On the Application of the Continuous-Time Theory of Finance to Financial Intermediation and Insurance*

by Robert C. Merton**

1. Introduction

The core of financial economic theory is the study of the micro behavior of agents in the intertemporal deployment of their resources in an environment of uncertainty. Economic organizations are regarded as existing primarily to facilitate these allocations and are, therefore, endogenous to the theory. From within the permeable and flexible boundaries of this core, I choose on this occasion to explore the risk-pooling and risk-sharing roles of financial intermediaries and to derive some of the operating technologies that can be used to fulfill those roles. The tool of analysis is the continuous-time model of finance. The focus is on the economic function of financial intermediaries, rather than on their specific institutional structure. Nevertheless, the institutional homes of the financial products and management techniques studied can be readily associated with current structures of banks, investment-management firms, and insurance companies.

Although surely coincidental, the choice of subject matter makes it especially fitting that this year's Geneva Association Lecture takes place in the environs of Paris. It was, after all, here at the Sorbonne in 1900 that Louis Bachelier wrote his magnificent dissertation on the theory of speculation. This work marks the simultaneous births of both the continuous-time mathematics of stochastic processes and the continuous-time economics of option and derivative-security pricing.[1] Although Bachelier's research was unknown in the

[1] In analyzing the problem of option pricing, Bachelier derives much of the mathematics of probability diffusions, and this, five years before Einstein's famous discovery of the mathematical theory of Brownian motion. What financial economist doesn't relish the thought of the great intellectual debt owed to this early option-pricing theorist by the mathematical physicists and probabilists? However, because Bachelier dedicates his thesis to Henri Poincaré, there may be more to the story.

finance literature for more than a half-century[2] and although from today's perspective, his economics and mathematics are flawed, the lineage from Bachelier to modern continuous-time finance is direct and indisputable. In this light, perhaps you will forgive me for making a few general remarks on the continuous-time model as both a synthesis and watershed of finance theory. This to be followed immediately by that promised discussion of financial intermediation.

Over the past two decades, the continuous-time model in which agents can revise their decisions continuously has proved to be a versatile and productive tool in the development of modern finance theory.[3] As exemplified by the works of Breeden, Cox, Huang, Ingersoll, Ross and my own studies of the lifetime consumption-portfolio-selection problem, the continuous-time model has frequently produced both more-precise theoretical solutions and more-refined empirical hypotheses than can otherwise be derived from its discrete-time counterpart.[4] As it provides new insights, the continuous-time analysis also reaffirms old ones. It shows us that those classic pillars of finance – the Markowitz-Tobin Mean-Variance Model, the Sharpe-Lintner-Mossin Capital Asset Pricing Model, the Arrow-Debreu Complete-Markets Model, and the Modigliani-Miller Theorems – are all far more robust than had been believed.[5] And, of course, there is the seminal contribution of Black and Scholes that, virtually on the day it was published, brought the field to closure on the subjects of option and corporate-liability pricing.[6] As the Black-Scholes work was closing the gates on fundamental research in these areas, it was simultaneously opening new ones: in applied and empirical study and in setting the foundation for a new branch of finance called Contingent Claims Analysis or as we say for compactness, "CCA". The applications of CCA range from the pricing of complex financial securities to the evaluation of corporate capital-budgeting and strategic decisions.[7] As we will see shortly, it also has an important place in the theory of financial intermediation.

Time and uncertainty are the central elements that influence financial economic behavior. It is the complexity of their interaction that provides intrinsic excitement to the study

---

[2] The rediscovery of his work in the early 1950s is generally credited to P. A. Samuelson via L. J. Savage. Samuelson's own work on warrant pricing had an important impact on the development of continuous-time finance [cf. Merton (1983, pp. 106-107; 128-134)].

[3] Development of the continuous-time model in finance was the work of many. See Merton (forthcoming) for an overview and an extensive bibliography. Cox and Rubinstein (1985) cover its application to option-pricing theory. Adler and Dumas (1983) provide an excellent survey on the application of the continuous-time model in international finance, an application pioneered by Solnik (1974).

[4] Breeden (1979), Cox and Huang (forthcoming), Cox, Ingersoll and Ross (1985), and Merton (1969; 1971; 1973b).

[5] Although I've made no scientific study, it appears that for a theory to move from the "seminal" to the "classic" category requires an unspecified, but appropriately-long, passage of time from its publication. For that reason alone, I do not mention the Arbitrage Pricing Theory of Ross (1976a) in the text. But, here too, the continuous-time model can add something to the robustness of that theory by providing a rigorous foundation for preserving the linearity of the return-generating process, even in the presence of derivative securities with nonlinear payoff functions.

[6] Black and Scholes (1973). See Black (1987) for a brief history on how he and Scholes came to arrive at their famous pricing formula.

[7] For an overview of the development of CCA and discussion of its multi-dimensional applications in finance, see Mason and Merton (1985) and Merton (forthcoming, Chapters 10, 13, and 14).

of finance. To capture the effects of this interaction often requires sophisticated analytical tools. Indeed, the mathematics of the continuous-time finance model contains some of the most beautiful applications of probability and optimization theory.[8] Of course, all that is beautiful in science need not be practical. And surely, not all that is practical in science is beautiful. But, here we have both. With all its seemingly abstruse mathematical complexity, the continuous-time model has nevertheless had a direct and significant influence on finance practice.[9] Although not unique, this conjoining of intrinsic intellectual interest with extrinsic application is a prevailing theme of research in financial economics, and especially so, in continuous-time finance.

While intended to exemplify this theme, my remarks today are not exclusively or even primarily aimed at practitioners. Nor, however, is my aim to break new ground in the theory of finance. Both practitioner and theorist will happily find their cups abundantly filled by the multifarious papers of these three days. Instead, in this salutatory session, I try my hand at providing a frame of reference for what is to follow, by shuttling back and forth between conceptual issues surrounding the continuous-time model in the theory of intermediation and potential applications of that model in the practice of intermediation.

In following this zig-zag course, I touch upon three categories of contributions of the continuous-time analysis to the theory and practice of financial intermediation: product identification, product implementation and pricing, and risk management and control for an intermediary's entire portfolio.

A commonplace result in the continuous-time model is that investors' optimal demands for securities exhibit a structure such that each and every investor's optimal portfolio can be duplicated by various combinations of a relatively small and select set of mutual funds.[10] That is, if individual securities are "pre-packaged" into a specified group of portfolios, then the theory holds that investors can achieve the same optimal-portfolio allocations by selecting from just this group as they could by choosing from the entire universe of available securities. This finding is generally used to help identify the various sources of systematic risk in

---

[8] Prime examples are the application of Itô's Lemma in the derivation of the Black-Scholes option-pricing theory and the application of Martingale theory of the French probability school in the elegant Cox-Huang (forthcoming) solution of the lifetime consumption-investment problem.

[9] The model has influenced the practices of asset allocation, risk analysis, and performance measurement. However, as exemplified by the papers of this conference, its most-direct impact on practice has been in the pricing and hedging of financial instruments, an area that has experienced an explosion of innovations over the last decade. The continuous-time model is the prime mode of analysis used for pricing options on equities, fixed-income securities, stock and bond futures, and a variety of commodities. It is also used to price and hedge mortgage-backed securities; default risk, seniority, call provisions and sinking-fund arrangements on debt; bonds convertible into stock, commodities, or different currencies; floor and ceiling arrangements on interest rates; stock and debt warrants; rights and stand-by agreements. Indeed, much of the applied research on using the continuous-time model in this area takes place within practicing financial organizations.

[10] Cf. Breeden (1979), Cox, Ingersoll and Ross (1985), Merton (1971; 1973b), and Solnik (1974). These analyses describe the economic function for each of the mutual funds as well as the explicit portfolio rules for their construction. They also establish the minimum information set required to implement each fund's portfolio strategy. Such findings are not, of course, unique to the continuous-time model. Examples in discrete time are the Markowitz-Tobin Mean-Variance Model, the Arrow-Debreu Model, and the Arbitrage-Pricing Model of Ross.

multi-dimensional versions of the Capital Asset Pricing Model. However, in the context of financial intermediation, these same mutual-fund theorems also serve to identify a class of investment products for which there would seem to be a natural demand.[11]

A *derivative security* is a security with payoffs that can be expressed as a function of other traded-securities' prices, and these traded securities are called the *underlying securities* (of the derivative security). Common examples of derivative securities are option and futures contracts. As I need hardly mention in this company, Contingent Claims Analysis has achieved both theoretical and practical success in the pricing of derivative securities. In brief, the core of CCA is that a dynamic trading strategy in the underlying securities can be used to create a portfolio with payoffs that exactly replicate the payoffs to the derivative security. If the derivative is itself traded, then by an arbitrage argument, its equilibrium price must equal the value of the replicating portfolio.

The contribution of CCA to the enrichment of the theory of intermediation and insurance is, however, deeper than just the pricing of financial products.[12] CCA can also contribute to product implementation by providing the "blueprints" or production technologies for intermediaries to manufacture these securities. The portfolio-replication process used to derive derivative-security prices applies whether or not the security actually exists.[13] Thus, the specified dynamic portfolio strategy used to create an arbitrage position against a traded derivative security is also a prescription for synthesizing an otherwise nonexistent security. The investment required to fund the replicating portfolio becomes, in this context, the production cost to the intermediary that creates the security. In Section 3, I expand on this theory of production for financial intermediaries.

The focus of CCA is on the hedging and pricing of an individual security or financial product. However, as shown in Section 4, CCA, together with general dynamic portfolio theory, can be used to measure and control the total risk of an intermediary's entire portfolio. Although few in the practice of intermediation would doubt the central importance of risk management, such doubts do arise in the theory. It is, after all, standard fare that the Modigliani-Miller Theorems hold (at least approximately) in economic models with well-functioning capital markets. It follows as a corollary that at most, only the systematic-risk component of the firm's total risk warrants first-order attention by the firm's managers. In Sections 4 and 5, we examine the issue of whether, as a theoretical matter, financial intermediaries are different in this respect from other types of business firms. We conclude that the management of total risk by intermediaries can be of significant importance even in an

---

[11] As in the case of mutual funds, financial-intermediation activities often involve the combining of diverse financial assets into a package and the issuing of a single class of securities as claims against the portfolio. However, it is also common to "reverse" the process and issue a diverse set of claims against a relatively-homogeneous package of financial assets. One real-world example is the Collateralized Mortgage Obligation in which the portfolio contains mortgages of the same expected duration. Several classes of securities (called "tranches") are issued that have claims to different components of the total cash flow generated by the portfolio. A theoretical foundation for such "stripping" of various parts of a financial asset is laid in Section 3 with the development of Arrow-Debreu pure securities. See also Hakansson (1976) for a theory of "stripping" in his development of the "superfund".

