TWO PARADIGMS FOR THE MARKET VALUE OF LIABILITIES

by Robert R. Reitano*

ABSTRACT

Asset/liability management (ALM) theory and practices of insurers have matured and developed from early applications to guaranteed investment contracts (GICs) to all annuity and insurance products today. An important and logical next step of inquiry is the definition of, and calculation procedures for, the market value of an insurance liability. Because all ALM strategies have as their goal the management of some value of assets in relation to some value of liabilities, this inquiry will provide at last a canonical basis for ALM: the management of relative market values.

To set the stage for this exploration, the theory and application of pricing in a complete market are reviewed, as are the practical limitations of this theory in the real, and far from complete, financial markets. The notion of an ad hoc pricing model is developed, and examples are reviewed and critiqued. These models, though imperfect compared with pricing in a complete market, bridge the gap between pricing theory and practice.

The current state of the liabilities market is also discussed, and this market is seen to naturally split into a ''long'' and a ''short'' submarket. Of particular interest is the theoretical possibility of these markets becoming broad-based, deep and active, and the conclusions are relevant to the issue of long/short price equalization.

Two paradigms are then explored for defining and subsequently calculating an insurance liability market value. A ''paradigm'' is a generalized model or framework for accomplishing the task at hand. Each paradigm reflects observable market trading activity, however infrequent, and each is based on methods of valuation consistent with finance-theoretic approaches that are routinely used for the market valuation of assets.

In addition, each paradigm allows for a sequence of ad hoc valuation methodologies, which differ in the extent to which various risks are explicitly modeled versus judgmentally reflected in a risk spread. These paradigms are discussed and contrasted, and arguments made for the potential evolution of the respective values if a ''liability'' market began trading actively. Practical constraints on the realization of this evolution are also noted.

The last section of this paper discusses a host of considerations related to the application of option-pricing theory to insurance company liabilities.

1. INTRODUCTION

The "market" value of liabilities (MVL) refers to the market value of liabilities as it might be defined and calculated if a deep and active market truly existed. Although this market does not exist in the real sense, that is, related trades are sparse and the market is very thin, this in no way precludes the theoretical exploration of market value pricing in a hypothetical active market.

After all, the financial markets repeatedly introduce new products, and both buyers and sellers are able to develop sensible approaches to their theoretical "market" prices, which in turn form the foundations for trading activities and, ultimately, real market prices.

The market is oftentimes able to evaluate new products in a rational manner because of the so-called "law of one price." Specifically, this law applies to pairs or groups of financial instruments that behave identically in all future states of the market, and it asserts that they must be priced identically today. Otherwise, market participants would recognize the "arbitrage" opportunity, buying the "cheap" instrument

*Robert R. Reitano, F.S.A., Ph.D., is Vice President, Investment Policy and Research, John Hancock Mutual Life Insurance Co., 200 Clarendon St., Boston, Mass. 02117.

and selling the "dear," and book an immediate profit with no future risk. Consequently, while departures from this law can occur, the market activities of arbitrageurs put upside pressure on the price of the cheap contract through buying, and downside pressure on the dear contract through selling, until the prices stabilize to market equilibrium prices. In practice, these prices do not converge exactly, but only to within a margin reflective of trading costs. At this point, arbitrage is stopped and so too is price convergence.

As an example, a European option on a stock can be priced in theory by using the law of one price because a stock option can be "replicated" by a dynamically managed portfolio of stock and risk-free bonds. Replicated means that the portfolio and option behave identically in all future states. Consequently, the price of the stock option is theoretically equal to the known price of the replicating portfolio. Black-Scholes (1973) and Cox-Ross-Rubinstein (1979) were the seminal papers for the continuous time and the discrete time developments of this notion, respectively.

While formally relating to identical contracts, this law can easily be applied to the pricing of "similar" contracts. That is, if two financial instruments behave similarly in all future states of the market, they must be priced similarly today. Also, the more similar is the behavior in these future states, the more similar are the current prices. For example, if in every future state the payoff on security B is in between the payoffs of securities A and C, it must be the case that the prices of the securities satisfy: $p^A \leq p^B \leq p^C$. If not, market participants would put pressure on the unacceptably disparate prices through an obvious arbitrage until the expected ordering of prices resulted. In addition, if the differences between the B and A state dependent payoffs were double the differences between the C and B payoffs, it must also be the case that $p^B - p^A = 2(p^C - p^B)$. Otherwise, a risk-free arbitrage is again possible.

As an example, consider private placements. In general, private placement pricing follows public bond pricing in a way that reflects their similarities—maturity, quality, and embedded options—and their differences—protective covenants and liquidity. While private placements are relatively easy to sell, the comparative thinness of the market makes such sales subject to higher bid/ask spreads to compensate the investment banker for either inventorying the bond or facilitating its sale to a third party. This extra cost, which results in a smaller expected sales payoff in any future state, is in theory built into the initial price.

Further, public bonds with thinner markets tend to trade at prices nearer to those of private placements of the same quality, maturity, and optionality. Private placements, which are issued under Section 144a of the SEC regulations and are correspondingly more liquid than traditional private placements, trade at prices closer to those of comparable public bonds.

On the other hand, private placements often have protective covenants that provide value vis-a-vis the corresponding public issue by enhancing the workout payoffs in any future default state. Once again, this value is in theory reflected in the price.

In summary, where there is financial market activity, there are applications of the law of one price. It forms the basis on which the market evaluates new products, perhaps by first decomposing them into simpler more identifiable parts, and it is also the basis by which, through active trading, the market maintains a discipline on the relative valuation of all products.

Consequently, while there is currently no active market for the trading of insurance company liabilities, this does not preclude the estimation of the market values that would in theory exist in an actively traded market. The market has already developed the necessary tools and discipline to accommodate similar situations on numerous previous occasions.

For a development of a measurement theory of integration between two markets, see Chen and Knez (1995). For a market valuation analysis of the property-liability insurance industry, see Babbel and Staking (1995). For more discussion and detail on the implementation of the methodologies defined below, see Becker (1993) and the American Academy of Actuaries Task Force Report, "Fair Valuation of Life Insurance Company Liabilities," dated October 30, 1995.

2. PRICING THEORY AND PRACTICE

The law of one price is perhaps the fundamental insight of financial pricing theory. This section addresses the issue of how this "law" can be used in pricing practice. For example, suppose that security A is identical to security B in terms of all future payoffs in all future states except that it pays an extra $1 at time 5 for certain. Clearly, the price of A must exceed that of B, otherwise an arbitrage is possible: buy A, sell B, and for no cost (or a profit) obtain a free $1 at time 5. But in practice we need to know more. We need to know how much the price of A should exceed the price of B.

The logical answer to this question is that this excess must equal the cost of a security C that pays a certain \$1 at time 5. However, not all such securities need exist. For example, rather than provide a certain extra payment, security A might pay only the extra \$1 in one or a fixed number of states of the world at time 5, and not otherwise. For the law of one price to provide the prices of all securities in practice, the market must contain many traded securities. Specifically, for any given future state, a security is needed that pays \$1 in that state and \$0 in all other states. Such a security is called an Arrow-Debreu security, and such a market a complete market. In theory, a complete market does not actually have to trade all Arrow-Debreu securities, but all such securities either are traded or can be replicated by other traded securities.

More formally, we can in theory parametrize all future states by the pair $(t, \mathbf{X})$, where t denotes time, and the vector $\mathbf{X}=(X_1, X_2, \ldots, X_n)$ has as components the factors that collectively define all future states of the world. In general, $\mathbf{X}$ evolves with time, $\mathbf{X}\equiv\mathbf{X}_t$, reflecting an evolution model of an information structure, but this dependency is notationally suppressed below. For example, the first several components of $\mathbf{X}$ might reflect yields of "on the run" Treasury securities, while others might reflect various nations' GDPs and CPIs, the value of the S&P 500 and other stock market indexes, the volatilities of various markets, foreign currency exchange rates, and so on. Each component X_j can assume a range of numerical values defined by an interval I_j, which can be taken to be bounded. Consequently, $\mathbf{X}$ has values in the product space $\mathbf{I}=\otimes I_j$. For each $\mathbf{X}_0\in\mathbf{I}$ and time t_0, an Arrow-Debreu security has payoff function $C(t, \mathbf{X})$, so that $C(t_0, \mathbf{X}_0)=1$, and $C(t, \mathbf{X})=0$, otherwise.

In complete markets, pricing theory and practice converge. Every security can be decomposed into, that is, "replicated" by, a sum of Arrow-Debreu securities, which in turn either are traded or can be replicated by other traded securities. By the law of one price, each such security must in turn be priced to equal the "sum" of its component prices; otherwise an arbitrage will correct the mispricing. Of course, in order for such a "sum" to make sense, it is sufficient to assume that $\mathbf{I}$ only contains a finite number of states and the time interval only a finite number of discrete times. However, the finiteness in the number of time-states is not as restrictive as it first appears.

For example, it seems reasonable to assume that there is an N so that the market cannot distinguish between two Arrow-Debreu securities with payoff states $\mathbf{X}$ and $\mathbf{X}'$ at times t and t' so that $|(t, \mathbf{X})-(t', \mathbf{X}')|<10^{-N}$ for any reasonable definition of norm, $|(t, \mathbf{X})|$. Consequently, the finiteness of the number of future time-states follows from this, the finiteness of the number factors, the boundedness of the I_j, and the boundedness of the time interval for all securities except perpetuities.

