Finance Stochast. 1, 229–238 (1997)

Finance and
Stochastics

An application of hidden Markov models to asset allocation problems*

Robert J. Elliott[1], **John van der Hoek**[2]

[1]Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1
[2]Department of Applied Mathematics, University of Adelaide, Adelaide, South Australia 5005

Abstract. Filtering and parameter estimation techniques from hidden Markov Models are applied to a discrete time asset allocation problem. For the commonly used mean-variance utility explicit optimal strategies are obtained.

Key words: Hidden Markov models, asset allocation, portfolio selection

JEL classification: C13, E44, G2

Mathematics Subject Classification (1991): 90A09, 62P20

1. Introduction

A typical problem faced by fund managers is to take an amount of capital and invest this in various assets, or asset classes, in an optimal way. To do this the fund manager must first develop a model for the evolution, or prediction, of the rates of returns on investments in each of these asset classes. A common procedure is to assert that these rates of return are driven by rates of return on some observable factors. Which factors are used to explain the evolution of rates of return of an asset class is often proprietary information for a particular fund manager. One of the criticisms that could be made of such a framework is the assumption that these rates of return can be adequately explained in terms of observable factors. In this paper we propose a model for the rates of return in terms of observable and non-observable factors; we present algorithms for the identification for this model, (using filtering and prediction techniques) and

* Robert Elliott wishes to thank SSHRC for its support. John van der Hoek wishes to thank the Department of Mathematical Sciences, University of Alberta, for its hospitality in August and September 1995, and SSHRC for its support.
Manuscript received: December 1995 / final version received: August 1996

indicate how to apply this methodology to the (strategic) asset allocation problem using a mean-variance type utility criterion.

Earlier work in filtering is presented in the classical work [4] of Liptser and Shiryayev. Previous work on parameter identification for linear, Gaussian models includes the paper by Anderson et al. [1]. However, this reference discusses only continuous state systems for which the dimension of the observations is the same as the dimension of the unobserved state. Parameter estimation for noisily observed discrete state Markov chains is fully treated in [2]. The present paper includes an extension of techniques from [2] to hybrid systems, that is, those with both continuous and discrete state spaces.

2. Mathematical model

A bank typically keeps track of between 30 and 40 different factors. These include observables such as interest rates, inflation and industrial indices. The observable factors will be represented by $x_k \in R^n$.

In addition we suppose there are unobserved factors which will be represented by the finite state Markov chain $X_k \in S$. We suppose X_k is not observed directly. Its states might represent in some approximate way the psychology of the market, or the actions of competitors. Therefore, we suppose the state space of the factors generating the rates of return on asset classes has a decomposition as a hybrid combination of $\mathbb{R}^n$ and a finite set $\widehat{S} = \{s_1, s_2, \ldots, s_N\}$ for some $N \in \mathbb{N}$.

Note that if $\widehat{S} = \{s_1, s_2, \ldots, s_N\}$ is any finite set, by considering the indicator functions $I_{s_i} : \widehat{S} \to \{0, 1\}$, ($I_{s_i}(s) = 1$ if $s = s_i$ and 0 otherwise), there is a canonical bijection of $\widehat{S}$ with $S = \{e_1, e_2, \ldots, e_N\}$. Here, the $e_i = (0, 0, \ldots, 1, \ldots, 0)$; $1 \le i \le N$ are the canonical unit vectors in $\mathbb{R}^N$. The time index of state evolution will be the non-negative integers $\mathbb{N} = \{0, 1, 2, \ldots\}$. Let $\{x_k\}_{k\in\mathbb{N}}$ be a process with state space $\mathbb{R}^n$ and $\{X_k\}_{k\in\mathbb{N}}$ be a process whose state space we can, without loss of generality, identify with S.