[12] For specific applications to insurance products and underwriting, see Brennan and Schwartz (1976) and Kraus and Ross (1982).

[13] See the general derivation of CCA in Merton (1977).

environment where the Modigliani-Miller Theorems obtain with respect to ordinary business firms. Although mainly of theoretical interest, the analysis does provide some foundation for real-world policies that selectively discriminate between intermediaries and other firms when deciding on government bailouts and loan guarantees and when setting regulations.

With all the continuous-time model seems to offer, its application to the theory of intermediation nevertheless carries with it an apparent paradox. In the standard model, investors are entirely indifferent as to whether or not derivative securities are created, because investors can themselves use the dynamic portfolio strategies of CCA to replicate the payoff patterns to these securities. Thus, each derivative security is redundant, and because it adds nothing new to the market, creation of such a security provides no social benefit. Of course, in the real world, the prescribed dynamic replications may not be feasible. But, the CCA methodology is valid only if the payoffs to the derivative security can be reproduced by trading in existing securities. Hence, we have the Hakansson paradox: CCA only provides the production technology and production cost for creating securities that are of no consequence.[14]

Much the same paradox applies to the mutual-fund theorems of the continuous-time model: investors are again simply indifferent between selecting their portfolios from a group of funds that span the optimal-portfolio set or from all available securities. It would thus seem that the rich menu of financial intermediaries and derivative securities observed in the real world has no important economic function in a frictionless environment where investors have the same information, can trade continuously, and face no transactions costs or taxes.

Such indifference is indeed the case if *all* investors can gather information and transact without cost. Hence, some type of information or transactions-cost structure in which financial intermediaries and market makers have a comparative advantage with respect to other investors and corporate issuers is required to provide a *raison d'être* for financial intermediation and markets for derivative securities.[15]

With this in mind, I begin the formal analysis in Section 2 by using the Cox-Ross-Rubinstein (1979) binomial model to derive the production technology and cost for creating a derivative security in the presence of transactions costs. The derived costs of hedging both long and short positions in the same derivative security provides an endogenous specification of the relation among bid-ask price spreads for derivatives and their underlying securities. For an empirically-relevant range of investor transactions-cost, we show by example that the induced spreads in derivative prices can be substantial, and thereby, suggest the prospect of significant benefits from efficient intermediation.

---

[14] In the context of option securities, Hakansson (1979, p. 722) calls this "The Catch 22 of Option Pricing." In discussing the CCA methodology, he writes, "So we find ourselves in the awkward position of being able to derive unambiguous values only for redundant assets and unable to value options which do have social value." (p. 723).

[15] For example, in Merton (1978), the cost of surveillance by the deposit insurer is, in equilibrium, borne by the depositors in the form of a lower yield on their deposits. If all investors can transact costlessly, then none would hold deposits and instead would invest directly in higher-yielding UST bills. Thus, to justify this form of intermediation, it is necessary to assume that at least some investors face positive transactions costs for such direct investments in the market.

Although analytically tractable, even the simple binomial pricing model is greatly complicated by the explicit recognition of transactions costs. As we know from the work of Constantinides (1986), Leland (1985), and Sun (1987), incorporation of such costs in the continuous-time model is considerably more difficult.[16] Moreover, development of a satisfactory equilibrium theory of allocations and prices in the presence of transactions costs promises still more complexity because it requires a simultaneous endogenous determination of prices, allocations, *and* the least-cost form of financial intermediation and market structures.

To circumvent all this complexity and also preserve a role for intermediation, I turn to a continuous-time model in which many investors cannot trade costlessly, but the lowest-cost transactors (by definition, financial intermediaries) can. In this model, standard CCA can be used to determine the production costs for financial products issued by intermediaries. However, unlike in the standard zero-transaction-cost model, these products can significantly improve economic efficiency. If the traded-security markets and financial-services industry are competitive, then equilibrium prices will equal the production costs of the lowest-cost producers.[17] It is shown in Section 5 that in this environment, a set of feasible contracts between investors and intermediaries exists that permit all investors to achieve optimal consumption-bequest allocations as if they could trade continuously without cost. Thus, this model provides a resolution of the Hakansson paradox by showing that mutual funds and derivative securities can provide important economic benefits to investors and corporate issuers, even though these securities are priced in equilibrium as if they were redundant.

In sum, the analysis shall demonstrate the versatility of the continuous-time model in applications ranging from a detailed micro production theory for individual products, to risk management and control for the entire intermediary, and on to broad functional roles for intermediation. All of this is a part of a larger agenda to use the model as a unifying framework for a general theory of the financial-services industry. Along that line of thought, I shall conclude with a brief afterword that touches upon application of the continuous-time model to policy and strategy issues in intermediation. With this overview as a guide, I turn now to the substantive analysis.

## 2. Derivative-Security Pricing with Transactions Costs

In this section, we examine the effects of transactions costs on derivative-security pricing by using the two-period version of the Cox-Ross-Rubinstein (1979) binomial option-

---

[16] With diffusion processes and proportional transactions costs, investors cannot trade continuously and therefore, cannot perfectly hedge derivative-security positions. The reason is that with continuous trading, transactions costs at each trade will be proportional to $|\,dz\,|$, where dz is a Brownian motion. However, for any non-infinitesimal T, $\int_0^T |\,dz\,| = \infty$, almost certainly and hence, with continuous trading, the total transactions cost is unbounded with probability one.

[17] More generally, standard CCA will provide a close approximation if the "mark-up" per unit required to cover the intermediary's transactions costs and profit is sufficiently small that from the perspective of its customers' behavior, the additional cost is negligible. Of course, a tiny margin applied to large volume can produce substantial total profits for the financial-intermediation industry.

pricing model.[18] As illustrated in Figure 1, the initial stock price $S(0)$ is given by S_0. At time 1, the stock price will equal either S_{11} or S_{12}. If $S(1) = S_{11}$, then at time 2, $S(2)$ will equal either S_{21} or S_{22}. If $S(1) = S_{12}$, then $S(2)$ will equal either S_{23} or S_{24}. R denotes the return per dollar invested in the riskless security and is constant over both periods. To capture the effect of transactions costs, we assume that a commission must be paid on each purchase or sale of the stock and that the commission rate is a fixed proportion τ of the dollar amount of the transaction. Equivalently,

Figure 1. Tree Diagram of Possible Stock-Price Paths
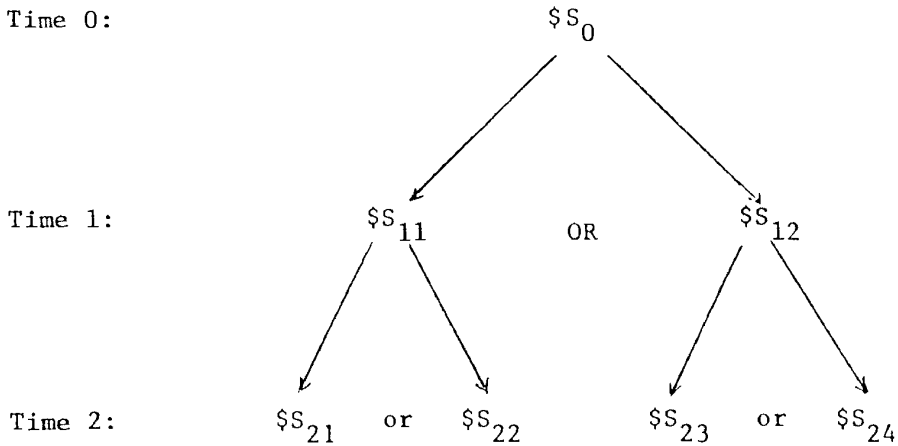
we could assume a bid-ask spread in which investors pay the *ask price* for the stock, $S^a(t) \equiv (1 + \tau)S(t)$, when they buy and receive the *bid price*, $S^b(t) \equiv (1 - \tau)S(t)$, when they sell. There are no costs for transacting in the riskless security.

As shown in Cox, Ross and Rubinstein (1979), the array of possible stock prices must satisfy certain conditions to rule out the possibility of arbitrage or dominance opportunities between the stock and the riskless security.[19] The corresponding set of restrictions in the presence of transactions costs can be written as:

---

[18] The binomial model is, of course, a discrete-time model. If, however, the time interval between successive price changes is h; the magnitudes of the price changes between successive periods, $S(t + h) - S(t)$, are proportional to $\sqrt{h}$; and the probability of each of the two possible changes is $.50 + 0(\sqrt{h})$, then Cox, Rubinstein, and Ross (1979) have shown that in the limit as $h \to dt$, continuous time, the binomial option-pricing model converges to the Black-Scholes continuous-time model. This same limiting process was used by Bachelier (1900) as one of his arguments to justify his option-pricing model. And, along the way, he also used it to derive the Fourier partial differential equation as the governing equation for the probabilities of diffusion processes.

[19] Satisfaction of these conditions ensures that one is never unambiguously worse off to pay the transactions costs necessary to exactly hedge the position instead of saving the costs and not hedging the position. However, under the hypothesized conditions of the preceding footnote with $(S_{12} - S_0)/S_0 \equiv \lambda\sqrt{h} > 0$, (1) will fail for all time intervals $h \leq \bar{h} \equiv 4\tau^2/[(1 - \tau)\lambda]^2$. This reflects the remark in footnote 16, that it never pays to trade continuously in the presence of transactions costs.

(1a) $\qquad S_{11} < S_0 R < (1 - \tau) S_{12}/(1 + \tau)$,

(1b) $\qquad S_{21} < S_{11} R < (1 - \tau) S_{22}/(1 + \tau)$,

(1c) $\qquad S_{23} < S_{12} R < (1 - \tau) S_{24}/(1 + \tau)$.

Consider an intermediary that sells to a customer a call option with exercise price E and expiration date two periods from now. The terms of the option require cash settlement in which the customer is paid the in-the-money value of the call, $S(t) - E$, if the call is exercised. In the case where prices are quoted as a spread, the stock price is determined by the average of the bid and ask prices, $\bar{S}(t) \equiv [S^a(t) + S^b(t)]/2 = S(t)$.

