \sigma_Y^2$. Then the density of the observed return at each stage j is

$$f(r \mid \theta_X^\star, \theta_Y^\star) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2\sigma^2}(r - (\theta_X^\star + \theta_Y^\star))^2\right], \tag{9}$$

where $-\infty < r < \infty$. A large positive value of r implies X and Y are both large and positive. A moderate positive value implies weakly positive X and a large positive Y, or vice versa. A large negative value implies large negative X and Y.

4.2. *Prior Distribution*

Our prior beliefs about the value of $\theta_X^\star$ and $\theta_Y^\star$ are contained in the random variables θ_X and θ_Y, respectively, for which we have a prior probability distribution. We assume the prior distribution of θ_X and θ_Y is a bivariate normal distribution with means μ_{X_0} and μ_{Y_0}, variances $\tau_{X_0}^2$ and $\tau_{Y_0}^2$, and correlation ρ_0. Hence, the prior density is

$$g(\theta_X, \theta_Y) = \frac{1}{2\pi\tau_{X_0}\tau_{Y_0}\sqrt{1-\rho_0^2}} \exp\left(-\frac{1}{2(1-\rho_0^2)}\left[\frac{(\theta_X-\mu_{X_0})^2}{\tau_{X_0}^2} + \frac{(\theta_Y-\mu_{Y_0})^2}{\tau_{Y_0}^2} - \frac{2\rho_0(\theta_X-\mu_{X_0})(\theta_Y-\mu_{Y_0})}{\tau_{Y_0}\tau_{X_0}}\right]\right) \tag{10}$$

where $-\infty < \mu_{X_0}, \mu_{Y_0} < \infty$, $\tau_{X_0} > 0$, $\tau_{Y_0} > 0$, and $-1 < \rho_0 < 1$.

4.3. *Posterior Distribution*

Each time we make an observation, we update our prior belief. In what follows, we find the joint posterior distribution of the mean returns of the assets.

The joint posterior distribution of θ_X and θ_Y is obtained using Bayes rule; hence, having the following form using Eqs. (9) and (10)

$$h(\theta_X, \theta_Y \mid R_1) \propto \exp\frac{-1}{2}\left[\frac{(R_1-(\theta_X+\theta_Y))^2}{\sigma^2} + \frac{1}{(1-\rho_0^2)}\left(\frac{(\theta_X-\mu_{X_0})^2}{\tau_{X_0}^2} + \frac{(\theta_Y-\mu_{Y_0})^2}{\tau_{Y_0}^2} - \frac{2\rho_0(\theta_X-\mu_{X_0})(\theta_Y-\mu_{Y_0})}{\tau_{Y_0}\tau_{X_0}}\right)\right] \tag{11}$$

Thus, the posterior distribution is also a bivariate normal and can be written as

$$\begin{aligned} h(\theta_X, \theta_Y \mid R_1) \propto \exp\frac{-1}{2}\Bigg[&\theta_X^2\left(\frac{1}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{X_0}^2}\right) \\ &- 2\theta_X\left(\frac{R_1}{\sigma^2} + \frac{\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}^2} - \frac{\rho_0\mu_{Y_0}}{(1-\rho_0^2)\tau_{X_0}\tau_{Y_0}}\right)\theta_Y^2\left(\frac{1}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{Y_0}^2}\right) \\ &- 2\theta_Y\left(\frac{R_1}{\sigma^2} + \frac{\mu_{Y_0}}{(1-\rho_0^2)\tau_{Y_0}^2} - \frac{\rho_0\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}\tau_{Y_0}}\right) \\ &- 2\theta_X\theta_Y\left(\frac{-1}{\sigma^2} + \frac{\rho_0}{(1-\rho_0^2)\tau_{X_0}^2\tau_{Y_0}^2}\right)\Bigg]. \end{aligned} \tag{12}$$

We need to find the parameters of this bivariate normal distribution. To do so, let a, b, c, d, and e be the coefficients in Eq. (12). Then the posterior distribution given by Eq. (12) can be written as

$$h(\theta_X, \theta_Y \mid R_1) \propto \exp\frac{-1}{2}\left[a\theta_X^2 - 2b\theta_X + c\theta_Y^2 - 2d\theta_Y - 2e\theta_X\theta_Y\right]. \tag{13}$$

Recall that the bivariate normal density with means (μ_X, μ_Y) and variance–covariance matrix

$$\begin{pmatrix} \pi_X^2 & \rho\pi_X\pi_Y \\ \rho\pi_X\pi_Y & \pi_Y^2 \end{pmatrix}$$

has the following form:

$$h(\theta_X, \theta_Y) \propto \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{\theta_X^2}{\pi_X^2} + \frac{\theta_Y^2}{\pi_Y^2} - \frac{2\rho\theta_X\theta_Y}{\pi_X\pi_Y} - \frac{2\theta_X}{\pi_X}\left(\frac{\mu_X}{\pi_X} - \frac{\rho\mu_Y}{\pi_Y}\right) - \frac{2\theta_Y}{\pi_Y}\left(\frac{\mu_Y}{\pi_Y} - \frac{\rho\mu_X}{\pi_X}\right)\right]\right\}. \tag{14}$$

The following lemma shows how the coefficients of the posterior distribution are computed.