Unfortunately, financial markets are far from complete because of the complexity of the necessary state space and the relative scarcity of traded securities. However, some "submarkets" are nearly complete, for example, the U.S. Treasury market. Because this market is in theory risk-free in U.S. currency, prices of option-free Arrow-Debreu securities are independent of the value of $\mathbf{X}$ and depend only on t. Since Treasury securities trade at virtually all maturities in monthly or quarterly step sizes and most are option-free, in theory option-free Arrow-Debreu securities can be created by mixing long and short positions of coupon-bearing securities. The amount of each security needed can be derived by solving a linear system of equations. In practice, these securities trade under the name Treasury strips or Treasury zeros. Each pays one U.S. dollar amount at a fixed time in all states, and \$0 otherwise.

Because Treasury security prices are independent of $\mathbf{X}$ and reflect only t, there is a natural one-to-one correspondence between the collection of prices, $P(t)$, and a collection of "implied spot yields," i_t, which provide the necessary discounting when defined by $P(t)\equiv\exp(-ti_t)$. This correspondence makes the pricing of both risk-free and certain cash flows easy. Given any collection of such flows $\{c_t\}$, one can in theory use the prices of Treasury strips, as required by the law of one price, or the associated Treasury spot rates to calculate price:

$$P=\sum c_t P(t)=\sum c_t \exp\,(-ti_t).$$

In practice, supply/demand pressures and investor tax effects prevent the prices of Treasury securities from exactly satisfying this identity in terms of Treasury strip prices even to within trading costs, so the market is only "nearly" complete. Similarly, the securities issued by many foreign governments are risk-free in local currency, and the respective markets nearly complete.

While the financial market overall is not complete, the law of one price has practical value for numerous applications in which a given traded security can be "replicated" by a collection of other traded securities. For example, a European call or put option on a stock can be replicated by the underlying stock and a risk-free bond; risk-free bonds with embedded options can also be replicated by long and short option-free and risk-free bonds; a European put option on a stock can be replicated by a European call option, a short position in the stock, and a risk-free bond (this is known as put-call parity); a Treasury security can be replicated by a Canadian government bond and a series of foreign currency swaps, and so on. Consequently, even though Arrow-Debreu securities do not in general exist, in many cases they are not needed. Prices in practice must equal the theoretical prediction of the law of one price as long as the respective securities can be replicated by traded securities.

Unfortunately, many securities also exist that cannot be replicated by traded securities. The simplest example is a fixed bond issued by a corporation. Unlike a risk-free bond with cash flows that depend on time only, a corporate bond's cash flows also depend on the company's financial health. Corporate bonds with embedded options are also problematic. More difficult yet are insurance company liability payments, which, like corporate bond payments, are made subject to an insurer's financial health but, unlike bonds, often depend on contingencies related to a third-party's mortality, morbidity, or other event risk.

Perhaps the most complicated example of security pricing relative to the law of one price is the equity of a corporation or, more generally, the purchase price of its distributable earnings. Like the corporate bond, future cash flows (here earnings) also depend on the health of the corporation, but unlike the bond, even in times of strength earnings can vary materially in different future states of the world.

For such valuations, practical ad hoc approaches have been developed. Unlike the Treasury bond market in which fixed bond prices can be reasonably calculated based on either the prices of zeros or by valuations on the Treasury spot rate curve, for these more complicated valuations no such one-to-one correspondence exists between the current market price and a practical mechanism for translating promised or expected future payments to that market price. Any ad hoc approach represents a stylistic blend of "market judgment"—the model sensibly reflects the nature and extent of the uncertainties in future cash flows—and "market fit"—when applied to observable market securities the model reasonably reproduces actual prices.

For example, the law of one price demands that a corporate bond sell at a discount relative to its risk-free valuation price: $\Sigma c_t \exp(-ti_t)$, since otherwise an arbitrage with the Treasury strip market is possible. Once sorted into default risk (that is, ratings) classes, these discounts tend to display two patterns:

- For a given risk class the discount increases with bond maturity
- For a given maturity the discount increases as default risk increases.

The first observation suggests that the above discounted-cash-flow model should be modified by multiplying the resulting price by a discount factor, say $\exp(-ns_n)$, where n reflects the bond's maturity. However, since the early cash flows of two similarly rated bonds are presumably equally risky, they should be priced independently of the bond's final maturities. Consequently, a better pricing model is one that discounts each cash flow separately by $\exp(-ts_t)$, which is equivalent to simply increasing the risk-free rates, i_t, by the "spread," s_t. Such inferred spreads are then seen to have a "term structure"; that is, s_t tends to increase with t. Also, reflecting the second observation above, these term structures increase as default risk increases.

Besides fitting traded prices well, this ad hoc spread-pricing model also satisfies the market judgment criterion above. Specifically, ignoring unanticipated shifts in either the term structure of i_t or s_t, that is, assuming that the combined term structure moves into its forward structure with time, a corporate bondholder earns an excess return vis-a-vis a Treasury bondholder, at a level reflective of the spread, in any year in which the corporate bond does not default. That this should occur seems logical because corporate bondholders inevitably face default losses, which dilute some of these excess returns. Empirically, these excess returns exceed losses on average for a well-diversified portfolio, and this too seems logical because investors should be compensated for the risk of the year-to-year volatility of loss experience.

An alternative ad hoc model for corporate bond pricing is one that explicitly recognizes the dependence of future cash flows on the event of default and the probabilities of default in each year. This is the "mortality" model, so named because of the obvious analogy with the pricing of life insurance. The problem with this approach is that while historical probabilities can be calculated, they cannot be translated into price without an explicit assumption about the market's utility function, that is, the market's mechanism for charging for risk. That the market indeed charges for risk—is risk averse—is apparent because corporate debt also trades at a discount to the risk-free value of "expected" cash flows using any historically reasonable prediction of defaults.

Besides corporate default, "spreads" are also used as the ad hoc explanatory variable for a host of other risks. For example, because an illiquid bond of a given quality often sells at a discount relative to the price implied by the spreads for comparable quality liquid bonds, a "liquidity spread" is inferred. Again, this model fits prices well and has the intuitive justification of "extra return for extra risk" until the loss associated with liquidity is incurred. Embedded options also are often quantified in terms of spreads. For example, the price of the callable bond is seen to equal the discounted value of fixed cash flows with an extra spread for optionality. For bonds with long embedded put options, this spread is negative.

Virtually any risk can be so quantified, even complicated risks such as those encountered with the valuation of corporate earnings. For example, given a corporation's current equity value and virtually any projection of earnings, there is obviously some risk spread that can be inferred. Equally obvious, the larger the spread, the smaller the current value of projected cash flows, so "spread" is at least a correlated proxy for risk. The question is not whether such a spread can be calculated, but rather whether such a spread is useful for the relative valuation of "similarly" uncertain cash flows.

More generally, the problem of pricing in practice is not one of model calibration, that is, the mapping of actual market prices to assumed pricing models in order to reproduce observable values, but one of "market prediction," that is, whether such a calibrated model can then predict the price at which the market trades a new security that appears comparable.

In the following sections, the spread valuation model is often referenced as the ad hoc methodology for valuing many uncertainties in cash flows. Because of the incompleteness of the markets, some approach is needed so that information contained in the prices of traded securities can be transferred to the pricing of securities that are neither actively traded nor can in general be replicated by actively traded securities. In a sense, these valuation models serve the purpose of approximately "completing" an otherwise incomplete market. Rather than replicating a given security by traded securities, which may be impossible, we instead attempt to replicate the market's mechanisms for generating the observable prices of traded securities and apply these assumed mechanisms to the valuation of the security in hand.

While providing a useful relative valuation tool, the spread valuation model is not without its shortcomings. In general, spreads are not additive, so, for example, the option spread for a given optionality structure depends on the quality of the bond. As another example, while quality and liquidity spreads are relatively stable across asset classes, minor differences in an optionality structure can vary the inferred spread materially.

Fortunately, the price effect of embedded options can usually be valued directly, if only approximately, through option-pricing methodologies by making an ad hoc spread adjustment to the yield structure for credit risk. Consequently, although an implied spread for any option can always be deduced from price, one does not need to predict that this spread will be similar to that of a "comparable security." In this case, predictions can be replaced by calculations.

As a final caveat on the spread valuation model, any observable market price for a stream of uncertain cash flows reflects the market's risk-adjusted distillation of all future cash-flow streams in all future states of the world. That such a risk distillation can be further distilled to spreads that are usable for pricing "comparable" securities cannot be assumed in theory and should not be assumed in practice without substantial market testing and validation.

For a far more complete analysis of pricing theory in the financial markets see Duffie (1988, 1992), Geanakoplos (1992), Huang and Litzenberger (1988), and Martin, Cox, and MacMinn (1988).

3. OTHER CONSIDERATIONS IN THE MARKET VALUATION OF LIABILITIES: THE "LONG" AND THE "SHORT"

An insurance company liability represents a financial contract between the insurer (the issuer) and a second party (the owner) to pay certain amounts based on certain contingencies typically related to a third party (the insured). When attempting to define the market value of such a liability, we must consider whether this market value should be the same for the short position (the issuer) as for the long (the owner).