The production cost for manufacturing the call option is determined by deriving a dynamic portfolio strategy in the stock and riskless security that exactly replicates the payoff to the option. By following this strategy, the intermediary can completely hedge all the risk of this liability. In determining the cost, we assume that the intermediary has no initial position in the underlying stock and that all stock held at the expiration date of the option is sold in the market.[20]

If $S(t) = S$ and the commission rate is τ, then let $N(S,t;\tau)$ denote the number of shares of stock held in the portfolio at time t after adjusting the portfolio to the desired position. If $N(S,t;\tau) < 0$, then the portfolio is short $| N(S,t;\tau) |$ shares. Let $B(S,t;\tau)$ denote the amount of the riskless security held in the portfolio after the payment of the transactions costs associated with adjustments to the portfolio at time t. If $B(S,t;\tau) < 0$, then the portfolio has borrowed \$ $| B(S,t;\tau) |$. Let $F(S,t;\tau)$ denote the value of the portfolio *before* payment of transactions costs incurred at time t.

We derive the replicating portfolio strategy by beginning at the expiration date and working backwards in time in a dynamic-programming-like fashion. If $S(1) = S_{11}$, then to exactly match the payoff to the option at $t = 2$, the portfolio composition must satisfy:

$N(S_{11},1;\tau)(1 - \tau)S_{21} + B(S_{11},1;\tau)R = H(S_{21})$ in the event $S(2) = S_{21}$, and $N(S_{11},1;\tau)(1 - \tau)S_{22} + B(S_{11},1;\tau)R = H(S_{22})$ in the event $S(2) = S_{22}$. $H(S) \equiv Max[0,S - E]$ is the schedule of payments to the customer at expiration and we have taken account of commissions paid on the sale of the stock in the portfolio. From the matching conditions, we have that:

(2a) $\qquad N(S_{11},1;\tau) = [H(S_{22}) - H(S_{21})]/[(1 - \tau)(S_{22} - S_{21})]$
$\qquad\qquad\qquad = N(S_{11},1;0)/(1 - \tau)$

and

(2b) $\qquad B(S_{11},1;\tau) = [H(S_{21})S_{22} - H(S_{22})S_{21}]/[R(S_{22} - S_{21})]$
$\qquad\qquad\qquad = B(S_{11},1;0)$.

Because $S_0 > S_{11}$, we tentatively assume (and verify later) that in the event $S(1) = S_{11}$, the portfolio holdings of the stock should be reduced from the initial position $N(S_0,0;\tau)$, established at $t = 0$. Hence, for $S(1) = S_{11}$, the intermediary will incur a transaction cost of

---

[20] Because commissions are paid for both purchases and sales of the stock, this assumption produces the most "conservative" (i. e., highest) estimate of the cost. In practice, an intermediary with an ongoing business of writing options on this stock would avoid the double costs of liquidating stock at the expiration of one option and repurchasing stock to hedge the new issue of another. Hence, the inventory of stock held by the intermediary will affect its marginal cost of producing options. Therefore, the prices derived here provide the outerbound on the bid-ask spread.

$\tau[N(S_0,0;\tau) - N(S_{11},1;\tau)]S_{11}$ to adjust the portfolio. Therefore, from (2a) and (2b), the total resources required in the portfolio at time 1 to support this strategy can be written as:[21]

(2c) $\quad F(S_{11},1;\tau) = N(S_{11},1;\tau)S_{11} + B(S_{11},1;\tau) + \tau[N(S_0,0;\tau) - N(S_{11},1;\tau)]S_{11}$
$\quad\quad\quad\quad\quad = F(S_{11},1;0) + \tau N(S_0,0;\tau)S_{11}.$

If instead at $t = 1$, $S(1) = S_{12}$, then at $t = 2$, $S(2)$ will equal either S_{23} or S_{24}. By the same analysis leading to (2a) and (2b), we have that:

(3a) $\quad N(S_{12},1;\tau) = [H(S_{24}) - H(S_{23})]/[(1 - \tau)(S_{24} - S_{23})]$
$\quad\quad\quad\quad\quad = N(S_{12},1;0)/(1 - \tau)$

and

(3b) $\quad B(S_{12},1;\tau) = [H(S_{23})S_{24} - H(S_{24})S_{23}]/[R(S_{24} - S_{23})]$
$\quad\quad\quad\quad\quad = B(S_{12},1;0) .$

Because $S_{12} > S_0$, we assume that in the event $S(1) = S_{12}$, the stock holdings in the portfolio should be increased from their level at $t = 0$. Hence, the intermediary will incur a transaction cost of $\tau[N(S_{12},1;\tau) - N(S_0,0;\tau)]S_{12}$ to adjust the portfolio. From (3a) and (3b), the total portfolio value required at $t = 1$ is:[22]

(3c) $\quad F(S_{12},1;\tau) = N(S_{12},1;\tau)S_{12} + B(S_{12},1;\tau) + \tau[N(S_{12},1;\tau) - N(S_0,0;\tau)]S_{12}$
$\quad\quad\quad\quad\quad = F(S_{12},1;0) + \tau[2N(S_{12},1;\tau) - N(S_0,0;\tau)]S_{12} .$

By inspection of (2a) and (3a), $N(S(1),1;\tau) = N(S(1),1;0)/(1 - \tau)$, and therefore, the number of shares of stock held is larger with transactions costs than in the no-cost case. From (2b) and (3b), $B(S(1),1;\tau) < 0$, and the amount borrowed is independent of the level of transactions costs. From (2c) and (3c), $F(S(1),1;\tau)$ exceeds $F(S(1),1;0)$, the amount required to fund the portfolio with no transactions costs.

To exactly replicate the return on the option from $t = 0$ until expiration, the portfolio strategy at $t = 0$ must produce a portfolio value of $F(S_{11},1;\tau)$ at $t = 1$ if $S(1) = S_{11}$ and a value of $F(S_{12},1;\tau)$ if $S(1) = S_{12}$. Because $F(S(1),1;\tau)$ includes the transactions costs for portfolio changes at $t = 1$, this funding requirement can be met if $N(S_0,0;\tau)S(1) + B(S_0,0;\tau)R = F(S(1),1;\tau)$. It follows that $N(S_0,0;\tau) = [F(S_{12},1;\tau) - F(S_{11},1;\tau)]/(S_{12} - S_{11})$ and that $B(S_0,0;\tau) = [F(S_{11},1;0)S_{12} - F(S_{12},1;0)S_{11}]/[R(S_{12} - S_{11})] - [N(S_0,0;\tau)(1 - \tau) - N(S_0,0;0)]S_{11}/R$. By substitution from (2c) and (3c) and the rearrangement of terms, we have that:[23]

(4a) $\quad N(S_0,0;\tau) = N(S_0,0;0) + 2\tau[N(S_{12},1;\tau)S_{12}$
$\quad\quad\quad\quad\quad\quad - N(S_0,0;0)(S_{11}+S_{12})/2]/[(1+\tau)S_{12}-(1-\tau)S_{11}]$

---

[21] As discussed in footnote 20, the resources required to fund the portfolio depend on the inventory of stock held by the intermediary. Thus, if the same option were just being created at $t = 1$, (2a) and (2b) would still apply, but the transaction cost paid would be $\tau N(S_{11},1;\tau)S_{11}$. Hence, the required funding for the portfolio would be $F(S_{11},1;0) + 2\tau N(S_{11},1;\tau)S_{11}$.

[22] As in the preceding footnote, if the portfolio were just now being created, (3a) and (3b) would still apply, but the transaction cost paid would be $\tau N(S_{12},1;\tau)S_{12}$. The corresponding funding for the portfolio would be $F(S_{12},1;0) + 2\tau N(S_{12},1;\tau)$.

[23] Using (2a) and (3a) with $H(S) = Max[0, S-E]$, we have from (4a) that $N(S_{11},1;\tau) < N(S_0,0;\tau) < N(S_{12},1;\tau)$. Hence, the direction of changes in the portfolio, tentatively assumed in the derivation of (2c) and (3c), is verified.

and

(4b) $B(S_0,0;\tau) = B(S_0,0;0) - [N(S_0,0;\tau)(1 - \tau) - N(S_0,0;0)]S_{11}/R$.

Because $N(S_{12},1;0) > N(S_0,0;0)$ and $S_{12} > S_{11}$, we have from (4a) and (4b) that $N(S_0,0;\tau) > N(S_0,0;0) > 0$ and that $B(S_0,0;\tau) < B(S_0,0;0) < 0$. Hence, the presence of transactions costs causes a larger long position in the stock and additional borrowing in the replicating portfolio.

The initial investment in the portfolio required to undertake these positions (including the transaction cost of $\tau N(S_0,0;\tau)S_0$) can be written as:

(4c) $F(S_0,0;\tau) = F(S_0,0;0)+[N(S_0,0;\tau)-N(S_0,0;0)][S_0-S_{11}/R]$
$+\tau N(S_0,0;\tau)[S_0+S_{11}/R]$.

Because $N(S_0,0;\tau) > N(S_0,0;0) > 0$ and $S_0 > S_{11}/R$, we have by inspection of (4c) that $F(S_0,0;\tau) > F(S_0,0;0)$. We thus verify that an increase in the cost of producing a call option caused by commissions charged in the stock market increases the option price charged by the intermediary. In a competitive financial-services industry, the ask price for the call option is its production cost, $F(S_0,0;\tau)$.

To explore further the spread in call-option prices induced by transactions costs in the stock market, consider a customer who would like to sell a call option to the intermediary. To determine the (maximum) price to bid for the call option, the intermediary solves for the portfolio strategy with a return that exactly hedges the payoffs it would receive from holding the call option. If there were no transactions costs, the replicating strategy would be the exact mirror-image of the one used to hedge a short position in the call option. That is, the intermediary would hold short positions in the stock given by $-N(S(t),t;0) < 0$, and hold positive amounts of the riskless security given by $-B(S(t),t;0) > 0$. The portfolio would require a negative initial investment of $-F(S_0,0;0)$ and this net cash flow to the intermediary at $t = 0$ would be the maximum amount that it would pay to the customer for the call option.

Although the qualitative features of the replicating portfolio will be the same with transactions costs, the magnitudes of the positions held will not be the same because the intermediary must pay the commissions no matter which side of the transaction it undertakes. Applying the same analysis used to derive (2)–(4), we have that:

(5a) $N^1(S_0,0;\tau) = -N(S_0,0;-\tau)$

(5b) $B^1(S_0,0;\tau) = -B(S_0,0;-\tau)$

(5c) $F^1(S_0,0;\tau) = -F(S_0,0;-\tau)$

where the prime on each variable denotes the positions and amounts required to hedge a long position in a call option. By inspection of (4a)–(4c), it is readily apparent that $N(S_0,0;\tau)$, $B(S_0,0;\tau)$ and $F(S_0,0;\tau)$ are not even-functions of τ. Therefore, it follows from (5a)–(5c) that for $\tau > 0$, the replicating strategy to hedge a long position in a call option is not simply the reverse of the replicating strategy to hedge a short position in a call option.