Suppose $(\Omega, \mathscr{F}, P)$ is a probability space upon which $\{w_k\}_{k\in\mathbb{N}}$, $\{b_k\}_{k\in\mathbb{N}}$ are independent, identically distributed (i.i.d.) sequences of Gaussian random variables, with zero means and non-singular covariance matrices Σ and Γ respectively; x_0 is normally distributed, X_0 is uniformly distributed, independently of each other and of the sequences $\{w_k\}_{k\in\mathbb{N}}$ and $\{b_k\}_{k\in\mathbb{N}}$. Let $\{\mathscr{F}_k\}_{k\in\mathbb{N}}$ be the complete filtration with $\mathscr{F}_k$ generated by $\{x_0, x_1, \ldots, x_k, X_0, X_1, \ldots, X_k\}$ and $\{\mathscr{X}_k\}_{k\in\mathbb{N}}$ the complete filtration with $\mathscr{X}_k$ generated by $\{X_0, X_1, \ldots, X_k\}$. The state $\{x_k\}_{k\in\mathbb{N}}$ is assumed to satisfy

$$x_{k+1} = Ax_k + w_{k+1} \tag{1}$$

that is, $\{x_k\}_{k\in\mathbb{N}}$ is a VAR(1) process, [5], and A is an $n \times n$ matrix. $\{X_k\}_{k\in N}$ is assumed to be a $\mathscr{X}$-Markov chain with state space S. The Markov property implies that

$$P\left(X_{k+1} = e_j \,|\, \mathscr{X}_k\right) = P\left(X_{k+1} = e_j \,|\, X_k\right).$$

Write $\pi_{ji} = P(X_{k+1} = e_j \,|\, X_k = e_i)$ and Π for the $N \times N$ matrix (π_{ji}). Then

$$\begin{aligned} E\left[X_{k+1}|\mathscr{X}_k\right] &= E\left[X_{k+1}|X_k\right] \\ &= \Pi X_k. \end{aligned}$$

Define $M_{k+1} := X_{k+1} - \Pi X_k$. Then

$$\begin{aligned} E\left[M_{k+1}|\mathscr{X}_k\right] &= E\left[X_{k+1} - \Pi X_k|\mathscr{X}_k\right] \\ &= 0 \end{aligned}$$

so M_{k+1} is a martingale increment and

$$X_{k+1} = \Pi X_k + M_{k+1} \tag{2}$$

is the semimartingale representation of the Markov chain.

This form of dynamics is developed in [2] and [3].

The process x_k with dynamics given by (1) will be a model for the observed factors, and the process X_k, with dynamics given by (2), will model the unobserved factors.

Suppose that there are m asset classes in which a fund manager can make investments. (For strategic asset allocation, the number of asset classes is usually between 5 and 10 and it includes general groups of assets such as bonds, manufacturing and energy securities.) Let $r_k \in \mathbb{R}^m$ denote the vector of the rates of return on the m asset classes in the kth period for, $k = 1, 2, 3, \ldots$.

We shall assume that the model for the "true" rates of return, which apply between times $k-1$ and k, is given by:

$$r_k = Cx_k + HX_k. \tag{3}$$

Here C and H are $m \times n$ and $m \times N$ matrices respectively. However, we shall assume that these rates are observed in Gaussian noise. This means that, if $\{y_k\}_{k\in\mathbb{N}}$ is the sequence of our observations of the rates of return, then

$$y_k = Cx_k + HX_k + b_k. \tag{4}$$

Let $\{\mathscr{Y}_k\}_{k\in\mathbb{N}}$ be the complete filtration with $\mathscr{Y}_k$ generated by $\{y_0, y_1, \ldots, y_k, x_0, x_1, \ldots, x_k\}$, for $k \in \mathbb{N}$, and denote by $\widehat{X}_k$ the conditional expectation under P of X_k given $\mathscr{Y}_k$.

The Markov chain X_k is noisily observed through (4) and so we have a variation of the Hidden Markov Model, HMM, discussed in [3].

Our first task will be to identify this model. This means that we shall provide best estimates for A, Π, C, H given the observations of $\{y_k\}_{k\in\mathbb{N}}$ and $\{x_k\}_{k\in\mathbb{N}}$. These estimates will then be used to predict future rates of return. Finally we shall provide algorithms for allocating funds to the asset classes based on these predictions.

3. A measure change

We first consider all processes initially defined in an 'ideal' probability space $(\Omega, \mathscr{F}, \overline{P})$; under a new probability P, to be defined, the dynamics (1), (2) and (4) will hold.