In the asset markets, traded securities like bonds represent long interests in a company's future cash flows, and it is these positions that are typically actively traded. Market participants can take long or short positions in such securities, purchasing the rights to these cash flows or incurring the liability to replace the value of these cash flows to the original owner in the future, respectively. Issuing companies that are originally in a short position for these contracts can effectively offset these obligations, that is, sell their short position and deleverage the company, by repurchasing their securities in the open market. In general, except in the case of a corporate merger or acquisition, one company does not offset its short position by "trading" its obligations to another company.

In the insurance liability "market" are examples of the trading of both long and short positions. The primary example of long market activity is the secondary GIC market, whereby a "stable value" fund manager sells the proceeds from a GIC to an independent third party. Here, the third party is purchasing the GIC issuer's cash flows for a given price. This transaction is effectively identical to the purchase of a corporate bond. However, because this market is relatively thin, only long positions are currently traded; that is, no reports have yet been made of investors selling a given insurer's GICs short.

The primary example of short market activity is the sale of a block of liabilities from one insurer to another. In this case, the buyer receives the price from the seller in cash or assets as compensation for assuming the responsibility of making cash-flow payments to contract-owners under the conditions of the given policies. Because there is not an active long market for insurers' liabilities other than GICs, an insurer has no way currently to offset most liabilities through open market repurchases. Consequently, short-market activities provide the only facility for an insurer to deleverage its balance sheet.

Because the goal of this paper is to explore two paradigms for the market value of a liability as it might be defined and calculated if an active market truly existed, we next consider to what extent long and short active markets can truly exist.

For short-market activities, while the market is still thin by asset market standards, virtually every type of contract has traded through block sales, corporate acquisitions, or mergers. As the level and pace of this activity increase, there do not appear to be any material obstacles to the development of a relatively deep and active short market.

On the other hand, the long-market activity of secondary market GICs, which could well develop into a deep and active market if insurer downgrades accelerate, appears to have no chance of expansion to the majority of other insurance company liabilities. That is, we cannot expect most liabilities to ever trade like bonds in an open market. Secondary market GICs are an exception because they represent simple financial contractual obligations between the insurer (the issuer) and buyer (the owner), which are virtually identical to a typical debt instrument.

Analogously, we can imagine similar markets in single- and flexible-premium deferred annuity (SPDA/FPDA) contracts and "market-value adjusted" (MVA) annuities, which are much like bonds or GICs. These contracts are complicated somewhat by the existence of an annuitization option, which involves the survivorship of one or more third parties to whom life payments are defined. However, other than the administrative complications for the issuer in making payments to the investor during the lifetime of an independent third party and possible regulatory restrictions, there does not appear to be any obstacle to the eventual evolution of an open market whereby annuity owners sell future proceeds to investors. Indeed, the market for insurance company fixed annuities associated with state lotteries is already well-developed.

For most other insurance liabilities, the implications of third parties identified in the contracts are very significant and financially material. For life insurance contracts, insurable interest between the owner and insured "third party" is required by state statutes and prudent for the issuer and insured in any case to avoid moral hazard. Similar arguments can be made for disability, health, and casualty contracts, whereby it is in the issuer's financial interest that the owner and recipient of benefits have an insurable interest in that which is insured.

While trusts have been used to facilitate the sale of life insurance benefits on individuals with AIDS, where proceeds have been used for medical care, this market is more akin to a secured financing market than a true long market in life insurance liabilities.

Consequently, while it is interesting to muse over the likely evolution of market pricings if long positions in all insurance liabilities were actively traded in an open market, the issuer's need for owners to have an insurable interest in the persons or objects insured precludes most insurance liabilities from ever achieving such a trading status.

Therefore, the market value of liabilities "as it might be defined and calculated if a deep and active market truly existed" must therefore refer to the value that would be observed in an active and deep market of short position trading. In particular, this value is the subject of this paper.

In the asset markets, the market value of a long position should equal that of a short position within trading expenses (that is, bid/ask), and market activities abound to reinforce this view. For the special cases of GICs and annuity contracts, which could in theory trade both as short positions or like bonds as long positions, price equality could occur or not.

State guarantee funds provide value to owners of insurance contracts and owners of insurance companies by providing insurer credit support at a cost typically unrelated to the risks assumed. Because of this, we would expect that the market value of a long position (to the owner) would in general exceed that of a short position (to the issuer), with the difference equal to the value of the issuer's put option to the fund. Similarly, we would expect that the market value of the company to its owners is enhanced to the extent that the benefit of a lower cost of funds is not offset by the cost of the guarantee fund assessments. See Babbel and Staking (1995) and the discussion below.

It is tempting to speculate that the presence of this put option and its implied credit enhancement of the insurer is the driving force behind the existence of a "secondary" GIC market. After all, any issuer would likely retire the GIC at its internally assessed market value, so the secondary market must place a higher value on the long position than the issuer places on the short. Although the value of the put option may be partly responsible for this valuation discrepancy, the more likely explanation is that most issuers discount payable proceeds on early GIC termination at rates above current market in order to fully recover initial expenses and pricing profit.

4. TWO PRICING PARADIGMS

Applying the framework above to an insurance company liability, two pricing paradigms emerge, both of which allow for a sequence of ad hoc pricing methodologies that differ in their implementation from simple yet judgmental to more complicated and yet more objective. The transition in each sequence is created by eliminating judgmental risk-based spreads in the discount yield curve that would be applied to relatively simplified cash flows and, instead, modeling the risks more explicitly through projected scenarios of future state-dependent cash flows.

The two pricing paradigms are direct and indirect and can be described as follows.

- *Direct Paradigm*. The direct pricing paradigm views an insurance company liability much like a corporate bond, whereby contractually defined payments are to be made to a third party subject to a variety of contingencies. Payments can be defined in fixed dollars or indexed relative to the performance of actual or market benchmark indexes or portfolios, and in the latter case minimal payments can be defined. In addition, payment contingencies include the financial health of the insurer, long and/or short positions in embedded options, as well as the mortality, morbidity, or casualty loss of a third or fourth party. For the direct paradigm, the market value of liabilities is then defined as the value of these contractually defined payments.
- *Indirect Paradigm*. The indirect pricing paradigm views an insurance company liability as a lien on insurance company assets, the net effect of which creates a stream of corporate earnings, that is, distributable earnings, which can be purchased by an investor. These earnings reflect the contractual provisions and contingencies of both assets and liabilities, as well as the various accounting practices and conventions generally accepted as appropriate for defining earnings that can be distributed to investors. For insurance companies, distributable earnings typically equal statutory income adjusted for changes in required risk capital, although there may at times be a statutory/GAAP net worth adjustment (see Becker 1993). For the indirect paradigm, the market value of liabilities is then defined as the market value of assets less the market value of distributable earnings and follows from the accounting identity: $L=A-E$.

These two pricing paradigms differ in a number of ways. The direct pricing paradigm evaluates liabilities directly as assets are valued and hence is constructive. The indirect pricing paradigm first evaluates the earnings of the insurance enterprise and then defines the value of liabilities as the known market value of assets less this value and hence is a deductive. The direct pricing paradigm values liabilities in terms of actual cash flows, as is the case for assets, while the indirect paradigm values payments not in terms of the insurer's cash flows but in terms of its "distributable" earnings, because these reflect the payments actually purchased by a given investor.

As a corollary to this, values produced by the direct pricing methodologies are absolute and independent of accounting standards, while the indirect pricing methodologies are fundamentally driven by the notions of statutory and GAAP earnings and statutory risk capital. Finally, direct pricing methodologies produce values that are independent of the supporting assets or at most dependent on their state-dependent cash flows, while the indirect pricing methodologies values reflect both the type and amount of the supporting assets, in addition to their state-dependent cash flows.

For these paradigms to produce results that are in theory comparable, three refinements must be made to the descriptions above and calculations below.

- *Franchise Value*. When a firm is valued, future distributable earnings reflect both earnings from contracts currently on the books and the firm's "franchise value," which reflects a capitalization of the firm's earnings on business expected to be booked in the future. The direct approach ignores this value by definition, so when applying the indirect paradigm, it is important that franchise value also be ignored. In other words, the firm's business must be valued as a closed block to produce a reasonable value for *current* insurance liabilities
- *Insurance Company Debt*. If an insurance company has market debt, the indirect paradigm yields the market value of insurance company total liabilities including this debt. Consequently, to be consistent with the direct paradigm valuation, this market value of liabilities must be reduced by the market value of outstanding debt to result in the market value of insurance liabilities.
- *Put Option to State Guarantee Funds*. As discussed above, an insurance company's option to put its liabilities to a state guarantee fund at the time of its insolvency has value to insureds that is expressed in their providing the insurer a lower cost of funds than might be appropriate given the insurer's actual credit quality. Depending on the calculation details, the direct paradigm may capture this cheapness and consequently understate the market value of liabilities that would exist in the absence of this option, although it can also be implemented in a manner (see below) that estimates this latter value.

