TRANSACTIONS OF SOCIETY OF ACTUARIES
1993 VOL. 45

MULTIVARIATE STOCHASTIC IMMUNIZATION THEORY

ROBERT R. REITANO

ABSTRACT

In the current paper, a new theory of immunization is introduced in which the approach is multivariate, and the goal is stochastic in the sense of minimizing stochastic risk. The risk measure utilized is reminiscent of the Markowitz [4] approach of variance minimization but generalized to also reflect a measure of worst-case risk. The approach is multivariate in that full yield curve risk to nonparallel shifts is reflected, as in Reitano [6–13], by modeling the yield curve as a vector of yield curve drivers. Explicit solutions to the risk minimization problems are developed, subject to constraints on portfolio returns and/or various portfolio directional durations. In addition, explicit solutions are determined that can be achieved by trading any given collection of assets. Applications are developed in detail and are exemplified by an analysis of the example introduced in Reitano [11].

1. INTRODUCTION

In Reitano [8], [11], [12], the classical theories of immunization are extended to a general multivariate framework, which allow their application to virtually any model of the dynamics by which the yield curve shifts. The full yield curve is modeled as a vector of yields, reflecting values throughout the yield curve maturity spectrum.

Denoting by $\mathbf{i}_0=(i_1, i_2, \ldots, i_m)$, the current value of an m-point yield curve vector, one model for yield curve shifts is the "directional shift" model, whereby a direction vector of interest, $\mathbf{N}=(n_1, n_2, \ldots, n_m)$, is specified and fixed in advance. The yield curve shift model is then:

$$\mathbf{i}_0 \rightarrow \mathbf{i}_0 + t\mathbf{N} = (i_1 + tn_1, i_2 + tn_2, \ldots, i_m + tn_m), \tag{1.1}$$

with t denoting the variable magnitude of the shift in the direction of $\mathbf{N}$. The classical parallel shift model is a well-known example, in which $\mathbf{N}=(1, 1, \ldots, 1)$.

Another example is the "key rate" model of Ho [2], in which $\mathbf{i}_0$ reflects a complete spot rate specification, and a number of pyramid-shaped direction vectors are specified, with values of 1 at the respective peaks and corresponding to the key rate maturities. The spot rates are initially specified via a regression model that best prices a large collection of bonds, subject to certain smoothness constraints. The "key rate direction vectors" have the property that they sum to the parallel shift direction vector. See also Reitano [10] for more details on this model.

Given any such directional shift model, one can consider duration measures: directional durations in Reitano [6], [10], [12], [13], or key rate durations given the above model in Ho [2], as well as directional convexities in the above Reitano references. In addition, one can then seek conditions under which a portfolio is immunized "in the direction of **N**" locally, or for relatively small shifts, or globally, for all shifts.

An alternative model for yield curve shifts is the full multivariate shift model, whereby $\mathbf{i}_0$ above is assumed to shift by $\mathbf{\Delta i}=(\Delta i_1, \ldots, \Delta i_m)$:

$$\mathbf{i}_0 \rightarrow \mathbf{i}_0 + \mathbf{\Delta i} = (i_1 + \Delta i_1, \ldots, i_m + \Delta i_m). \tag{1.2}$$

Here, no explicit relationships between the various Δi_j are assumed. In this context, partial durations and partial convexities are defined (Reitano [6], [7], [9], [10]) and nondirectional immunization is explored, whereby conditions are developed that ensure immunization against all yield curve shifts, $\mathbf{\Delta i}$, in a local or global context.

In general, these multivariate models can be applied within the context of any yield curve basis—bond yields, spot yields, forward yields—and using any nominal rate basis—semiannual, annual, continuous, and so on. The basis of choice, due to its tractability, accuracy and ease of use, and recommended in the various Reitano references is the "yield curve driver" model. Here, the bond yield curve is modeled in terms of market-based yields at the actively traded maturities, for example: 0.5, 1, 2, 3, 4, 5, 7, 10, 20, and 30 years. Other bond yields then can be interpolated from these 10 or so market observations, and a complete spot yield curve derived in the usual way and used for all valuations of fixed and contingent claims. Consequently, with this approach the complete yield curve can be easily and accurately modeled as a vector with 10 or so components.

In practice, it is just as easy to locally immunize in any direction **N** as it is to locally immunize against parallel shifts. In general, one constraint is required on the respective portfolio directional duration and one constraint on the directional convexity. Unfortunately, as for classical immunization, local directional immunization in any given direction **N** leaves the portfolio vulnerable to shifts in other directions, as well as to large shifts in the given direction. For example, nonparallel yield curve shifts often can lead to the failure of a classical immunization strategy that protects against parallel shifts (Reitano [11], [12]).

On the other hand, local nondirectional immunization, while effective over the full spectrum of conceivable yield curve shifts, may again be limited in the magnitude of shifts protected against. Also, it is difficult and usually expensive to implement because it requires the portfolio to satisfy a number of durational and convexity constraints. For instance, in the case of fixed cash-flow assets and liabilities, the constraints are effectively equivalent to a cash-matching strategy when a complete spot rate model is used. In the context of the "yield curve driver" model, nondirectional immunization requires average cash-matching over "sections" of the yield curve.

Two approaches currently exist for overcoming the implementation difficulties and/or impracticalities of nondirectional immunization, while improving portfolio protection vis-a-vis directional immunization.

In Litterman and Scheinkman [3], a principal component analysis (Theil [14], Wilks [15]) is conducted on historic yield curve vector shifts over discrete time periods (monthly, for instance). In this analysis, the first principal component, $\mathbf{N}_1$, is the vector shift such that the historical shifts are best approximated by its multiples, $t\mathbf{N}_1$, in the least-squares sense. Multiples of the second principal component, $\mathbf{N}_2$, then best approximate the residuals left over from the first principal components, again in the least-squares sense, and so on for other principal components.

In other words, this analysis decomposes historic shifts, $\Delta\mathbf{i}_j$, into a linear combination of vector shifts, $\mathbf{N}_1, \mathbf{N}_2, \mathbf{N}_3, \ldots, \mathbf{N}_m$, so that for each $k=1, \ldots, m-1$, the collection of residuals:

$$\{\Delta\mathbf{i}_j - \Sigma t_{ij}\, \mathbf{N}_i\},\ j = 1, 2, \ldots \tag{1.3}$$

is as small as possible in the sense of sums of squares, where the summation runs from $i=1$ to $i=k$.

Using as few as three components, the above authors show that a significant proportion of the total actual return on certain bonds is explained, implying that these direction vectors may be used effectively to anticipate bond performance and hence can perhaps be used in an immunization strategy to control performance.

Once these components are identified, the above referenced papers (Reitano [8], [11], [12]) provide the necessary criteria for immunization in all such principal component directions simultaneously.

In the current paper, a new approach is introduced that again has as its basis an historic database of yield curve vector shifts over discrete periods and the associated covariance matrix. It is shown that one can then approximate the variance of surplus, or of any net hedged portfolio, using this portfolio's partial durations and the historic covariance matrix. Consequently, reflecting the Markowitz [4] approach to risk minimization, one goal of multivariate stochastic immunization is to structure the partial durations of the portfolio so that its variance is minimized, subject to given portfolio constraints on expected return, and/or directional duration values in various directions. More generally, what can be minimized subject to such constraints is a risk measure defined as the weighted average of the variance of the portfolio and a measure of worst-case risk.

Explicit solutions to the constrained risk minimization problems are obtained based on a general result proved in the Appendix. This proposition provides an explicit solution to the problem of minimizing a positive definite quadratic form subject to an arbitrary number of compatible linear constraints (see also Martin et al. [5, p. 683]).

The theory and applications are first explored in the case of many tradable assets. That is, it is assumed that sufficient assets exist to allow all targeted partial duration structures to be realized by trading. Such target total duration vectors are developed in detail, as is the method for determining the trades of the given assets that will produce the desired result.

These methods are then generalized to the case of fewer tradable assets. Here, the problem is to minimize variance and/or worst-case risk subject not only to the above constraints but also to the constraint that the resulting target total duration vector is achievable by trading the given assets.

The example introduced in Reitano [11] is then analyzed in detail, illustrating the application of a variety of results.

For a review of other stochastic approaches to duration-matching and dedication, see Hiller and Schaak [1]. In general, these methods were developed to handle asset/liability management with stochastic cash flows over long time periods, given a model of future yield curve dynamics and a model for how future cash flows will vary with the levels of prevailing yields. These models often reflect the classical approach to duration in terms of parallel shifts and are implemented by using linear programming methods.

2. STOCHASTIC IMMUNIZATION WITH MANY ASSETS

In this section, we investigate stochastic immunization for the case in which sufficient tradable assets exist to pose no additional constraints on the durational structures achievable, that is, in the case in which any targeted partial duration structure can be obtained by a feasible market trade.

A. General Model: A Priori Estimates

Let $P(\mathbf{i})$ be a price function defined on an m-point yield curve vector, $\mathbf{i}=(i_1, \ldots, i_m)$, where $\mathbf{i}_0$ denotes its current value. We assume for convenience that $P(\mathbf{i}_0)\neq 0$, although, as shown below (Section 2-E), this restriction can be circumvented in applications. As developed in Reitano [6], [7], [9], [10], the ratio $P(\mathbf{i}_0+\Delta\mathbf{i})/P(\mathbf{i}_0)$ can be linearly approximated by $R(\Delta\mathbf{i})$:

$$R(\Delta\mathbf{i}) = 1 - \mathbf{D}(\mathbf{i}_0) \cdot \Delta\mathbf{i}. \tag{2.1}$$

In (2.1), $\Delta\mathbf{i}$ represents an arbitrary yield curve shift, $\Delta\mathbf{i}=(\Delta i_1, \ldots, \Delta i_m)$, and $\mathbf{D}(\mathbf{i}_0)=[D_1(\mathbf{i}_0), D_2(\mathbf{i}_0), \ldots, D_m(\mathbf{i}_0)]$ is the total duration vector whose components are partial durations:

$$D_j(\mathbf{i}_0) = -d_jP(\mathbf{i}_0)/P(\mathbf{i}_0), \tag{2.2}$$

where the $d_jP(\mathbf{i}_0)$ are the corresponding partial derivatives of $P(\mathbf{i})$. For matrix calculations, $\mathbf{D}(\mathbf{i}_0)$ is treated as a row vector, while all other vectors are treated as column vectors. In (2.1), $\mathbf{D}(\mathbf{i}_0)\cdot\Delta\mathbf{i}$ denotes the standard dot or inner product of two vectors: $\mathbf{x}\cdot\mathbf{y}=\Sigma x_j y_j$.

Assume that $\Delta\mathbf{i}$ has a given probability density function, $f(\Delta\mathbf{i})$, which may depend on $\mathbf{i}_0$, and let $\mathbf{E}(\mathbf{i}_0)$ and $\mathbf{K}(\mathbf{i}_0)$ be defined as the mean value vector of $\Delta\mathbf{i}$ and the associated covariance matrix, which are also assumed to exist:

$$E_j(\mathbf{i}_0) = E[\Delta i_j], \qquad j = 1, \ldots, m \tag{2.3}$$

$$K_{jk}(\mathbf{i}_0) = E[(\Delta i_j - E_j(\mathbf{i}_0))(\Delta i_k - E_k(\mathbf{i}_0))], \qquad j, k = 1, \ldots, m \tag{2.4}$$

where E denotes the usual expectations operator. Note that $\mathbf{K}(\mathbf{i}_0)$ has the important property of positive semi-definiteness. That is, $\mathbf{x}^T\mathbf{K}\mathbf{x}\geq 0$ for all vectors $\mathbf{x}$, where $\mathbf{x}^T$ denotes the row vector transpose of the column vector $\mathbf{x}$.

While positive semi-definiteness is a useful property, it is in general too weak to ensure the invertibility of $\mathbf{K}(\mathbf{i}_0)$, which the methods of this paper require. It is straightforward to show that if $\mathbf{K}(\mathbf{i}_0)$ is positive semi-definite, then $\mathbf{x}^T\mathbf{K}\mathbf{x}=0$ if and only if $\mathbf{K}\mathbf{x}=0$. Hence, in the current context, $\mathbf{K}(\mathbf{i}_0)$ will be invertible if and only if it is positive definite; that is, $\mathbf{x}^T\mathbf{K}\mathbf{x}=0$ if and only if $\mathbf{x}=\mathbf{0}$.

Henceforth, the positive definiteness of $\mathbf{K}(\mathbf{i}_0)$ is assumed. What this assumption implies about the random vector $\mathbf{\Delta i}$ is that no linear combination of the Δi_j variates is degenerate, or nonstochastic. In practice, this condition is virtually ensured for any reasonable yield curve model, as a simple analysis of historic yields demonstrates.

Proposition 1:

Given the above definitions:

$$E[R(\mathbf{\Delta i})] = 1 - \mathbf{D}(\mathbf{i}_0) \cdot \mathbf{E}(\mathbf{i}_0), \tag{2.5}$$

$$\text{Var}[R(\mathbf{\Delta i})] = \mathbf{D}(\mathbf{i}_0)\,\mathbf{K}(\mathbf{i}_0)\,\mathbf{D}^T(\mathbf{i}_0), \tag{2.6}$$

where $\mathbf{D}^T(\mathbf{i}_0)$ represents the column vector transpose of the row vector $\mathbf{D}(\mathbf{i}_0)$.

Proof: By the linearity of the operator E, (2.5) is immediate. Further,

$$\begin{aligned}\text{Var}(R(\mathbf{\Delta i})) &= \text{Var}[\mathbf{D}(\mathbf{i}_0) \cdot \mathbf{\Delta i}] \\ &= E[(\mathbf{D}(\mathbf{i}_0) \cdot (\mathbf{\Delta i} - \mathbf{E}(\mathbf{i}_0))^2] \\ &= E[\Sigma\, \Sigma\, D_j(\mathbf{i}_0) D_k(\mathbf{i}_0)(\Delta i_j - E_j(\mathbf{i}_0))(\Delta i_k - E_k(\mathbf{i}_0))] \\ &= \mathbf{D}(\mathbf{i}_0)\,\mathbf{K}(\mathbf{i}_0)\,\mathbf{D}^T(\mathbf{i}_0),\end{aligned}$$

by the linearity of E and (2.4). □

Expression (2.6) provides the exact variance of $R(\mathbf{\Delta i})$, where $R(\mathbf{\Delta i})$ is a linear approximation to $P(\mathbf{i}_0+\mathbf{\Delta i})/P(\mathbf{i}_0)$ given in (2.1). It is natural to inquire, therefore, how well this expression approximates the actual variance of this price function ratio.

Using methods of multivariate stochastic calculus, where the Δi_j values are assumed to follow correlated generalized Ito processes, it is possible to show that (2.6) equals the instantaneous variance of $P(\mathbf{i}_0+\mathbf{\Delta i})/P(\mathbf{i}_0)$, at $t=0$.

Proposition 2:

Given the definitions above:

$$|E[R(\mathbf{\Delta i})] - 1| \leq |\mathbf{D}(\mathbf{i}_0)||\mathbf{E}(\mathbf{i}_0)|, \tag{2.7}$$

$$\text{Var}[R[\mathbf{\Delta i})] \leq |\mathbf{D}(\mathbf{i}_0)|^2 \, tr[\mathbf{K}(\mathbf{i}_0)], \tag{2.8}$$

where $tr[\mathbf{K}(\mathbf{i}_0)]$ is the trace of $\mathbf{K}(\mathbf{i}_0)$, or sum of the diagonal elements, and $|\mathbf{x}|$ is the usual Euclidean norm, $|\mathbf{x}|^2=\mathbf{x}\cdot\mathbf{x}$.

Proof: By the Cauchy-Schwarz inequality,

$$|\mathbf{D}(\mathbf{i}_0) \cdot \mathbf{E}(\mathbf{i}_0)| \leq |\mathbf{D}(\mathbf{i}_0)|\,|\mathbf{E}(\mathbf{i}_0)|,$$

from which (2.7) is immediate. Similarly, we have that

$$[\mathbf{D}(\mathbf{i}_0) \cdot (\mathbf{\Delta i} - \mathbf{E}(\mathbf{i}_0))]^2 \leq |\mathbf{D}(\mathbf{i}_0)|^2|\mathbf{\Delta i} - \mathbf{E}(\mathbf{i}_0)|^2,$$

and hence,

$$\text{Var}[R(\mathbf{\Delta i})] \leq |\mathbf{D}(\mathbf{i}_0)|^2 \, E[|\mathbf{\Delta i} - \mathbf{E}(\mathbf{i}_0)|^2],$$

from which we obtain (2.8). □

It is clear from Proposition 2 that $E[R(\mathbf{\Delta i})]$ will be close to 1 and $\text{Var}[R(\mathbf{\Delta i})]$ close to 0, if the length of the total duration vector, $|\mathbf{D}(\mathbf{i}_0)|$, is made small. This length was shown to be related to another measure of risk in Reitano [6], [7], [9], [10]. Specifically, if $\mathbf{N}$ is a direction vector, $\mathbf{N}\neq\mathbf{0}$, and $D_N(\mathbf{i}_0)$ the directional duration of $P(\mathbf{i})$ in the direction of $\mathbf{N}$, $D_N(\mathbf{i}_0)=\mathbf{D}(\mathbf{i}_0)\cdot\mathbf{N}$, and we have by (2.1) that:

$$|D_N(\mathbf{i}_0)| = |R(\mathbf{N}) - 1| \leq |\mathbf{D}(\mathbf{i}_0)|\,|\mathbf{N}|. \tag{2.9}$$

Also, all values of $D_N(\mathbf{i}_0)$ in this implied interval are possible in theory, for various $\mathbf{N}$ of the given length.

Based on (2.9), $|\mathbf{D}(\mathbf{i}_0)|$ is seen to be a measure of extreme durational sensitivity for $P(\mathbf{i})$, for shift vectors $\mathbf{N}$ of fixed length. Consequently, when $|\mathbf{D}(\mathbf{i}_0)|$ is made small, first-order sensitivity of $P(\mathbf{i})$ is made relatively small in even the most severe directions, independent of their potential likelihood as implied by $f(\mathbf{\Delta i})$. Equivalently, using (2.9), an upper bound for $|R(\mathbf{\Delta i})-E[R(\mathbf{\Delta i})]|$ can be easily seen to depend on $|\mathbf{D}(\mathbf{i}_0)|$:

$$|R(\mathbf{\Delta i}) - E[R(\mathbf{\Delta i})]| \leq |\mathbf{D}(\mathbf{i}_0)||\mathbf{\Delta i} - \mathbf{E}(\mathbf{i}_0)|.$$

In the following, two portfolio price risk measures are considered for minimization: $\text{Var}(R(\mathbf{\Delta i}))$ and $|\mathbf{D}(\mathbf{i}_0)|^2$, subject to additional constraints.