Suppose that under $\overline{P}$ $\{y_k\}_{k\in\mathbb{N}}$ is an i.i.d. $N(0,\Gamma)$ sequence of Gaussian random variables. For $y \in \mathbb{R}^m$, set

$$\phi(y) = \phi_\Gamma(y) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\det \Gamma|^{1/2}} \exp[-\frac{1}{2} y' \Gamma^{-1} y]. \tag{5}$$

With

$$\overline{\lambda}_\ell := \phi(y_\ell - Cx_\ell - HX_\ell)/\phi(y_\ell), \quad \ell = 0, 1, \ldots \tag{6}$$

write

$$\overline{\Lambda}_k = \prod_{\ell=0}^{k} \overline{\lambda}_\ell. \tag{7}$$

Define $\{\mathscr{G}_k\}_{k\in\mathbb{N}}$ to be the complete filtration with terms generated by $\{x_0, \ldots, x_k, y_0, \ldots, y_k, X_0, \ldots, X_{k-1}\}$, and define P on $(\Omega, \mathscr{F})$ by setting the restriction of the Radon-Nikodym derivative $dP/d\overline{P}$ to $\mathscr{G}_k$ equal to $\overline{\Lambda}_k$. One can check, (see [3], page 61), that on $(\Omega, \mathscr{F})$ under P, $\{b_k\}_{k\in\mathbb{N}}$ are i.i.d. $N(0,\Gamma)$, where

$$b_k := y_k - Cx_k - HX_k.$$

Furthermore, by Bayes' Theorem ([3], page 23)

$$E[X_k | \mathscr{Y}_k] = \overline{E}[\overline{\Lambda}_k X_k | \mathscr{Y}_k] / \overline{E}[\overline{\Lambda}_k | \mathscr{Y}_k]$$

and

$$\overline{E}[\overline{\Lambda}_k | \mathscr{Y}_k] = \langle \overline{E}[\overline{\Lambda}_k X_k | \mathscr{Y}_k], \mathbf{1} \rangle,$$

where $\mathbf{1} = (1, 1, \ldots, 1)$. Here, $\langle \, , \, \rangle$ denotes the scalar product in R^N.

Write

$$q_k = \overline{E}[\overline{\Lambda}_k X_k | \mathscr{Y}_k].$$

Then, by a simple modification of the arguments in [3]:

$$\begin{aligned}
q_{k+1} &= \overline{E}[\overline{\Lambda}_{k+1} X_{k+1} | \mathscr{Y}_{k+1}] \\
&= \frac{1}{\phi(y_{k+1})} \overline{E}[\overline{\Lambda}_k \phi(y_{k+1} - Cx_{k+1} - HX_{k+1}) X_{k+1} | \mathscr{Y}_{k+1}] \\
&= \frac{1}{\phi(y_{k+1})} \sum_{i=1}^{N} \overline{E}[\overline{\Lambda}_k \phi(y_{k+1} - Cx_{k+1} - HX_{k+1}) X_{k+1} \langle X_{k+1}, e_i \rangle | \mathscr{Y}_{k+1}] \\
&= \frac{1}{\phi(y_{k+1})} \sum_{i=1}^{N} \phi(y_{k+1} - Cx_{k+1} - He_i) \overline{E}[\overline{\Lambda}_k \langle X_{k+1}, e_i \rangle | \mathscr{Y}_{k+1}] e_i \\
&= \sum_{i=1}^{N} \frac{\phi(y_{k+1} - Cx_{k+1} - He_i)}{\phi(y_{k+1})} \overline{E}\Big[\sum_{j=1}^{N} \overline{\Lambda}_k \langle X_k, e_j \rangle \langle \Pi X_k, e_i \rangle | \mathscr{Y}_k \Big] e_i
\end{aligned}$$

$$= \sum_{i,j=1}^{N} \frac{\phi(y_{k+1} - Cx_{k+1} - He_i)}{\phi(y_{k+1})} \overline{E}\,[\overline{\Lambda}\langle X_k, e_j\rangle\,\langle \Pi X_k, e_i\rangle | \mathscr{Y}_k] e_i$$

$$= \sum_{i,j=1}^{N} \frac{\phi(y_{k+1} - Cx_{k+1} - He_i)}{\phi(y_{k+1})} \pi_{ij} \overline{E}\,[\overline{\Lambda}_k\langle X_k, e_j\rangle | \mathscr{Y}_k] e_i$$

$$= \sum_{i,j=1}^{N} \frac{\phi(y_{k+1} - Cx_{k+1} - He_i)}{\phi(y_{k+1})} \pi_{ij}\,\langle q_k, e_j\rangle e_i\,.$$