  On the other hand, the indirect paradigm captures this cheapness only to the extent that the value of this option exceeds the cost to the insurer as reflected in future guarantee fund assessments. Because it is easy to modify distributable earnings to eliminate such assessments, the respective liability market values can be made to be comparable and consistent and equal to the market value of liabilities for a firm with such a put option. Philosophically, these paradigms produce liability market values in a market in which purchasers of short positions continue to have access to guarantee funds.

As an example of the put option effect, assume that a Aaa/AAA-rated company can issue debt at 9% and an A-rated company at 9.50%. If the A-rated company obtains credit support to borrow at 9%, there are two ways to evaluate the market value of its liabilities. As an A-rated company, the value of a 9% borrowing is, say, 97.50, discounting at 9.5%. This is the value of the bond to the company and understates the value of 100 to investors who recognize the value of the credit support and discount at 9%.

Distributable earnings are enhanced by cheap borrowing and decreased by the cost of the credit support. If this cost is eliminated from the earnings analysis, distributable earnings are overstated relative to the company with the bond and no credit support by the present value of the 0.50% coupon savings. The indirect paradigm consequently values liabilities (that is, the bond with credit support) as 100, the increment to assets, less the value of an extra 0.50% coupon income in distributable earnings.

This value equals the direct value of the bond to the company of 97.50 only if distributable earnings are risk-adjusted as are flows from an A-rated bond. Since earnings hold a subordinated position to bond payments, it is logical that they would be risk-adjusted more conservatively and that the indirect paradigm would produce a liability in excess of 97.50.

While both pricing paradigms are theoretically defensible and appealing, even with the calculation refinements above, they in general give different results for the market value of insurance liabilities and do not yield identical results except in the simplest, most contrived hypothetical case. Beyond the issue of risk-adjustment exemplified above, an important reason for this difference is that many insurance markets are relatively inefficient and there is often little relationship between an insurer's "real" cost of funds, as implied by the market price at the time of issue of a contract, and an insurer's "hypothetical" cost of funds, as implied by its true financial quality. One exception to this observation is the secondary GIC market where GICs trade like bonds at spreads consistent with the financial quality of the insurer.

As evidence of this general inefficiency, we know that insurers of various financial qualities can sell comparable contracts at comparable prices. In theory, this pricing anomaly may be consistent with an efficient market hypothesis in a market that values different levels of service or perhaps the insurer's put option to state guarantee funds in a way that offsets the financial quality differentials. However, we can also find comparable contracts sold by insurers of similar financial quality and service reputations for materially different prices. In this case, market inefficiency appears to be the most logical conclusion. Of course, subsectors of the insurance market are fiercely competitive and efficient, but the conclusion of inefficiency appears valid in many subsectors as well.

As a consequence of this inefficiency, the direct market value of many insurance liabilities at the time of issue is relatively insensitive to the insurer's financial quality and hence also relatively insensitive to the level, pattern, or variability of the insurer's profits. Hence, these direct market values must also be relatively insensitive to the accounting standards by which the insurer's profits are released.

Conversely, the value of an insurance enterprise fundamentally reflects the accounting standards underlying the definition of distributable earnings, and so too must the indirect market value of the insurer's liabilities. For example, if statutory profits were redefined or statutory risk reserves increased materially, the values of insurance enterprises would change reflecting differentials in the stream of distributable earnings available, as would the implied values of their liabilities as calculated with the indirect paradigm.

Although it could be countered that the market would alter its implied pricing mechanisms to exactly offset the change in distributable earnings and thereby obtain the same value of the firm, this would imply that equity investors in general are indifferent to the level of actual dividends received. This in turn implies that an enterprise can produce the same favorable risk-adjusted level of return on capital independent of the level of capital it retains. Although no definitive proof of the impossibility of this conclusion can be provided, the dividend management activities of publicly traded companies suggest its implausibility.

Note, however, that it is not necessary that these market values be identical. In the financial markets discipline is maintained by arbitrageurs and their ability to buy the cheap contract and sell the dear. When the market does not allow easy trading—is not complete—actual pricings can easily diverge from theoretical pricings. Currently, there is a relatively thin "short" market for trading an insurance block or enterprise in which pricings reflect valuations of distributable earnings, as well as a thin "long" market for some individual liability contracts, such as GICs and lottery annuities, where pricings reflect valuations of cash flows.

As noted above, while the short market currently encompasses all types of insurance liabilities, the long is unlikely to expand much beyond the liabilities currently traded. With such a limited long market there is no possibility that investors can take short positions in most insurance liabilities at prices consistent with the direct paradigm's long position valuations. Consequently, there is no possibility that investors can arbitrage short market values produced by the indirect paradigm with comparable positions established by "shorting" the long market at values produced by the direct paradigm. In conclusion, there does not appear to be a market-based mechanism that forces the convergence of the respective paradigm's values in the current environment.

In summary, the methodologies of the direct pricing paradigm, as described further in Section 5A, provide valuations of liabilities that are consistent with valuations of assets currently and that are similar to the values at which such contracts trade in the "long" markets that currently exist, GICs and lottery annuities. However, the methodologies of the indirect pricing paradigm, as described further in Section 5B, also have significant merit because, like the direct, they too reflect valuations observed in today's "short" market.

In theory then, both paradigms produce results consistent with current market activities and finance-theoretic valuation practices. Consequently, either paradigm could be "declared" as providing reasonable estimates of the true market value of liabilities today. However, the direct pricing methodologies are preferred in practice because they produce values that are absolute and independent of accounting standards, and this has the advantage of consistency as accounting practices evolve. In addition, the direct pricing methodologies can be implemented with reasonable confidence far easier than the indirect methodologies, because distributable earnings are the most difficult flows to value in what is fundamentally an incomplete market.

5. MARKET VALUE OF LIABILITIES

A. Direct Methodologies

The direct pricing methodologies for the MVL are based on the law of one price and reflect the methods used in the thinly traded insurance liability markets of secondary GICs and lottery annuities. That is, we seek similar instruments currently trading in the financial markets and value liabilities in a way that reflects the valuation of those contracts.

Viewed from the perspective of the issuer, that is, the short position, an insurance contract is simply a financial instrument whereby for a given price, defined in terms of a single or periodic payments, the insurance company agrees to pay given amounts based on the realization of a certain contingency or contingencies, and whereby in addition to this basic structure the issuer and/or owner has certain financial options that can be elected and that will in general at least modify the basic structure, if not cancel it altogether (see Section 6 for examples).

From the asset market's perspective, valuing an insurance company liability is equivalent to valuing a series of payments contractually defined, as well as the long and short embedded options of both parties. For all such features, comparable features exist today in actively traded securities that trade at prices consistent with those implied by the theoretical models, as motivated by the law of one price, and implemented through various ad hoc pricing models. Consequently, the direct pricing methodologies apply these well-utilized techniques in this new setting of liabilities.

The simplest methodology possible under this paradigm is the risk-adjusted present value (RAPV) approach. Here, we "simply" define a series of cash flows that are feasible under the contract and discount these flows at interest rates reflecting both the risk-free time value of money (Treasury rates) and risk spreads adequate to compensate the investor for the risk that the actual cash flows may be different from those initially defined.

Actual cash flows can vary, for example, due to contingency risk, a prepayment or extension (embedded options), issuer default (credit risk) or loss relative to fair value on sale prior to maturity (liquidity risk). Of course, this methodology involves significant judgment, but by using various market comparables this approach can operate efficiently in an actively traded market even in the absence of a theoretical model to substantiate the risk pricing.

For example, callable bonds have existed longer than fixed-income option-pricing models. A 10-year bond, say, callable from year 5 to 10, might be sold at par (100) with a coupon of 9.50%, where the net spread to an 8.00%, 10-year Treasury of 1.50% would reflect call risk, default risk, and liquidity risk. In this example, the market defines the "feasible" cash flows as those realizable if the bond is not sold and neither calls nor defaults. The spread of 1.50% is then a measure of the risk that an event not explicitly reflected may indeed occur.

While it is entirely logical and conventional in the bond markets to define cash flows for spread definition purposes on the basis of "no-call, no default, no sale," we could have hypothetically assumed that the cash flows would end in a call in year 5 with a principal repayment of 104.75. Based on these cash flows, we would infer that the bond was selling at a discount, that the actual yield was higher than 9.50% (10.25%), and that the risk spread to the 7.25%, 5-year Treasury was larger than 1.50% (3.00%).

This illustration shows that hypothetically any set of feasible cash flows can be used initially when calibrating a spread model, although not all approaches are equally useful in price predictions because of a lack of comparability and the instability of results. To the extent that assumed flows reflect a worst-case scenario, risk spreads and total yields can even be negative. For example, consider the yield and spread implications of pricing the above bond to an assumed default in the second year with a 50% recovery.

While this method is apparently simple to use when performing calculations, its weakness is the difficulty of judgmentally determining the necessary risk spread or, more generally, the relationship between the risk spread and the feasible cash-flow stream assumed. One simple solution commonly used in the asset markets is to seek out "similar" securities prices and determine the market's risk spreads. For example, by sorting noncallable public bonds into credit risk classes, clear patterns of risk spreads to like-maturity Treasuries emerge. This comparative approach has the advantage of simplicity and works well for relative valuation applications; for example, which of security A or B is cheaper?