For economy of argument, we combine these measures using a weighting parameter w, $0 \leq w \leq 1$, as follows:

$$w \operatorname{Var}[R(\mathbf{\Delta i})] + (1 - w)|\mathbf{D}(\mathbf{i}_0)|^2. \qquad (2.10)$$

Intuitively, for w near 1, the results below minimize the variance of $R(\mathbf{\Delta i})$ subject to the given constraints. For w close to 0, $|\mathbf{D}(\mathbf{i}_0)|^2$ is minimized subject to constraints, thereby minimizing "worst-case" durational sensitivity.

Proposition 3:

There exists a positive definite matrix, $\mathbf{K}_w(\mathbf{i}_0)$, such that:

$$w \operatorname{Var}(R(\mathbf{\Delta i})) + (1 - w)|\mathbf{D}(\mathbf{i}_0)|^2 = \mathbf{D}(\mathbf{i}_0)\, \mathbf{K}_w(\mathbf{i}_0)\, \mathbf{D}(\mathbf{i}_0)^T. \qquad (2.11)$$

Proof: By (2.6) and the fact that $|\mathbf{D}(\mathbf{i}_0)|^2 = \mathbf{D}(\mathbf{i}_0)\, \mathbf{I}\, \mathbf{D}(\mathbf{i}_0)^T$, with $\mathbf{I}$ equal to the identity matrix, it is clear that:

$$\mathbf{K}_w(\mathbf{i}_0) = w\, \mathbf{K}(\mathbf{i}_0) + (1 - w)\, \mathbf{I}. \qquad (2.12)$$

To see that $\mathbf{K}_w(\mathbf{i}_0)$ is positive definite, assume that $0<w<1$, as the result is obvious by definition if $w=0$ or 1. If there is a vector $\mathbf{x}$ so that $\mathbf{x}^T\, \mathbf{K}_w(\mathbf{i}_0)\mathbf{x} \leq 0$, we then have by linearity and (2.12):

$$w\, \mathbf{x}^T\, \mathbf{K}(\mathbf{i}_0)\, \mathbf{x} \leq -(1 - w)|\mathbf{x}|^2.$$

Because the covariance matrix $\mathbf{K}(\mathbf{i}_0)$ is positive definite, the above inequality is satisfied only when $\mathbf{x}=\mathbf{0}$. □

Two types of constraints are considered for the minimization problem below:

$$\mathbf{D}(\mathbf{i}_0) \cdot \mathbf{E}(\mathbf{i}_0) = r, \qquad (2.13)$$

$$\mathbf{D}(\mathbf{i}_0) \cdot \mathbf{N} = D, \qquad (2.14)$$

where $\mathbf{N} \neq \mathbf{0}$ is a given direction vector. Constraint (2.13) fixes the expected value of $R(\mathbf{\Delta i})$ to equal $1-r$ by (2.5), while (2.14) constrains a given directional duration, $D_N(\mathbf{i}_0)$, to equal D. When $\mathbf{N}=(1, \ldots, 1)$, the traditional parallel shift (that is, modified) duration is constrained by (2.14).

For notational convenience, we will often suppress the dependence of the various functions on $\mathbf{i}_0$.

B. Single-Constraint Minimization

Proposition 4:

Given $\mathbf{N} \neq \mathbf{0}$, the solution to:

$$\text{Min}(\mathbf{D}\,\mathbf{K}_w\,\mathbf{D}^T), \text{ subject to: } \mathbf{D} \cdot \mathbf{N} = D, \tag{2.15}$$

is given by the total duration vector:

$$\mathbf{D}_0^T = \frac{D}{\mathbf{N}^T\,\mathbf{K}_w^{-1}\,\mathbf{N}}\,\mathbf{K}_w^{-1}\,\mathbf{N}. \tag{2.16}$$

Further, the minimum value in (2.15) is given by:

$$\mathbf{D}_0\,\mathbf{K}_w\,\mathbf{D}_0^T = D^2/(\mathbf{N}^T\,\mathbf{K}_w^{-1}\,\mathbf{N}). \tag{2.17}$$

Proof: Since $\mathbf{K}_w$ is positive definite, we can apply Proposition A of the Appendix directly. For this single constraint case, $\mathbf{C}$ in (A.4) is the one-element matrix, $\mathbf{C}=\mathbf{N}^T\,\mathbf{K}_w^{-1}\,\mathbf{N}$. Hence, (2.16) is given by (A.3). Similarly, (2.17) follows directly from (A.21) of Corollary A.1. □

Although it is not true in practice, consideration of the special case in which $\mathbf{K}(\mathbf{i}_0)$ is a diagonal matrix provides insight to the above formulas. In this case of uncorrelated $\{\Delta i_j\}$, the inverse of $\mathbf{K}_w(\mathbf{i}_0)$ also is a diagonal matrix, which is given by:

$$K_w^{-1}(\mathbf{i}_0)_{jj} = 1/(w\sigma_j^2 + (1 - w)), \tag{2.18}$$

where $\sigma_j^2 = K_{jj}(\mathbf{i}_0)$ is the variance of Δi_j. Hence, the partial durations of the risk-minimizing total duration vector given in (2.16) are:

$$D_{0j} \approx n_j/(w\sigma_j^2 + (1 - w)), \tag{2.19}$$

with proportionality constant: $D/\mathbf{N}^T\,\mathbf{K}_w^{-1}\,\mathbf{N}$.

Consequently, for $w=1$ and variance minimization, the components of $\mathbf{D}_0(\mathbf{i}_0)$ vary in inverse proportion to σ_j^2. For example, when $\mathbf{N}=(1, \ldots, 1)$ and traditional duration is constrained in (2.15), less partial duration exposure is allowed to the more variable yield curve points, and more exposure is allowed to less variable yield curve points.

When $w=0$ and worst-case risk is minimized, $\mathbf{D}_0(\mathbf{i}_0)$ in (2.19) is proportional to $\mathbf{N}$. This result generalizes the conclusion in Reitano [6], [10] that if $\mathbf{N}=(1, \ldots, 1)$, $|\mathbf{D}(\mathbf{i}_0)|$ is minimized subject to $\mathbf{D}\cdot\mathbf{N}=D$ if all partial durations equal D/m, that is, when $\mathbf{D}_0(\mathbf{i}_0)$ is proportional to $\mathbf{N}$.

Returning to general $\mathbf{K}_w(\mathbf{i}_0)$, note that because of positive definiteness, $\mathbf{D}_0(\mathbf{i}_0)$ in (2.16) is the zero vector if and only if $D=0$. Also note that the total duration vector given in (2.16) produces an expected return, $E[R(\mathbf{\Delta i})]=1-r$, where:

$$r = \frac{\mathbf{E}^T \mathbf{K}_w^{-1} \mathbf{N}}{\mathbf{N}^T \mathbf{K}_w^{-1} \mathbf{N}} D, \tag{2.20}$$

which varies proportionately with D. Since $\mathbf{K}_w^{-1}$ is also positive definite, the denominator in (2.20) must be positive since $\mathbf{N}\neq\mathbf{0}$ by assumption. However, the numerator is clearly 0 when $\mathbf{E}=\mathbf{0}$ and also can attain either sign.

To see this, consider again the simple case in which $\mathbf{K}_w$ is diagonal and $\mathbf{N}=(1, \ldots, 1)$. We then have:

$$\mathbf{E}^T \mathbf{K}_w^{-1} \mathbf{N} = \Sigma E_j/(w\sigma_j^2 + 1 - w), \tag{2.21}$$

which can clearly assume positive or negative values, depending on the signs of the E_j values. In addition, it is clear from (2.21) that $\mathbf{E}^T \mathbf{K}_w^{-1} \mathbf{N}$ may equal 0 even if $\mathbf{E}\neq\mathbf{0}$.

Finally, from (2.17) one notion of an "efficient frontier" can be defined in the duration management context. For fixed $\mathbf{N}$, (2.17) can be regarded as defining the minimum risk measure as a quadratic function $D=D_N(\mathbf{i}_0)$. Consequently, for any such value of $D_N(\mathbf{i}_0)$, all portfolios with the given directional duration will have risk measures equal to or exceeding the value on this curve. From this perspective, therefore, the portfolios with durational structures given by (2.16) are "efficient," in that they are risk minimizing for each value of $D=D_N$. Clearly, these efficient frontiers depend on the direction vector constrained.

Given $\mathbf{N}$ and $\mathbf{E}$, if $\mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{N}\neq 0$, it is possible to combine (2.20) with (2.17) and also express the risk measure of these portfolios as functions of their implied returns:

$$\mathbf{D}_0 \mathbf{K}_w \mathbf{D}_0^T = \frac{\mathbf{N}^T \mathbf{K}_w^{-1} \mathbf{N}}{(\mathbf{E}^T \mathbf{K}_w^{-1} \mathbf{N})^2} r^2. \tag{2.17$'$}$$

That is, for each r this formula expresses the value of the risk measure of a certain portfolio with $\mathbf{D}\cdot\mathbf{E}=r$, where this portfolio is chosen to be risk minimizing given its value of $\mathbf{D}\cdot\mathbf{N}=D$.

It is natural to ask whether the above equation also represents an efficient frontier in "risk-r" space. That is, given r, does this formula provide the minimal risk of all portfolios with $\mathbf{D}\cdot\mathbf{N}=r$? We will see below that the answer is, in general, no.

By simply interchanging **E** and **N**, we have the following:

Proposition 5:

Given $\mathbf{E}\neq\mathbf{0}$, the solution to:

$$\text{Min}(\mathbf{D}\,\mathbf{K}_w\,\mathbf{D}^T),\ \text{subject to: } \mathbf{D}\cdot\mathbf{E}=r, \tag{2.22}$$

is given by:

$$\mathbf{D}_0^T=\frac{r}{\mathbf{E}^T\,\mathbf{K}_w^{-1}\,\mathbf{E}}\,\mathbf{K}_w^{-1}\,\mathbf{E}. \tag{2.23}$$

Further, the minimum value in (2.22) is given by:

$$\mathbf{D}_0\,\mathbf{K}_w\,\mathbf{D}_0^T=r^2/(\mathbf{E}^T\,\mathbf{K}_w^{-1}\,\mathbf{E}).\quad\square \tag{2.24}$$

Returning to the efficient frontier discussion above, (2.24) defines the frontier as a function of r, in the same way that (2.17) defines this frontier as a function of $D=D_N(\mathbf{i}_0)$. Comparing (2.24) with (2.17)′, the question raised earlier can now be addressed. Specifically, the curve in (2.17)′, derived from the efficient frontier in "risk-D_N" space, is not, in general, an efficient frontier in "risk-r" space; that is:

$$\frac{\mathbf{N}^T\,\mathbf{K}_w^{-1}\,\mathbf{N}}{(\mathbf{E}^T\,\mathbf{K}_w^{-1}\,\mathbf{N})^2}\geq\frac{1}{\mathbf{E}^T\,\mathbf{K}_w^{-1}\,\mathbf{E}},$$

by the Cauchy-Schwarz inequality applied to the inner product: $(\mathbf{x},\mathbf{y})_K=\mathbf{x}\mathbf{K}_w^{-1}\mathbf{y}$, discussed in (A.17) of the Appendix. Also, equality holds if and only if **E** and **N** are colinear.

C. Double-Constraint Minimization

Before considering a minimization problem subject to both constraints (2.13) and (2.14), or two constraints of type (2.14), note that such conditions need not be compatible. For example, if $\mathbf{N}=(1,\ \ldots,\ 1)$, and $\mathbf{E}=\Delta i\ \mathbf{N}$, it is clear that this problem can be solved only if $r=D\Delta i$, since otherwise the constraint set is empty. In this case, only one constraint can be formally used as in Section 2 above.

In general, compatibility is ensured if **E** and **N** are not proportional, that is, if they are linearly independent.

Proposition 6:

Given $\mathbf{N}\neq\mathbf{0}$, $\mathbf{E}\neq\mathbf{0}$, and **N**, **E** linearly independent, the solution to:

$$\text{Min}(\mathbf{D}\,\mathbf{K}_w\,\mathbf{D}^T) \text{ subject to: } \mathbf{D}\cdot\mathbf{N} = D,\ \mathbf{D}\cdot\mathbf{E} = r, \tag{2.25}$$

is given by the total duration vector:

$$\mathbf{D}_0^T = \lambda_1\,\mathbf{K}_w^{-1}\,\mathbf{N} + \lambda_2\,\mathbf{K}_w^{-1}\,\mathbf{E}, \tag{2.26}$$

where:

$$\lambda_1 = \frac{(\mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{E})D - (\mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{N})r}{(\mathbf{N}^T\mathbf{K}_w^{-1}\mathbf{N})(\mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{E}) - (\mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{N})^2} \tag{2.27}$$

$$\lambda_2 = \frac{(\mathbf{N}^T\boldsymbol{K}_w^{-1}\mathbf{N})r - (\mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{N})D}{(\mathbf{N}^T\mathbf{K}_w^{-1}\mathbf{N})(\mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{E}) - (\mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{N})^2}.$$

Further, we have for this total duration vector:

$$\mathbf{D}_0\,\mathbf{K}_w\,\mathbf{D}_0^T = \lambda_1^2\,\mathbf{N}^T\mathbf{K}_w^{-1}\mathbf{N} + 2\lambda_1\,\lambda_2\,\mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{N} + \lambda_2^2\,\mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{E}. \tag{2.28}$$

Proof: This result is immediate from Proposition A and Corollary A.1 since

$$\mathbf{C} = \begin{pmatrix} \mathbf{N}^T\mathbf{K}_w^{-1}\mathbf{N} & \mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{N} \\ \mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{N} & \mathbf{E}^T\mathbf{K}_w^{-1}\mathbf{E} \end{pmatrix}$$

in this case. □

Note that $\mathbf{D}_0(\mathbf{i}_0)$ in (2.26) reduces to that in (2.16) of Proposition 4 when $\lambda_2=0$, since this occurs exactly when r is equal to the value produced in (2.20) and implied by Proposition 4. An analogous statement holds regarding the restriction on D for which the result of Proposition 6 reduces to that of Proposition 5.

Note also that Proposition 6 can be applied easily in the case of two directional duration constraints as in (2.14), via a simple change in notation. The only requirement is that the direction vectors utilized be linearly independent.

While not apparent from (2.28), the efficient frontier implied by this formula in "risk-(r,D)" space is the three-dimensional analogue of the parabolas of (2.17) and (2.24), or a paraboloid in 3-space. In general, the efficient frontier implied by any model encompassed by Proposition A is a paraboloid in the $(p+1)$ dimensional "risk-**r**" space, as can be demonstrated from (A.21)′ of the Appendix.

To see this, first note that $\mathbf{B}^T\mathbf{K}^{-1}\mathbf{B}$ of (A.21)′ is positive definite in p-dimensional space. This is because $\mathbf{K}$ and hence $\mathbf{K}^{-1}$ are positive definite in m-dimensional space, and $\mathbf{B}$ has rank $p \leq m$ since the column vectors of $\mathbf{B}$, the $\mathbf{B}_j$, are p linearly independent m-vectors. Consequently, $(\mathbf{B}^T\mathbf{K}^{-1}\mathbf{B})^{-1}$ is also positive definite, and the conclusion follows. However, as seen above in the one-dimensional case, the efficient frontier implied by (A.21)′ depends on the collection of vectors in which directions the minimization problem is constrained.

D. Multiple-Constraint Minimization

Clearly, the above propositions can be extended to reflect more constraints by an immediate application of Proposition A in the Appendix. For example, a variety of directional durations can be specified for independent direction vectors, in addition to a period return constraint.

Details are left to the interested reader.

E. Applications

We first consider in some detail the application of minimizing the volatility (or more generally, risk) of surplus. If $P(\mathbf{i})=S(\mathbf{i})$ denotes the price function of surplus or net worth, where $S(\mathbf{i})=A(\mathbf{i})-L(\mathbf{i})$ and $S(\mathbf{i}_0)\neq 0$, the above propositions can be applied directly to minimize its risk. As above, denoting by $\mathbf{D}_0(\mathbf{i}_0)$ the resulting risk-minimizing total duration vector for surplus, a target duration structure for assets can be derived by the identity (see Reitano [6], [8], [10], [11]):

$$\mathbf{D}_0^A(\mathbf{i}_0) = r^s\,\mathbf{D}_0(\mathbf{i}_0) + (1 - r^s)\,\mathbf{D}^L(\mathbf{i}_0), \tag{2.29}$$

where $r^s=S(\mathbf{i}_0)/A(\mathbf{i}_0)$ is the surplus-to-asset ratio or net worth-to-asset ratio, and $\mathbf{D}^L(\mathbf{i}_0)$ is the total duration vector of liabilities.

While it may be tempting to try to extend (2.29) to the case $r^s=0$ by simple substitution, since then $\mathbf{D}_0^A(\mathbf{i}_0)=\mathbf{D}^L(\mathbf{i}_0)$ results and perhaps seems plausible, two problems are encountered:

(1) $r^s\,\mathbf{D}_0(\mathbf{i}_0)$ need not equal 0 since $\mathbf{D}_0(\mathbf{i}_0)$ is not even defined.

(2) Since $\mathbf{D}_0(\mathbf{i}_0)$ totally drops out of the solution, the original constraints need not be satisfied.

To circumvent this difficulty in the case in which $S(\mathbf{i}_0)=0$, we consider a new ratio function:

$$R'(\Delta\mathbf{i}) = -(\mathbf{D}^A(\mathbf{i}_0) - \mathbf{D}^L(\mathbf{i}_0)) \cdot \Delta\mathbf{i}, \tag{2.30}$$

which equals the first-order approximation to the ratio $S(\mathbf{i}_0+\Delta\mathbf{i})/A(\mathbf{i}_0)$. The variance of $R'(\Delta\mathbf{i})$ is then given by (2.6) with $\mathbf{D}=\mathbf{D}^A-\mathbf{D}^L$. Consequently, the propositions can be applied to minimize the weighted average of the variance of $R'(\Delta\mathbf{i})$ and $|\mathbf{D}^A(\mathbf{i}_0)-\mathbf{D}^L(\mathbf{i}_0)|^2$ as in (2.10) subject to constraints on:

$$(\mathbf{D}^A(\mathbf{i}_0) - \mathbf{D}^L(\mathbf{i}_0)) \cdot \mathbf{N},$$

which equals the difference between asset and liability directional durations in the given direction $\mathbf{N}$, and/or:

$$-(\mathbf{D}^A(\mathbf{i}_0) - \mathbf{D}^L(\mathbf{i}_0)) \cdot \mathbf{E}(\mathbf{i}_0),$$

which equals the expected value of $S(\mathbf{i}_0+\Delta\mathbf{i})$ as a percentage of current assets $A(\mathbf{i}_0)$.