Lemma 4.1. *The distribution of* (θ_X, θ_Y) *on the plane with density proportional to*

$$\exp\frac{-1}{2}\left[a\theta_X^2 - 2b\theta_X + c\theta_Y^2 - 2d\theta_Y - 2e\theta_X\theta_Y\right]$$

is a bivariate normal distribution with variances,

$$\pi_X^2 = Var(\theta_X) = \frac{c}{ac-e^2} \quad and \quad \pi_Y^2 = Var(\theta_Y) = \frac{a}{ac-e^2},$$

correlation coefficient

$$\rho = Corr(\theta_X, \theta_Y) = \frac{e}{\sqrt{ac}},$$

and means

$$\mu_X = E(\theta_X) = \frac{bc+ed}{ac-e^2} \quad and \quad \mu_Y = E(\theta_Y) = \frac{da+eb}{ac-e^2}.$$

Proof. Comparing Eqs. (13) and (14) results in the following equations:

$$a = \frac{1}{(1-\rho^2)\pi_X^2}, \quad c = \frac{1}{(1-\rho^2)\pi_Y^2} \quad \text{and} \quad e = \frac{\rho}{\pi_X\pi_Y(1-\rho^2)} \tag{15}$$

and

$$b = \frac{1}{(1-\rho^2)\pi_X}\left[\frac{\mu_X}{\pi_X} - \frac{\rho\mu_Y}{\pi_Y}\right] \quad \text{and} \quad d = \frac{1}{(1-\rho^2)\pi_Y}\left[\frac{\mu_Y}{\pi_Y} - \frac{\rho\mu_X}{\pi_X}\right]. \tag{16}$$

Using the above equations we solve for the correlation, variances and the means of the bivariate distribution. This completes the proof. □

We have made the assumption that $-1 < \rho_0 < 1$ and the formulas above depend on this. If there is perfect negative correlation, however, and $\rho = -1$, then the posterior distribution of (θ_X, θ_Y) is a bivariate normal distribution over a straight line with slope -1. With a perfect positive correlation, $\rho = 1$, you know that a value of θ_X will result in the same value of θ_Y. In both cases, variances are zero.

We will shortly illustrate the limiting properties of the parameters of the posterior distribution as the number of observations from the return from the investment increases.

The next theorem formulates the parameters of the posterior distribution in terms of the prior information.

Theorem 4.1. *The posterior distribution of θ_X and θ_Y after observing $R_1 = X_1 + Y_1$ is a bivariate normal distribution with the following parameters. The updated variances are*

$$\tau_{X_1}^2 = \frac{\frac{1}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{Y_0}^2}}{\frac{1}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}\tau_{Y_0}}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)}, \tag{17}$$

$$\tau_{Y_1}^2 = \frac{\left(\frac{1}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{X_0}^2}\right)}{\frac{1}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}\tau_{Y_0}}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)}; \tag{18}$$

the updated correlation is

$$\rho_1 = \frac{\frac{-1}{\sigma^2} + \frac{\rho_0}{(1-\rho_0^2)\tau_{X_0}\tau_{Y_0}}}{\sqrt{\left(\frac{1}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{X_0}^2}\right)\left(\frac{1}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{Y_0}^2}\right)}}; \tag{19}$$

the updated means are

$$\begin{aligned}\mu_{X_1} = {} & \frac{\left(\frac{R_1}{\sigma^2} + \frac{\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}^2} - \frac{\rho_0\mu_{Y_0}}{(1-\rho_0^2)\tau_{X_0}\tau_{Y_0}}\right)\left(\frac{1}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{Y_0}^2}\right)}{\frac{1}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}\tau_{Y_0}}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)} \\ & + \frac{\left(\frac{R_1}{\sigma^2} + \frac{\mu_{Y_0}}{(1-\rho_0^2)\tau_{Y_0}^2} - \frac{\rho_0\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}\tau_{Y_0}}\right)\left(\frac{-1}{\sigma^2} + \frac{\rho_0}{(1-\rho_0^2)\tau_{X_0}^2\tau_{Y_0}^2}\right)}{\frac{1}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}\tau_{Y_0}}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)},\end{aligned} \tag{20}$$