We can however say more. From (1a), $(1 - \tau)S_{12} > (1 + \tau)S_{11}$ and from (3a) and (4a), $N(S_{12},1;0) > N(S_0,0;0)$. Using this condition in (4a), we have that:

(6) $0 < N(S_0,0;-\tau) < N(S_0,0;0) < N(S_0,0;\tau)$.

Hence, the number of shares held short to hedge a long call position is fewer than the number held long to hedge a short call position. From manipulation of (4c), it can also be shown that:

(7) $0 < F(S_0,0;-\tau) < F(S_0,0;0) < F(S_0,0;\tau)$.

That is, the minimum price at which the intermediary would sell a call option exceeds the maximum price at which the intermediary would buy a call option. The zero-transactions-cost price of the option is between the two.

At the outset of our analysis, we saw that with symmetric proportional transactions costs, the average of the bid and ask prices of the stock, $\bar{S}(t)$, is equal to the stock price with no transactions costs, $S(t)$. In a competitive financial-services industry, the bid price for the call option is given by $F(S_0,0;-\tau)$. In this environment, the average of the bid and ask prices of the call option, $\bar{F}(S_0,0;\tau)$, is equal to $[F(S_0,0;\tau) + F(S_0,0;-\tau)]/2$. From the conditions leading to (6) and (4c), it can be shown that $0 < F(S_0,0;0) - F(S_0,0;-\tau) < F(S_0,0;\tau) - F(S_0,0;0)$.

It follows that:

(8) $\bar{F}(S_0,0;\tau) > F(S_0,0;0)$.

That is, the average of the bid and ask prices of the option is a biased-high estimate of its zero-transactions-cost price.[24] Thus, symmetry of the bid and ask prices of the stock around its zero-transactions-cost price does not imply a corresponding symmetry for the bid and ask prices of the call option.

To provide some indication of the size of the effect on option prices induced by transactions costs in the underlying stock, consider the following numerical example of an at-the-money, two-period call option: the exercise price is $ 100; the interest rate is 5 percent; and the array of stock prices is $S_0 = \$ 100$; $S_{11} = \$ 90$; $S_{12} = \$ 115$; $S_{21} = \$ 70$; $S_{22} = \$ 110$; $S_{23} = \$ 90$; $S_{24} = \$ 140$. The bid and ask prices for the option, $F(100,0;-\tau)$ and $F(100,0;\tau)$, along with the initial number of shares required to hedge the position, are presented in Table 1 for transactions costs of a tenth of a percent, a half of a percent, and one percent. With transactions costs in that range, the effect on the initial hedge ratio of shares of stock per option is small. For $\tau = .01$ versus the zero-cost case, the difference in the number of shares is less than 3 percent. The asymmetry between the bid and ask prices is also small. For $\tau = .01$, the average of the bid and ask prices is $ 15.63 versus $ 15.61 for the $\tau=0$ case. However, the effect of transactions costs on the levels of the bid and ask prices is substantial. The percentage premium of the ask price above the zero-cost price is approximately linear in τ and equal to 10τ. Thus, the percentage premium is about 1 percent for $\tau = .001$ and 10 percent for $\tau = .01$. Similar results hold for the percentage discount of the bid price below the zero-cost price. Hence, the percentage spread between the bid and ask price is approximately 20τ. Although the price of the stock is much larger than the option price, the *dollar* spread between the bid and ask prices of the option is larger than the corresponding spread for the stock. Dollar spreads of $ 0.20, $ 1.00, and $ 2.00 for the stock induce respective spreads in the option prices of $ 0.31, $ 1.53, and $ 3.07.

Care should always be exercised in drawing strong inferences from a single example of a simple model. Nevertheless, these findings seem to indicate that for investors facing high, but empirically-relevant, levels of transactions costs for trading stocks, the cost of synthetically creating their own option contracts and other derivative securities can be prohibitively high.

---

[24] The magnitude of the bias is $0(\tau^2)$, and therefore, typically quite small.

Table 1. Bid and Ask Call-Option Prices with Transactions Costs

Assumed Interest Rate and Stock Price Environment

$R = 1.05$	$S_{11} = \$90$	$S_{21} = \$70$	$S_{23} = \$90$
$S_0 = \$100$	$S_{12} = \$115$	$S_{22} = \$110$	$S_{24} = \$140$

Intermediary Sells Call Option (E = $ 100)

Commission Rate, τ	Initial Hedge Ratio $N(100,0;\tau)$	Ask Price $F(100,0;\tau)$	Percentage Above $F(100,0;0)$
0.000	0.7038	$ 15.61	0.00 %
0.001	0.7054	15.76	0.98
0.005	0.7116	16.38	4.95
0.010	0.7192	17.17	9.96

Intermediary Buys Call Option (E = $ 100)

Initial Hedge Ratio $-N(100,0;-\tau)$	Bid Price $F(100,0;-\tau)$	Percentage Below $F(100,0;0)$
-0.7038	$ 15.61	0.00 %
-0.7022	15.46	0.98
-0.6957	14.85	4.88
-0.6873	14.10	9.69

One would, of course, expect that the costs for financial intermediaries to transact in stocks are considerably lower than for most investors. Further, even the lower transaction-cost numbers in Table 1 may significantly overstate the actual costs to intermediaries that create options as an ongoing business. As discussed in footnotes 20-22, the calculations leading to Table 1 assume that the intermediary carries no inventory of either stocks or options and that each position is perfectly hedged. As we discuss in Section 4, a more-complete analysis would take account of the opportunities for intermediaries to create and optimally manage such inventories to net out many of the transactions otherwise required to hedge individual option exposures. Moreover, for controlling the risks of intermediaries with derivative-security liabilities contingent on many different stocks, diversification may provide a cost-reducing alternative to a complete hedge of each position.

In summary, the two-period binomial model illustrates how bid and ask prices for derivative securities can be endogenously determined from the transaction-cost structure of their underlying securities. The analysis shows that the percentage spreads in the production costs of derivative securities can be many times larger than the spreads in their underlying securities. Hence, even with modest transactions costs for investors in traded securities, there is an economic function for financial intermediaries that specialize in the creation of derivative securities and take advantage of economies of scale to produce them at a greatly reduced cost.

3. Production Theory for Zero-Transaction-Cost Financial Intermediaries

As we have seen, transactions costs among investors and corporate issuers are virtually a requirement to justify an important economic role for financial intermediation.[25] However, as we have also seen, explicit recognition of such costs can cause even the most simple of models to become extraordinarily complex. As suggested in the introduction, a happy compromise is to assume that some agents face significant transactions costs, but that financial intermediaries, as the lowest-cost transactors in the financial markets, do not. Hence, for the balance of the paper, we proceed under that assumption and develop the general theory for production and pricing of derivative securities by intermediaries.

An *Arrow-Debreu pure state-contingent security* is a security that pays its holder $ 1 if a particular state of the world obtains at a particular point in time and otherwise, pays nothing. More than a generation ago, Arrow demonstrated that the payoff structure for any state-contingent security can be exactly replicated by a portfolio combination of pure securities. Hence, to avoid arbitrage, the equilibrium price of the state-contingent security can be expressed as a weighted sum of the prices of these pure securities.[26] As I have discussed

---

[25] Information costs together with agency problems that prohibit the direct sale of information to investors can, of course, justify certain types of financial intermediation such as mutual funds. Regulation and special features of the tax laws may also induce the creation of specialized financial instruments and institutions. However, in the absence of transactions costs (which may themselves be induced by information-gathering costs), it is difficult to explain the complex structure of financial intermediaries and their wide scope of activities that we observe in the real world.

[26] See the classical works of Arrow (1953; 1964) and Debreu (1959) for the original development of pure state-contingent securities. There is an enormous literature based on the Arrow-Debreu model [cf. Radner (1972) and Merton (1982, Section 5)].

elsewhere,[27] it was recognized from the outset that the Black-Scholes approach to option pricing could also be applied to the pricing of derivative securities in general. However, it was not until later that these two theories of valuation were explicitly connected. Ross (1976b), Hakansson (1976), Banz and Miller (1978), and Breeden and Litzenberger (1978) were among the first to show that combinations of options could be used to create pure securities and that these pure securities could be used to price derivative securities. In this section, we derive the natural analog to Arrow-Debreu pure securities in the continuous-time model and demonstrate their application to the pricing of contingent-claim securities.

Merton (1977, Section 2) derives the price of a general derivative security with payoffs that are a function of a traded asset's price and time. Under the usual frictionless-market assumptions [(A.1)-(A.6) there], it is shown that the price of the derivative security, $F[V,t]$, will satisfy:

$$(9) \qquad 0 = \frac{1}{2}\,\sigma^2 V^2 F_{11} + [rV - D_1]F_1 - rF + F_2 + D_2,$$

for $0 \leq \underline{V}(t) < V < \bar{V}(t) \leq \infty$ and $t < T$, subject to the boundary conditions

$$(10a) \qquad F[\bar{V}(t),t] = f[\bar{V}(t),t]$$

$$(10b) \qquad F[\underline{V}(t),t] = g[\underline{V}(t),t]$$

$$(10c) \qquad F[V,T] = h[V].$$

Subscripts on F in (9) denote partial derivatives with respect to V and t. $D_1(V,t)$ and $D_2(V,t)$ denote the cash-flow rates paid to the holders of the traded asset and its derivative security, respectively. σ^2 is the instantaneous variance rate of the return on the traded asset and r is the riskless interest rate (assumed to be constant over time). The limited liability of the traded asset implies that $V(t) = 0$ only if $V(t + \tau) = 0$ and $D_1(0,t + \tau) = 0$ for all $\tau > 0$. Without loss of generality, we can assume that $D_2(0,t) = 0$.[28] The functions D_2, f, g and h specify the payoff schedule for the derivative security.

In the analysis to follow, we assume that the stochastic process governing the dynamics of the traded asset is such that prob $\{ V(\tau) = 0 \mid V(t) > 0\} = 0$ for $t \leq \tau < \infty$.[29] Let $\pi[V(t),t;E,T]$ denote the price at time t of the particular derivative security with a payoff structure given by: $D_2 = 0$, and for $V(T) = V$ and $E > 0$:

$$(11) \qquad \pi[V,T;E,T] = \delta(E - V)$$

where $\delta(x)$ denotes the Dirac delta function with the properties that $\delta(x) = 0$ for $x \neq 0$ and $\delta(0)$ is infinite in such a way that $\int_a^b \delta(x)dx = 1$ for any $a < 0 < b$.[30] We now show that the

---

[27] Merton (forthcoming, Chapters 10 and 14).