As in (2.29), the resulting target total duration vector, $\mathbf{D}_0(\mathbf{i}_0)$, then can be translated to a target for assets, only here using:

$$\mathbf{D}^A(\mathbf{i}_0) = \mathbf{D}_0(\mathbf{i}_0) + \mathbf{D}^L(\mathbf{i}_0). \tag{2.31}$$

We next consider the question of trading assets to achieve the total duration vector target. For this purpose, it is irrelevant whether this target duration vector applies to assets or surplus. That is, we simply assume that $P(\mathbf{i})$ is given with total duration vector $\mathbf{D}(\mathbf{i}_0)=(D_1(\mathbf{i}_0), \ldots, D_m(\mathbf{i}_0))$, and we wish to implement the necessary trades of a collection of assets: $P_1(\mathbf{i}), \ldots, P_n(\mathbf{i})$, so that the transformed price function, $P'(\mathbf{i})$, has total duration vector equal to the given target, $\mathbf{D}_0(\mathbf{i}_0)=(D_{01}(\mathbf{i}_0), \ldots, D_{0m}(\mathbf{i}_0))$.

Let a_j be the amount traded of the asset with price function $P_j(\mathbf{i})$, so that $a_j>0$ corresponds to a purchase and $a_j<0$ to a sale or short position.

Based on the linearity property of total duration vectors:

$$P'(\mathbf{i}_0) = P(\mathbf{i}_0) + \Sigma a_j, \tag{2.32}$$

$$\mathbf{D}'(\mathbf{i}_0) = [P(\mathbf{i}_0)\mathbf{D}(\mathbf{i}_0) + \Sigma a_j \mathbf{D}_j(\mathbf{i}_0)]/P'(\mathbf{i}_0), \tag{2.33}$$

where $\mathbf{D}_j(\mathbf{i}_0)=(D_{j1}(\mathbf{i}_0), \ldots, D_{jm}(\mathbf{i}_0))$ denotes the total duration vector of $P_j(\mathbf{i}_0)$.

Setting the resulting total duration vector in (2.33), $\mathbf{D}'(\mathbf{i}_0)$, equal to the target vector $\mathbf{D}_0(\mathbf{i}_0)$, a system of equations for $\{a_j\}$ results, which is readily seen to equal:

$$\sum a_j[D_{jk}(\mathbf{i}_0) - D_{0k}(\mathbf{i}_0)] = P(\mathbf{i}_0)[D_{0k}(\mathbf{i}_0) - D_k(\mathbf{i}_0)], \tag{2.34}$$

where $k=1, \ldots, m$, and the summations run from $j=1$ to $j=m$.

In general, (2.34) can be solved only if the collection of n total duration vectors: $\{\mathbf{D}_j(\mathbf{i}_0)-\mathbf{D}_0(\mathbf{i}_0)\}$, which form the columns of the coefficient matrix, have rank m. This rank requirement implies that $n \geq m$. That is, the number of assets traded (n) is greater than or equal to the number of yield points in the yield curve vector (m). If $n=m$, the rank requirement means that this collection of vectors above is linearly independent, and in this case, the solution to (2.34) is unique. If $n>m$, there will be infinitely many solutions.

If $(a_1, \ldots, a_n)$ is a solution of (2.34), we will in general have $\Sigma a_j \neq 0$. That is, additional funds will need to be invested if $\Sigma a_j > 0$, while a net divestment will be necessary if $\Sigma a_j < 0$. Alternatively, the net residual investment of Σa_j could be invested or borrowed at overnight REPO rates, introducing virtually no additional interest rate sensitivity to the price function.

Because we require that $n \geq m$ to solve (2.34) in the general case, it is clear that an additional "cash neutral" constraint, $\Sigma a_j = 0$, can be added to this system only if $n \geq m+1$. In this case, we obtain the system:

$$\begin{aligned} \sum a_j D_{jk}(\mathbf{i}_0) &= P(\mathbf{i}_0)[D_{0k}(\mathbf{i}_0) - D_k(\mathbf{i}_0)], \quad k = 1, \ldots, m \\ \sum a_j &= 0. \end{aligned} \tag{2.35}$$

The cash neutral constraint in (2.35) can be explicitly incorporated into the first m equations by setting $a_n = -\Sigma_{j=1}^{n-1} a_j$, for example, producing:

$$\sum_{j=1}^{n-1} a_j[D_{jk}(\mathbf{i}_0) - D_{nk}(\mathbf{i}_0)] = P(\mathbf{i}_0)[D_{0k}(\mathbf{i}_0) - D_k(\mathbf{i}_0)], \tag{2.36}$$

where $k=1, \ldots, m$.

This system can be expressed in matrix notation by:

$$\mathbf{A}\mathbf{a}' = P(\mathbf{i}_0)[\mathbf{D}_0(\mathbf{i}_0) - \mathbf{D}(\mathbf{i}_0)]^T, \tag{2.37}$$

where $\mathbf{a}'=(a_1, \ldots, a_{n-1})$ is the truncated trade vector, and $\mathbf{A}$ is the $m \times (n-1)$ matrix with column vectors equal to $[\mathbf{D}_j(\mathbf{i}_0)-\mathbf{D}_n(\mathbf{i}_0)]^T$, $j=1, \ldots, n-1$.

Recall that total duration vectors are identified with row vectors by convention, and hence the presence of the matrix transpose symbol T to convert them into column vectors in (2.37).

In order for (2.37) to be solvable in general, **A** must have rank m, which as above implies that $n-1 \geq m$, or $n \geq m+1$. If $n=m+1$, this implies that the column vectors, $[\mathbf{D}_j(\mathbf{i}_0)-\mathbf{D}_n(\mathbf{i}_0)]^T$, are linearly independent, or equivalently, that if:

$$\sum_{j=1}^{n} \beta_j \mathbf{D}_j(\mathbf{i}_0) = 0,$$

and

$$\sum_{j=1}^{n} \beta_j = 0,$$

then all $\beta_j=0$.

In practice, the restriction $\Sigma a_j=0$ is usually desirable, and consequently, (2.36) will represent the system of equations to be solved. For example, if $P(\mathbf{i})$ is a non-zero surplus function with current total duration vector $\mathbf{D}(\mathbf{i}_0)$, Proposition 4 could be applied to preserve the current modified duration $D(\mathbf{i}_0)$, that is, with the constraint, $\mathbf{D}\cdot\mathbf{N}=D(\mathbf{i}_0)$, where $\mathbf{N}=(1, 1, \ldots, 1)$ and $D(\mathbf{i}_0)$ is equal to the given modified duration. The resulting target total duration vector, $\mathbf{D}_0(\mathbf{i}_0)$, then could be obtained with a cash-neutral trade via (2.36), eliminating the need to add or remove investments or borrow/invest in overnight funds.

Alternatively, one might choose to determine a target total duration vector so that surplus will have the same modified duration and expected return as a 5-year bond, say. Letting $\mathbf{D}^B(\mathbf{i}_0)$ represent the total duration vector of the given security, Proposition 6 could then be applied with $D=\mathbf{D}^B(\mathbf{i}_0)\cdot\mathbf{N}$, $\mathbf{N}=(1, \ldots, 1)$, and $r=\mathbf{D}^B(\mathbf{i}_0)\cdot\mathbf{E}(\mathbf{i}_0)$. To do this, all that is required is that $\mathbf{N}$ and $\mathbf{E}(\mathbf{i}_0)$ are linearly independent.

Naturally, other constraints could be imposed, fixing a given directional duration and/or expected period return. Two directional durations in the direction of $\mathbf{N}_1$ and $\mathbf{N}_2$ could also be fixed with Proposition 6, by changing notation, $\mathbf{N}=\mathbf{N}_1$, $\mathbf{E}=\mathbf{N}_2$, and with D and r denoting the desired constrained values. For example, one might fix the modified duration and a "tilt" duration to equal those of a 10-year bond, whereby $\mathbf{N}_1=(1, \ldots, 1)$, and $\mathbf{N}_2=(n_1, n_2, \ldots, n_m)$, $n_j \geq n_{j-1}$ (see Reitano [6], [10]). To fix more directional durations, Proposition A in the Appendix can be applied directly.

The above surplus portfolio optimization also can be accomplished simultaneously with the hedging of new liabilities or the establishment of an initial asset portfolio. For example, let $P(\mathbf{i})$ represent the surplus function obtained when liabilities are augmented by the new contract sales, say, and assets are increased by an associated cash position, modeled as if held in 6-month instruments. The application of the above propositions then will simultaneously target the initial duration structure and/or return requirements, as defined via the problem's constraints, as well as optimize the portfolio for minimal risk as implied by $\mathbf{K}_w(\mathbf{i}_0)$.

For the application of creating a hedge for a given security, one creates a price function, $P(\mathbf{i})=H(\mathbf{i})-A(\mathbf{i})$, where $H(\mathbf{i})$ equals the price function of the asset or liability to be hedged and $A(\mathbf{i})$ the price function of a 6-month investment, initially "proposed" as the hedging asset. Because $P(\mathbf{i}_0)=0$ by definition, the optimization problem is applied to $R'(\Delta\mathbf{i})$ in (2.30). Constraints then are defined as above, and a target duration vector $\mathbf{D}_0(\mathbf{i}_0)$ is determined. The desired hedging asset has target duration vector given in (2.31), with $\mathbf{D}^H=\mathbf{D}^L$. Asset trades then are determined by using (2.36) and represent trades from the initial 6-month investments. Because $\Sigma a_j=0$, the resultant hedged portfolio price function, $P(\mathbf{i})$, will again satisfy $P(\mathbf{i}_0)=0$ as desired.

The above techniques also can be applied to the problem of structuring a minimum variance asset pool, subject to constraints on its period return, and/or directional durational(s). Such an application might arise in establishing an asset portfolio for traditional life products or in structuring a separate account vehicle or mutual fund.

As a final point, note that if the resulting trade vector has negative components, a variety of approaches can be investigated. First and most directly, the associated asset can be sold "short." If impractical or not allowed by investment policy or statute, an alternative approach would be to develop a comparable "synthetic" position. For example, if the associated asset is a 10-year bond, one might "short" a financial note futures position. For total duration vector purposes, this asset is approximated by a short Treasury note and long cash position with zero initial market value, thereby ignoring the yield effect of the cheapest-to-deliver option. To be more exact, this option could be hedged out of the futures contract by using a financial options contract.

Another alternative is to approximate the short asset by an interest rate swap in which the floating rate is received and fixed rate paid.

Finally, if possible, a liability could be sold that approximates the durational structure of the asset to be shorted. This liability may be similar to those already in the account, such as in the case of a GIC or annuity pool, or simply be a traditional debt instrument issued by the account to other accounts in the corporation, or directly to the financial markets.

3. STOCHASTIC IMMUNIZATION WITH FEWER ASSETS

A. *General Model: Change of Variables*

As in Section 2, we again seek to minimize the objective function in (2.11). The difference here is that due to the more limited number of tradable assets assumed, not all target duration vectors developed earlier will necessarily be achievable. Consequently, additional constraints need to be introduced so that the resulting target total duration vector, $\mathbf{D}_0(\mathbf{i}_0)$, can be achieved by trading the given assets. Naturally, such limitations on the target vectors will have as a necessary by-product that the resultant portfolios will be suboptimal from the perspective of the results of Section 2.

Assume that $n \geq 2$ assets are given, with associated total duration vectors, $\mathbf{D}_1(\mathbf{i}_0), \ldots, \mathbf{D}_n(\mathbf{i}_0)$. For a general trade of a_j units of the j-th asset, denoted in vector form by $\mathbf{a}=(a_1, \ldots, a_n)$, the resulting total duration vectors satisfy (2.33), where again $\mathbf{D}(\mathbf{i}_0)$ is the initial portfolio total duration vector.

As noted above for a range of applications, the additional constraint, $\Sigma a_j=0$, is desirable and we impose this restriction here. Consequently, the resulting total duration vector in (2.33) can be expressed:

$$\mathbf{D}'(\mathbf{i}_0) = \mathbf{D}(\mathbf{i}_0) + \Sigma a_j \, \mathbf{D}_j(\mathbf{i}_0), \tag{3.1}$$

where a_j in (3.1) equals $a_j/P(\mathbf{i}_0)$ in (2.33) to simplify notation.

The restriction $\Sigma a_j=0$ again can be reflected in expression (3.1) by setting

$$a_n = \sum_{j=1}^{n-1} a_j,$$

and hence:

$$\mathbf{D}'(\mathbf{i}_0) = \mathbf{D}(\mathbf{i}_0) + \sum_{j=1}^{n-1} a_j[\mathbf{D}_j(\mathbf{i}_0) - \mathbf{D}_n(\mathbf{i}_0)] \tag{3.2}$$

As in (2.37), let $\mathbf{A}$ denote the $m\times(n-1)$ matrix with columns equal to the vectors: $[\mathbf{D}_j(\mathbf{i}_0)-\mathbf{D}_n(\mathbf{i}_0)]^T$, $j=1, \ldots, n-1$. Clearly, the rank of $\mathbf{A}$, denoted $\rho(\mathbf{A})$, cannot exceed the lesser of $n-1$ and m. Also, let $\mathcal{R}$ denote the collection of vectors in (3.2) for all $\mathbf{a}'=(a_1, \ldots, a_{n-1})$, which is easily seen to be an affine space in E^m, m-dimensional Euclidean space. An affine space is a linear subspace of E^m, translated by a fixed vector (here, equal to $\mathbf{D}(\mathbf{i}_0)$).

In the next proposition, $\mathcal{R}$ is characterized in a manner that is ideal for use with Proposition A.

***Proposition** 7:*

If $\rho(\mathbf{A})=m$, then $\mathcal{R}=E^m$. If $\rho(\mathbf{A})=m-\upsilon<m$, then there exist υ linearly independent vectors in E^m, $\mathbf{N}_1, \ldots, \mathbf{N}_\upsilon$, so that $\mathbf{A}^T\mathbf{N}_j=0$, and:

$$\mathcal{R} = \{\mathbf{D}|\mathbf{D}\cdot\mathbf{N}_j = \mathbf{D}(\mathbf{i}_0)\cdot\mathbf{N}_j, \quad j = 1, \ldots, \upsilon\}. \tag{3.3}$$

That is, $\mathcal{R}$ equals the intersection of υ hyperplanes.

Proof: Denoting by $\mathbf{a}'$ the vector $(a_1, \ldots, a_{n-1})$, we have by definition that:

$$\mathcal{R} = \{\mathbf{D}|\mathbf{D} = \mathbf{D}(\mathbf{i}_0) + (\mathbf{A}\mathbf{a}')^T, \quad \mathbf{a}' \in E^{n-1}\}. \tag{3.4}$$

Now if $\mathbf{A}$ has rank m, the result is obvious since $\mathcal{R}\subset E^m$ by definition, and the rank of a matrix equals the dimension of its range.

Next, assume that $\rho(\mathbf{A})=m-\upsilon$. Then since $\mathbf{A}^T$, the transpose of $\mathbf{A}$, also has rank equal to $m-\upsilon$, there exists υ independent vectors in E^m, $\mathbf{N}_1$, $\ldots$, $\mathbf{N}_\upsilon$, which span the null space of $\mathbf{A}^T$. That is, $\mathbf{A}^T\mathbf{N}_j=0$ for all j, and $\mathbf{A}^T\mathbf{N}=0$ if and only if $\mathbf{N}=\Sigma c_j\mathbf{N}_j$.

Now if $\mathbf{D}\in\mathcal{R}$ and hence $\mathbf{D}=\mathbf{D}(\mathbf{i}_0)+(\mathbf{A}\mathbf{a}')^T$:

$$\mathbf{D}\cdot\mathbf{N}_j = \mathbf{D}(\mathbf{i}_0)\cdot\mathbf{N}_j, \tag{3.5}$$

since $(\mathbf{A}\mathbf{a}')^T\mathbf{N}_j=(\mathbf{a}')^T\mathbf{A}^T\mathbf{N}_j=0$, by assumption.

On the other hand, assume that $\mathbf{D}\cdot\mathbf{N}_j=\mathbf{D}(\mathbf{i}_0)\cdot\mathbf{N}_j$ for $j=1, \ldots, \upsilon$. Then the vector $\mathbf{D}-\mathbf{D}(\mathbf{i}_0)$ is orthogonal to the null space of $\mathbf{A}^T$. Consequently, $\mathbf{D}-\mathbf{D}(\mathbf{i}_0)$ is in the range of $\mathbf{A}$ by a simple consequence of the definition of transpose, and there exists $\mathbf{a}'\in E^{n-1}$ so that:

$$\mathbf{D} - \mathbf{D}(\mathbf{i}_0) = [\mathbf{A}\mathbf{a}']^T. \ \square \tag{3.6}$$

Note that the rank of $\mathbf{A}$, $\rho(\mathbf{A})$, does not depend on the representation in (3.2), although the components of $\mathbf{A}$ do. That is, if the substitution

$$a_1 = \sum_{j=2}^{n} a_j$$

was made in (3.1), the resulting affine space would clearly be identical to that in (3.2). Consequently, while the matrix $\mathbf{A}$ would have changed, its rank and null space would remain constant. Naturally, in any given problem, the collection of vectors $\{\mathbf{N}_j\}$ will not be unique, although their linear span will be unique.

From Proposition 7, we see that constraining potential target total duration vectors to be those achievable by trades of given assets as in (3.2) can be equivalently formulated as constraining the target duration vector to be in the intersection of the υ hyperplanes as in (3.3). Equivalently, this constraint can be articulated as specifying υ direction vectors, $\mathbf{N}_1$, ..., $\mathbf{N}_\upsilon$, so that the directional durations of the resulting portfolio, $\mathbf{D}\cdot\mathbf{N}_j$, equal those of the original portfolio $\mathbf{D}(\mathbf{i}_0)\cdot\mathbf{N}_j$.

Consequently, the minimization problem that results is:

$$\min\,(\mathbf{D}\;\mathbf{K}_w\;\mathbf{D}^T), \tag{3.7}$$

subject to:

$$\mathbf{D}\cdot\mathbf{N}_j = r_j, \tag{3.8}$$

where:

$$r_j = \mathbf{D}(\mathbf{i}_0)\cdot\mathbf{N}_j, \tag{3.9}$$

for $j=1$, ..., υ. Of course, if $\upsilon=0$, (3.8)–(3.9) provide no constraint and this problem reduces to that in Section 2, since in this case, the rank of $\mathbf{A}$ in (2.37) equals m and hence, $n\geq m+1$.