and

$$\begin{aligned}\mu_{Y_1} = {} & \frac{\left(\frac{R_1}{\sigma^2} + \frac{\mu_{Y_0}}{(1-\rho_0^2)\tau_{Y_0}^2} - \frac{\rho_0\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}\tau_{Y_0}}\right)\left(\frac{1}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{X_0}^2}\right)}{\frac{1}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}\tau_{Y_0}}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)} \\ & + \frac{\left(\frac{R_1}{\sigma^2} + \frac{\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}^2} - \frac{\rho_0\mu_{Y_0}}{(1-\rho_0^2)\tau_{X_0}\tau_{Y_0}}\right)\left(\frac{-1}{\sigma^2} + \frac{\rho_0}{(1-\rho_0^2)\tau_{X_0}^2\tau_{Y_0}^2}\right)}{\frac{1}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}\tau_{Y_0}}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)}.\end{aligned} \tag{21}$$

Proof. From Eqs. (12) and (13), we find the values of a, b, c, d, and e as shown in Eqs. (15) and (16). Putting these values into the equations of Lemma 4.1 gives the result. □

Now suppose that we have observed the values of $R_i = (X_i + Y_i)$ for j periods. The density of these observations after j periods is

$$\begin{aligned} f(R_1, R_2, \ldots, R_j \mid \theta_X^\star, \theta_Y^\star) &\propto \exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{j}\left(R_i - (\theta_X^\star + \theta_Y^\star)\right)^2\right\} \\ &= \exp\left\{\frac{-j}{2\sigma^2}\left[\frac{1}{j}\sum_{i=1}^{j}R_i^2 - 2(\theta_X^\star + \theta_Y^\star)\overline{R}_j + (\theta_X^\star + \theta_Y^\star)^2\right]\right\}, \end{aligned} \tag{22}$$

where $-\infty < R_1, R_2, \ldots, R_j < \infty$. Note that $\overline{R}_j$ is a sufficient statistic for (θ_X, θ_Y) and $\overline{R}_j \in N\big((\theta_X^\star + \theta_Y^\star), \frac{\sigma^2}{j}\big)$. Thus, the posterior distribution after j periods can be obtained by replacing R_1 and σ^2 in Eq. (12) by $\overline{R}_j$ and $\frac{\sigma^2}{j}$, respectively.

Theorem 4.2. *The posterior distribution of θ_X and θ_Y after j observations is a bivariate normal distribution with parameters as follows. The updated variances are*

$$\tau_{X_j}^2 = \frac{\left(\frac{j}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{Y_0}^2}\right)}{\frac{j}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}^2\tau_{Y_0}^2}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)}, \tag{23}$$

$$\tau_{Y_j}^2 = \frac{\left(\frac{j}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{X_0}^2}\right)}{\frac{j}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}^2\tau_{Y_0}^2}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)}; \tag{24}$$

the updated correlation after j periods is

$$\rho_j = \frac{\frac{-j}{\sigma^2} + \frac{\rho_0}{(1-\rho_0^2)\tau_{X_0}^2\tau_{Y_0}^2}}{\sqrt{\left(\frac{j}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{X_0}^2}\right)\left(\frac{j}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{Y_0}^2}\right)}}, \tag{25}$$

Finally, the updated means of the distribution are

$$\begin{aligned}\mu_{X_j} = {} & \frac{\left(\frac{j\overline{R}_j}{\sigma^2} + \frac{\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}^2} - \frac{\rho_0\mu_{Y_0}}{(1-\rho_0^2)\tau_{X_0}\tau_{Y_0}}\right)\left(\frac{j}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{Y_0}^2}\right)}{\frac{j}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}^2\tau_{Y_0}^2}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)} \\ & + \frac{\left(\frac{j\overline{R}_j}{\sigma^2} + \frac{\mu_{Y_0}}{(1-\rho_0^2)\tau_{Y_0}^2} - \frac{\rho_0\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}\tau_{Y_0}}\right)\left(\frac{-j}{\sigma^2} + \frac{\rho_0}{(1-\rho_0^2)\tau_{X_0}^2\tau_{Y_0}^2}\right)}{\frac{j}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}^2\tau_{Y_0}^2}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)},\end{aligned} \tag{26}$$