[28] Because $V(t) = 0$ implies that $V(t + \tau) = 0$ for $\tau > 0$, the capitalized value of any promised payments, $D_2(0,t + \tau)$, can be incorporated into the termination function, $g[0,t]$. Thereby, we can formally set $D_2(0,t + \tau) = 0$ without loss of generality.

[29] That is, we assume that $V(t) = 0$ is an *inaccessible boundary* for finite t. For example, this assumption is satisfied if the underlying asset has a proportional payout-rate policy, $D_1(V,t) = \rho(t)V$, $0 \leq \rho(t) \leq M$, where $M < \infty$ for all t and σ is a constant. General mathematical conditions for inaccessible boundaries are given in Karlin and Taylor (1981, pp. 226-250).

[30] $\delta(x)$ is not a function in the usual sense and instead, belongs to an extended class called "generalized functions." Cf. Dettman (1969, pp. 228-229).

derivative securities defined by (11) for various values of E and T are essentially Arrow-Debreu pure securities, extended to an environment with a continuum of states defined by the price of the traded asset and time.[31]

Let $dE > 0$ denote the infinitesimal differential of the parameter E. Consider a portfolio strategy that at time t purchases dE units of each of the continuum of derivative securities with parameter values $E\varepsilon[E_1,E_2]$ and $0 < E_1 < E_2$. If the portfolio is held until time T and $V(T) = V$, then from (11), the value of the portfolio is given by:

$$(12) \qquad \int_{E_1}^{E_2} \delta(E - V)dE = 1 \qquad \text{if} \qquad E_1 < V < E_2$$
$$= 0 \qquad \text{otherwise.}$$

The cost of acquiring this portfolio at time t is $\int_{E_1}^{E_2} \pi[V(t),t;E,T]dE$.

If we undertake the particular limiting strategy where $E_2 = \bar{E} + dE/2$ and $E_1 = \bar{E} - dE/2$, then from (12), the value of the portfolio is, in the limit, \$ 1 if $V(T) = \bar{E}$ and \$ 0, otherwise. By the Mean-Value Theorem, the cost of the portfolio at time t is $\pi[V(t),t;\bar{E},T]dE$. Thus, for $E > 0$, $\pi[V(t),t;E,T]dE$ is the price at time t of an Arrow-Debreu state-contingent security that pays \$ 1 at time T if $V(T) = E$ and nothing otherwise.[32]

Consider the limiting portfolio strategy in which $E_2 \to \infty$ and $E_1 \to 0$. From (12) and the assumption that $V(T) > 0$ with probability one, the value of the portfolio at time T will be \$ 1 for all possible values of $V(T)$. Therefore, to rule out arbitrage between the riskless security and the derivative securities, their prices must for all $V(t)$ and $t < T$, satisfy:[33]

$$(13) \qquad e^{-r(T-t)} = \int_0^{\infty} \pi[V(t),t;E,T]dE.$$

Because all possible payoffs to each of the derivative securities are nonnegative, the no-arbitrage condition requires that $\pi[V,t;E,T] \geq 0$. Hence, from (13), we have that $\pi[V,t;E,T]$ is a bounded function in the limit as $V \to \infty$.

If at time t, we construct a portfolio that holds EdE units of each of the continuum of derivative securities with parameter values $E\varepsilon(0,\infty)$, then from (11), the value of the portfolio at T is given by $\int_0^{\infty} E\delta(E - V)dE = V$ for $V(T) = V$. An investment in one unit of the

---

[31] In the usual applications of the Arrow-Debreu model, the number of states is countable, and the payoffs to the pure securities are contingent on "states of nature." That is, which state is realized is assumed to be unaffected by the actions of economic agents, either individually or collectively. We do assume here that the actions of any one agent have no effect on the time path of $V(t)$. However, $V(t)$ is a price and hence, is endogenously determined within the economic system. Thus, unlike the usual case, the state-contingent securities here depend on a state-space description that is controlled by the collective actions of economic agents.

[32] As is well known for state-contingent prices in the standard Arrow-Debreu model, the $\{\pi[V,t;E,T]\}$ have a functional structure like that of a conditional probability density function. As with the probability density for a diffusion process, π is an order-one function, but the probability that $V(T) = E$ is infinitesimal, if $V(t) = E$ is not an absorbing-barrier point for $t \leq T$. Hence, the value of a security that pays a finite amount only in that state is also infinitesimal.

[33] This result is a well-known property of pure securities in the Arrow-Debreu model. Cf. Merton (1982, Section 5).

traded asset at time t will also be worth $V(T)$ at T. However, in addition, that investment will also receive all payouts to the asset, $\{D_l\}$, between t and T. Hence, to avoid arbitrage opportunities between the traded asset and the derivative securities, their prices must satisfy:

(14)
$$V(t) \geq \int_0^\infty \pi[V(t),t;E,T]E\,dE.$$

From (14) and the nonnegativity of π, we have that $\pi[V,t;E,T]E/V$ is a bounded function in the limit as $V \to 0$. Therefore, for $E > 0$ and fixed, $\pi[V,t;E,T]/V$ is bounded in the limit as $V \to 0$.

From the general derivation in Merton (1977), to avoid arbitrage opportunities, $\pi[V,t;E,T]$ must satisfy (9) with $D_2 = 0$, $\underline{V}(t) = 0$, and $\bar{V}(t) = \infty$ for all $E > 0$ and $t < T$. The boundary conditions required for a unique solution are that $|\pi[V,t;E,T]/V|$ is bounded and that (11) is satisfied. Thus, under the hypothesized frictionless-market conditions [(A.1)–(A.6) in Merton (1977)], we have determined the prices of a complete set of pure state-contingent securities, where the state space is defined by the price of the traded asset and time.

In the context of this section, $\pi[V,t;E,T]$ is the production cost to a zero-transaction-cost financial intermediary for creating this security at time t when $V(t) = V$. From Merton (1977), the production technology for replicating the payoffs to the security calls for a total portfolio investment at time t of $\pi[V,t;E,T]$ dollars with $(\partial\pi[V,t;E,T]/\partial V)V$ dollars in the traded asset and the balance in the riskless security.

The connection between these pure state-contingent securities and the theory of financial intermediation can be made apparent by examining the general class of derivative securities with payoff structures given by (10a)–(10c), with $\underline{V}(t) = 0$ and $\bar{V}(t) = \infty$ for all t. It is well known from the Green's functions method of solving linear differential equations, that the solution to (9), subject to these boundary conditions can be written as :[34]

(15)
$$F[V,t] = \int_t^T \int_0^\infty D_2(E,\tau)\pi[V,t;E,\tau]\,dE\,d\tau + \int_0^\infty h[E]\pi[V,t;E,T]\,dE.$$

Just as in the standard Arrow-Debreu model, the payoff structure to this derivative security can be expressed as a linear combination of the payoff structures of the pure state-contingent securities. Thus, these securities provide the fundamental "building blocks" for constructing more-complex securities. The term "building blocks" is apt because none of these pure securities, taken individually, is likely to be demanded by any customer of the intermediary. Nevertheless, once an intermediary has determined the production costs for the complete set of these state-contingent securities, it can use simple quadrature in (15) to calculate the production cost for any derivative security with $\underline{V}(t) = 0$ and $\bar{V}(t) = \infty$ for all t.

Equation (15) is a powerful tool for the evaluation of derivative-security prices. It cannot, however, be applied to securities for which there is a positive probability that either $V(t) \leq \underline{V}(t)$ or $V(t) \geq \bar{V}(t)$ for some $t < T$. To see why, consider the example of an American put option. As shown in Merton (1973a, Theorem 13), there always exists at each point

---

[34] See Dettman (1969, Chapter 5).

in time, a sufficiently small, but positive stock price, such that it pays to exercise the put immediately. Let $\{\underline{V}(t) > 0\}$ denote the early-exercise schedule of stock prices, such that the put option is exercised at t if $V(t) \leq \underline{V}(t)$. From (10b), $g[\underline{V}(t),t] = X - \underline{V}(t)$ and from (10c), $h[V] = Max\ [0, X - V]$, where X is the exercise price of the put. If $\underline{V}(t)$ is a continuous function of t, then a naive application of (15) might suggest that the put price can be written as:

(16) $\qquad F[V,t] = \int\limits_{t}^{T}\ [X - \underline{V}(\tau)]\pi[V,t;\underline{V}(\tau),\tau]d\tau\ +\ \int\limits_{0}^{X}\ [X - E]\pi[V,t;E,T]dE$.

However, (16) gives an incorrect evaluation. For times t_1 and t_2 such that $t \leq t_1 < t_2 \leq T$, the events that $V(t_1) = \underline{V}(t_1)$ and $V(t_2) = \underline{V}(t_2)$ are not mutually exclusive. Equation (16) implies that if both events occur, the put holder would receive payments of $[X - \underline{V}(t_1)]$ at $t = t_1$ and $[X - \underline{V}(t_2)]$ at $t = t_2$. But, of course, if the put is exercised at $t = t_1$, it cannot also be exercised at $t = t_2$. Thus, when early exercise is possible, (16) overstates the value of the put.[35]

The American put option is an example of a contingent-claim security with a *path-dependent* payoff structure. That is, looking forward from the perspective of date t, the payoff at time $\tau > t$ depends not only on the price of the underlying asset at τ, $V(\tau)$, but also on the time path followed by the asset's price between t and τ. In contrast, the payoff to the pure state-contingent security given by (11) depends only on the price of the traded asset at date T and therefore, is not path-dependent.[36] Thus, to have the right to a payment of $[X - \underline{V}(\tau)]$ dollars at time τ, contingent on both $V(\tau) = \underline{V}(\tau)$ and $V(s) > \underline{V}(s)$ for all $s\epsilon[t,\tau)$, is not the same as owning $[X - \underline{V}(\tau)]d\tau$ units of a pure security that, collectively , pay $[X - \underline{V}(\tau)]$ dollars at time τ, contingent only on $V(\tau) = \underline{V}(\tau)$. It is that lack of equivalence that causes (16) to fail as a valuation formula.