The disadvantages of this simple solution are: nonexistence of comparables (most liabilities), the sparseness of comparables for even some simple risks (very low credit quality), the lack of a clear pattern of pricing for more complicated risks (embedded options), and the inability to determine whether the price is in some sense "correct." For example, while security A may be the better priced Baa/BBB bond, is it cheap enough to be preferred to a Treasury bond?

Consequently, an alternative solution to the problem of risk spread determination, given assumed cash flows, is to model the more complicated risks explicitly, so that they are reflected in the variable nature of future cash flows rather than in only the discount rate applied to the single fixed set of cash flows. Because this represents an extension of the ad hoc spread pricing procedure beyond the calibration of the market's spreads on some risks, to the estimation of the spreads necessary on a wider set of risks, there is no single or perhaps even best way to implement this extension.

The intuitive framework for this extension is the stochastic generation of the various cash flows possible under the contract, reflecting the possible outcomes from various risks, and then the discounting of these flows with rates reflecting:

- Risk-free time value of money (Treasury rates)
- Risk aversion spread for the risks modeled
- Risk spread for risks not modeled.

In this framework, a given risk has two effects on price and hence on the necessary risk spread, as can be observed in the financial markets for risks such as issuer default. First, the investor must be compensated for the losses expected (the average loss), and this discounting of price can be either implicitly reflected in the spread or explicitly in the cash-flow scenarios. These values are often estimated or modeled based on past experience on the risk. However, this approach is meaningful for this purpose only if the portfolio of risks priced is similar to, and as diversified as, the portfolio of historical risks on which loss estimates are made.

Second, the investor must be compensated for taking the risk that actual losses may differ from this average. This second component reflects the buyer's utility function in theory and provides the appropriate mechanism for calculating price, given the level of risk, to equalize the investor's expected utility, that is, satisfy $u(w)=E[u(w-P+C)]$, where w is current wealth, C is the risky cash flow, and P is the equilibrium price. For a risk-averse investor, this price is always less than the expected value of the risky cash flows, $E[C]$, implying an additional spread component.

In practice, we can also identify this risk with capital risk, for which the investor must be paid an appropriate return for holding the capital necessary to assume this risk. That is, compared with a risk-free asset, a risky asset requires the investor to hold risk capital to absorb period cash-flow fluctuations and allow the investor to realize a given fixed consumption opportunity defined in terms of expected cash flows. As compensation for holding this capital in relatively risk-free assets, the investor must in turn earn an extra return on the risky asset.

Using either perspective, this risk aversion spread charge or price discount increases with risk variability and decreases with risk diversification.

One risk that can typically be evaluated outside this intuitive framework is interest-rate contingent claim risk, that is, embedded option risk. As noted above for options on stocks, risk-free bonds with embedded options can also be replicated but by a dynamically managed portfolio of long and short risk-free and option-free bonds. Consequently, the price of such a security does not reflect the market's utility function beyond the extent to which the prices of bonds in the replicating portfolio reflect this function. That is, no additional explicit provision for option risk can be reflected in the price.

In practice, a yield curve evolution or scenario model is required that is internally consistent in that it precludes the formation of risk-free portfolios that outperform the risk-free rate; that is, the model precludes risk-free arbitrage. Such evolution models can always be reparametrized so that the price of a given security equals the simple average of the prices obtained scenario-by-scenario. For risky bonds, while it is not technically correct to do so, it is common practice to apply yield curve evolution techniques to term structures with spreads appropriate for the credit risk.

For specific evolution models see Heath, Jarrow, and Morton (1992), Ho and Lee (1986), and Pedersen, Shiu, and Thorlacius (1989). For a general review of the theory and applications see Gerber and Shiu (1995), Hull (1993), Smith (1976, 1990), and Tilley (1992). In Section 6, considerations are given for making the transition to pricing embedded liability options.

For the above bond example, an option-pricing model that projects the various payoffs obtainable from the bond would identify that the option is worth 50 basic points (bp) so that 100 bp of the risk spread represents compensation for credit risk and liquidity risk. That is, if 9.00% is used to discount the option-adjusted cash flows in the model, a price of 100 would result, which is the same as applying 9.50% to the noncallable cash flows.

In this example, as for insurance company liabilities generally, the spread for credit risk can be interpreted as having two components: one reflecting average losses and the other being compensation for taking risk. In the bond example, losses of 35 bp on average might be expected based on historical experience for the quality, with much of the remaining 65 bp of risk spread reflecting compensation for taking the risk that losses are variable and can be much worse than this average. Of course, some of the residual 65 bp reflects liquidity risk because even the deepest bond markets have wider bid/ask spreads than the Treasury market.

In theory, credit risk can also be modeled explicitly so that the corresponding charge for expected losses can be removed from the discount rate. For example, from each path in the option-pricing model for the bond, 100 paths can be constructed, say, each reflecting randomly generated occurrences of default and loss. On average, if modeled correctly, these losses would average about 35 bp, so a discount rate of about 8.65% (9.00–0.35) would be used in the model. Further, by explicitly developing an annual charge to coupon income in every scenario of, say 55 bp, reflecting the cost of the risk capital assumed to be held to allow the assumption of credit risk, the discount rate needed in the model would reduce to about 8.10%, for a 10 bp residual risk charge over Treasury rates for liquidity risk.

Stochastically modeling the credit risk of a security, or mortality/morbidity/property and casualty risk on a liability, has one very useful purpose, but this purpose can usually be achieved without further complicating the option-pricing modeling as indicated above. The purpose alluded to is the development of the loss probability distributions from which average losses and the necessary risk capital bases can be estimated. However, these estimates can in general be made outside the yield curve scenario model used in pricing options, because most such risks are independent of the level of interest rates projected. In these cases, it is sufficient to reflect ''average'' losses in each scenario of the option-pricing model as cash outflows, as well as the cost of capital charges needed in the underlying term structure.

For some insurance risks, such as health benefits or casualty payments, it is often assumed that losses are positively correlated with interest rates because of the effects of inflation on such payments. In these cases, expected losses on each yield curve scenario, and potentially the capital charges as well, must be varied to reflect this effect. In essence, this loss ''path dependency'' is equivalent to two embedded options on inflation, one long and one short, which in turn is equivalent to a forward contract. However, even in these cases it is possible and desirable to separate the modeling of these loss parameters from the basic liability valuation within the yield curve scenario generator to avoid having a cumbersome stochastic model.

In principle, this process can be followed for insurance company liabilities just as for assets. However, it is reasonable and prudent not to model all risks but only those that are so complex or variable that they preclude the judgmental development of the necessary component of the risk spread from market comparables. For example, embedded options are routinely modeled explicitly in the asset markets because their values are quite variable and no simple rules of thumb for the equivalent spreads suffice. Explicitly modeling liability options is also desirable for the same reason. In addition, interest-rate-dependent insurance losses, being in effect options, also require explicit option-pricing modeling using path-dependent average losses and capital charges.

On the other hand, credit risk on assets is effectively never modeled because spreads appropriate to financial ratings are readily available from hosts of comparables in the financial markets and seem to suffice for most applications. For liability valuations, credit spreads corresponding to the issuer's financial ratings seem appropriate for ''short'' position market values, while in theory spreads corresponding to a rating of Aaa/AAA may be preferred for ''long'' position market values to reflect the implicit credit support of state guarantee funds. However, this is not observed in practice, where implied spreads are reflective of insurers' qualities.

Finally, asset liquidity spreads again are never modeled explicitly, but generally judgmentally established based on market comparables. In the case of secondary market GICs, spreads typically reflect the liquidity of private placements. For most other liability valuations, however, liquidity spreads must be largely judgmental because little long market activity currently exists and, as noted above, is ever anticipated to exist.

B. Indirect Methodologies

The indirect or deductive pricing methodologies for the market value of liabilities reflect actual valuations in the relatively thin market of insurance company or insurance block sales transactions. In this setting the investor-buyer is typically ''buying'' a block of liabilities, often together with a block of supporting assets, for a given sum, which may be negative. In this latter case the buyer receives additional funds from the seller equal to the negative price. In effect, the price of the block or enterprise equals the value of future earnings (distributable earnings).

One distinction here is that for estimating the market value of liabilities, the franchise value of the enterprise, that is, the value of future sales, must be omitted from earnings, as is common practice for closed-insurance-block transactions, rather than being included, as is common practice for insurance company acquisitions. The reason for this exclusion is that ''franchise'' or ''brand'' value, while important to both a buyer and a seller, has nothing to do with the value of outstanding liabilities that we seek.

Given the market value of distributable earnings and the market value of assets, the indirect paradigm market value of liabilities is defined:

$$MVL = MVA - MVE,$$

that is, the market value of liabilities equals the market value of assets less the market value of the insurance block or enterprise. This value of liabilities is self-evident, because we can in theory realize it through a market trade. Specifically, we can purchase an insurance block or enterprise, sell the assets, and for the above net proceeds assume responsibility for the block of liabilities. Of course it must be the case that $MVA - MVE > 0$, because liabilities require payouts.

The market value of the enterprise reflects not only the market value of assets but also the actual assets reflected in the *MVA* calculation. Consequently, in theory we can have two *MVE* values corresponding to one *MVA* value, and correspondingly the above formula can produce two *MVL* values.