Alternatively, for the i^{th} component of q_k

$$q_k^i = \overline{E}\,[\overline{\Lambda}_k\langle X_k, e_i\rangle | \mathscr{Y}_k], \quad i = 1, \ldots, N$$

we have the recurrence relationship:

$$q_{k+1}^i = \sum_{j=1}^{N} \frac{\phi(y_{k+1} - Cx_{k+1} - He_i)}{\phi(y_{k+1})} \pi_{ij} q_k^j, \tag{8}$$

for $i = 1, 2, \ldots, N$. If we define the normalized conditional expectation p_k by

$$p_k^i = E[\langle X_k, e_i\rangle | \mathscr{Y}_k], \quad i = 1, \ldots, N \tag{9}$$

then

$$p_k^i = q_k^i \Big[\sum_{i=1}^{N} q_k^i\Big]^{-1}. \tag{10}$$

If the vector p_0 is given, or estimated, then (8) can be initialized with

$$\begin{aligned} q_0 &= \overline{E}\,[\overline{\Lambda}_0 X_0 | \mathscr{Y}_0] \\ &= E[\overline{\Lambda}_0^{-1}\overline{\Lambda}_0 X_0 | \mathscr{Y}_k] \big/ E[\Lambda_0^{-1} | \mathscr{Y}_0] \\ &= p_0 \big/ E[\overline{\Lambda}_0^{-1} | \mathscr{Y}_0] \end{aligned}$$

where

$$\begin{aligned} E[\overline{\Lambda}_0^{-1} | Y_k] &= E\Big[\frac{\phi(y_0)}{\phi(y_0 - Cx_0 - HX_0)} | x_0, y_0\Big] \\ &= \sum_{i=1}^{N} \frac{\phi(y_0)}{\phi(y_0 - Cx_0 - He_i)} p_0^i. \end{aligned} \tag{11}$$

This discussion leads to the following proposition:

Proposition 1. *For given* A, Π, C, H *in equations* (1), (2), (4) *the estimate of the* H.M.M. *is given by*

$$\widehat{X}_k = E[X_k | \mathscr{Y}_k] = p_k$$

where the p_k *are determined by* (10).

□

Now consider the issue of estimating the coefficient matrices A, C, Π, H. We estimate A by maximizing an appropriate log-likelihood function. As this is relatively standard procedure, we only outline the details. To this end suppose that under probability Q, $\{x_k\}_{k\in\mathbb{N}}$ is an i.i.d. $N(0,\Sigma)$ sequence of Gaussian random variables. Choose $\phi = \phi_\Sigma$ and define

$$\begin{aligned} \gamma_\ell^A &:= \tfrac{\phi(x_\ell - Ax_{\ell-1})}{\phi(x_\ell)} \\ \Gamma_k^A &:= \textstyle\prod_{\ell=1}^k \gamma_\ell^A. \end{aligned}$$

If $\{\mathscr{H}_k\}_{k\in\mathbb{N}}$ is the complete filtration with $\mathscr{H}_k$ the sigma-algebra generated by $\{x_0, x_1, \ldots, x_k\}$, define Q^A on $(\Omega, \mathscr{F})$ by setting the Radon-Nikodym derivative dQ^A/dQ, restricted to $\mathscr{H}_k$, equal to Γ_k^A. Then $\{w_k\}_{k\in\mathbb{N}}$ defined by

$$w_{k+1} := x_{k+1} - Ax_k$$

will, under Q^A, be an i.i.d. sequence of $N(0,\Sigma)$ Gaussian random variables. For some interim estimate A_1 of A we choose $A = \widehat{A}$ to maximimize

$$\log\left[\frac{dQ^A}{dQ^{A_1}}|\mathscr{H}_k\right] = -\frac{1}{2}\sum_{\ell=1}^k (x_\ell - Ax_{\ell-1})'\Sigma^{-1}(x_\ell - Ax_{\ell-1}) + R$$

where R does not depend on A. This leads to

$$\sum_{\ell=1}^k x_\ell x_{\ell-1}' = A\sum_{\ell=1}^k x_{\ell-1}x_{\ell-1}'$$

or

$$\widehat{A} = \widehat{A}_k = \Big(\sum_{\ell=1}^k x_\ell x_{\ell-1}'\Big)\Big(\sum_{\ell=1}^k x_{\ell-1}x_{\ell-1}'\Big)^{-1}. \tag{12}$$

Here we have used the result that for $k \geq n$, $\sum_{\ell=1}^k x_{\ell-1}x_{\ell-1}'$ is, almost surely, invertible [4]. This will be our estimate for A, given observations of $x_0, x_1, \ldots, x_k$.