The pure-securities approach to valuation can be modified to accommodate derivative securities with path-dependent payoffs. To do so, one constructs a set of state-contingent securities that are also path-dependent by replacing the payoff function (11) with the condition that $\pi[V,T;E,T] = \delta(E - V)$ if $\underline{V}(t) < V(t) < \overline{V}(t)$ for *all* $t < T$ and $\pi[V,T;E,T] = 0$ otherwise.[37] With this modified set of pure securities, one can use quadrature in an equation like (15) or (16) to determine the prices of all derivative securities with the *same* specified schedules, $\{\underline{V}(t)\}$ and $\{\overline{V}(t)\}$. Note, however, that a different set of pure securities is required for each specification of $\{\underline{V}(t)\}$ and $\{\overline{V}(t)\}$. Unfortunately, these schedules tend to be specific to each particular derivative security.[38] Hence, with path-dependent payoff structures,

---

[35] Equation (16), with $\underline{V}(t) \equiv 0$ does provide the correct price for an *European* put option, because $\pi[V,t;0,\tau] = 0$ for $t \leq \tau \leq T$ and $V > 0$.

[36] Path-independent and path-dependent processes are directly related to Markov processes. The returns on a security with path-dependent payoffs will not be a Markov process in $V(t)$. Because their payoffs are path-independent, pure state-contingent securities' returns are Markov processes in $V(t)$.

[37] Formally, we define an indicator variable $X(t)$ such that $X(t) = 1$ if $\underline{V}(s) < V(s) < \overline{V}(s)$ for all $s < t$ and $X(t) = 0$, otherwise. The path-dependent pure-security payoff function becomes $\pi[V,X,T;E,T] = X\delta(E - V)$. Although not Markov in $V(t)$ alone, $\pi[V(t),X(t),t;E,T]$ does follow a Markov process in $V(t)$ and $X(t)$. The general method of expanding the number of variables to convert a non-Markov process to a Markov process is discussed in Cox and Miller (1968, p. 262).

[38] Consider, for example, a collection of American put options on the same traded asset. From Merton (1973a, Section 8), the optimal early-exercise boundary will depend on both the exercise price and the expiration date. Thus $\{\underline{V}(t)\}$ will be different for each one of these derivative securities.

there will generally be no computational advantage for this technique over the direct solution of (9), subject to boundary conditions (10), for each individual derivative security.[39]

Whether or not the derivative security is path-dependent, the intermediary requires only two traded assets, the risky asset and the riskless security, to synthesize the derivative security. It thus needs only two traded assets to create any one of the uncountable number of pure Arrow-Debreu securities. Hence, we see that a market structure that permits frequent trading can substantially reduce the number of different security markets required to service a given set of allocations.[40]

4. Risk Management for Financial Intermediaries

Financial intermediaries, like other firms, issue stock and other liability instruments to investors in order to raise capital for operations. However, intermediaries are different from other business firms because they create explicit liabilities whenever they sell their products. Although intermediaries do act as agents in some transactions, their primary function is to act as principals and provide financial instruments and products that cannot be efficiently supported by trading in organized secondary markets.[41] The purchasers of these products are therefore *de facto* liabilityholders of the intermediary. Indeed, as we all know, the vast bulk of a typical intermediary's liabilities are held by its customers.

In general, customers, unlike investors in the firm, prefer to have the payoffs on their contracts as insensitive as possible to the fortunes of the firm itself. For example, a customer who buys a warranty on a new car from an automobile manufacturer wants the repairs paid for in the event that the car is defective. In fact, the customer's contract pays for repairs in the joint contingency that the car is defective *and* the automobile manufacturer is financially solvent. Even if an actuarily-fair reduction in the price of the warranty were made to reflect the risk of insolvency, it is still likely that the customer would prefer the warranty with the least default risk.[42] As with customers of automobile manufacturers, so with

---

[39] Breeden and Litzenberger (1978) demonstrate the direct connection between the pricing theories of options and pure securities. They show that a pure security is achieved as the limit of a "butterfly" spread in call options as the exercise prices of the options converge to each other. Also see Merton (forthcoming, Chapter 14).

[40] Radner (1972) was the first to formally show that dynamic trading could in principle reduce the number of different securities required to achieve an Arrow-Debreu complete-markets equilibrium allocation. Duffie and Huang (1985) provide a general proof that a Radner-type equilibrium can be achieved with a finite number of securities in a continuous-trading environment with continuous sample-path processes.

[41] In effect, the test is whether the bid-ask spread charged by the intermediary is smaller than the one that market makers would charge the customer if the instrument were traded in an organized market. A stereotypical financial intermediary purchases and issues nontraded financial instruments to its customers, and uses the financial markets to raise capital and to hedge its positions. As in the case of mutual funds, the intermediary may also create "pools" of traded assets for its customers if their individual demands for these assets are not an economic lot size for direct market purchases and sales.

[42] Further discussion of this assertion is provided in Section 5. Much the same point can be made with respect to the viability of organized futures and forward markets as discussed in Merton (forthcoming, Chapter 10). Regulations on capital requirements, collateral, guarantees and insurance purchased from external sources, and escrows of the intermediary's assets are among the devices used to insulate customers from the business risk of the intermediary.

customers of financial intermediaries. Thus, the success of an intermediary depends not only on charging adequate prices to cover its production costs, but also on providing adequate assurances to its customers that promised payments will be made. Hence, even in theory, an important part of the management of financial intermediaries is the measurement and control of the risk exposures created by issuing their financial products.

As we have seen, Contingent Claims Analysis can be used to determine the production process and cost for an intermediary to create virtually any financial product that has the properties of a derivative security. To match any specific payoff function, the intermediary need only set down the appropriate boundary conditions in (10), solve (9), and follow the prescribed rules. Therefore, it is in principle no more difficult to create derivative securities with specialized payoff patterns than it is to create ones with standard patterns (e. g., call options). CCA thus provides the means for intermediaries to create custom financial products in an "assembly-line" fashion. In such an environment, intermediaries will have a diverse array of complex financial assets and liabilities. This diversity and complexity makes measurement of the risk exposure for intermediaries a difficult task, even in a theoretical model. CCA can be used to significantly reduce the difficulty of that task.

Consider an intermediary that issues derivative securities on n different traded assets. Let V_k denote the price of asset k with dynamics specified as:

(17) $$dV_k = [\alpha_k V_k - D_j^k]dt + \sigma_k V_k dz_k ,$$

where α_k and σ_k are the expected rate of return and standard deviation of return, respectively and D_j^k is the payout rate on asset k, $k = 1,\ldots,n$. For $i = 1,\ldots,m_k$ and $k = 1,\ldots,n$, let $F^{ki}[V,t]$ denote the solution to (9) subject to boundary conditions (10) that are appropriate for the *ith type* of derivative security, which is contingent on the price of traded asset k. If $V_k(t) = V_k$ at time t, then $F^{ki}[V_k,t]$ is the unit production cost to the intermediary and it is also the value to the intermediary of owning one unit of derivative security i. From Itô's Lemma and (9), we have that the dollar return to the intermediary from owning one unit of the derivative security between t and $t + dt$ can be written as:[43]

(18) $$dF^{ki}+D_2^{ki}dt = (F_1^{ki}[V_k,t](\alpha_k-r)V_k+rF^{ki}[V_k,t])dt + F_1^{ki}[V_k,t]\sigma_k V_k dz_k$$
$$= F_1^{ki}[V_k,t](dV_k+D_j^k dt) + (F^{ki}[V_k,t]-F_1^{ki}[V_k,t]V_k)rdt ,$$

where D_2^{ki} is the payout rate on derivative security i and subscripts on F denote partial derivatives as in (9). By inspection of (18), owning one unit of derivative security i between t and $t + dt$ produces the same dollar return as holding $F_1^{ki}[V_k,t]$ units of traded asset k and $(F^{ki}[V_k,t] - F_1^{ki}[V_k,t]V_k)$ dollars of the riskless security over the same period.

If at time t the intermediary owns M_{ki} units of this derivative security,[44] then its exposure from this position between t and $t + dt$ is equivalent to owning N_{ki}^{ℓ} units of the traded asset k and B_{ki}^{ℓ} dollars of the riskless security where $N_{ki}^{\ell} \equiv M_{ki}F_1^{ki}[V_k,t]$ and $B_{ki}^{\ell} \equiv M_{ki}(F^{ki}[V_k,t] - F_1^{ki}[V_k,t]V_k)$. Adding up the exposures from all types of derivative securities (contingent on the price of asset k), we have that the total exposure from these positions between t and $t + dt$ is equivalent to owning $N_k^{\ell}(\equiv \sum_1^{m_k}N_{ki})$ units of asset k and $B_k^{\ell}(\equiv \sum_1^{m_k}B_{ki})$ dollars in the riskless security. Thus, the intermediary can express its entire risk exposure on the m_k

---

[43] By Itô's Lemma, $dF^{ki} = (F_2^{ki} + \sigma_k^2 V_k^2 F_{11}^{ki}/2)dt + F_1^{ki} dV_k$. By substituting for $(F_2^{ki} + \sigma_k^2 v_k^2 F_{11}^{ki}/2)$ from (9) and rearranging terms, we arrive at (18).

[44] If $M_{ki} < 0$, then the intermediary has a short position (i. e., a liability) of $| M_{ki} |$ units of the derivative security.

different types of derivative securities in terms of the risk exposure to just two traded securities: asset k and the riskless security.[45]

The intermediary can, of course, make this conversion for all derivative securities in its portfolio. If $W(t)$ denotes the net value of the financial-security holdings of the intermediary at time t, then we have that:

(19) $\qquad W(t) = \sum_1^n \sum_1^{m_k} M_{ki}(t)F^{ki}[V_k,t] + \sum_1^n N_k(t)V_k + B(t) \,,$

where $B(t)$ is the dollar amount of the riskless security held at t and $N_k(t)$ is the number of units of traded asset k held at t, $k = 1,\ldots,n$. Thus, from (19), CCA provides the means for continuous valuation of the intermediary's portfolio, even if the bulk of its financial products and liabilities are not traded in organized secondary markets.