For example, a buyer would instinctively pay less for the distributable earnings of a universal life insurance block if all assets were S&P call options than if assets equaled a well-diversified bond portfolio of the same market value, assuming that the buyer was inexplicable forced to hold the given assets. However, here again we would see market arbitrage at work. Savvy buyers know that they do not have to hold the given assets, but can instead sell them at the market price and replace them with a more appropriate asset portfolio. In doing so they reduce the original *MVA* somewhat due to transaction costs, but should increase the theoretically fair *MVE* value as well because the insurance block is now hedged by a more appropriate asset portfolio. The goal of the trade, of course, is to create more value in the *MVE* calculation through improved asset/liability matching than is destroyed in the *MVA* calculation through transaction costs.

In a competitive market of savvy investors, therefore, *MVE* need not necessarily explicitly reflect the actual assets in the block if an asset arbitrage, defined in terms of the effect on *MVE* relative to the effect on *MVA*, is possible. That is, in theory *MVE* can be calculated in a competitive market for a block sale as the maximum of all values produced by asset portfolios that can be purchased with proceeds from the sale of the original assets. Of course, a savvy buyer can do this calculation to evaluate potential value-added and profit or to justify an initially excessive asked price. Similarly, a savvy seller can do this calculation, and indeed implement an asset portfolio restructuring, to avoid leaving too much money "on the table."

In this hypothetical model of a competitive market, the implied value of *MVL* is therefore unique and effectively minimized. In practice, however, blocks do not necessarily trade at this maximum *MVE* price, and consequently, the implied *MVL* on purchase may well be larger than that achievable after an asset portfolio restructuring. But this does not contradict the market model. After all, that an insurance block buyer can restructure the asset portfolio and resell the block at a profit is no more counterintuitive than that an investment bank can buy GNMA portfolios and resell them at a profit as tranches of a CMO. In both cases the potential profit is a reflection of a value-added by the buyer and not an initial mispricing in the market.

For the application of the indirect pricing paradigm to the market value of liabilities, however, it makes sense to use the actual asset portfolio associated with these liabilities. If the above arbitrage is feasible, the insurer can implement the corresponding asset portfolio restructuring and then reap the rewards of a smaller *MVL* value.

Similar to the direct methodologies for *MVL*, the indirect methodologies again form a sequence of approaches from simplest yet most judgmental, to most complicated yet objective. In contrast to the direct paradigm in which cash flows are the basis of value, for evaluating an insurance block it is the so-called distributable earnings. These earnings, which are defined as after-tax statutory income adjusted for capital gains and losses and changes in required risk capital but are further decreased if GAAP equity is smaller than statutory equity, form the basis of value in this context. It is precisely these values that can be removed from the block annually and distributed to the buyer as returns on the initial investment.

The simplest approach for the indirect pricing paradigm is analogous to the risk-adjusted present value approach of the direct pricing paradigm. That is, a feasible and likely set of distributable earnings is projected and then discounted by rates that reflect both the risk-free time value of money and appropriate risk spreads to compensate the buyer for the host of risks and contingencies that can significantly alter the original earnings projected. Given the nature of the implied acquisition, risk spreads are typically reflective of those implied by the capital-asset-pricing model for equity markets.

In contrast to the RAPV method for direct liability valuations, however, in practice this initial valuation is supplemented with sensitivity tests by which, albeit on a simple deterministic basis, other sets of feasible distributable earnings are explicitly modeled and in which one or a number of the risks of the block are modeled conservatively. From these "stress test" valuations, whereby all risks are charged in the assumed risk spread, we can again use a sequence of pricing methodologies to explicitly model individual risks and capital costs and correspondingly eliminate charges for these risks from the discount rate.

For example, we can option-adjust this initial valuation (see Becker 1993) to produce the option-adjusted value of distributable earnings (OAVDE). The modeling here would largely be the same as that for an option-adjusted valuation in the direct pricing paradigm except that, rather than focus on cash flows of liabilities and assets, we model the effects of embedded options on distributable earnings. As is the case for the direct pricing paradigm, once the option risk had been explicitly modeled, it can be eliminated from the risk spread. The spread would then need to reflect risks such as mortality, morbidity, and other life and casualty insurance contingencies, asset credit and liquidity risk, as well as tax and expense risk. More simply, these risks can be viewed collectively as creating equity risk to the buyer of the distributable earnings.

As is the case for the direct valuation methodologies, judgment is required to identify which risks must be explicitly modeled and which risks can be valued based on the simpler method of defining the corresponding risk spread to equal expected losses plus a return on required risk capital. Financial market comparables are again of value here. As for the direct pricing paradigm, because option risk is one of the most complex risks, it should be modeled explicitly in option-rich insurance company liabilities. For term life and some health and casualty products, option risk is nonexistent and contingency risk may be deemed most important for explicit modeling.

Finally, required risk capital is explicitly incorporated into these indirect methodologies because the adjustments to statutory earnings include an adjustment for changes in this risk capital. Consequently, once defined, returns on risk capital invested in the block are easily modeled by correspondingly adjusting statutory earnings for these capital costs. In this way only expected losses due to risks need be modeled because the corresponding risk premiums are captured through this required risk capital assumption.

Although the NAIC model for risk-based capital (RBC) is a logical starting point for required capital, it is only that. To maintain high financial quality, some multiple of RBC in excess of 100% should be built into the model at a minimum. In addition, because RBC was developed as a solvency standard and not a comparative measure of strength for well-capitalized companies, risk capital assumptions for a given company may well vary relative to RBC values based on the particular risk or reflect risks perhaps not reflected in the NAIC formulation.

6. OPTION-PRICING CONSIDERATIONS FOR INSURANCE LIABILITIES

Many insurance company liabilities are rich in option structures that for the most part put the insurer in the short position. For example, universal life (UL) and single- or flexible-premium deferred annuity (SPDA/FPDA) contracts grant the contract-holders both put and call options. Specifically, a partial or full withdrawal of funds from the contract at "book" value is an exercise of a put option whereby the policyholder "sells" back to the company part or all of its contractual obligations at a price indexed by the book value method of fund valuations. In the case of bank- or corporate-owned life insurance (COLI/BOLI) programs, the client company is the owner of this put option and as such potentially exposes the insurer to greater risk because of the dollar size of such cases and the possibility of greater election efficiency.

Similarly, policyholders also have call options or rights to buy additional amounts of the insurer's contractual benefits by contributing additional premiums or considerations. Minimum-interest-rate guarantees are formally equivalent to an embedded "interest rate floor" contract, which in turn is equivalent to a series of call options for the contract-holder. For example, a 4% floor is equivalent to a series of call options on 1-year, 4% bonds, whereby the long can buy these bonds at par.

Insurers also have short positions in put options to plan participants in the stable value funds to which GICs, PICs, and synthetic GICs are issued, and are also short "options" on health and casualty products with interest-rate- or inflation-dependent payments.

One example of a liability in which an insurer holds a long option position is a so-called "callable" GIC. A callable GIC is largely identical to a callable bond in structure, only here the insurer is the borrower rather than the investor. Because the borrower on a callable bond is long the option, so too is an insurer in this case. Two subtle differences between callable GICs and callable bonds are that GICs can be coupon-bearing (simple-interest GICs) or not (compound GICs), and often there is no call premium (that is, GICs are callable at book value).

While an option is a right but not an obligation to act, option-pricing theory often treats it as an obligation. That is, the long position is assumed to be 100% "efficient" in exercising its right, which is to say that the long always elects when it is most financially advantageous to do so and never elects when it is financially disadvantageous. Stock options and embedded call options in bonds are typically treated this way because in such cases the anticipated level of sophistication of the long position demands this assumption and market experience reinforces this view. However, many counterexamples to the 100% efficiency rule exist even for sophisticated investors/borrowers.

For example, residential and commercial mortgage loans can be called independent of the interest-rate-based financial merits simply because the property has been sold and the mortgage note is not assignable. Similarly though less frequently, a bond can be called even if "out of the money" because the corporation is restructuring or consolidating its debt. Of course, if the bond trades in the open market, it may be more economical for this issuer to simply repurchase the issue at the presumed discount price than to call at "par plus."

In addition, many mortgages are far "in the money" but not called because the borrower is unsophisticated or the value of the property and/or financial status of the borrower has deteriorated sufficiently to preclude refinancing. As another example, the election of options on many insurance company liabilities is also visibly less than 100% efficient. Again, contract-holder sophistication can be a cause, as can tax implications, a change in insurability, or even a strong agent/contract-holder relationship.

One counterexample to this inefficiency is the implied option on some health and casualty products that effectively index losses to inflation. By definition, we assume that this option will be "elected" with 100% efficiency in both cases, that is, where the insurer is short (inflation rises) and where the insurer is long (inflation falls).

Because options are "contingent claims," which is to say that their future value is currently unknown and contingent on future events, the problem of "pricing" is highly nontrivial. In contrast, while the pricing of yearly renewable term life insurance also involves contingencies, it is simplified by the "law of large numbers," which guarantees that, if properly underwritten, a large pool of similar risks has fairly predictable contingency costs even though any given individual's cost in the pool is entirely unpredictable.