We now provide analogous estimates for C. Again use the measure change, as in (5), (6), (7), but instead write $P = P^C$, $\Lambda = \Lambda^C$ to indicate that we are focusing on the dependence of these variables of the matrix C.

Given C_1, an interim estimate for C, we choose $\widehat{C}$ to maximize

$$E_1\Big[\log\frac{dP^C}{dP^{C_1}}|\mathscr{Y}_k\Big] \tag{13}$$

(see [2], p. 36). Here E_1 denotes expectation taken with respect to P^{C_1}, given the observations of $x_0, x_1, \ldots, x_k$ and $y_0, y_1, \ldots, y_k$. The expression in (13) is

$$E_1\Big\{-\frac{1}{2}\sum_{\ell=0}^k (y_\ell - Cx_\ell - HX_\ell)'\Gamma^{-1}(y_\ell - Cx_\ell - HX_\ell)|\mathscr{Y}_k\Big\} + \widehat{R}$$

where $\widehat{R}$ does not depend on C. Using the notation $\langle A, B\rangle = \sum_{ij} A_{ij}B_{ij}$ for matrices A, B, we can write the expression (13) in the form:

$$E_1\Big\{ -\frac{1}{2}\sum_{\ell=0}^{k} \big[y'_\ell \Gamma^{-1} y_\ell \quad -2\langle C\,,\Gamma^{-1}y_\ell x'_\ell\rangle - 2\langle H\,,\Gamma^{-1}y_\ell X'_\ell\rangle + \langle C\,,\Gamma^{-1}Cx_\ell x'_\ell\rangle$$

$$+\langle H\,,\Gamma^{-1}HX_\ell X'_\ell\rangle + 2\langle C\,,\Gamma^{-1}HX_\ell x'_\ell\rangle\big] \,|\mathscr{Y}_k \Big\} + \widehat{R}.$$

The first order optimality condition for C yields:

$$C\sum_{\ell=0}^{k} x_\ell x'_\ell + H\sum_{\ell=0}^{k} E_1[X_\ell|\mathscr{Y}_k]x'_\ell = \sum_{\ell=0}^{k} y_\ell x'_\ell. \tag{14}$$

This implies that

$$\widehat{C} = \widehat{C}_k = \Big[\sum_{\ell=0}^{k} y_\ell x'_\ell - H\sum_{\ell=0}^{k} E_1[X_\ell|\mathscr{Y}_k]x'_\ell\Big]\Big[\sum_{\ell=0}^{k} x_\ell x'_\ell\Big]^{-1} \tag{15}$$

where we have again used the fact that $\sum_{\ell=0}^{k} x_\ell x'_\ell$ is almost surely invertible if $k \geq n$. In order to compute $\widehat{C}_k$ we need $\sum_{\ell=0}^{k} E_1[X_\ell|\mathscr{Y}_\ell]x'_\ell$.

Now $\sum_{\ell=1}^{k} E_1[X_\ell|\mathscr{Y}_\ell]x'_\ell = q_k x'_k + \sum_{\ell=1}^{k-1} E_1[X_\ell|\mathscr{Y}_\ell]x'_{\ell-1}$. Write

$$T_k := \sum_{\ell=0}^{k-1} X_\ell x'_\ell$$

$$T_k^{rs} := \sum_{\ell=0}^{k-1} \langle X_\ell, e_r\rangle\, x_\ell^s$$

where $x'_\ell = (x'_\ell, x_\ell^2, \ldots, x_\ell^n) \in \mathbb{R}^n$.

We need to compute: $E_1[T_k^{rs}|\mathscr{Y}_k]$.