From (17), (18) and (19), the dollar return to the intermediary's portfolio between t and $t + dt$ can be written as:

(20) $\qquad dW = \sum_1^n \sum_1^{m_k} M_{ki} \, (dF^{ki} + D_2^{ki}dt) + \sum_1^n N_k \, (dV_k + D_k^s dt) \; + rBdt$

$\qquad\qquad = \sum_1^n \, (N_k^e + N_k)(dV_k + D_k^s dt) + r[B + \sum_1^n B_k^e] dt$

$\qquad\qquad = [\sum_1^n \, (N_k^e + N_k)V_k \alpha_k + (B + \sum_1^n B_k^e)r]dt + \sum_1^n \, (N_k^e + N_k) \, \sigma_k V_k \, dz_k.$

By inspection of (20), we have that the intermediary's risk exposures to the $\sum_1^n m_k$ different derivative securities and the n traded assets can be expressed in terms of risk exposures to the n traded assets alone. Hence, by the application of CCA, risk management of the intermediary's complex portfolio can be reduced to the management of a relatively-simple portfolio containing only traded assets.

The analysis leading to (20) assumes no transactions costs. It can nevertheless be used to identify management policies that reduce the volume of hedging transactions and thereby, reduce the impact of such costs on financial intermediation. The derivation of bid-ask price spreads in Section 2 overstates the effect of transactions costs on an active financial intermediary, because it assumes no inventories of either traded assets, or other derivative securities. To perfectly hedge its entire portfolio, the intermediary need only hold inventories of traded assets that are sufficient to hedge the *net* exposures created by its financial products. In the context of (20), a perfect hedge requires only that $N_k = -N_k^e$, $k = 1,\ldots,n$.

To further illustrate the point, consider an intermediary with current exposures that are completely hedged. Now, suppose it issues call options on asset k with an equivalent exposure of N_{ki} and issues put options on asset k with an equivalent exposure of N_{kj}. Because for these instruments, $N_{ki} < 0$ and $N_{kj} > 0$, the dollar transaction in asset k required to jointly hedge these new exposures, $| (N_{ki} + N_{kj}) | V_k$, is less than $(|N_{ki}| + |N_{kj}|)V_k$, the dollar amount required if each new exposure is treated as an isolated transaction. Thus, by offering a mix of "bullish" and "bearish" products, the intermediary can reduce the volume of hedging transactions in traded assets. If necessary, the intermediary can adjust its bid and ask prices for selected financial products so as to induce its customers to buy ones that reduce the overall need for hedging transactions.

---

[45] In practice, restating the risk exposures in these terms may be beneficial, even if $m_k = 1$. For example, many managers may have difficulty in assessing the exposure of a short position of a six-month call option on 1000 shares of stock, with an exercise price of \$ 110 per share, when the stock is currently selling for \$ 100. Assuming that $F = \$ 8$ and $F_1 = 0.40$, they would perhaps understand better the equivalent risk-exposure statement, that they are short 400 shares of stock and have \$ 32,000 in the riskless security.

Even in the absence of new business, transactions in traded assets are required if the intermediary is to maintain a hedged position against its current inventory of financial products. If the policy is to maintain $N_k = -N_k^c$, then by Itô's Lemma, we have that:

$$(21) \qquad dN_k = -[\,\frac{1}{2}\sigma_k^2 V_k^2 \frac{\partial^2 N_k^c}{\partial V_k^2} + \frac{\partial N_k^c}{\partial t}\,]\,dt - \frac{\partial N_k^c}{\partial V_k}\,dV_k\,.$$

By inspection of (21), the volume of transactions induced by unexpected changes in traded-asset prices is proportional to the absolute magnitude of $\partial N_k^c/\partial V_k$. Hence, the size of transactions caused by unanticipated changes in traded-asset prices can be minimized if the intermediary can maintain a hedged portfolio with $\partial N_k^c/\partial V_k$ close to zero. If we neglect the effect of new business (i.e., $dM_{ki} = 0$, $i = 1,\ldots,m_k$, $k = 1,\ldots,n$), then from the definition of N_k^c, we have that:

$$(22) \qquad \frac{\partial N_k^c}{\partial V_k} = \sum_1^{m_k} M_{ki} F_{11}^{ki}[V_k,t]\,.$$

If the production cost for a derivative security of type i is a strictly convex function of V_k, then $F_{11}^{ki}[V_k,t] > 0$. If it is a strictly concave function, then $F_{11}^{ki}[V_k,t] < 0$. Thus, by offering a mix of "convex" and "concave" products, the intermediary can reduce the volume of transactions in traded assets that is required to maintain a hedged portfolio.

In the parlance of option-pricing analysis,[46] the hedge ratio, $N_{ki} = F_1^{ki}[V_k,t]$, is called the "delta" of the derivative security i and the change in the hedge ratio, $\partial N_{ki}/\partial V_k = F_{11}^{ki}[V_k,t]$, is called the "gamma" of derivative security i. The gamma of the intermediary's aggregate position, $\partial N_k^c/\partial V_k$, characterizes the degree of local convexity or concavity of the position with respect to V_k. If $N_k = -N_k^c$, then the intermediary is perfectly hedged against any small ("local") changes in V_k. If, however, this locally-hedged position is such that $\partial N_k^c/\partial V_k > 0$, then the intermediary will gain from any large ("nonlocal") changes in V_k, whether up or down.[47] Similarly, if $\partial N_k^c/\partial V_k < 0$, the intermediary will lose from any large change in V_k. Thus, for $N_k = -N_k^c$, the gamma of the aggregate position measures the intermediary's exposure to large moves in V_k. Our assumptions about the dynamics of traded-asset prices are such that nonlocal movements cannot occur. Such assurances cannot, of course, be given for the behavior of real-world asset prices. Hence, the assessment of one's exposure to convexity is an important element of risk management for practitioners.[48] Our analysis also neglects taxes. However, CCA analysis and the resulting prescriptions for risk management by intermediaries can be modified to include taxes along the lines of Scholes (1976) and Constantinides and Scholes (1980).

---

[46] Cf. Cox and Rubinstein (1985, pp. 222-235).

[47] If $H[V_k,t]$ denotes the value of the intermediary's position, then by Taylor's Theorem, $H[V_k + X,t] - H[V_k,t] = H_1[V_k,t]X + H_{11}[V,t]X^2/2$, where V is some number satisfying $|\,V - V_k\,| \le |\,X\,|$. If $N_k = -N_k^c$, then $H_1[V_k,t] = 0$. Hence, the change in the value is positive (negative), if $H_{11}[V,t] > 0$ (< 0) within the region of change.

[48] The risklessness of the derived hedges is, of course, contingent on a correct specification of the asset dynamics. As in Cox and Ross (1976) and Merton (1976), CCA techniques can be adopted to formally recognize the possibility of nonlocal changes (i.e., "jumps" or "gaps") in asset prices. As is often the case, real-world applications of CCA are a compromise: practitioners typically assume the model presented here and then using (22), adjust their holdings to create the desired "gamma" exposure to nonlocal movements in prices.

As noted at the outset, risk management is important in large part because customers are better served by products from intermediaries that have negligible default risk. However, especially (but not exclusively) because of transactions costs, perfect hedging of an intermediary's entire portfolio can be a suboptimal policy. In such cases, diversification can be used as a risk-management alternative to transacting in the traded assets. From (20), the instantaneous variance of the dollar return on an intermediary's portfolio can be expressed as $\sum_i^n \sum_j^n (N_k^e + N_k)(N_j^e + N_j)\sigma_k\sigma_j\rho_{kj}V_kV_j$, where ρ_{kj} is the instantaneous correlation coefficient between returns on asset k and asset j. Diversity among the products offered and diversity among the underlying traded assets are essentially the two ways by which diversification can reduce the variance of the portfolio. As we have seen, a diverse product mix can decrease the absolute magnitude of the net exposure to each traded asset. Among traded assets with positively correlated returns $(\rho_{kj} > 0)$, a product mix that leads to $N_k^e N_j^e < 0$ will also reduce the volatility of the intermediary's portfolio. For product mixes that lead to $N_k^e N_j^e > 0$, the intermediary can reduce portfolio variance by offering these products on traded assets with relatively small correlations among their returns (i.e., $\rho_{kj} \ll 1$).

Regulation can be beneficial to both financial intermediaries and their customers if it can reduce expenditures by customers on information gathering and monitoring. Regulation, whether beneficial or not, can significantly influence risk-management policy, even for an intermediary that can transact without cost. One common form of regulation is to require the intermediary to maintain a minimum level of net worth or capital if it is to continue in business.[49] If the intermediation franchise has a positive net present value, then regulations of this type can cause value-maximizing financial intermediaries to exhibit risk-averse-like behaviour in the management of their financial assets. In choosing among financial assets and liabilities with the same net present values, an intermediary will prefer the mix that minimizes the probability of violating the regulatory constraint, because that choice maximizes the value of its franchise.[50]

Given the "cliff-edge" nature of the regulatory constraint, one risk-management strategy that meets this objective is to systematically maintain a put-option position on the entire portfolio of net financial assets. By selecting an exercise price larger than the minimum net-worth requirement and a time until expiration of sufficient duration, the intermediary can protect itself against violation of the regulation and provide adequate time to raise any necessary additional capital. Although occasionally referred to as an *insured-equity*, or *protective-put* strategy, this risk-management strategy is most commonly called *portfolio insurance*.[51]

To implement this strategy, the intermediary could purchase insurance from another firm. However, as a minimum-cost transactor, it may find it more cost-effective to synthetically create its own portfolio insurance by using CCA techniques. By pursuing this route, the intermediary, in effect, creates a captive, portfolio-insurance subsidiary that sells the

---

[49] Such provisions are common in the regulation of insurance companies.

[50] See, for example, Merton (1978) on the investment behavior of banks covered by deposit insurance.

[51] Cf. Leland (1980), Brennan and Solanki (1981), and Rubinstein (1985). There are a number of variations of the basic portfolio-insurance idea. For instance, having the exercise price of the put option grow at a specified rate over time can guarantee a minimum positive rate of return on the portfolio.

put to the parent and undertakes dynamic strategies to hedge its liability. One advantage of this approach is that the intermediary retains any mark-up or spread that would otherwise be paid to an external issuer of the insurance. This saving includes any non-competitive profit to the issuer and the "deadweight" costs of redundant information gathering and surveillance that often arise in arms-length transactions. An external insurer of the intermediary's portfolio would surely place restrictions on the types of assets and liabilities held in the portfolio. Thus, self-insuring also allows greater flexibility in the intermediary's investment policy. The main disadvantage is, of course, that the self-insurer retains the risk that the dynamic hedging strategies fail to replicate the payoffs to a put option on the portfolio.[52] But, given their presumed financial expertise and continuous involvement in the financial markets, intermediaries are, almost surely, the best-suited among individuals and institutions to bear this risk.