and

$$\begin{aligned}\mu_{Y\,j} = {} & \frac{\left(\frac{j\overline{R}_j}{\sigma^2} + \frac{\mu_{Y_0}}{(1-\rho_0^2)\tau_{Y_0}^2} - \frac{\rho_0\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}\tau_{Y_0}}\right)\left(\frac{j}{\sigma^2} + \frac{1}{(1-\rho_0^2)\tau_{X_0}^2}\right)}{\frac{j}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}^2\tau_{Y_0}^2}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)} \\ & + \frac{\left(\frac{j\overline{R}_j}{\sigma^2} + \frac{\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}^2} - \frac{\rho_0\mu_{Y_0}}{(1-\rho_0^2)\tau_{X_0}\tau_{Y_0}}\right)\left(\frac{-j}{\sigma^2} + \frac{\rho_0}{(1-\rho_0^2)\tau_{X_0}^2\tau_{Y_0}^2}\right)}{\frac{j}{\sigma^2(1-\rho_0^2)}\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}^2\tau_{Y_0}^2}\right) + \frac{1}{(1-\rho_0^2)^2\tau_{X_0}^2\tau_{Y_0}^2}\left(1 - \frac{\rho_0^2}{\tau_{X_0}^2\tau_{Y_0}^2}\right)}.\end{aligned} \tag{27}$$

Proof. The posterior distribution after j periods is a bivariate normal distribution. Consequently, the updated parameters of the posterior distribution after j observations have the form of those in Eqs. (17)–(21) with R_1 replaced by $\overline{R_j}$ and σ^2 replaced by $\frac{\sigma^2}{j}$. □

Thus, when the observations follow a normal distribution and our prior belief follows a bivariate normal distribution, the posterior distribution follows a bivariate normal distribution where the updated parameters are given by Theorem 4.2. It can also be seen that even though both the updated variances and updated correlation do not depend on the observations, the updated means do.

When the mean returns from the assets are independent; that is $\rho_0 = 0$, the equations given by Theorem 4.2 simplify. The posterior distribution has the form of that in Sec. 4.3 with ρ_0 replaced by 0.

4.4. *Asymptotic Behavior of Parameter Estimates*

The values of X and Y are drawn from Normal distributions with parameters $\mu_X^\star$ and $\mu_Y^\star$. We start with a prior distribution with means μ_{X_0} and μ_{Y_0}, variances $\tau_{X_0}^2$ and $\tau_{Y_0}^2$, and correlation ρ_0. Now we formulate the limiting values of the updated parameters as $j \longrightarrow \infty$.

Recall that in general, when we start with a correlated prior, the updated parameters are given by Theorem 4.2. Taking a limit of Eqs. (23)–(25) as $j \longrightarrow \infty$ results in the following:

$$\lim_{j\to\infty} \tau_{X_j}^2 = \frac{(1-\rho_0^2)}{\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}^2\tau_{Y_0}^2}\right)}, \tag{28}$$

$$\lim_{j\to\infty} \tau_{Y_j}^2 = \frac{(1-\rho_0^2)}{\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}^2\tau_{Y_0}^2}\right)}, \tag{29}$$

$$\lim_{j\to\infty} \rho_j = -1 \tag{30}$$

Recall from Sec. 4.3 that the updated variances and correlation do not depend on the observations. It can be seen that as $j \longrightarrow \infty$ the variances smoothly decrease, but never reach zero, which in effect increases the precision of the current point estimates. As j increases the correlation approaches -1—this implies that as $j \longrightarrow \infty$ the posterior distribution of (θ_X, θ_Y) gets closer to a bivariate normal distribution over a straight line with slope -1. The limiting variances of the posterior distribution determines the precision of the estimates. In other words, the smaller the variances, the smaller the spread of the distribution over the line with slope of -1; hence, the closer the estimates to the true values.

The limiting values of the posterior means depend on $\overline{R_j}$. By the Law of Large Numbers, as j increases $\overline{R_j}$ converges to the true mean of the distribution, $\theta_X^\star + \theta_Y^\star$. Then the limiting values of the means of the posterior distribution are

$$\lim_{j\to\infty} \mu_{X_j} = \frac{\frac{(\theta_X^\star+\theta_Y^\star)}{(1-\rho_0^2)\tau_{Y_0}^2}\left(1 - \frac{\rho_0}{\tau_{X_0}^2}\right) + \frac{\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}}\left(\frac{1}{\tau_{X_0}} + \frac{\rho_0}{\tau_{Y_0}}\right) - \frac{\mu_{Y_0}}{(1-\rho_0^2)\tau_{Y_0}}\left(\frac{1}{\tau_{Y_0}} + \frac{\rho_0}{\tau_{X_0}}\right)}{\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}^2\tau_{Y_0}^2}\right)}. \tag{31}$$