We proceed as in Elliott [2] using Theorem 5.3 etc. viz.,

$$E_1[T_k^{rs}|\mathscr{Y}_k] = \overline{E}[\Lambda_k^{C_1} T_k^{rs}|\mathscr{Y}_k]/\overline{E}[\Lambda_k^{C_1}|\mathscr{Y}_k]$$
$$\equiv \sigma_k(T_k^{rs})/\sigma_k(1)$$

where

$$\sigma_k(H_k) := \overline{E}[\Lambda_k^{C_1} H_k|\mathscr{Y}_k].$$

With $H_k \equiv T_k^{rs}$, for Theorem 5.3 we have $\alpha_k = \langle X_{k-1}, e_r\rangle\, x_{k-1}^s, \quad \beta_k = \delta_k = 0$. Then

$$\sigma_k(T_k^{rs}X_k) = \sum_{j=1}^{N} \Gamma^j(y_k)\,\langle \sigma_{k-1}(T_{k-1}^{rs}X_{k-1}), e_j\rangle\,\pi_j$$
$$+\Gamma^r(y_k)\,\langle \sigma_{k-1}(X_{k-1}, e_r\rangle\, x_{k-1}^s \pi_r$$
$$\sigma_k(X_k) = \sum_{j=1}^{N} \Gamma^j(y_k)\,\langle \sigma_{k-1}(X_{k-1}), e_j\rangle\,\pi_j$$

where $\pi_j = \Pi e_j, \quad \Gamma^j(y_k)$ is given by $\phi(y_k - Cx_k - He_j)/\phi(y_k)$, with ϕ as in (5). We then have

$$\sigma_k(T_k^{rs}) = \langle \sigma_k(T_k^{rs} X_k), 1 \rangle$$
$$\sigma_k(1) = \langle \sigma_k(X_k), 1 \rangle$$

as usual.

Summarizing the results so far we can state the following:

Proposition 2. *Given Π, H the log-likelihood estimates $\widehat{A}$ and $\widehat{C}$, given observations $x_0, x_1, \ldots, x_k$ and $y_0, y_1, \ldots, y_k$, are determined by* (12) *and* (15).

□

Once C has been estimated, we can estimate Π, H using the HMM methodology of [3]. In fact setting $z_k = y_k - Cx_k$ we can estimate Π, H with the observations

$$z_k = HX_k + b_k \tag{16}$$

using the procedures of [3], pages 68-70.

Thus, estimates of the model parameters can be made in the order: A, C, Π, H, and then updated in this order given new observations. An issue for further investigation is the estimation of N. We expect that estimates of N will be smaller for the case $n \geq 1$.

4. Asset allocation

Suppose we have a vector $r_{k+1} \in \mathbb{R}^m$ of rates of return given by (3). We wish to determine at time k a vector $w \in \mathbb{R}^m$ with $w'\mathbf{1} = 1$, $\mathbf{1} = (1, 1, \ldots, 1)'$ which maximizes

$$J(w) = \nu E[w' r_{k+1} | \mathscr{Y}_k] - \text{var}[w' r_{k+1} | \mathscr{Y}_k], \tag{17}$$

for some $\nu > 0$. This expresses the utility of the rate of return of a portfolio in which wealth is distributed among the m asset classes in ratios expressed by w. Writing $\widehat{X}_k = E[X_k | \mathscr{Y}_k]$ we can estimate this objective function and compute the optimal w. Note such a choice $w = w_k$ makes $\{w_k\}_{k \in \mathbb{N}}$ a predictable sequence of decision variables. In fact

$$\begin{aligned} J(w) &= \nu E[w' r_{k+1} | \mathscr{Y}_k] - E[(w' r_{k+1})^2 | \mathscr{Y}_k] \\ &\quad + (E[w' r_{k+1} | \mathscr{Y}_k])^2 \\ &= \nu w' \widehat{r}_{k+1} + w' \widehat{r}_{k+1} \widehat{r}'_{k+1} w \\ &\quad - w' E[r_{k+1} r'_{k+1} | \mathscr{Y}_k] w \end{aligned} \tag{18}$$

where $\widehat{r}_{k+1} = E[r_{k+1} | \mathscr{Y}_k] = CAx_k + H\Pi\widehat{X}_k$, and $\widehat{X}_k = E[X_k | \mathscr{Y}_k]$. Now by (1), (2)

$$\begin{aligned} E[r_{k+1} r'_{k+1} | \mathscr{Y}_k] &= E[CAx_k x'_k A'C' + C w_{k+1} w'_{k+1} C' + H\Pi X_k X'_k \Pi' H' \\ &\quad + HM_{k+1} M'_{k+1} H' + 2CAx_k X'_k \Pi' H' | \mathscr{Y}_k] \end{aligned} \tag{19}$$

where we have used $E[w_{k+1} | \mathscr{Y}_k] = 0 \in R^n$,

$$E[M_{k+1}|\mathscr{Y}_k] = 0 \in R^N \quad \text{and} \quad E[w_{k+1}M'_{k+1}|\mathscr{Y}_k] = 0 \in R^{n\times N}.$$