In the general case, additional constraints on $\mathbf{D}$, such as in (2.13) and (2.14), also may be added depending on the size of υ, since the total number of constraints must not exceed m. The general solution to such problems is given by Proposition A in the Appendix, so it is not repeated here. Note, however, that if the constraints in (3.8) are added to, the resulting constraints will not necessarily be compatible. For example, the various constraining direction vectors:

$$\mathbf{N}_1, \ldots, \mathbf{N}_\upsilon, \mathbf{N}, \mathbf{E}(\mathbf{i}_0), \tag{3.10}$$

need not be independent, as Proposition A requires, unless $\mathbf{N}$ and $\mathbf{E}(\mathbf{i}_0)$ are outside the span of $\{\mathbf{N}_1, \ldots, \mathbf{N}_\upsilon\}$, the null space of $\mathbf{A}^T$, and are non-colinear.

B. Two-Asset Minimization

Given $n=2$ assets, the matrix **A** will then be the $m \times 1$ column vector: $[\mathbf{D}_1(\mathbf{i}_0)-\mathbf{D}_2(\mathbf{i}_0)]^T$, which clearly has rank 1 for $\mathbf{D}_1(\mathbf{i}_0) \neq \mathbf{D}_2(\mathbf{i}_0)$. Consequently, by Proposition 7, there exist $m-1$ independent vectors: $\mathbf{N}_1, \ldots, \mathbf{N}_{m-1}$, which can be used in (3.8) to restrict the resulting target total duration vectors, $\mathbf{D}_0(\mathbf{i}_0)$, to those actually obtainable by trading the given two assets.

As noted above, these $m-1$ vectors are in the null space of $\mathbf{A}^T$, a $1 \times m$ row vector. That is, we seek $m-1$ vectors that satisfy:

$$[\mathbf{D}_1(\mathbf{i}_0) - \mathbf{D}_2(\mathbf{i}_0)] \cdot \mathbf{N}_j = 0, \quad j = 1, \ldots, m-1. \tag{3.11}$$

For the associated constrained minimization problem, therefore, at most one constraint can be added to those in (3.8). For example, the expected period return can be constrained as in (2.13) if $\mathbf{E}(\mathbf{i}_0)$ is independent of the $\mathbf{N}_j$, or one directional duration can be constrained as in (2.14), again subject to an independence criterion.

In this extremely constrained case, however, adding this last constraint will make the problem trivial, because the resultant constraint set will contain a unique vector, so no real risk minimization will occur.

C. Multiple-Asset Minimization

As the number, n, of assets increases, in general the number, υ, of constraints given in (3.8) decreases. For example, given three assets for which $\mathbf{D}_1(\mathbf{i}_0)-\mathbf{D}_3(\mathbf{i}_0)$ and $\mathbf{D}_2(\mathbf{i}_0)-\mathbf{D}_3(\mathbf{i}_0)$ are linearly independent, the rank of **A** equals 2, and there will be $m-2$ constraints in (3.8), where the $\mathbf{N}_j$ satisfy:

$$\begin{aligned} &[\mathbf{D}_1(\mathbf{i}_0) - \mathbf{D}_3(\mathbf{i}_0)] \cdot \mathbf{N}_j = 0 \\ &[\mathbf{D}_2(\mathbf{i}_0) - \mathbf{D}_3(\mathbf{i}_0)] \cdot \mathbf{N}_j = 0 \quad \text{for } j = 1, \ldots, m-2. \end{aligned} \tag{3.12}$$

Given $n=m+1$ assets, such that the vectors $\mathbf{D}_j(\mathbf{i}_0)-\mathbf{D}_n(\mathbf{i}_0)$ are linearly independent for $j=1, \ldots, n-1$, the matrix **A** will be square and have rank equal to m. Consequently, the associated system:

$$\mathbf{A}^T \mathbf{N} = 0, \tag{3.13}$$

will be solved only by $\mathbf{N}=\mathbf{0}$, and the constraints in (3.8) will be without effect. That is, limiting trades to reflect only the given assets does not further constrain the minimum value of the objective function in (3.8).

Viewed from a different perspective, had any of the constrained minimization problems of Section 2 been solved, a trade of the given assets could always have been implemented to achieve the resulting target total duration vector, $\mathbf{D}_0(\mathbf{i}_0)$. That is, (2.37) could always be solved in this case because the rank of $\mathbf{A}$ is equal to m.

D. *Applications*

Subject to only the possibility of additional constraints as in (3.8), the applications of this approach are identical to those of Section 2.

4. EXAMPLES OF PORTFOLIO VARIANCE MINIMIZATION

A. *The Portfolio*

In this section, we apply the results of the preceding sections to the portfolio exemplified in Reitano [11]. Recall that the given liability is a \$100-million GIC payment at the end of year 5. Based on a simplified three-point bond yield curve of $\mathbf{i}_0=(0.075, 0.090, 0.100)$ at maturities of 0.5 years, 5 years, and 10 years and linear interpolation for other yields, this GIC payment has a market value of \$63.97 million and a duration of 4.855. In practice, the yield curve would typically reflect more "yield curve drivers" (Reitano [6], [8], [10], [11]), for example, at maturities of 1, 3 and 7 years, as well as beyond 10 years.

Available assets total \$71.08 million and are divided between a 10-year, 12 percent coupon bond and 6-month commercial paper in such a way that the surplus portfolio at time $k=\frac{1}{2}$ will be immunized against parallel yield curve shifts. Consequently, the current surplus of $S(\mathbf{i}_0)=\$7.11$ million requires a duration of 0.482, the duration of a 6-month zero-coupon bond. To achieve this, \$49.35 million of the \$71.08 million total is invested in the 6.151 duration bond, purchasing \$43.75 million par, while the remainder of \$21.73 million is invested in the 0.482 duration commercial paper, purchasing \$22.54 million par.

The forward value of surplus, $S_k(\mathbf{i}_0)=\$7.37$ million at time $k=\frac{1}{2}$, then has a duration of 0 and therefore is immunized against parallel shifts. The total duration vector of $S_k(\mathbf{i}_0)$ equals: $\mathbf{D}(S_k)=(5.26, -46.21, 40.95)$. Here, as defined in Reitano [8], [11], the forward value of surplus is given by: $S_k(\mathbf{i})=S(\mathbf{i})/Z_k(\mathbf{i})$, where $Z_k(\mathbf{i})$ denotes the price of a \$1 par value k-period zero-coupon bond.

B. Yield Curve Dynamics Model

Using treasury data and overlapping 6-month periods from August 1984 to June 1990, the following mean vector and covariance matrix were estimated:

$$\mathbf{E} = (-0.002904, -0.003648, -0.003606) \tag{4.1}$$

$$\mathbf{K} = \begin{pmatrix} 8.58211 & 8.02453 & 6.79183 \\ 8.02453 & 10.26390 & 9.30600 \\ 6.79183 & 9.30600 & 8.94903 \end{pmatrix} \times 10^{-5} \tag{4.2}$$

For our example, we assume that these values represent estimates of the mean vector and covariance matrix for the distribution of 6-month yield curve vector changes, $\mathbf{\Delta i}$, from the initial value of $\mathbf{i}_0=(0.075, 0.090, 0.100)$. That is, we assume $\mathbf{E}(\mathbf{i}_0)=\mathbf{E}$, $\mathbf{K}(\mathbf{i}_0)=\mathbf{K}$.

Admittedly, in practice one may be more confident of the stability of $\mathbf{K}$ than $\mathbf{E}$ through time, but we make the above assumptions for illustrative purposes. Other numerical estimates for $\mathbf{E}(\mathbf{i}_0)$ and $\mathbf{K}(\mathbf{i}_0)$ can be readily modeled and easily applied, including what might be considered the logical ex ante estimator for $\mathbf{E}(\mathbf{i}_0)$, $\mathbf{E}(\mathbf{i}_0)=\mathbf{0}$.

C. Statistics Associated with $S_k(\mathbf{i}_0)$, $k=\frac{1}{2}$

Because we are interested in the forward value of surplus, the ratio function of interest may not fit exactly into the model of (2.1). That is, we may be interested in $S_k(\mathbf{i}_0+\mathbf{\Delta i})/S(\mathbf{i}_0)$, or the volatility of the forward value as a percentage of today's value. However, this is readily accommodated by considering: $R_k(\mathbf{\Delta i})=c_k(1-\mathbf{D}(\mathbf{S}_k)\cdot\mathbf{\Delta i})$, $c_k=S_k(\mathbf{i}_0)/S(\mathbf{i}_0)$.

That is, we can approximate this ratio of interest by a simple multiple of the ratio function in (2.1), applied to $S_k(\mathbf{i}_0+\mathbf{\Delta i})/S_k(\mathbf{i}_0)$. In the above example,

$$\begin{aligned} c_k &= S_k(\mathbf{i}_0)/S(\mathbf{i}_0) \\ &= 1/Z_k(\mathbf{i}_0) \\ &= 1.0375, \end{aligned}$$

where $Z_k(\mathbf{i}_0)$ is the price (0.96386) of a \$1 par value 6-month zero-coupon bond and c_k is 1 plus the period return.

Readily adapting Proposition 1, we have:

$$
\begin{aligned}
E[R_k(\mathbf{\Delta i})] &= c_k[1 - \mathbf{D}(S_k) \cdot \mathbf{E}(\mathbf{i}_0)] \\
&= 1.0375\,(1 - 0.005633) \\
&= 1.0317. \qquad (4.5)
\end{aligned}
$$

That is, the expected 6-month return on surplus of 3.17 percent is less than the return on a 6-month zero-coupon bond of 3.75 percent because the durational structure of $S_k(\mathbf{i}_0)$, that is, $\mathbf{D}(S_k)$, generates relative losses of 0.56 percent given the expected shift in yields.

It is interesting to compare the expected annualized return from (4.5) of 6.43 percent to that produced by Formula (5.8) in Reitano [11]. That is, letting $I_k(\mathbf{i}_0)$ denote the annualized return, the formula there is:

$$E[I_k(\mathbf{i}_0)] \approx j(k) - [1 + j(k)]\,\mathbf{D}(S_k) \cdot \mathbf{E}(\mathbf{i}_0)/k, \qquad (4.6)$$

where $j(k)$ equals the annualized return on a k-period zero-coupon bond. Applying (4.6) to the above problem, with $k={}^1\!/_2$, we obtain:

$$E[I_k(\mathbf{i}_0)] \approx 0.0764 - 2(1.0764)(0.005633) = 0.0643, \qquad (4.7)$$

for the same annualized expected return of 6.43 percent.

The theoretical relationship between these two approximation methods can be readily understood. Equation (4.6) is based on the linear approximation to the annualized period return:

$$[S_k(\mathbf{i}_0 + \mathbf{\Delta i})/S(\mathbf{i}_0)]^{1/k} - 1 \approx j(k) - [1 + j(k)]\,\mathbf{D}(S_k) \cdot \mathbf{\Delta i}/k,$$

while the annualized return from (4.5) reflects a linear approximation to the period return,

$$
\begin{aligned}
S_k(\mathbf{i}_0 + \mathbf{\Delta i})/S(\mathbf{i}_0) &\approx c_k(1 - \mathbf{D}(S_k) \cdot \mathbf{\Delta i}) \\
&= [1 + j(k)]^k\,(1 - \mathbf{D}(S_k) \cdot \mathbf{\Delta i}),
\end{aligned}
$$

which is then annualized. It is easy to show that if this latter annualized formula is linearly approximated, the formula in (4.6) is produced.

Turning next to the variance of $R_k(\mathbf{\Delta i})$, we obtain:

$$
\begin{aligned}
\mathrm{Var}[R_k(\mathbf{\Delta i})] &= c_k^2[\mathbf{D}(S_k)\,\mathbf{K}(\mathbf{i}_0)\,\mathbf{D}(S_k)^T] \\
&= (1.0764)\,[0.009667] \\
&= 0.01041. \qquad (4.8)
\end{aligned}
$$

The associated standard deviation is then 0.10201, which translates to a value of about 10.2 percent of the initial surplus value of $7.11 million per half-year.

D. Surplus Optimization with Many Assets

Dropping the constant c_k above for notational simplicity, we now focus on the variability of the ratio function in (2.1) applied to $P(\mathbf{i})=S_k(\mathbf{i})$. That is, we are interested in the variability of $S_k(\mathbf{i}_0+\Delta\mathbf{i})$, or the forward value of surplus on the future actual yield curve, as a percentage of the forward value of surplus, $S_k(\mathbf{i}_0)$, based on today's yield curve $\mathbf{i}_0$. We see that, for this portfolio with total duration vector $\mathbf{D}(S_k)=(5.26, -46.21, 40.95)$:

$$\mathrm{Var}[R(\Delta\mathbf{i})] = \mathbf{D}(S_k)\,\mathbf{K}(\mathbf{i}_0)\,\mathbf{D}(S_k)^T$$

$$= 0.009667, \tag{4.9a}$$

$$\mathbf{D}(S_k)\cdot\mathbf{N} = 0, \tag{4.9b}$$

$$\mathbf{D}(S_k)\cdot\mathbf{E}(\mathbf{i}_0) = 0.005633, \tag{4.9c}$$

where $R(\Delta\mathbf{i})$ is the approximation for $S_k(\mathbf{i}_0+\Delta\mathbf{i})/S_k(\mathbf{i}_0)$ as in (2.1), and $N=(1,1,1)$ is the parallel shift direction vector. From (4.9), we see that the standard deviation of the given portfolio is 9.83 percent per half-year, while the expected period return is -0.56 percent.

Applying Proposition 4, with constraint $\mathbf{D}\cdot\mathbf{N}=0$, and using $\mathbf{K}_1\equiv\mathbf{K}$, it is clear from (2.17) and general reasoning that the minimum variance of $S_k(\mathbf{i}_0)$ is 0, and this occurs by (2.16) with $\mathbf{D}_0(S_k)=\mathbf{0}$, the zero total duration vector. For the more general constraint on the modified duration, $\mathbf{D}\cdot\mathbf{N}=D$, we obtain:

$$\mathbf{D}_0(S_k) = (1.002, -2.055, 2.052)\,D, \tag{4.10a}$$

$$\mathbf{D}_0\,\mathbf{K}\,\mathbf{D}_0^T = 0.000061\,D^2. \tag{4.10b}$$

Equation (4.10a) implies that given the yield curve dynamics implied by $\mathbf{K}(\mathbf{i}_0)$, a specific "barbell" configuration in the total duration vector is optimal for any positive constraint on the modified duration, while a negative barbell configuration is optimal for negative modified durations. While the given $\mathbf{D}(S_k)=(5.26, -46.21, 40.95)$ has a barbell configuration, its profile is quite different from that in (4.10a) and hence its suboptimality. Based on (4.10b), the given portfolio's variance of 0.009667 in (4.9a) equals that of an optimal portfolio on the efficient frontier with duration $D=12.64$. In addition, for this optimal portfolio, $r=\mathbf{D}_0(S_k)\cdot\mathbf{E}(\mathbf{i}_0)=-0.035594$, implying a positive expected yield curve return for 6 months of 3.56 percent, in contrast to that of the given portfolio of -0.56 percent by (4.9c).

Next applying Proposition 5 with constraint $\mathbf{D}\cdot\mathbf{E}=r$, the optimum portfolio is characterized by:

$$\mathbf{D}_0(S_k) = (-117.935, 306.711, -492.622)\, r, \qquad (4.11a)$$

$$\mathbf{D}_0\, \mathbf{K}\, \mathbf{D}_0^T = 6.531447\, r^2. \qquad (4.11b)$$

In general, (4.11a) implies that a negative barbell profile is optimal for negative yield curve returns ($r>0$), while a positive barbell is optimal for positive yield curve returns ($r<0$). In particular, for the given portfolio's value of $r=0.005633$, the associated optimal variance equals 0.000207, for a standard deviation of 1.44 percent per half-year, and an associated portfolio duration of -1.71. Looked at another way, the current portfolio has the same variance (0.009667) as an optimal portfolio with $r=+0.038472$ and associated duration of $+11.69$.

Finally, applying Proposition 6 with general constraints $\mathbf{D}\cdot\mathbf{N}=D$, $\mathbf{D}\cdot\mathbf{E}=r$, the optimum portfolio is given by:

$$\mathbf{D}_0(S_k) = (4.643, -8.251, 4.608)\, D + (1292.857, -2200.337, 907.480)\, r, \qquad (4.12a)$$

$$\mathbf{D}_0\, \mathbf{K}\, \mathbf{D}_0^T = 0.000419\, D^2 + 0.254863\, D\, r + 45.251100\, r^2. \qquad (4.13b)$$

Substituting the actual portfolio constraints, $D=0$, $r=0.005633$, the optimal portfolio is seen to have a variance of 0.001436, for a standard deviation of 3.79 percent per half-year, compared with that of the original portfolio of 9.83 percent.

Subject only to $D=0$, (4.13b) implies that the actual portfolio variance of 0.009667 corresponds to an optimal portfolio with $r=+0.014616$. Subject to only $r=0.005633$, the actual portfolio variance is seen to correspond to an optimal portfolio with $D=-6.46$ or 3.04.

The above analysis shows that the given portfolio is far from optimal. Even in the double constraint case, the actual standard deviation of 9.83 percent per half-year is significantly greater than the optimal value of 3.79 percent. However, no goal for r was targeted in the original portfolio development; the only constraint used was $\mathbf{D}(S_k)=0$. Consequently, the actual portfolio variance should be compared with that of an optimal portfolio with that given constraint; that is, it should be compared to an optimal portfolio variance of 0.

E. Asset Trading for Optimization

To ensure that any target total duration vector developed above can in fact be achieved, a minimum of three assets is required, since three is the number of yield points. To also reflect the constraint in (2.35) that the trade be cash-neutral, $\Sigma a_j=0$, a minimum of four assets is required. For (2.36) to be solvable, these assets also must have total duration vectors $\mathbf{D}_j(\mathbf{i}_0)$, such that $\{\mathbf{D}_j(\mathbf{i}_0)-\mathbf{D}_4(\mathbf{i}_0)\}$, $j=1,2,3$ are linearly independent.

Even though the goal of trading is to change the durational characteristics of $S_k(\mathbf{i}_0)$, or the future surplus portfolio, the actual trading is to be implemented in $S(\mathbf{i}_0)$, or the current surplus portfolio. Formally, the target total duration vectors for $S_k(\mathbf{i}_0)$, or $\mathbf{D}_0(S_k)$, must be translated to targets for $S(\mathbf{i}_0)$, or $\mathbf{D}_0(S)$. Similarly, the current total duration vector for $S_k(\mathbf{i}_0)$, $\mathbf{D}(S_k)$, must be translated to that for $S(\mathbf{i}_0)$, or $\mathbf{D}(S)$. Trades then are implemented in $S(\mathbf{i}_0)$ to convert $\mathbf{D}(S)$ to $\mathbf{D}_0(S)$ using (2.36).