In the financial markets, no such paradigm exists. While holding options on many stocks eliminates "specific" or "diversifiable" risk, the investor is ultimately left with an option on a market index, say, the S&P 500, the future behavior of which is entirely unknown. Similarly, holding many callable bonds barely decreases the contingency risk of holding only one because here "contagion" risk is significant. That is, all calls are expected to be elected in relatively similar conditions and with little spread of risk.

Even in a pool of mortgage-backed securities (MBS) or a group of UL contracts, only specific risk is eliminated, that is, the risk of one mortgagor or one contract-holder electing. However, the nondiversifiable risk in such a pool is significant because, unlike mortality, it is generally not known to what "average experience" the pool should converge. The reason for this risk is that such pools are typically priced at far less than 100% efficiency, so the major nondiversifiable risk is that the efficiency of the pool changes or, more specifically, moves closer to 100%.

Many option-pricing models assume 100% exercise efficiency on the part of the long position. There are at least two simple ways to modify option-pricing models to reflect long position type 1 inefficiency, where type 1 inefficiency means that the long will not necessarily exercise when in the money. Exercising when out of the money, or type 2 inefficiency, is discussed below.

First, we can hypothesize that the cause of this type 1 inefficiency is "inertia" caused by a lack of perfect information or, even in the presence of such information, caused by the fact that the long position may need to incur hard costs plus expend significant effort to capture the option's value. For simple options on securities, imperfect information affects the timing of investors' actions, while the need to expend significant effort to elect can affect even sophisticated investors' propensity to act. For embedded call options, the long must also incur the effort and hard costs associated with refinancing, because "refinancing" is the logical expectation in the case of an interest-rate-driven call rather than simply a payoff of the debt. This reluctance to call or in general elect can be easily modeled by altering the payoff functions in the option-pricing model.

For example, if a bond call payoff is par plus 4%, or 104, and 2% refinancing costs are estimated, the bond could be priced as if the payoff were 106. To reflect "effort," an extra "point" or so could be added, resulting in an increase to the call premiums of 3% to 107. Instead of calling at 104, the long will then be assumed to delay exercise until rates fall further, to increase the security price to 107. This technique reduces call exercises and the option's value. Logically, this type 1 inefficiency assumption reduces the value of the long and cost to the short of the embedded option compared with the 100% assumption.

For more complicated securities such as MBSs and CMOs and analogously for UL/SPDA contracts, we can also model the prepayment/lapse activity to explicitly reflect type 1 inefficiency. In the bond example, the call decision is driven by the relationship between the call price and market value of future bond payments. For more complicated securities, however, decisions are based on the relationship between mortgage interest rates paid (that is, the weighted average coupon, or WAC) or UL interest rates credited and the rates currently available in the marketplace.

To reflect type 1 inefficiency here, we typically model little or no incremental election activity for relatively small favorable shifts in rates, perhaps reflecting pool age, past experience, or financial disincentives such as taxes or surrender charges. For larger shifts, incremental activity reflects significant "inertia" of contract-holders or mortgage-holders relative to 100% efficiency and an eventual capping off of all incremental activities. In the MBS/CMO markets, such models have been estimated from complex econometric studies of past experience, while for insurance company applications judgmental models are typical. Even so, there is no compelling proof that econometrics beats judgment in predicting future behavior, although econometric models are preferred for calibrating option-pricing models and replicating current prices.

A second method for reflecting "in the money" type 1 inefficiency is to simply "discount" the option values produced by the model vis-a-vis option-free values by 10–40%, thereby intuitively reducing "efficiency" by 10–40%. In this simple discount method the intuitive "behavioral" model is that, for example, 80% of the long positions are 100% efficient and 20% do not elect at all. This behavioral model produces the same price as discounting the option value by 20% directly, because the weighted average of the option-free and option-adjusted prices is mathematically equivalent to the simple discounting of the embedded option.

There is another type of "inefficiency," here called type 2, whereby in addition to some failure to exercise when "in the money," we can often expect some level of exercise when the long is "out of the money." Both forms of inefficiency reduce the value of the option compared with the full efficiency model. However, while qualitatively similar to type 1, type 2 inefficiency must be modeled with extra care, because, rather than reducing only the value of the option, to 0 in the limit of type 1 inefficiency type 2 inefficiency, provides an implied "profit" to the short. This can ultimately create an option with "negative" value, that is, an option for which the "short" both receives a premium for selling and obtains a contract with positive economic value. Equivalently, this type of inefficiency modeling can compel the short to pay the long to accept the option.

This pricing anomaly defies the law of one price in the most elementary way and cannot occur in actively traded markets. For example, while MBSs are priced with both types of inefficiency, the "option" always retains net positive value. If pricing assumptions implied otherwise, market arbitrageurs would likely "take the bet" and trade to capture this value, thereby correcting the mispricing.

When developing and pricing contracts with embedded options that are not actively traded, such as UL or SPDA/FPDA contracts, it is important to understand the effect of the inefficiency assumptions and to ascertain that, on net, the option has been modeled to have real positive value even if current experience dictates otherwise. To do otherwise is to risk future profitability, and possibly solvency, on a very risky and counter-market assumption.

For example, an assumption that a baseline level of lapsation exists on an SPDA/FPDA/UL contract independent of the level of credited versus market rates creates both type 1 and 2 option election inefficiencies. This is because it is posited that some longs (contract-holders) put (lapse) each year even when credited rates are above market rates, and many do not put independent of how relatively attractive market rates become.

When credited rates are based on asset portfolio earnings, the needed asset liquidations at market rates then generate gains or losses to the short (insurer), in addition to the incremental gains or losses reflecting the relationship between surrender charges and unamortized acquisition expenses. Baseline recurring deposit or premium assumptions on FPDA and UL contracts, respectively, create the analogous inefficiencies for this embedded call option.

On the other hand, typical "adjustment" formulas to the baseline assumptions provide a partial "efficiency adjustment," which limits type 1 inefficiencies in that the propensity to put is posited as being negatively correlated to the attractiveness of the credited rate vis-a-vis the market rate, while the propensity to call is modeled as being positively correlated. This adjustment formula still allows for substantial contract-holder type 1 inefficiency, however, because the adjustment is only a partial one and 100% efficiency is not produced when the market reference rate prevails, even by a substantial amount. Typically, this inefficiency is both explicitly modeled by recognizing the logical dampening effects of taxes, surrender, or sales charges, and indirectly reflected by assuming that even when such charges are valued, less than 100% will elect when "in the money."

Another important implication of type 1 and type 2 inefficiencies is that the theoretical bounds for option prices may well be violated. To put this in perspective, consider a stock with current price X, an American call option with strike at 90, and an American put option with strike at 110. What is apparent and easily proved with an arbitrage argument is that in theory the call option can never be worth less than the lesser of 0 and $X-90$, while the put can never be worth less than the lesser of 0 and $110-X$. Based on the assumption of 100% efficiency, which market arbitrage ensures, these minimums are absolute.

Similarly, a bond currently callable at 110 has in theory an absolute maximum market price of 110, since in this case the bond is short a call option with a minimum value of min(0, $X-110$), where X is the market value of the future cash flows of the bond. In addition, if this bond is currently putable at 100, its market value can never be less than this put price, because in this case the bond is long the option.

Without 100% efficiency these theoretical bounds are easily violated within any option-pricing model. In addition, there are examples in the market of actual traded prices that "violate" these relationships. The most obvious examples of securities that are visibly priced with substantial type 1 inefficiency are MBSs/CMOs. Such securities often trade at a premium, when in theory such securities could be called immediately at par by the collective exercise of many homeowners electing their options. Clearly, this example of a market reality reflects a view far from this hypothetical possibility. Of course, the market does indeed recognize that the embedded call options have positive market value, but this value is measurably less than the 100% efficiency value.

Analogously, MVLs of insurance and annuity contracts can easily be calculated that are below guaranteed cash surrender values. While seemingly illogical on a contract-by-contract basis, as would be the above example mortgage-by-mortgage, this calculated value is justified on a block basis by reasonable and even conservative models of policyholder behavior. Certainly, acquiring individual residential mortgages at a premium may seem equally illogical, but the behavioral dynamics of a large block are sufficiently compelling for the market to trade on this basis.

We next consider the applicability of lattice-based option-pricing methods, such as Ho-Lee (1986), to more complex contracts such as MBS, UL, and SPDA contracts. One immediate complication on liability contracts but not on assets is the presence of death claims or other contingencies in addition to option elections. The effect of these events can be approximated by modeling cash flows related to the assumed claim probabilities and the benefits payable at each future time, and treating these as coupon payments.

As a more exact but also more computer-intensive approach, we could model the individual time-state payments "stochastically" as random variables reflecting the assumed claim distributions. Each lattice-based calculation would then provide a "conditional market value" of the liability, conditional on the values of claims generated by the claims model. The actual market value is then given by the "law of total probability" as the average of these conditional values.

For option election, a review of the lattice-based literature confirms that to be used for a given security, in each time-state the decision to call or not must be based on information available only in that time-state or in the connected future time-states.

More specifically, the decision to call cannot be based on information available only in earlier time-states. This is not a problem for a callable GIC or bond, because by the replicating portfolio argument all that is needed to decide on option election is the one-period spot rate in each time-state and the values of the bond in the two connecting future time-states. These future values effectively summarize all pertinent information in all future connected time-states. For more complicated securities, however, this decision paradigm is generally not possible.