Furthermore, $X_k X'_k = \text{diag}\, X_k$, $E[w_{k+1}w'_{k+1}|\mathscr{Y}_k] = D$, and $E[M_{k+1}M'_{k+1}|\mathscr{Y}_k] = \text{diag}\, \Pi\widehat{X}_k - \Pi\, \text{diag}\, \widehat{X}_k \Pi'$, (see [3], page 20). Consequently,

$$\begin{aligned} E[r_{k+1}r'_{k+1}|\mathscr{Y}_k] &= CAx_k x'_k A'C' + CDC' + H\Pi[\text{diag}\, \widehat{X}_k]\Pi' H' \\ &\quad + H[\text{diag}\, \Pi\widehat{X}_k]H' + 2CAx_k \widehat{X}'_x \Pi' H'. \end{aligned} \tag{20}$$

Noting

$$\begin{aligned} (w'\widehat{r}_{k+1})^2 &= w'\widehat{r}_{k+1}\widehat{r}'_{k+1}w \\ &= CAx_k x'_k A'C' + H\Pi\widehat{X}_k\widehat{X}'_k\Pi' H' \\ &\quad + 2CAx_k\widehat{X}'_k\Pi' H', \end{aligned}$$

then

$$J(w) = w'K - w'Vw$$

where

$$K = \nu\widehat{r}_{k+1} = \nu[CAx_k + H\Pi\widehat{X}_k] \tag{21}$$

and

$$\begin{aligned} V &= CDC' + H(\text{diag}\, \Pi\widehat{X}_k)H' \\ &\quad + H\Pi E[(X_k - \widehat{X}_k)(X_k - \widehat{X}_k)'|\mathscr{Y}_k]\Pi' H'. \end{aligned} \tag{22}$$

We can assume that C has rank m, and CDC' is positive definite (as D is). The other terms in V are non-negative definite so V is positive definite and hence invertible. The maximum of $J(w)$, subject to $w' \cdot \mathbf{1} = 1$, is given by the first order (Kuhn-Tucker) conditions:

$$K - 2Vw + \lambda\mathbf{1} = 0 \tag{23}$$

$$w' \cdot \mathbf{1} = 1 \tag{24}$$

for some Lagrange multipler λ. Solving (23) and (24) we obtain

$$w = \frac{1}{2} V^{-1}[K + \lambda\mathbf{1}], \tag{25}$$

with

$$\lambda = [2 - K'V^{-1}\mathbf{1}]/(\mathbf{1}'V^{-1}\mathbf{1}). \tag{26}$$

This solves the one-period asset allocation problem. In subsequent periods, we have the option to update estimates for Π, A, C, H. We can update $K = K_k$ and $V = V_k$, (given by (21) and (22)), in terms of updates on $\widehat{X}_k = p_k$, (see (10)), since

$$E[(X_k - \widehat{X}_k)(X_k - \widehat{X}_k)'|\mathscr{Y}_k] = \text{diag}\, \widehat{X}_k - (\widehat{X}_k)^2.$$

References

1. Anderson, W.N. Jr., Kleindorfer, G.B., Kleindorfer, P.R., Woodroofe, M.B.: Consistent estimates of the parameters of a linear system. Ann. Math. Stat. **40**, 2064–2075 (1969)
2. Elliott, R.J.: Exact adaptive filters for Markov chains observed in Gaussian noise. Automatica **30**, 1399–1408 (1994)
3. Elliott, R.J., Aggoun, L., Moore, J.B.: Hidden Markov models: Estimation and control. Berlin, Heidelberg, New York: Springer 1995
4. Liptser, R.S., Shiryayev, A.N.: Statistics of random processes. Vols. I and II. Berlin, Heidelberg, New York: Springer 1977
5. Lütkepohl, H.: Introduction to multiple time series analysis. Berlin, Heidelberg, New York: Springer 1991