For both translations, we have from Reitano [8], [11] that:

$$\mathbf{D}(S) = \mathbf{D}(S_k) + \mathbf{D}(Z_k), \tag{4.14}$$

and hence, $\mathbf{D}_0(S)-\mathbf{D}(S)=\mathbf{D}_0(S_k)-\mathbf{D}(S_k)$. Consequently, in (2.36)–(2.37), it is irrelevant whether this translation is made. However, the importance of recognizing that trading is performed on current surplus stems from the definition of $P(\mathbf{i}_0)$ in these equations. That is, we must have $P(\mathbf{i}_0)=S(\mathbf{i}_0)=\$7.11$ million, and not equal to $S_k(\mathbf{i}_0)=\$7.37$ million.

The four assets to be used here are: 6-month commercial paper; a 5-year, 9-½ percent coupon note; a 5-year, 8 percent coupon note with equal annual sinking fund payments; and the original 10-year, 12 percent coupon bond. The market values per 100 of par and durational structures are as follows:

	C.P.	5 year	5 year S.F.	10 year
Market Value	96.39	102.00	98.64	112.80
D_1	0.48	0.02	0.79	0.04
D_2	0	3.95	1.76	0.22
D_3	0	0	0	5.90
D	0.48	3.97	2.55	6.16

As an example, assume that the target total duration vector desired is $\mathbf{D}_0(S_k)=(0,0,0)$, the optimizing value subject to $\mathbf{D}\cdot\mathbf{N}=0$. A priori, it is expected that the trade implied by (2.36)–(2.37) would require the sale of 100 percent of the 10-year bonds in the portfolio, since this is the only security in the trade group above or in the surplus portfolio that has a yield sensitivity to the 10-year bond yield.

Solving (2.37) and setting

$$a_4 = -\sum_{j=1}^{3} a_j,$$

the following trade is derived:

$$\mathbf{a} = (-15.20, 103.00, -38.45, -49.35), \tag{4.15}$$

where the components are in the order of the securities in the above table.

Based on (4.15), the trade that will produce $\mathbf{D}(S_k)=0$ is one whereby all 10-year bonds in the portfolio are sold, as is \$15.20 million of the commercial paper, as is \$38.45 million of the 5-year notes (short sale), and the \$103.00 million of total proceeds is invested in the 5-year sinking fund note. The resultant portfolio is then:

Assets		Liabilities	
C.P.	6.53	GIC	63.97
5-year S.F. Notes	103.00	5-year Note (short)	38.45
	109.53		102.45

F. Surplus Optimization with Fewer Assets

In this section, we consider two applications of Section 3. Consider first the case in which only trades between the original 10-year bond and the 5-year note above are allowed. Here $m=3$ and $n=2$, and the $m\times(n-1)$ matrix $\mathbf{A}$ in (2.37) and Proposition 7 is given by:

$$\mathbf{A} = [\mathbf{D}_1(\mathbf{i}_0) - \mathbf{D}_2(\mathbf{i}_0)]^T = (0.02, -3.73, 5.90)^T. \tag{4.16}$$

Here $\mathbf{D}_1(\mathbf{i}_0)$ and $\mathbf{D}_2(\mathbf{i}_0)$ are the total duration vectors of the 10-year bond and 5-year note, respectively, and the transpose is used since the $\mathbf{D}_j(\mathbf{i}_0)$ are identified with row matrices by convention.

The rank of $\mathbf{A}$ is clearly equal to 1, and hence by Proposition 7 we seek $\upsilon=2$ independent vectors, so that $\mathbf{A}^T\mathbf{N}_j=0$. Two such vectors are:

$$\mathbf{N}_1 = (0, 1.581769, 1), \quad \mathbf{N}_2 = (-295, 0, 1). \tag{4.17}$$

Finally, to specify the trading constraints as in (3.8), we require $r_j=\mathbf{N}_j \cdot \mathbf{D}(S_k)$, and the following minimization problem is produced as in (3.7)–(3.9):

$$\min\,(\mathbf{D}\,\mathbf{K}_w\,\mathbf{D}^T), \tag{4.18a}$$

subject to:

$$\mathbf{D}\cdot\mathbf{N}_1 = -32.143545, \tag{4.18b}$$

$$\mathbf{D}\cdot\mathbf{N}_2 = 1510.75.$$

The solution to (4.18) then will equal the variance-minimizing total duration vector that can be obtained by trading only between the given two assets.

Applying Proposition 6, the target total duration vector and associated variance are given by:

$$\mathbf{D}_0(S_k) = (5.249, -44.142, 37.678), \tag{4.19a}$$

$$\mathbf{D}_0\,\mathbf{K}\,\mathbf{D}_0^T = 0.009528. \tag{4.19b}$$

Comparing (4.19b) to the original portfolio variance of 0.009667 in (4.9a), it is clear that even the best such trade has little effect. To determine the actual trade, we again use (2.37). As noted in Section 4-E above, it is immaterial whether $\mathbf{D}$ and $\mathbf{D}_0$ are based on $S_k(\mathbf{i}_0)$ or translated to $S(\mathbf{i}_0)$ via (4.14); however, the price function on which we are trading is $P(\mathbf{i}_0)=S(\mathbf{i}_0)$. A calculation based on (2.37) produces $a'=-\$3.94$ million, which implies from (4.16) that \$3.94 million of the 10-year bond is sold, and a like amount of the 5-year note purchased.

In (4.18), it is also possible to add a constraint on the modified duration, such as that of the original value, $\mathbf{D}\cdot\mathbf{N}=0$, where $\mathbf{N}=(1,1,1)$. This is because the vector $\mathbf{N}$ is linearly independent of the $\mathbf{N}_j$ in (4.17). However, because $m=3$, the constraints are satisfied by a unique vector

$\mathbf{D}_0$, determined by the intersection of the three constraining planes. Because the original $\mathbf{D}(S_k)$ satisfies these three constraints by construction, it is clear that we must then have $\mathbf{D}_0(S_k)=\mathbf{D}(S_k)$. That is, the current portfolio is optimum if the modified duration is fixed at 0 and if trades only between the two given securities are allowed.

Note that the above conclusion also could have been predicted from first principles, since any amount of trading between the given instruments would change $D(S_k)$ from its original value of 0.

Next assume that a third asset is added to the allowable trading group, say, commercial paper. That is, we now seek a cash-neutral trade between commercial paper, the 5-year note, and original 10-year bond. The matrix $\mathbf{A}$ of Proposition 7 is now 3×2, and given by:

$$\mathbf{A} = \begin{pmatrix} \mathbf{D}_1(\mathbf{i}_0) - \mathbf{D}_3(\mathbf{i}_0) \\ \mathbf{D}_2(\mathbf{i}_0) - \mathbf{D}_3(\mathbf{i}_0) \end{pmatrix}^T = \begin{pmatrix} -0.46 & -0.44 \\ 3.95 & 0.22 \\ 0 & 5.90 \end{pmatrix}, \tag{4.20}$$

where $\mathbf{D}_i(\mathbf{i}_0)$, $i=1, 2, 3$ denote the total duration vectors of the 10-year bond, 5-year note and commercial paper, respectively. With a rank equal to 2, we seek $\upsilon=1$ vector, so that $\mathbf{A}^T\mathbf{N}_1=0$.

A calculation produces the following vector, as well as associated constraint constant, $r_1=\mathbf{N}_1\cdot\mathbf{D}(S_k)$, as in (3.3):

$$\mathbf{N}_1 = (1, 0.116456, 0.070234),$$

$$r_1 = 2.7154610. \tag{4.21}$$

With no additional constraints, the solution to the minimization problem in (3.7)–(3.9) is given by Proposition 4:

$$\mathbf{D}_0(S_k) = (3.118, -4.623, 2.493), \tag{4.22a}$$

$$\mathbf{D}_0\, \mathbf{K}\, \mathbf{D}_0^T = 0.000182. \tag{4.22b}$$

Comparing (4.22b) to the original variance in (4.9a), we observe a significant reduction.

Solving for the associated trade vector using (2.37), we obtain:

$$\mathbf{a} = (-46.33, 77.41, -31.08). \tag{4.23}$$

That is, \$46.33 million of the original \$49.35 million 10-year bond is to be sold, as is \$31.08 million of the \$21.73 million commercial paper (netting a short position), and the total proceeds of \$77.41 million invested in the 5-year note.

G. Establishing the Original Asset Portfolio

In the above sections, the methodologies of Sections 2 and 3 were applied to the problem of optimizing an existing portfolio. However, they also can be applied easily to the problem of establishing the initial asset portfolio.

To this end, we begin with a surplus portfolio that reflects the GIC liability and the assumption that 100 percent of the assets are held in commercial paper. Naturally, $S(\mathbf{i}_0)$ and $S_k(\mathbf{i}_0)$ are now quite different from the perspective of the profile of their total duration vectors. For example, we now have:

$$\mathbf{D}(S_k) = (8.36, -47.70, 0), \tag{4.24a}$$

$$D(S_k) = -39.34. \tag{4.24b}$$

A calculation shows that the variance of the above portfolio is given by:

$$\mathbf{D}\,\mathbf{K}\,\mathbf{D}^T = 0.175532, \tag{4.25}$$

for an implied standard deviation of 41.9 percent per half-year. Not surprisingly, this value is far in excess of the 9.83 percent standard deviation of the original portfolio as given in (4.9a).

Starting with this portfolio, the analyses of the above sections are readily repeated. We leave the details to the interested reader.

REFERENCES

1. HILLER, RANDALL S., AND SCHAAK, CHRISTIAN. "A Classification of Structured Bond Portfolio Modeling Techniques," *The Journal of Portfolio Management* 16 (Fall 1990): 37–48.
2. HO, T.S.Y. *Strategic Fixed Income Management*. Homewood, Ill.: Dow Jones-Irwin, 1990.
3. LITTERMAN, ROBERT, AND SCHEINKMAN, JOSE. "Common Factors Affecting Bond Returns," *The Journal of Fixed Income* 1 (June 1991): 54–61.
4. MARKOWITZ, HARRY. *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons, 1959.

5. MARTIN, JOHN D., COX, SAMUEL H., AND MACMINN, RICHARD D. *The Theory of Finance: Evidence and Applications.* New York: The Dryden Press, 1988.
6. REITANO, ROBERT R. "A Multivariate Approach to Duration Analysis," *ARCH* 1989.2: 97–181.
7. REITANO, ROBERT R. "Non-parallel Yield Curve Shifts and Durational Leverage," *The Journal of Portfolio Management* 16 (Summer 1990): 62–67.
8. REITANO, ROBERT R. "A Multivariate Approach to Immunization Theory," *ARCH* 1990.2: 261–312.
9. REITANO, ROBERT R. "Non-parallel Yield Curve Shifts and Spread Leverage," *The Journal of Portfolio Management* 17 (Spring 1991): 82–87.
10. REITANO, ROBERT R. "Multivariate Duration Analysis," *TSA* XLIII (1991): 335–91.
11. REITANO, ROBERT R. "Multivariate Immunization Theory," *TSA* XLIII (1991): 392–438.
12. REITANO, ROBERT R. "Non-parallel Yield Curve Shifts and Immunization," *The Journal of Portfolio Management* 18 (Spring 1992): 36–43.
13. REITANO, ROBERT R. "Nonparallel Yield Curve Shifts and Convexity," *TSA* XLIV (1992): 479–507.
14. THEIL, HENRI. *Principles of Econometrics.* New York: John Wiley & Sons, 1971.
15. WILKS, SAMUEL S. *Mathematical Statistics.* New York: John Wiley & Sons, 1962.

APPENDIX

Proposition A:

Let $\mathbf{K}$ be an $m \times m$ symmetric, positive definite matrix and $\{\mathbf{B}_j\}$ a collection of p independent m-vectors, so $p \leq m$. Given $\{r_j\}_{j=1}^{p}$, consider the problem:

$$\text{Min } (\mathbf{x}^T \mathbf{K} \mathbf{x}) \tag{A.1}$$

subject to:

$$\mathbf{x}^T\mathbf{B}_j = r_j \quad j = 1, \ldots, p. \tag{A.2}$$

Then a solution $\mathbf{x}_0$ exists and is given by:

$$\mathbf{x}_0 = \Sigma \lambda_j \mathbf{K}^{-1} \mathbf{B}_j, \tag{A.3}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$ is the unique solution of:

$$\mathbf{C}\, \boldsymbol{\lambda} = \mathbf{r}, \tag{A.4}$$

$C_{jk} = \mathbf{B}_j^T \mathbf{K}^{-1} \mathbf{B}_k$ and $\mathbf{r} = (r_1, \ldots, r_p)$.

Proof: Because **K** is symmetric, it has real eigenvalues, $\{a_i\}$, and m-independent unit eigenvectors, $\{\mathbf{A}_i\}$, which are mutually orthogonal. Let **P** denote the $m \times m$ matrix with $\{\mathbf{A}_i\}$ as column vectors. The matrix **P** is then an orthogonal matrix, $\mathbf{P}^T = \mathbf{P}^{-1}$. Changing coordinates to the $\{\mathbf{A}_i\}$ basis, let $\mathbf{x} = \mathbf{P}\mathbf{y}$; that is, the components of **y** are the coordinates of **x** in the $\{\mathbf{A}_i\}$ basis.

Substituting into (A.1) and recalling that $(\mathbf{P}\mathbf{y})^T = \mathbf{y}^T\ \mathbf{P}^T$, we obtain:

$$\begin{aligned} \mathbf{x}^T\,\mathbf{K}\,\mathbf{x} &= \mathbf{y}^T(\mathbf{P}^T\,\mathbf{K}\,\mathbf{P})\,\mathbf{y} \\ &= \Sigma\, a_i y_i^2, \end{aligned} \tag{A.5}$$

since $\mathbf{P}^T\ \mathbf{K}\ \mathbf{P}$ is a diagonal matrix with eigenvalues $\{a_i\}$ along the diagonal. Similarly:

$$\begin{aligned} \mathbf{x}^T\,\mathbf{B}_j &= \mathbf{y}^T(\mathbf{P}^T\mathbf{B}_j) \\ &= \sum_i b_{ji}\, y_i, \end{aligned} \tag{A.6}$$

where $b_{ji} = \mathbf{A}_i^T\mathbf{B}_j$.

In the new coordinates, the above problem reduces to:

$$\text{Min}\ \Sigma\, a_i y_i^2, \tag{A.7}$$

subject to:

$$\sum_i b_{ji} y_i = r_j, \quad j = 1, \ldots, p. \tag{A.8}$$

Let $F(\mathbf{y}) = \Sigma\, a_i y_i^2$ and $G_j(\mathbf{y}) = \Sigma_i\, b_{ji} y_i$. By the method of Lagrange multipliers, if $\mathbf{y}_0$ is a critical value of $F(\mathbf{y})$ subject to the constraints $G_j(\mathbf{y}_0) = r_j$, then there exists constants $\{\lambda_k\}$, the Lagrange multipliers, so that:

$$\nabla\mathbf{F}(\mathbf{y}_0) = \Sigma\, \lambda_k\, \nabla\mathbf{G}_k(\mathbf{y}_0), \tag{A.9}$$

where $\nabla\mathbf{F}$ denotes the gradient of $F(\mathbf{y})$:

$$\nabla\mathbf{F} = \left(\frac{\partial F}{\partial y_1}, \ldots, \frac{\partial F}{\partial y_m}\right).$$

Further, if the function:

$$H(\mathbf{y}) = F(\mathbf{y}) - \Sigma\, \lambda_k G_k(\mathbf{y}), \tag{A.10}$$

has a Hessian matrix $\mathbf{H}[H(\mathbf{y}_0)]$, which is positive definite on the vector subspace:

$$\mathbf{Y} = \{\mathbf{y} | \mathbf{y} \cdot \nabla \mathbf{G}_k(\mathbf{y}_0) = 0, \quad k = 1, \ldots, p\}, \tag{A.11}$$

where:

$$\mathbf{H}[H(\mathbf{y}_0)]_{ij} = \frac{\partial^2 H}{\partial y_i \partial y_j}, \tag{A.12}$$

then $\mathbf{y}_0$ is a minimum of $F(\mathbf{y})$, subject to the above constraints.

For the above problem, a calculation produces:

$$\nabla \mathbf{F}(\mathbf{y}) = 2(a_1 y_1, \ldots, a_m y_m),$$

$$\nabla \mathbf{G}_k(\mathbf{y}) = \mathbf{P}^T \mathbf{B}_k = (b_{k1}, \ldots, b_{km}). \tag{A.13}$$

Hence, from (A.9) we see that if $\mathbf{y}_0$ is a critical point of $F(\mathbf{y})$ under the above constraints, we must have:

$$y_i = \frac{\sum_k \lambda_k b_{ki}}{a_i}, \tag{A.14}$$

which is well-defined since $\mathbf{K}$ is positive definite, and hence $a_i > 0$ for all i. For simplicity, the factor of 2 in (A.13) has been absorbed into the unknown λ_k.

The unknowns $\{\lambda_k\}$ in (A.14) are determined so that $\mathbf{y}_0$ satisfies the constraints in (A.8):

$$\sum_{i,k} \frac{b_{ji} b_{ki} \lambda_k}{a_i} = r_j, \quad j = 1, \ldots, p. \tag{A.15}$$

The system in (A.15) is equivalent to that in (A.4) because the coefficient matrix, $\mathbf{C}$, has components:

$$C_{jk} = \sum_i \frac{b_{ji} b_{ki}}{a_i} = (\mathbf{P}^T \mathbf{B}_j)^T (\mathbf{P}^T \mathbf{K} \mathbf{P})^{-1} (\mathbf{P}^T \mathbf{B}_k)$$

$$= \mathbf{B}_j^T \mathbf{K}^{-1} \mathbf{B}_k. \tag{A.16}$$

The last step in (A.16) follows because $\mathbf{P}$ is an orthogonal matrix and hence $\mathbf{P}^T = \mathbf{P}^{-1}$.

To show that (A.15) has a unique solution, it is sufficient to show that $\mathbf{C}$ has a non-zero determinant. To this end, consider the bilinear function:

$$(\mathbf{x},\mathbf{y})_K = \mathbf{x}^T \mathbf{K}^{-1} \mathbf{y}. \tag{A.17}$$

Because $\mathbf{K}^{-1}$ is a positive definite matrix, $(\mathbf{x},\mathbf{y})_K$ is a well-defined inner product on R^m. Consequently, the matrix $\mathbf{C}$ in (A.16) is seen to be the Gram matrix associated with this inner product:

$$C_{jk} = (\mathbf{B}_j, \mathbf{B}_k)_K. \tag{A.18}$$

As is well-known, a Gram matrix has a positive determinant when the vectors $\{\mathbf{B}_j\}$ are linearly independent.