For MBSs, for example, mortgage prepayment models are complex and reflect not only the current but also the past relationship of market yields to the given mortgage rate. For example, if market rates are currently relatively low, the prepayment model generates more prepayments if this is the first time this event has occurred than if this event has also occurred earlier.

Based on experience, this model posits (1) "hot money," or mortgagors who prepay (refinance) as soon as it is profitable, (2) "warm money," or mortgagors who do not act quickly but are more prone to act as favorable conditions prevail or improve, and (3) "cold money," or mortgagors who most likely do not prepay under any circumstances. For example, after a long period of relatively low rates, a mortgage pool can be said to be "burned out," in that virtually nothing but cold money is left.

Options that exhibit this election behavior are often called "path-dependent" options, because their exercise reflects not only current but also past circumstances. That is, their exercise depends on the "path" of interest rates from time 0 to the time-state of interest.

As another example, policyholder behavior with respect to withdrawals (puts) and optional premium payments (calls) reflects the relationship between market rates and credited rates. If both the market rate and credited rate reflect only the interest rate environment "known" in each time-state, that is, "new money" rates, a lattice-based valuation of the contract is possible. However, the credited rate often reflects a "portfolio" of assets either held, as in the "portfolio-based" crediting strategy, or modeled as an index, as in the "average of past new money rates" strategy. In either case, each time-state's crediting rate depends on interest rates prevailing at prior time-states. In addition, it is often assumed that the "market rate" is that of a competitor, hypothetical or real, and so reflects the competitor's crediting rate strategy, which may again reflect interest rate experience in earlier time-states.

Consequently, even when insurance company contract-holder behavioral models are simple and not apparently path-dependent, their dependence on credited and market rates usually creates path dependence. In general then, such insurance liabilities, like complex assets, are often unsuitable for lattice-based option valuation methodologies.

As it turns out, well-known results from the literature imply that either backward substitution on lattice-based models or forward generation on scenario-based models can be used to value any security for which the cash flows are fixed and known or are contingent and definable in any time-state. Consequently, European options are easy to handle because the option decision is straightforward, depending only on information in the time-state.

On the other hand, MBS, UL, or SPDA contracts can be valued using scenario-based models because cash flows are predicted in every time-state using an option election "model." When this scenario-based approach is used, the answer is the same as if these cash flows were valued by a replicating portfolio argument in a yield curve lattice. In general, American options embedded in bonds are problematic in a scenario-based model if 100% election efficiency is sought, although Tilley (1993) has introduced an approximate approach.

Because of the theoretical equivalence between lattice- and scenario-based models, the model best suited to the problem can be chosen. In either case, of course, the yield curves generated within the model must preclude the opportunity for risk-free arbitrage. If option exercise can be determined based on information available only at a given time-state or later, lattice-based calculations on commutative lattices such as Ho-Lee work well. If information prior to the given time-state is needed, scenario-based models are generally required.

Lattice-based calculations on noncommutative lattices, for which a two-period shift sequence down-up is not equal to up-down, are never done unless the number of periods, n, is relatively small. On such a lattice the number of backward substitution formula applications is of the order of magnitude of 2^{n+1}. In a commutative lattice these calculations grow only as n^2+n, a far more manageable number. Because noncommutative lattices require enormous calculation time for even moderate values of n, it is common to choose the scenario-based alternative for simplicity.

One significant advantage of scenario-based approaches not available to lattice-based methods is the ability to "sample" paths. In theory, an average of 2^n present values is needed for an exact answer, whether the yield curve dynamics are commutative or not. On the surface, this is no better than what the above formula implies for a noncommutative lattice. However, we can view the theoretical formula for price as stating that the value at time 0 is the mean of a distribution of 2^n values, each generated by one of 2^n possible paths. Viewed this way, it is natural to "estimate" this mean value by sampling from this distribution rather than generating all possible paths.

In general, the price estimate can be improved by increased sampling, but improved more efficiently by judicious sampling methods. One such method, called the antithetic method, involves creating a second path from each sampled path, the "mirror" path, whereby downs and ups are interchanged. This method automatically generates paths symmetric about the hypothetical "mean path" and improves accuracy compared to random sampling. That is, antithetic sampling substantially reduces the standard error of the price estimate for a given number of paths.

Another method involves "partitioning" all paths into equivalence groups, sampling from each group separately, and combining results with appropriate probability-based weightings. This method avoids the major problem of both random and antithetic sampling whereby many redundant paths are generated near the "mean path," and too few paths are generated in the "tails", which may be more important.

See Tilley (1992) for a survey of sampling methods, and Ho (1992) and Pohlmann and Wolf (1993) for the "linear path space" equivalence class approach.

ACKNOWLEDGMENT

The author gratefully acknowledges the contributions of several anonymous reviewers. The concepts described in this paper first appeared as part of a report written for the American Academy of Actuaries Task Force on Fair Valuation of Liabilities, James E. Hohmann, chair.

REFERENCES

BABBEL, D., AND STAKING, K. 1995. "The Relation Between Capital Structure, Interest Rate Sensitivity, and Market Value in the Property-Liability Insurance Industry," *Journal of Risk and Insurance* 62, no. 4 (December):690–718.

BECKER, D.N. 1993. "Use of Cash Flow Testing in Product Development," *Record of the Society of Actuaries* 19:2421–52.

BLACK, F., AND SCHOLES, M. 1973. "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81: 637–59.

CHEN, Z., AND KNEZ, P.J. 1995. "Measurement of Market Integration and Arbitrage," *The Review of Financial Studies* 8, no. 2:287–325.

COX, J.C., ROSS, S.A., AND RUBINSTEIN, M. 1979. "Option Pricing: A Simplified Approach," *Journal of Financial Economics* 7:229–63.

DUFFIE, D. 1988. *Security Markets, Stochastic Models*. San Diego: Academic Press.

DUFFIE, D. 1992. *Dynamic Asset Pricing Theory*. Princeton, N.J.: Princeton University Press.

GEANAKOPLOS, J. 1992. "Arrow-Debreu Model of General Equilibrium," in *The New Palgrave Dictionary of Money and Finance*, edited by P. Newman, M. Milgate, and J. Eatwell. New York: The Macmillan Press Limited (Vol. 1):59–68.

GERBER, H.U., AND SHIU, E.S.W. 1995. "Actuarial Approach to Option Pricing," *Actuarial Research Clearing House* 1995.1:301–36.

HEATH, D., JARROW, R., AND MORTON, A. 1992. "Bond Pricing and the Term Structure of the Interest Rates: A New Methodology," *Econometrica* 60:77–105.

HO, T.S.Y., AND LEE, S.-B. 1986. "Term Structure Movements and Pricing Interest Rate Contingent Claims," *Journal of Finance* 41:1011–29.

HO, T.S.Y. 1992. "Managing Illiquid Bonds in Linear Path Space," *Journal of Fixed Income* (June), 2, no. 1:80–93.

HUANG, C., AND LITZENBERGER, R.H. 1988. *Foundations for Financial Economics*. New York: Elsevier Science Publishing Co.

HULL, J.S. 1993. *Options, Futures, and Other Derivative Securities*. 2nd ed. Englewood Cliffs, N.J.: Prentice Hall.

MARTIN, J.D., COX, S.H., AND MACMINN, R.D. 1988. *The Theory of Finance*. New York: The Dryden Press.

PEDERSEN, H.W., SHIU, E.S.W., AND THORLACIUS, A.E. 1989. "Arbitrage-Free Pricing of Interest Rate Contingent Claims," *Transactions of the Society of Actuaries* XLI:231–80.

POHLMANN, L., AND WOLF, A. 1993. "Note on Managing Illiquid Bonds in Linear Path Space," *Journal of Fixed Income* 3, no. 3 (December):80–87.

SMITH, C.W. 1976. "Option Pricing: A Review," *Journal of Financial Economics* 3:3–51.

SMITH, C.W. 1990. "Applications of Option Pricing Analysis," in *The Modern Theory of Corporate Finance*. 2nd ed. edited by C.W. Smith. New York: North-Holland Publishing Co.

TILLEY, J.A. 1992. "An Actuarial Layman's Guide to Building Stochastic Interest Rate Generators," *Transactions of the Society of Actuaries* XLIV:509–64.

TILLEY, J.A. 1993. "Valuing American Options in a Path Simulation Model," *Transactions of the Society of Actuaries* XLV:499–550.

Discussions

DAVID F. BABBEL*

When it comes to the valuation of Insurance liabilities, the driving intuition behind the two most common valuation approaches—arbitrage and comparables—fails us. This is because, for the vast majority of insurance liabilities, there are neither liquid markets where prices can be disciplined by the forces of arbitrage and continuous trading, nor are there close comparables in this market.

We are left in a predicament, but not an impasse. If we can refocus our attention from "market value" to "present value," progress can be made. In doing so we need not descend the slippery slopes that surround the quagmire of equity valuation. The pseudo-scientific methods typically used there impart only a thin veneer of respectability. Moreover, there are many economic risks associated with insurance liabilities that are interesting to study but have little

*David F. Babbel, Ph.D., is Associate Professor in the Insurance and Finance Departments, The Wharton School, University of Pennsylvania, 303 Colonial Penn Center, Philadelphia, Pa. 19104-6218.