Similarly,

$$\lim_{j\to\infty} \mu_{Y_j} = \frac{\frac{(\theta_X^\star+\theta_Y^\star)}{(1-\rho_0^2)\tau_{X_0}^2}\left(1 - \frac{\rho_0}{\tau_{Y_0}^2}\right) + \frac{\mu_{Y_0}}{(1-\rho_0^2)\tau_{Y_0}}\left(\frac{1}{\tau_{Y_0}} + \frac{\rho_0}{\tau_{X_0}}\right) - \frac{\mu_{X_0}}{(1-\rho_0^2)\tau_{X_0}}\left(\frac{1}{\tau_{X_0}} + \frac{\rho_0}{\tau_{Y_0}}\right)}{\left(\frac{1}{\tau_{Y_0}^2} + \frac{1}{\tau_{X_0}^2} + \frac{2\rho_0}{\tau_{X_0}^2\tau_{Y_0}^2}\right)}. \tag{32}$$

Similarly, we can formulate the limiting values of the parameters estimate in case of independent prior returns by simply replacing $\rho_0 = 0$ in Eqs. (28)–(32).

4.5. *Consistency of the Bayes Estimate Despite the Inconsistency of the Posterior Means*

The lack of consistency of the estimates when the prior information follows a normal distribution is similar to that as in Sec. 2.5. That is, because of the assumption that we can only observe the sum of X and Y, θ_X and θ_Y are not identifiable. Hence, no method of estimation gives consistent estimates. However, the Bayes estimate of the sum of the means is consistent; that is, $\mu_{X_j} + \mu_{Y_j} \xrightarrow{a.s.} (\theta_X^\star + \theta_Y^\star)$ as $j \longrightarrow \infty$.

5. Conclusion

We have provided formulas for updating the joint prior distribution of random variables X and Y when only the sum $(X + Y)$ can be observed. When the distributions of X and Y are independent Bernoulli random variables with parameters $p^{x\star}$ and $p^{y^\star}$, respectively, the conjugate distribution for the parameters is shown to be a mixture of products of independent Beta distributions. When the distributions of X and Y are independent Normal random variables with means $\theta_X^\star$ and $\theta_Y^\star$, respectively, the conjugate distribution for the parameter is shown to be jointly Normal. In both cases, the means of the posterior distributions which are obtained via Bayesian updating are shown to be inconsistent — that is, the posterior means do not necessarily approach the true values. This lack of consistency is due to the fact that the separate values of X and Y cannot be observed. Our mixture models for investment decisions with the Normal and Bernoulli specifications are just two examples of conjugacy. Similar conjugate results should obtain in other standard cases.

References

Bernardo, A. E., Chowdhry, B. (2002). Resources, real options, and corporate strategy. *J. Fin. Econ.* 63:211—234.

Cao, G., West, M. (1996). Practical bayesian inference using mixtures of mixtures. *Biometrics* 52(4):1334–1341.

DeGroot, M. H. (2004). *Optimal Statistical Decisions*. Hoboken, NJ: John Wiley & Sons, Inc.

Doob, J. (1949). Application of the theory of martingales. *Colloques Internationaux du Centre National de la Recherche Scientique* 13:23–27.

Escobar, M. D., West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* 90(430):577–588.

Farzinnia, N. (2006). A Bayesian Scale-Up Policy for Stochastic Investments. PhD thesis, Department of Statistics, University of California, Los Angeles.

Jovanovic, B., Nyarko, Y., Griliches (1995). A Bayesian learning model fitted to a variety of empirical learning curves. *Brookings Papers on Economics Activity. Microeconomics* 1995:247–305.

Labbe, A., Thompson, M. (2005). Hierarchical inverse gaussian models and multiple testing: application to gene expression data. *Berkeley Electronic Press: Statistical Applications in Genetics and Molecular Biology* Vol. 4, Iss. 1, Article 23.

Lo, K., Gottardo, R. (2007). Flexible empirical bayes models for differential gene expression. *Bioinformatics* 23:328–335.

McCardle, K. F. (1985). Information acqusition and the adoption of new technology. *Manage. Sci.* 31(11):1379–1382.

Oikonomou, K. (1997). Prediction with the dynamic bayesian gamma mixture model. *Syst. Man Cybernet. Part A, IEEE Trans.* 27:529–542.

Rangarajan, A., Hsiao, I.-T., Gindi, G. (2000). A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction. *Colloques Internationaux du Centre National de la Recherche Scientique* 12(3):199–217.

Wiper, M., Insua, D. R., Ruggeri, F. (2001). Mixtures of gamma distributions with applications. *J. Comput. Graph. Statist.* 10:440–454.