Consequently, $\boldsymbol{\lambda}=(\lambda_1, \ldots, \lambda_p)$ is uniquely determined by (A.4), and the critical point $\mathbf{y}_0$ is given by (A.14):

$$\begin{aligned}\mathbf{y}_0 &= (\mathbf{P}^T\mathbf{K}\,\mathbf{P})^{-1}(\Sigma\,\lambda_k \mathbf{P}^T\mathbf{B}_k)\\ &= \Sigma\,\lambda_k \mathbf{P}^T\mathbf{K}^{-1}\mathbf{B}_k.\end{aligned} \tag{A.19}$$

Since $\mathbf{x}_0=\mathbf{P}\mathbf{y}_0$, (A.3) follows.

To show that the critical point $\mathbf{y}_0$ is in fact a minimum of $F(\mathbf{y})$ subject to the $G_j(\mathbf{y})$ constraints, consider $H(\mathbf{y})$ in (A.10). A calculation shows that:

$$\frac{\partial^2 H}{\partial y_i \partial y_j} = \begin{cases} 2a_i & i = j \\ 0 & i \neq j. \end{cases} \tag{A.20}$$

Consequently, the Hessian matrix of $H(\mathbf{y})$ in (A.12) is equal to $2(\mathbf{P}^T\mathbf{K}\,\mathbf{P})$, which is positive definite everywhere, so in particular, it is positive definite on the vector subspace given in (A.11). □

Corollary A.1:

Under the assumptions of Proposition A,

$$\begin{aligned}\text{Min}\,(\mathbf{x}_0^T\mathbf{K}\mathbf{x}_0) &= \boldsymbol{\lambda}^T\,\mathbf{C}\,\boldsymbol{\lambda}\\ &= \mathbf{r}^T\,\mathbf{C}^{-1}\,\mathbf{r},\end{aligned} \tag{A.21}$$

which equals 0 if and only if $\mathbf{r}=\mathbf{0}$.

Proof: By (A.3), $\mathbf{Kx}_0 = \Sigma\lambda_j\mathbf{B}_j$. Hence, since $\mathbf{K}$ is symmetric,

$$\begin{aligned}\mathbf{x}_0^T\,\mathbf{K}\,\mathbf{x}_0 &= (\Sigma\,\lambda_k\mathbf{K}^{-1}\mathbf{B}_k)^T\,\Sigma\,\lambda_j\mathbf{B}_j\\ &= \Sigma\,\Sigma\,\lambda_j\,\lambda_k\,\mathbf{B}_j\,\mathbf{K}^{-1}\,\mathbf{B}_k\\ &= \boldsymbol{\lambda}^T\,\mathbf{C}\,\boldsymbol{\lambda}.\end{aligned}$$

Now by (A.4),

$$\begin{aligned}\boldsymbol{\lambda}^T\,\mathbf{C}\,\boldsymbol{\lambda} &= (\mathbf{C}^{-1}\mathbf{r})^T\,\mathbf{r}\\ &= \mathbf{r}^T\mathbf{C}^{-1}\,\mathbf{r}.\end{aligned}$$

Because $\mathbf{C}$ has non-zero determinant, $\mathbf{C}^{-1}$ is well-defined, proving (A.21).

Finally, $\mathbf{C}$ is also positive definite, since all principal minors have positive determinants. This is due to the observation that every principal minor of $\mathbf{C}$ is again a Gram matrix and hence has positive determinant as before. Hence, $\mathbf{C}^{-1}$ is also positive definite, so $\mathbf{x}_0^T\mathbf{K}\,\mathbf{x}_0$ equals 0 if and only if $\mathbf{r}=\mathbf{0}$. □

Corollary A.2:

Assume that the $\{\mathbf{B}_j\}$ of Proposition A are linearly independent unit eigenvectors of $\mathbf{K}$, $\mathbf{B}_j=\mathbf{A}_j$, as in the proof of Proposition A. Then:

$$\mathbf{x}_0 = \Sigma\,r_j\,\mathbf{B}_j, \tag{A.22}$$

and

$$\mathbf{x}_0^T\,\mathbf{K}\,\mathbf{x}_0 = \Sigma\,a_j\,r_j^2, \tag{A.23}$$

where $\{a_j\}$ denotes the associated eigenvalues.

Proof: Because $\mathbf{B}_j$ is also an eigenvector of $\mathbf{K}^{-1}$, only with eigenvalue $1/a_j$, we have $\mathbf{K}^{-1}\mathbf{B}_j=\mathbf{B}_j/a_j$, and from (A.3):

$$\mathbf{x}_0 = \Sigma\,\frac{\lambda_j\mathbf{B}_j}{a_j}. \tag{A.24}$$

Also, because the eigenvectors of $\mathbf{K}$ are orthogonal:

$$C_{jk} = \mathbf{B}_j^T\,\mathbf{K}^{-1}\,\mathbf{B}_k = \begin{cases} 1/a_k & j = k,\\ 0 & j \neq k,\end{cases}$$

and the matrix $\mathbf{C}$ in (A.4) is a diagonal matrix. Consequently, $\mathbf{C}^{-1}$ is also diagonal with diagonal elements $\{a_j\}$, so it is clear from (A.4) that $\lambda_j = r_j a_j$. Hence, (A.22) follows from (A.24). Similarly, (A.23) follows from (A.21). □

Note: For a somewhat different proof of Proposition A, generalized to include a linear term in $\mathbf{x}$, see Martin et al. [5, p. 683].

Also, denote by $\mathbf{B}$ the $m \times p$ matrix with columns equal to the $\mathbf{B}_j$ vectors. Then (A.2) can be written:

$$\mathbf{x}^T\mathbf{B} = \mathbf{r}. \tag{A.2$'$}$$

In addition, the matrix $\mathbf{C}$ in (A.4) is seen to equal:

$$\mathbf{C} = \mathbf{B}^T\mathbf{K}^{-1}\mathbf{B},$$

and hence, the solution $\mathbf{x}_0$ in (A.3) can be expressed:

$$\begin{aligned} \mathbf{x}_0 &= \mathbf{K}^{-1}\mathbf{B}\boldsymbol{\lambda} \\ &= \mathbf{K}^{-1}\mathbf{B}(\mathbf{B}^T\mathbf{K}^{-1}\mathbf{B})^{-1}\mathbf{r}. \end{aligned} \tag{A.3$'$}$$

Similarily, the minimum value of the quadratic form in (A.1) as given in (A.21) of Corollary A.1 can be expressed:

$$\mathbf{x}_0^T\mathbf{K}\mathbf{x}_0 = \mathbf{r}^T(\mathbf{B}^T\mathbf{K}^{-1}\mathbf{B})^{-1}\mathbf{r}. \tag{A.21$'$}$$

However, the notational convention followed in the paper was chosen to better highlight the relationship between the optimizing target total duration vector, $\mathbf{D}_0$, the constraining direction vectors, $\mathbf{N}$, $\mathbf{N}_1$, …, $\mathbf{E}(\mathbf{i}_0)$, and the various constraining values, D, r_1, …, r.

DISCUSSION OF PRECEDING PAPER

ELIAS S.W. SHIU:

Dr. Reitano is to be congratulated for publishing this paper, which proposes a new theory of immunization. The following are my comments, some mathematical and some philosophical.

1. Different Approximation Formulas

This paper assumes that the price of an asset or a liability can be given as a smooth function of m variables. Thus, the price can be denoted as $P(\mathbf{i})$, $\mathbf{i} \in R^m$. For much of this discussion, there is no need to assume that $\mathbf{i}$ is an "m-point yield curve vector." The key assumption is that several variables or factors exist that determine the prices of assets and liabilities and that such price functions are at least twice differentiable. The assets need not be fixed-income securities as long as one is willing to assume the existence of such variables or factors determining their prices.

Given the current-status vector $\mathbf{i}_0$, we are interested in the distribution of the changed-price random variable

$$g(\Delta\mathbf{i}) = P(\mathbf{i}_0 + \Delta\mathbf{i}), \tag{1.1}$$

where $\Delta\mathbf{i}$ denotes the instantaneous-shock random vector. If the current price is not zero, $P(\mathbf{i}_0) \neq 0$, then the problem of determining the distribution of $g(\Delta\mathbf{i})$ is equivalent to that of determining the distribution of the relative-change-in-price random variable

$$h(\Delta\mathbf{i}) = \frac{g(\Delta\mathbf{i})}{P(\mathbf{i}_0)} = \frac{P(\mathbf{i}_0 + \Delta\mathbf{i})}{P(\mathbf{i}_0)}. \tag{1.2}$$

Equations (2.5) and (2.6) of the paper are approximations to the mean and variance of $h(\Delta\mathbf{i})$ in terms of the first two moments of $\Delta\mathbf{i}$, respectively.

To estimate $E[h(\Delta\mathbf{i})]$, the paper approximates $h(\Delta\mathbf{i})$ with the first two terms of the multivariate Maclaurin series,

$$h(\Delta\mathbf{i}) \approx h(0) + h'(0)^T\Delta\mathbf{i}, \tag{1.3}$$

and then takes expectation, resulting in Formula (2.5). However, the standard way is to approximate $h(\Delta\mathbf{i})$ with the first three terms of the multivariate Taylor series expanded at the mean of the random vector $\Delta\mathbf{i}$, $E(\Delta\mathbf{i})$ [which the paper denotes as $\mathbf{E}(\mathbf{i}_0)$], and then take expectation. The resulting approximation formula is

$$E[h(\Delta \mathbf{i})] \approx h(\mathbf{E}(\mathbf{i}_0)) + {}^1\!/\!_2\, tr[h''(\mathbf{E}(\mathbf{i}_0))\mathbf{K}(\mathbf{i}_0)], \tag{1.4}$$

where h'' denotes the Hessian matrix and tr, as in the paper, denotes the trace operator.

Formula (1.4) is a generalization of the following one-dimensional result: If X is a random variable with $E(X)=\mu$ and $\text{Var}(X)=\sigma^2$, and φ is a smooth function of one variable, then

$$E[\varphi(X)] \approx \varphi(\mu) + {}^1\!/\!_2\, \varphi''(\mu)\sigma^2. \tag{1.5}$$

On the other hand, Formula (2.5) of the paper corresponds to

$$E[\varphi(X)] \approx \varphi(0) + \varphi'(0)\mu. \tag{1.6}$$

Formula (1.5) is a better approximation formula than (1.6) because it is a higher-order approximation and X is probably closer to its mean μ than to 0. The standard approximation formula for the variance of $\varphi(X)$ is

$$\text{Var}[\varphi(X)] \approx [\varphi'(\mu)]^2\sigma^2, \tag{1.7}$$

while Formula (2.6) of the paper corresponds to

$$\text{Var}[\varphi(X)] \approx [\varphi'(0)]^2\sigma^2. \tag{1.8}$$

Formula (1.7) can also be generalized to the multivariate case; see (3.90) on page 72 of the Elandt-Johnson and Johnson book [1], which for several years was a textbook for one of the Society Associateship examinations. (Note that Formulas (3.87) and (3.88) of [1] contain typographical errors.) Consider the random variable

$$q(\Delta \mathbf{i}) = \frac{P(\mathbf{i}_0 + \Delta \mathbf{i})}{P(\mathbf{i}_0 + \mathbf{E}(\mathbf{i}_0))} = \frac{g(\Delta \mathbf{i})}{P(\mathbf{i}_0 + \mathbf{E}(\mathbf{i}_0))} = \frac{P(\mathbf{i}_0)h(\Delta \mathbf{i}_0)}{P(\mathbf{i}_0 + \mathbf{E}(\mathbf{i}_0))}; \tag{1.9}$$

it follows from (3.90) of the Elandt-Johnson and Johnson book [1] that

$$\text{Var}[q(\Delta \mathbf{i})] \approx \mathbf{d}^T\mathbf{K}(\mathbf{i}_0)\mathbf{d}, \tag{1.10}$$

where $\mathbf{d}$ denotes the transpose of the total duration vector evaluated at $\mathbf{i}_0+\mathbf{E}(\mathbf{i}_0)$,

$$\mathbf{d} = \mathbf{D}(\mathbf{i}_0 + \mathbf{E}(\mathbf{i}_0))^T. \tag{1.11}$$

(In this discussion lowercase boldface letters denote column vectors.) The difference between (1.10) and (2.6) of the paper is essentially the same as that between (1.7) and (1.8).

2. *Asset/Liability Management Strategies*

Formulas (1.4) and (1.10) imply the following asset/liability management strategy, if we believe that asset and liability values are smooth functions of several variables. As in the paper, let $S(\mathbf{i})$ denote the surplus function. By structuring the assets and liabilities such that the gradient of S at $\mathbf{i}_0+\mathbf{E}(\mathbf{i}_0)$ is the zero vector,

$$S'(\mathbf{i}_0 + \mathbf{E}(\mathbf{i}_0)) = \mathbf{0}, \tag{2.1}$$

we have

$$\mathrm{Var}[S(\mathbf{i}_0 + \Delta\mathbf{i})] \approx 0. \tag{2.2}$$

At the same time, because

$$E[S(\mathbf{i}_0 + \Delta\mathbf{i})] \approx S(\mathbf{E}(\mathbf{i}_0)) + \tfrac{1}{2}\, tr[S''(\mathbf{i}_0 + \mathbf{E}(\mathbf{i}_0))\mathbf{K}(\mathbf{i}_0)],$$

we would want to maximize

$$tr[S''(\mathbf{i}_0 + \mathbf{E}(\mathbf{i}_0))\mathbf{K}(\mathbf{i}_0)]. \tag{2.3}$$

It can be shown that the trace of the product of two positive semidefinite matrices is non-negative and that the trace of the product of two positive definite matrices is positive. Since a variance-covariance matrix is always positive semidefinite, we have

$$tr[S''(\mathbf{i}_0 + \mathbf{E}(\mathbf{i}_0))\mathbf{K}(\mathbf{i}_0)] \geq 0$$

if the Hessian matrix $S''(\mathbf{i}_0+\mathbf{E}(\mathbf{i}_0))$ is positive semidefinite.

The strategy above should be compared with the classical free-lunch strategy implied by the assumption that asset and liability values are smooth functions of several variables. The assets and liabilities are to be structured such that (i) $S(\mathbf{i}_0)=0$, (ii) $S'(\mathbf{i}_0)=\mathbf{0}$, and (iii) the Hessian matrix $S''(\mathbf{i}_0)$ is positive definite. Condition (i) means that there is zero net investment, and therefore the rate of return due to an instantaneous shock is infinite. Condition (ii) is a duration-matching condition similar to (2.1). Assuming that we know the first two moments of the shock random vector $\Delta\mathbf{i}$, we can improve condition (iii) by maximizing

$$E[\Delta\mathbf{i}^T S''(\mathbf{i})\Delta\mathbf{i}] = tr[S''(\mathbf{i}_0)E(\Delta\mathbf{i}^T\Delta\mathbf{i})]. \tag{2.4}$$

Note that

$$E(\Delta\mathbf{i}^T\Delta\mathbf{i}) = \mathbf{K}(\mathbf{i}_0) + E(\Delta\mathbf{i})^T E(\Delta\mathbf{i})$$
$$= \mathbf{K}(\mathbf{i}_0) + \mathbf{E}(\mathbf{i}_0)^T\mathbf{E}(\mathbf{i}_0). \quad (2.5)$$

3. A Puzzle

It might be useful to future readers if I describe a confusion I encountered when I first read this paper. I think I have now solved the puzzle, but Dr. Reitano might want to further clarify the point in his author's review. The random variable

$$h(\Delta\mathbf{i}) = \frac{P(\mathbf{i}_0 + \Delta\mathbf{i})}{P(\mathbf{i}_0)}$$

gives the relative change in price due to an instantaneous shock $\Delta\mathbf{i}$, and the random variable

$$h(\Delta\mathbf{i}) - 1$$

represents the rate of return due to an instantaneous shock. The expectation

$$E[h(\Delta\mathbf{i}) - 1] = E[h(\Delta\mathbf{i})] - 1$$

is the expected rate of return due to price changes, not interest income. One does not earn interest without the passage of time. The model described above is a model at a single point of time.

Let me reformulate condition (2.13) of the paper as

$$E[h(\Delta\mathbf{i})] - 1 \approx -r.$$

I was puzzled that, in Section 2-E of the paper, terms such as "expected return of a five-year bond" and "expected period return" seem to be used to describe r. I assumed that the expected return of a five-year bond meant the yield rate of a five-year bond. However,

$$E[h(\Delta\mathbf{i})] - 1$$

merely measures the expected rate of return due an instantaneous shock. In other words, it measures only the expected rate of return of capital gain and loss, without interest income. I was relieved when I reached Section 4 in which I saw the expression for the forward value of surplus

$$S_k(\mathbf{i}) = S(\mathbf{i})/Z_k(\mathbf{i}), \quad (3.1)$$

and the term "expected yield curve return" used to describe r.

4. Forward Value

Much of Section 4 of the paper is based on the distribution or moments of the random variable $S_k(\mathbf{i}_0+\Delta\mathbf{i})$, $k>0$. [Since $\mathbf{i}_0$ is fixed, the problem of determining the distribution of $S_k(\mathbf{i}_0+\Delta\mathbf{i})$ is essentially equivalent to determining the distribution of $S_k(\mathbf{i}_0 + \Delta\mathbf{i})/S(\mathbf{i}_0)$ or $S_k(\mathbf{i}_0 + \Delta\mathbf{i})/S_k(\mathbf{i}_0)$.] Now, $S_k(\mathbf{i}_0+\Delta\mathbf{i})$ is a forward-value random variable. I do not think that the distribution of the forward value of the surplus at time k is important. What really matters is the distribution of the surplus value at time k. By (3.1),

$$S_k(\mathbf{i}_0 + \Delta\mathbf{i}) = S(\mathbf{i}_0 + \Delta\mathbf{i})/Z_k(\mathbf{i}_0 + \Delta\mathbf{i}), \tag{4.1}$$

which means that, if we want to interpret $S_k(\mathbf{i}_0 + \Delta\mathbf{i})$ as the surplus value at time k, we would make the assumption that, after the shock to the yield curve at time 0, the yield curve movement up till time k is governed exactly by the pure expectations hypothesis, which is not a satisfactory hypothesis. However, to build a continuous-time model that would give the distribution or moments of the surplus value at time k may not be easy because of the Harrison-Pitbladdo-Schaefer theorem.

5. Proposition A

Proposition A of the paper is a very useful result. Its applications in other areas of actuarial science can be found in papers such as those by Gerber and Jones [2], Shiu [4], and Tilley [6]. I would like to present an alternative derivation.

Lemma:
Given $\mathbf{a}\in R^m$, consider the function

$$\phi(\mathbf{x}) = \mathbf{a}^T\mathbf{x} = \mathbf{x}^T\mathbf{a}, \qquad \mathbf{x} \in R^m;$$

then the gradient of $\phi(\mathbf{x})$ is $\mathbf{a}$,

$$\nabla\phi(\mathbf{x}) = \mathbf{a}. \tag{5.1}$$

Corollary:
Given an $m\times m$ matrix $\mathbf{A}$, consider the function

$$\phi(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}, \qquad \mathbf{x} \in R^m;$$

then

$$\nabla\phi(\mathbf{x}) = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}. \tag{5.2}$$

The problem posed in Proposition A of the paper is to maximize

$$\mathbf{x}^T\mathbf{K}\mathbf{x}$$

subject to the constraint

$$\mathbf{B}\mathbf{x} = \mathbf{r}. \tag{5.3}$$

Consider the Lagrangean function

$$H(\mathbf{x}) = \mathbf{x}^T\mathbf{K}\mathbf{x} - \boldsymbol{\lambda}^T\mathbf{B}\mathbf{x},$$

where $\boldsymbol{\lambda}$ is the Lagrange multiplier vector. Applying the corollary and the lemma yields

$$\begin{aligned}\nabla H(\mathbf{x}) &= \nabla(\mathbf{x}^T\mathbf{K}\mathbf{x}) - \nabla(\boldsymbol{\lambda}^T\mathbf{B}\mathbf{x})\\ &= (\mathbf{K} + \mathbf{K}^T)\mathbf{x} - (\boldsymbol{\lambda}^T\mathbf{B})^T\\ &= 2\mathbf{K}\mathbf{x} - \mathbf{B}^T\boldsymbol{\lambda},\end{aligned}$$

from which we get the equation

$$2\mathbf{K}\mathbf{x}_0 = \mathbf{B}^T\boldsymbol{\lambda},$$

or

$$\mathbf{x}_0 = \tfrac{1}{2}\,\mathbf{K}^{-1}\mathbf{B}^T\boldsymbol{\lambda}. \tag{5.4}$$

Substituting (5.4) in (5.3) yields

$$\tfrac{1}{2}\,\mathbf{B}\mathbf{K}^{-1}\mathbf{B}^T\boldsymbol{\lambda} = \mathbf{r},$$

or

$$\boldsymbol{\lambda} = 2(\mathbf{B}\mathbf{K}^{-1}\mathbf{B}^T)^{-1}\mathbf{r}. \tag{5.5}$$

The substitution of (5.5) in (5.4) gives

$$\mathbf{x}_0 = \mathbf{K}^{-1}\mathbf{B}^T(\mathbf{B}\mathbf{K}^{-1}\mathbf{B}^T)^{-1}\mathbf{r},$$

which is (A.3)′ at the end of the Appendix.

The lemma and the corollary are two simple, yet very useful results. In fact, much of the current Course 165, "Mathematics of Graduation," can be simplified by applying these multivariate differentiation formulas. I would also add that the theory of Lagrange multipliers is still an actively researched area; see Rockafellar [3].

6. *Two Inequalities*

I would like to suggest an alternative for Inequality (2.8) in the paper. The eigenvalues of a real symmetric matrix are real. Let m and M denote the minimum and maximum, respectively, of the eigenvalues of the real symmetric matrix $\mathbf{K}$. Then

$$m\mathbf{x}^T\mathbf{x} \leq \mathbf{x}^T\mathbf{K}\mathbf{x} \leq M\mathbf{x}^T\mathbf{x}. \tag{6.1}$$

One way to prove (6.1) is to invoke the *principal axis theorem*, as in the Appendix of the paper, that there is an orthonormal matrix $\mathbf{P}$ such that

$$\mathbf{P}^T\mathbf{K}\mathbf{P} = \mathbf{D},$$

a diagonal matrix. Thus

$$\mathbf{x}^T\mathbf{K}\mathbf{x} = \mathbf{x}^T\mathbf{P}\mathbf{D}\mathbf{P}^T\mathbf{x} = (\mathbf{P}^T\mathbf{x})^T\mathbf{D}\mathbf{P}^T\mathbf{x}.$$

Inequalities (6.1) now follow from

$$m\mathbf{I} \leq \mathbf{D} \leq M\mathbf{I}$$

and

$$(\mathbf{P}^T\mathbf{x})^T\mathbf{P}^T\mathbf{x} = \mathbf{x}^T\mathbf{x}.$$

Since the trace of a matrix is equal to the sum of its eigenvalues, the second inequality in (6.1) gives a tighter bound than (2.8) in the paper. Also, the first inequality in (6.1) proves the last part of Corollary A.1 in the Appendix of the paper. Since $\mathbf{K}$ is positive definite, the matrix

$$\mathbf{C}^{-1} = \mathbf{B}\mathbf{K}^{-1}\mathbf{B}^T$$

is also positive definite and hence its minimum eigenvalue is positive. Consequently, the first inequality in (6.1), with $\mathbf{K}=\mathbf{C}^{-1}$, shows that

$$\mathbf{r}^T\mathbf{C}\mathbf{r} \leq 0$$

if and only if $\mathbf{r}=\mathbf{0}$.

7. *A Lawsuit*

As I read that the Litterman and Scheinkman study would imply that three direction vectors might be used effectively to anticipate bond performance, I was reminded of a 1973 lawsuit, as told by the late Nobel laureate George J. Stigler [5]:

In that year a young man named Dascomb Henderson, a graduate of Harvard Business School (1969) and recently discharged as assistant treasurer of a respectable-sized corporation, sued his alma mater for imparting instruction since demonstrated to be false. This instruction—we may omit here its explicit and complex algebraic formulation—concerned the proper investment of working capital. One of Henderson's teachers at the Harvard school, a Professor Plessek, had thoroughly sold his students upon a sure-fire method of predicting short-term interest rate movements, based upon a predictive equation incorporating recent movements of the difference between high- and low-quality bond prices, the stock of money (Plessek had a Chicago Ph.D.), the number of "everything is under control" speeches given by governors of the Federal Reserve Board in the previous quarter, and the full-employment deficit. It was established in the trial that the equation had worked tolerably well for the period 1960–68 (and Henderson was exposed to this evidence in Plessek's course in the spring of 1969), but the data for 1969 and 1970, once analyzed, made it abundantly clear that the equation was capable of grotesquely erroneous predictions. Assistant Treasurer Henderson, unaware of these later results, played the long-term bond market with his corporation's cash, and in the process the cash lost its surplus character. He was promptly discharged, learned of the decline of the Plessek model, and sued.

This was a new area of litigation, and Henderson's attorney deliberately pursued several lines of attack, in the hope that at least one would find favor with the court:

1. Professor Plessek had not submitted his theory to sufficient empirical tests: had he tried it for the decade of the 1950s, he would have had less confidence in it.
2. Professor Plessek did not display proper scientific caution. Henderson's class notes recorded the sentence: "I'll stake my reputation as an econometrician that this model will not [engage in intercourse with] a portfolio manager." This was corroborated with a different verb by a classmate's notes.
3. Professor Plessek should have notified his former students once the disastrous performance of his theory in 1969 and 1970 became known.
4. Harvard University was grossly negligent in retaining (and hence certifying the professional competence of) an assistant professor whose work had received humiliating professional criticism (*Journal of Business* [April 1971]). Instead he had been promoted to associate professor in 1972.

> The damages asked were $500,000 for impairment of earning power and $200,000 for humiliation.

I leave it to the interested reader to look up the journal (which also contains the celebrated Black-Scholes paper) to find out the ending.

REFERENCES

1. ELANDT-JOHNSON, R.C., AND JOHNSON, N.L. *Survival Models and Data Analysis*. New York: Wiley, 1980.
2. GERBER, H.U., AND JONES, D.A. "Applications of Linear Algebra in Graduation and Other Disciplines of Actuarial Science," *ARCH* 1978.1.
3. ROCKAFELLAR, R.T. "Lagrange Multipliers and Optimality," *SIAM Review* 35 (1993): 183–238.
4. SHIU, E.S.W. "Minimum-R_z Moving-Weighted-Average Formulas," *TSA* XXXVI (1984): 489–500.
5. STIGLER, G.J. "A Sketch of the History of Truth in Teaching," *Journal of Political Economy* 81 (1973): 491–95.
6. TILLEY, J.A. "Achieving Consistency between Investment Practice and Investment Assumptions for Single-Premium New Money Products," *TSA* XXXI (1979): 63–91.

TIMOTHY C. CARDINAL:

I congratulate Dr. Reitano on his efforts in examining immunization that allows rate changes to be multidimensional (that is, multidirectional). Multivariate does not necessarily mean multidimensional. Earlier papers [6], [7], [8], [9], [10], [11], [12], [13] were essentially one-dimensional, albeit in direction **N**. From a practical perspective, one may consider a portfolio immunized if it is immunized in a given finite number of directions. However, from a theoretical perspective, the portfolio is not immunized. In this paper, the analysis is not limited to a finite number of directions.

Immunization theory by its very nature is dependent on the assumption about interest rate movements. It is my perception that Dr. Reitano's analysis depends on the assumptions that **K** is stationary, that interest rate movements are independent, and that second moments are finite; that is, **K** exists. Independence is needed only if more than one time period is analyzed.

Part I of my discussion starts with known consequences in using the above assumptions. I then inquire what happens if one of these assumptions is relaxed. Part II consists of a statistical examination of empirical data, which suggests that one should not assume that all three assumptions hold simultaneously. Becker's results [1] imply a similar conclusion.

I. Assumptions

Consequences of Stationarity, Independence, and Finite Second Moments

Brockett and Witt [2] show that in the traditional stationary independent increments model assuming finite variance, continuous trading and continuous price changes, prices *must* follow a Brownian motion (or a lognormal distribution for prices). The modern approach is to use the principle of no arbitrage. They relate the historical notions of no arbitrage and the use of martingale models. The main result of their paper is very interesting:

Theorem 3

Let $X(t)$ denote the price and $Y(t)=\exp\{-\delta(t)\}X(t)$ the properly discounted present value of $X(t)$. If the market is efficient in the sense that the return process $\xi(t)=\ln[Y(t)/Y(t-1)]$ is a square integrable, stationary, ergodic process and $\ln Y(t)$ has the martingale property, then $X(t)$ is approximately lognormal. Moreover, the entire price process $X(t)$, $0\leq t\leq T$ is simultaneously approximated by a geometric Gaussian process over the entire time interval.

Finite Second Moments

If one assumes increments are stationary and independent, then it can be shown empirically that a non-normal stable process fits the data better than a normal process (see Part II). A consequence is that, theoretically, second moments are infinite. As a result, one must use a bounded utility function to produce finite market prices. An alternative, suggested by the empirical data, is to truncate the distribution, resulting in finite market prices without the overhead of utility functions. This model is not considered to be theoretically attractive.

Stationarity

Section 4.B states ". . . one may be more confident of the stability of **K** than **E** through time. . . ." Although estimates of components may vary between small positive and negative values for different time periods, most studies find that **E** is not significantly different from **0**. Usually it is the stationarity of **K** that is questioned. The "lack of confidence" in the stability of **E** is due to sensitivity to the time period and the interval used to measure changes. One would expect that **E** is not significantly different from **0** for a shorter interval or longer time period.

Implications

Dr. Reitano's series of papers on multivariate duration, convexity, and immunization have successively solved increasingly more difficult problems in more general settings. A process with stochastic volatility (see Hull and White [5]) may explain the magnitudes of the second moments observed in empirical data. However, immunization then becomes a difficult endeavor. I do not see how to extend your results assuming a nonstationary process. Interestingly enough, your results may be generalized to a stable process (see Part II).

The assumptions of stationarity, independence, and finite second moments imply a normal process. Empirical evidence suggests that changes are not governed by a stationary normal process. Stationarity seems to be the logical assumption to relax. Of course, it may be *reasonable* to assume stationarity holds for short time periods. If so, your methodology will provide valuable insights for financial management. Further empirical research on the structure of nonstationarity is needed.

II. Statistical Analysis

Stable Distributions

The reader is referred to Feller [4] and Press [7] for more details. The normal distribution belongs to a larger class known as stable distributions, defined as the class of distributions that are invariant under convolution. In order that $F(\mathbf{x})$ be a limiting distribution of the sum of independent identically distributed (iid) summands, it is necessary and sufficient that $F(\mathbf{x})$ is stable. This is known as the generalized central limit theorem. Note that the variance of the summands is not necessarily finite. Thus, stable distributions retain the properties that make the normal distribution attractive for use in modeling.

Explicit expressions for the probability density function (pdf) or cumulative distribution function (cdf) of a general stable distribution are not known, so expansions must be used. Therefore, most analysis is done through its characteristic function, defined as

$$\phi_x(\mathbf{t}) = E(\exp(i\mathbf{t}'\mathbf{X})),$$

where $\mathbf{X}$ is a $p \times 1$ random vector with multivariate pdf $f(\mathbf{x})$ and $\mathbf{t}$ is a $p \times 1$ vector. In the multivariate case, F is symmetric stable if and only if $\phi(t)$ can be uniquely expressed as

$$\ln \phi(\mathbf{t}) = i\delta'\mathbf{t} - \tfrac{1}{2} \sum_{k=1}^{m} (\mathbf{t}'\Omega_k\mathbf{t})^{\alpha/2},$$

where each

$$\Omega_k = (\omega_{ij}(k))$$

is a positive semidefinite matrix of order p and rank r_j, $1 \le r_j \le p$. The distribution is nonsingular if and only if

$$\sum_{k=1}^{m} \Omega_k$$

is positive definite. It will be assumed that $m=1$. The matrix Ω is called the codispersion matrix. For $\alpha<2$, moments of order less than α exist, but moments of order greater than or equal to α do not exist. For a normal distribution, $\alpha=2$, $\delta=\mu$, and $\Sigma_{j=1}^{m} \Omega_j=\Sigma$. Marginal distributions are given by the appropriate submatrices. In the univariate case, c_i is used where

$$\omega_{ii} = 2^{2/\alpha}c_i^2.$$

The parameter c (dimension i understood) is called the scale parameter and takes a role in measuring risk similar to that of standard deviation.

Let $\mathbf{Y}=\mathbf{AX}+\mathbf{b}$, where $\mathbf{X}$ is a symmetric stable $p \times 1$ vector, $\mathbf{A}$ is a $q \times p$ matrix with $rank(\mathbf{A}_j)=q$, and $\mathbf{b}$ is a $p \times 1$ vector. Then it is easy to show via characteristic functions that $\mathbf{Y}$ is symmetric stable with $\delta_Y=\mathbf{A}\delta_X+\mathbf{b}$ and $\Omega_Y=\mathbf{A}\Omega_X\mathbf{A}'$. The concept of covariance and correlation can be generalized. The codispersion between X_i and X_j is ω_{ij}, and the association parameter ρ_{ij} between X_i and X_j is $\rho_{ij}=\omega_{ij}/(\omega_{ii}\omega_{jj})^{1/2}$. The familiar properties of ρ hold except that when $|\rho|=1$, the functional relationship is not necessarily linear.

Empirical Results

For estimation techniques, the reader is referred to Koutrouvelis's paper [6] for the univariate case and to Cardinal [3] for the multivariate case. Bootstrapping was used to obtain estimates for standard errors. Here the distribution of $\Delta \mathbf{i}$ for the three-point yield curve of 0.5-, 5-, and 10-year maturities is examined. Results for other "key" maturities (0.25, 1, 2, 3, 7, and 30 years) are similar. The data set consists of weekly yields provided by Salomon Brothers, Inc. for the period January 2, 1981 to December 23, 1993. To have a larger data size, the data set is extended roughly 3½ years before and after the one considered in the paper.

A common statistical test is testing the stability-under-addition property. Form data sets by taking nonoverlapping sums of size n and label the estimates with subscript n. Since stable distributions are invariant under convolution, one expects that

$$\alpha_n = \hat{\alpha}_1, \; c_n = n^{1/\alpha}\,\hat{c}_1, \quad \text{and} \quad \hat{\delta}_n = n\hat{\delta}_1.$$

For a nonstable distribution, the $\hat{\alpha}_n$ should have increasingly larger values tending towards 2. Since stability-under-addition holds only for iid variables, a second test is employed by first randomizing the original data set and then forming sums. Randomization should remove any effects of serial correlation and dependency. Failing either test is evidence against a distribution being stationary stable, but passing both tests does not mean that the distribution must be stable. A mixture of normals may pass the first test but fail the second test as $\hat{\alpha}_n$ tends to 2 as n increases.

Tables 1 and 2 present the results of the estimated univariate parameters using weekly changes for four different sum sizes. Table 3 presents the results of using overlapping 6-month periods. Standard errors are given in parentheses. Note that the standard errors became large as the data sets became very small (67 data points for a sum size of 10). For the original data set, estimates of $10^5\hat{\delta}$ are -7, -13, and -9, respectively. In the multivariate case, $\alpha=1.65$ and the estimate of the codispersion matrix is given in Table 4.

For each maturity, the estimates of α for weekly changes are more than 5 standard errors from 2. Thus the hypothesis of normality ($\alpha=2$) can be rejected at the 99 percent level using a t test. For each sum size, the hypotheses that $\alpha_n=\hat{\alpha}_1$ and $c_n=n^{1/\alpha}\hat{c}_1$ are not rejected for both the original and randomized data sets. The hypothesis that $\hat{\delta}=0$ cannot be

TABLE 1

WEEKLY CHANGES—ORIGINAL DATA

Maturity	Sum Sizes 1 $\hat{\alpha}$	1 $10^2\hat{c}$	2 $\hat{\alpha}$	2 $10^2\hat{c}$	3 $\hat{\alpha}$	3 $10^2\hat{c}$	6 $\hat{\alpha}$	6 $10^2\hat{c}$	10 $\hat{\alpha}$	10 $10^2\hat{c}$
6 Months	1.361	0.1134	1.485	0.1781	1.445	0.2094	1.353	0.3131	1.599	0.4608
	(0.062)	(0.006)	(0.077)	(0.018)	(0.113)	(0.018)	(0.174)	(0.046)	(0.174)	(0.064)
5 Years	1.681	0.1208	1.736	0.1819	1.807	0.2323	1.572	0.3288	1.824	0.4355
	(0.052)	(0.004)	(0.079)	(0.010)	(0.092)	(0.015)	(0.156)	(0.047)	(0.123)	(0.068)
10 Years	1.695	0.1150	1.750	0.1723	1.803	0.2183	1.592	0.3076	1.744	0.3706
	(0.046)	(0.004)	(0.070)	(0.010)	(0.089)	(0.015)	(0.121)	(0.041)	(0.169)	(0.067)

TABLE 2

WEEKLY CHANGES—RANDOMIZED DATA

Maturity	Sum Sizes 1 $\hat{\alpha}$	1 $10^2\hat{c}$	2 $\hat{\alpha}$	2 $10^2\hat{c}$	3 $\hat{\alpha}$	3 $10^2\hat{c}$	6 $\hat{\alpha}$	6 $10^2\hat{c}$	10 $\hat{\alpha}$	10 $10^2\hat{c}$
6 Months	1.352	0.1125	1.423	0.1881	1.514	0.2539	1.526	0.3624	1.540	0.5603
	(0.067)	(0.005)	(0.077)	(0.011)	(0.083)	(0.017)	(0.165)	(0.051)	(0.181)	(0.077)
5 Years	1.672	0.1186	1.767	0.1904	1.807	0.2279	1.817	0.3163	1.720	0.4045
	(0.054)	(0.005)	(0.057)	(0.010)	(0.090)	(0.018)	(0.106)	(0.029)	(0.188)	(0.064)
10 Years	1.686	0.1133	1.817	0.1785	1.839	0.2168	1.790	0.2923	1.820	0.4132
	(0.060)	(0.005)	(0.069)	(0.010)	(0.102)	(0.017)	(0.121)	(0.028)	(0.137)	(0.055)

TABLE 3

SIX-MONTH CHANGES

Maturity	$\hat{\alpha}$	$10^2\hat{c}$	$10^2\hat{\delta}$
6 Months	1.660	0.8007	−0.410
	(0.060)	(0.036)	(0.056)
5 Years	1.801	0.7680	−0.375
	(0.062)	(0.040)	(0.056)
10 Years	1.730	0.6677	−0.315
	(0.066)	(0.034)	(0.048)

TABLE 4

WEEKLY CHANGES

$$10^6\hat{\Omega} = \begin{pmatrix} 3.659 & 1.407 & 1.228 \\ 1.407 & 3.378 & 1.917 \\ 1.228 & 1.917 & 3.032 \end{pmatrix} \qquad 10^7\hat{\sigma}_{\hat{\Omega}} = \begin{pmatrix} 2.4 & 2.5 & 3.0 \\ 2.5 & 2.4 & 3.5 \\ 3.0 & 3.5 & 2.4 \end{pmatrix}$$

rejected for either the original or randomized data sets and any sum size—most estimates are within one standard error of 0, a few within two. Finally, a chi-square goodness-of-fit test was performed by using 13 intervals. The 10-year maturity test used 12 intervals; two intervals had to be merged since one interval had less than 5 observations. The results are given in Table 5. Normality may be rejected at any reasonable significance level. The stable hypothesis cannot be rejected at the 97.5 percent significance level for the 6-month or at the 95 percent significance level for the 5- and 10-year securities.

TABLE 5

CHI SQUARE VALUES
WEEKLY CHANGES

Maturity	Stable	Normal
6 Months	23.44	250.08
5 Years	8.88	210.41
10 Years	11.15	232.23

For 6-month changes using overlapping intervals, once again the hypothesis of normality ($\alpha=2$) can be rejected. The stability-under-addition property cannot be directly tested since the 6-month changes are not independent sums of weekly changes. However, the hypothesis that $\hat{c}=26^{1/\alpha}\hat{c}_1$ cannot be rejected at a 95 percent significance level.

Consequences

Empirical evidence demonstrates that the simultaneous assumption of stationarity, independence, and normality is not appropriate. If interest rates are governed by a stable process, measures such as sample standard deviation and covariance are meaningless since the second moments behave as if they were nonfinite. These measures can be replaced by the scale parameter c and the codispersion matrix Ω. Risk is measured by the scale parameter c (replacing standard deviation) or by c^α (replacing variance). Press [7] shows that a portfolio of assets following a multivariate symmetric stable distribution has an efficient frontier. This is done by maximizing

$$\lambda \mathbf{w}'\delta - \frac{1}{2}(\mathbf{w}'\Omega\mathbf{w})^{\alpha/2},$$

where $\mathbf{w}$ is a vector denoting the allocation of assets and λ is a weight determined by the investor.

To generalize Dr. Reitano's results to a stable process, make the following notational changes:

$$\mathbf{E}(\mathbf{i}) \rightarrow \delta_i, \qquad \mathrm{Var}[R(\Delta\mathbf{i})] \rightarrow \Omega_R,$$

$$\mathbf{E}(R(\Delta\mathbf{i})) \rightarrow \delta_R, \qquad \mathbf{E}(\mathbf{i}_0) \rightarrow \delta_0,$$

and

$$\mathbf{K}(\mathbf{i}_0) \rightarrow \Omega_0.$$

To avoid singular distributions, it is assumed Ω_0 is positive definite. Proposition 1 follows from the relationship between parameters for linear transformations given above. Since Ω_0 is positive definite, the rest of the paper's results follow *mutatis mutandis*. The minimization function becomes

$$w\Omega_R^\alpha + (1 - w)|\mathbf{D}(\mathbf{i}_0)|^2,$$

where $\Omega_R = \mathbf{D}\Omega_0\mathbf{D}'$ is one-dimensional and is a convex function in $\mathbf{D}$ for $1 \leq \alpha \leq 2$.

REFERENCES

1. Becker, D.N. "Statistical Tests of the Lognormal Distribution As a Basis for Interest Rate Changes," *TSA* XLIII (1991): 7–57.
2. Brockett, P., and Witt, R. "Relevant Distributions for Insurance Prices in an Arbitrage Free Equilibrium," *Journal of Risk and Insurance* 58 (1991): 13–29.
3. Cardinal, T.C. "A Multivariate Logstable Model of Interest Rate Changes," working paper; copies available upon request from the author at his *Directory* address.
4. Feller, W. *An Introduction to Probability Theory and Its Applications II*. New York: John Wiley, 1971.
5. Hull, J., and White, A. "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance* 42 (1987): 281–300.
6. Koutrouvelis, I. "An Iterative Procedure for the Estimation of the Parameters of Stable Laws," *Communications in Statistics. Part B. Simulation and Computations* 10 (1981): 17–28.
7. Press, S.J. *Applied Multivariate Analysis*. 2nd ed. New York: Holt, Rinehart, & Winston, Inc., 1972.

(AUTHOR'S REVIEW OF DISCUSSIONS)

ROBERT R. REITANO:

I thank Dr. Shiu and Mr. Cardinal for their interesting discussions.

Dr. Shiu is quite right that, in general, one is better off expanding Taylor series about the mean of the random vector, in this case $\mathbf{\Delta i}$, and then performing mean/variance analyses, rather than simply expanding about $\mathbf{\Delta i}=\mathbf{0}$ as I did in my paper. My motivation for this bit of carelessness is my personal conviction that virtually all analyses should be performed assuming $\mathbf{E}(\mathbf{i}_0)=\mathbf{0}$, which in the case of the forward price function model is equivalent to assuming that the future evolution of yields proceeds along today's forward structure. The only exception to this rule for me are simple what-if analyses.

I perhaps should have been clearer about my personal bias (I got close in Section 4-B), especially since some of my formulas contain $\mathbf{E}(\mathbf{i}_0)$; however, I convinced myself that practitioners with other biases should be presented with the more general results. Alas, Dr. Shiu notes that one cannot go halfway on this issue; if we believe $\mathbf{E}(\mathbf{i}_0)\neq\mathbf{0}$, the Taylor series expansions should have reflected this and been handled as he suggests.

Interestingly, Dr. Shiu's formula for $E[R(\mathbf{\Delta i})]$ differs from mine even in the case when $\mathbf{E}(\mathbf{i}_0)=\mathbf{0}$ [see his Formula (1.4) and my Formula (2.5)], in contrast to the case of variance, for which our formulas would agree. This is because of his convexity adjustment, which in this case becomes:

$$\tfrac{1}{2}\, tr[\mathbf{C}(\mathbf{i}_0)\, \mathbf{K}(\mathbf{i}_0)].$$

In general, I think that this refinement in estimates of $E[R(\mathbf{\Delta i})]$ is worthwhile if $\mathbf{C}(\mathbf{i}_0)$ has been estimated, independent of the practitioner's bias, and I have advocated this in another context (see Formula (A.14) in Reitano [5]). However, as noted in Elandt-Johnson and Johnson's book [2, p. 72], "often the approximation $E[g(x)]\approx g(E(x))$ is used," implying that his recommended refinement may not be as commonly used in practice as suggested.

Regarding Dr. Shiu's free-lunch analysis, I have little to add to this discussion beyond what I have addressed in my previous author's discussions [3, 4].

Dr. Shiu's puzzle brings out an important point about the definition of return in this analysis; that point being, the definition comes from the model. If $P(\mathbf{i}_0+\mathbf{\Delta i})$ represents the current price of a security after a yield curve shift of $\mathbf{\Delta i}$, then $E[R(\mathbf{\Delta i})]-1$, with $R(\mathbf{\Delta i})$ defined in my Formula (2.1), approximates the expected instantaneous yield curve shift return and will not equal the total return of the security over the given instantaneous time interval.

However, to represent total returns over a fixed period, a simple change in the model is sufficient. Specifically, as in Section 4C, we need to redefine the ratio function, $R(\mathbf{\Delta i})$, to $R_k(\mathbf{\Delta i})$, where k is the length of the period of interest. This new ratio function is the ratio of the forward shifted price to current price, $P_k(\mathbf{i}_0+\mathbf{\Delta i})/\mathrm{P}(\mathbf{i}_0)$, and it can be factored into the product of c_k, which reflects the period's "fixed return," and the ratio function $R(\mathbf{\Delta i})$ applied to the forward price function, $P_k(\mathbf{i}_0+\mathbf{\Delta i})$, which reflects the period's expected yield curve shift return. The expression, $E[\mathrm{R}_k(\mathbf{\Delta i})]$, then captures total return in the usual sense of the phrase. Expressions for annualized total return are also readily derivable (see Section 4-C).

Regarding my notion of forward value, Dr. Shiu is right that this is not a perfect model of the future value of surplus at time k, since after the shock, the evolution on the forward structure is explicitly assumed. However, this notion works very well in practice, especially when k reflects one to a few months, and it can be adapted to yield good insights on returns over one year. It also works compatibly with an active management strategy as discussed in my paper. It is especially gratifying that such a simple model works well in light of the Harrison-Pitbladdo-Schaefer theorem, which virtually assures us that the real $S_k(\mathbf{i}_0+\mathbf{\Delta i}(k))$ is of unbounded variation as a function of k, where $\mathbf{\Delta i}(k)$ is the shift from time 0 to time k.

Dr. Shiu provides a compact proof of my Proposition A based on the gradient formula (5.2). This is an especially nice approach since the cumbersome change of basis methodology I employed is circumvented.

His inequalities (6.1) also imply a better upper bound for variance than does my trace result in Formula (2.8). However, even though I am quite familiar with (6.1.) (see Proposition 12 in Reitano [3]), I could not resist the result given since the trace of $\mathbf{K}$ is the sum of the individual variances of the various yields, whereas the smallest and largest eigenvalues have no such pleasing statistical interpretation (of course, the sum of all eigenvalues equals the sum of all variances). It is also worth noting that, in practice, the largest eigenvalue of $\mathbf{K}$ is 80·percent or so of the trace of $\mathbf{K}$, so my result is not as crude as one may first imagine.

Finally, Dr. Shiu provides an interesting review of a 1973 lawsuit, which he was reminded of by my reference to the work of Litterman and Scheinkman. Regarding these authors' principal component analysis, I agree with Dr. Shiu's suggestion that more work needs to be done, especially with regard to out of sample testing. I share Dr. Shiu's skepticism and do not believe I could have hedged my current endorsement of this method further. Of course, I rather prefer the methodology I introduce in this paper.

Nevertheless, I genuinely enjoyed the excerpt from George J. Stigler's review of the *Henderson vs. Harvard* case, enough to look up the rest of the story. It is a delight to read, full of wonderful anecdotes. For example, as it turned out, Harvard's defense prevailed in the lower court, and "Judge MacIntosh (Harvard, LL.B. 1938) asserted that university instruction and publication were preserved from such attacks by the First Amendment, the principle of academic freedom, an absence of precedent for such a complaint, and the established unreliability of academic lectures."

Many more amusements follow with the appeal and aftermath, and I join Dr. Shiu in encouraging the reader "to find out the ending."

Mr. Cardinal begins with a delineation of multivariate versus multidimensional, which I believe is more theoretical than real. In earlier papers, I modeled the multivariate yield curve shift by either $\Delta i\mathbf{N}$ or $\mathbf{\Delta i}$, where in the former $\mathbf{N}$ is considered fixed and Δi variable, while in the latter, $\mathbf{\Delta i}$ is a variable vector. However, this delineation was introduced entirely for pedagogical reasons, to help bridge the gap between a truly multidimensional/multivariate representation of yield curve shifts and the truly one-dimensional representation within the parallel shift model. In a sense, I tried to trick the nontechnical reader into thinking of the multidimensional model as just a generalized parallel shift model. Of course, once results were developed for a given $\mathbf{N}$, all the generality of the $\mathbf{\Delta i}$ model was obtained, since one could now define $\Delta i = 1$ and let $\mathbf{N}$ vary. The only missing piece was the general mathematical formulas relating the two approaches, which can be developed only within the somewhat more challenging $\mathbf{\Delta i}$ model and which I hoped the reader would then be motivated to pursue.

Except for the case of nondirectional immunization (Reitano [4]), which requires extremely restrictive conditions on one's portfolios, Mr. Cardinal is quite right that by their very nature, immunization theories depend materially on the assumptions one makes about yield curve movements. For the current paper, the only assumption needed for the theory to apply is the existence of the covariance matrix, **K**. That is, the theory in my paper does not require **K** to be stationary, nor does it require that interest rate movements be independent. Explicitly, I assume that the practitioner has a multivariate probability distribution function, $f(\mathbf{\Delta i})$, believed applicable over the succeeding fixed period of time and that this probability distribution has two finite moments: **E** and **K**.

In practice, such a distribution is often based on some historical experience, although the theory does not require it and works equally well on hypothetical distributions, or distributions on which one wants to take a speculative position. For history to be meaningful, complete stationarity is not needed, only what might be called "relative" stationarity. That is, if we assume that the variability of the covariance structure is smooth, we can develop a reasonable estimate of the future covariance matrix from that of the recent past. On the other hand, if one posits that this variability follows a jump process, not even recent history is of much value.

Similarly, the use of historical information does not necessitate the assumption that $\{\mathbf{\Delta i}\}$ are independent period to period, but only the weaker assumption that some transformation of these shifts are independent random vectors.

For example, if one subscribes to a log-type model (although the work of Becker [1] and others argues against lognormal), one does not require the independence of $\mathbf{\Delta i}$, but of $\mathbf{ln}(\mathbf{i}_1/\mathbf{i}_0)$, defined by

$$\mathbf{ln}(\mathbf{i}_1/\mathbf{i}_0) \equiv (\ln(i_{11}/i_{01}), \ln(i_{12}/i_{02}), \ldots, \ln(i_{1m}/i_{0m}))$$

where $\mathbf{\Delta i}=\mathbf{i}_1-\mathbf{i}_0$. Assuming that the above vectors are independent and that the implied distribution has a finite second moment matrix, my analysis can easily be adapted to a model for $\mathbf{\Delta i}$ over the forthcoming period. Specifically, one defines:

$$\begin{aligned}\mathbf{\Delta i} &= \mathbf{i}_0(\mathbf{e}^{\mathbf{X}} - \mathbf{1})\\ &\equiv (i_{01}(e^{X_1} - 1), i_{02}(e^{X_2} - 1), \ldots, i_{0m}(e^{X_m} - 1)),\end{aligned}$$

where $\mathbf{i}_0$ is the beginning of period yield curve.

We then define **K** on the data developed by drawing **X** from its assumed distribution, based on historical observations. Again, we require only that this distribution be relatively stationary.

Regarding the probability distribution of the yield vector, $\mathbf{\Delta i}$, Mr. Cardinal extends the earlier work of Becker [1] and others to the multivariate case: that yield vectors are not multivariate normally distributed (that is, the affirmative hypothesis can be rejected with high confidence). He further shows that the hypothesis that such **Δi** follow a multivariate stable distribution cannot be so rejected.

While stable distributions in general have no covariance matrix, **K**, they do have coassociation matrices, which apparently can be used in an analogous way in my paper's results, subject to various other reinterpretations. Unfortunately, Mr. Cardinal does not provide sufficient details for me to understand clearly all the substitutions made. For example, the proposed minimization function in his last paragraph cannot in general be expressed as a quadratic form in **D**, that is, $\mathbf{DKD}^T$, so the application of my Proposition A is not apparent. However, this appears to be an interesting area for future research.

On the other hand, I am troubled by Mr. Cardinal's implication that my paper makes the assumption that **Δi** follows a multivariate normal distribution, when it only assumes the far weaker condition of the existence of a covariance matrix **K** as noted above. Implicitly, Mr. Cardinal's hypothesis testing, which could not reject the stable distribution but could reject normality, leads him to the conclusion that **Δi** must have a stable nonnormal distribution, and hence, **K** cannot exist.

In practice, however, sample covariance matrices for the random vectors, **Δi**, have been developed by many researchers, on many bases, and over many time periods, and these matrices have proven to give valuable insights into many problems related to yield vector variability. Given that sample **K**'s have proven to be so useful, I am at a loss to believe that, in theory, these **K**'s do not exist. Perhaps the models used more generalization to bring better closure between theory and practice. One possibility is Mr. Cardinal's referenced work of Hull and White, which deals with stochastic volatilities.

As noted above, I wholeheartedly agree with Mr. Cardinal's assertion that the mean vector $\mathbf{E}(\mathbf{i}_0)$ largely reflects the economic cycle underlying the data and that, in general, "one would expect that **E** is not significantly different than zero." I always assume this in practice, since anytime I assume otherwise, I find that I am able to construct tremendously profitable portfolios with modest risks. This conclusion provides me with proof positive that a mistake has been made.

In closing, let me again thank Dr. Shiu and Mr. Cardinal for their challenging and thought-provoking discussions.

REFERENCES

1. BECKER, DAVID N. "Statistical Test of the Lognormal Distribution as a Basis for Interest Rate Changes," *TSA* XLIII (1991): 7–57.
2. ELANDT-JOHNSON, R.C., AND JOHNSON, N.L. *Survival Models and Data Analysis*. New York: Wiley, 1980.
3. REITANO, ROBERT R. "Multivariate Duration Analysis," *TSA* XLIII (1991): 335–91.
4. REITANO, ROBERT R. "Multivariate Immunization Theory," *TSA* XLIII (1991): 392–438.
5. REITANO, ROBERT R. "Nonparallel Yield Curve Shifts and Convexity," *TSA* XLIV (1992): 479–507.