5. On the Role of Efficient Financial Intermediation in the Continuous-Time Model

As is evident from the preceding analyses, the continuous-time model provides a rich analytical framework for developing a theory of financial intermediation that encompasses both the general functional structure and the specific operational procedures of financial intermediaries. Perhaps less apparent is that the existence of a well-functioning financial-intermediation sector in the economy can do much to justify the continuous-time model with its frictionless-market assumptions as a relevant mode of analysis for the study of general financial economic behaviour. We thus close our formal analysis with an investigation of this more-subtle side of the symbiotic relationship between the continuous-time model and financial intermediation.

The optimal consumption and investment rules for individual investors derived in Breeden (1979), Cox, Ingersoll and Ross (1985), Merton (1971; 1973b) and Solnik (1974) are valid only if investors can trade continuously with no transactions costs. Real-world financial markets in well-developed economies are open virtually all the time, and therefore, the assumption of continuous trading is a reasonable approximation to real-world trading opportunities. However, it is also readily apparent that in the real world, individual investors generally face significant transactions costs for trading in the asset markets. This observation surely raises a question about the robustness of the continuous-time model's prescriptions as an approximation to feasible behavior in the real world. In this section, we show that with efficient financial intermediation, all investors can achieve the identical consumption-bequest allocations that they would have chosen if they could have traded continuously without cost. We also show that the aggregate demands for traded assets are the same as the ones that obtain in the frictionless-market model. Thus, the consumption

---

[52] Cf. the discussion in footnote 48. In recent years, several real-world institutional investors have adopted portfolio insurance as a risk-management strategy for both their equity and fixed-income portfolios. Although some purchased the insurance from intermediaries, the great majority chose to create their own by trading in stock- and bond-index futures and traded-options on these futures. The skills of both the insurers and the self-insurers in implementing these strategies were given an extreme test on Black Monday, October 19, 1987, when the Dow Jones Industrial Average fell 508 points (22.6%). By any standard, this unprecedented decline represented a "nonlocal" movement in traded-asset prices. The facts on how they did are not as yet known [cf. Rubinstein (1988)]. It is, however, a safe prediction that those with a positive gamma on their otherwise hedged portfolios did well, and those with a negative gamma did poorly.

and asset demands derived in these cited earlier models are shown to apply in a more-realistic one with transactions costs.

In Merton (1977), the pricing function for a derivative security was derived by using the techniques of dynamic portfolio theory, developed in Merton (1971). Let $C(t)$ denote the consumption rate of the investor and let $w(t)$ be the fraction of the investor's portfolio allocated to the risky traded asset. If $W(t)$ denotes the value of the investor's portfolio at time t, then from Merton (1977, eqn. (4)), we have that:

$$(23) \qquad dW = \{[w(\alpha - r) + r]W - C\}\, dt + wW\sigma dz\,,$$

where α and σ are assumed to be functions of V and t only. To derive the replicating portfolio for a particular derivative security requires that we choose: (i) $w(t) = F_1[V(t),t]V(t)/W(t)$; (ii) $C(t) = D_2(V(t),t)$; and (iii) $W(0) = F[V(0),0]$. It follows that to rule out arbitrage, $W(t) = F[V(t),t]$ for all t, where $F[V,t]$ is the solution to the partial differential equation (9).

In this section, we "reverse" this process. Instead of searching for a feasible portfolio that exactly replicates the payoffs to a particular derivative security, we search for a feasible derivative security with payoffs that exactly match the payoffs to a particular investor's optimal portfolio. Suppose that a blueprint for constructing such a security can be found and that the price charged by a financial intermediary for the security does not exceed the investor's budget constraint. Then, as an alternative to continuously trading assets to achieve his optimal consumption-bequest allocation, the investor can buy the appropriate derivative security from the intermediary and achieve his optimal allocation without ever trading again.[53] If information-acquisition and transactions costs are smaller for the intermediary than for the investor, then the availability of such derivative securities makes it possible for the investor to achieve welfare-improving allocations that would not otherwise be feasible if the investor had to trade directly in the asset markets.

To derive the terms of the derivative security that provides the optimal consumption-bequest allocation for a particular investor, we use the Cox-Huang (forthcoming) approach to the optimal consumption-investment problem. Although the explicit development here assumes a single risky asset and a riskless asset, the results can be generalized to the case of n assets.[54]

Let $X(t)$ denote the price per share of a mutual fund with an investment policy to follow the growth-optimum portfolio strategy and pay no dividends to its shareholders.[55] The dynamics of $X(t)$ for a single risky asset can be written as:

---

[53] That investors need only transact once to achieve their lifetime optimal allocation is a well-known feature of the intertemporal version of the Arrow-Debreu model with complete markets. That this result obtains here is no coincidence because the proposed set of derivative securities is equivalent to a complete market. See Merton (forthcoming, Chapter 16) for analysis and discussion.

[54] Cf. Merton (forthcoming, Chapter 16).

[55] The growth-optimum portfolio maximizes the expected logarithmic rate of growth of the portfolio. In the continuous-time model, this portfolio is instantaneously mean-variance efficient [cf. Merton (forthcoming, Chapter 6)]. The mutual-fund theorems of the continuous-time theory predict that investment companies with growth-optimum portfolio strategies are natural financial products to be offered by intermediaries. In the single risky-asset case, an unambiguous statement of the investment policy for such funds is: to undertake continuous trading so as to hold the portfolio fraction $w(V(t),t)$ in the risky asset, where $w(V(t),t) \equiv (\alpha(V(t),t) - r)/\sigma^2(V(t),t)$. If the funds are "open-ended" and thereby, required to either redeem or issue shares at net asset value, then the no-dividend policy of the funds does not matter to their shareholders.

(24) $dX = (\mu^2 + r)Xdt + \mu Xdz$,

where $\mu = \mu(V,t) \equiv (\alpha - r)/\sigma$. If $C^*(t)$ denotes the investor's optimal consumption rate at time t and $W^*(T)$ denotes his optimal bequest of wealth, then Cox and Huang (forthcoming) have shown that there exist functions G and H such that:

(25a) $C^*(t) = G(X(t),V(t),t)$

and

(25b) $W^*(T) = H(X(T),V(T))$,

where G and H depend parametrically on $X(0)$, $V(0)$, and the investor's initial wealth, $W(0)$. Moreover, for any feasible consumption plan, G and H are nonnegative, and $G(0,V,t) = G(X,0,t) = H(0,V) = H(X,0) = 0$ for all $t \leq T$.

It follows from (25a) and (25b), that to match the optimal consumption-bequest allocation of the investor, the derivative security must pay out at a continuous dividend rate of $G(X(t),V(t),t)$ for $t < T$ and make a lump-sum payment of $H(X(T),V(T))$ at time T. The derivation of the production technology and production cost to create this security follows along the general lines presented in Merton (1977). However, the development here is somewhat more complicated, because the value of the derivative security depends on the prices of two risky traded assets in addition to the riskless one.

Define $F[X,V,t]$ to be the solution to the linear partial differential equation:

(26) $\dfrac{1}{2} [\mu^2 X^2 F_{11} + 2\mu\sigma XVF_{12} + \sigma^2 V^2 F_{22}] + rXF_1 + rVF_2 + F_3 - rF + G(X,V,t) = 0$,

subject to the boundary conditions:

(27a) $F[0,V,t] = 0$

(27b) $F[X,0,t] = 0$

(27c) $F[X,V,T] = H(X,V)$,

where subscripts on F in (26) denote partial derivatives with respect to X, V, and t.

Let $P(t)$ denote the value of a portfolio that makes continuous payouts at the rate $G(X(t),V(t),t)$ and follows an investment strategy of allocating $\$w_1(t)P(t)$ to the growth-optimum portfolio, $\$w_2(t)P(t)$ to the risky asset, and the balance of the portfolio to the riskless asset, where:[56]

(28a) $w_1(t)P(t) = F_1[X(t),V(t),t]X(t)$

and

(28b) $w_2(t)P(t) = F_2[X(t),V(t),t]V(t)$.

It follows that the dynamics of $P(t)$ can be written as:

(29) $dP = F_1 dX + F_2 dV + \{r(P - F_1 X - F_2 V) - G(X,V,t)\} dt$.

---

[56] In the single risky-asset case here, the growth-optimum fund is just a prescribed mix of the traded risky asset and the riskless asset. Therefore, as is evident from (24), the instantaneous return on the fund is perfectly correlated with the return on the risky asset. Thus, in this case, we can rewrite (28a) and (28b) in terms of a single combined position in the risky asset: namely, hold $\$[w(V(t),t)w_1(t)P(t) + w_2(t)P(t)]$ in the risky asset (and the balance of the portfolio in the riskless asset) where $w(V(t),t)$ is defined in footnote 55.

Because F is twice-continuously differentiable, we can use Itô's Lemma to express the stochastic process for $F[X(t),V(t),t]$ as:

(30) $$dF = F_1dX + F_2dV + \{\frac{1}{2}\ [\mu^2X^2F_{11} + 2\mu\sigma XVF_{12} + \sigma^2V^2F_{22}] + F_3\}\ dt\ .$$

But, F satisfies (26). Hence, we can rewrite (30) as:

(31) $$dF = F_1dX + F_2dV + \{r(F - F_1X - F_2V) - G(X,V,t)\}\ dt\ .$$

Let $Q(t) \equiv P(t) - F[X(t),V(t),t]$. From (29) and (31), we have that:

(32) $$\begin{aligned}dQ &= dP - dF\\ &= r(P - F)dt\\ &= rQdt\ .\end{aligned}$$

The solution to (32) for any $t \le T$ is $Q(t) = Q(0)exp[rt]$. Therefore, if we choose the initial investment in the portfolio so that $P(0) = F[X(0),V(0),0]$, then $Q(0) = 0$ and $Q(t) \equiv 0$ for all $t \le T$. It follows that for all $t \le T$:

(33) $$P(t) = F[X(t),V(t),t]\ .$$

Thus, we have derived a dynamic portfolio strategy that exactly replicates the payoffs of a derivative security that makes continuous payments at the rate $G(X(t),V(t),t)$ for $t < T$ and has a final payout of $H(X(T),V(T))$ at time T. The production technology for creating this security is given by (28a) and (28b). The cost of producing the security at time t is $F[X(t),V(t),t]$.

The payoffs of the derivative security exactly match the consumption-bequest allocation of the investor's optimal continuous-trading strategy. Hence, the investor can achieve this allocation by simply buying the security, provided that the price of the security does not exceed his budget constraint. But, Cox and Huang (forthcoming) have shown that the investor's optimally-invested wealth, $W(t)$, can be expressed as a function of $X(t)$, $V(t)$, and t and that this function satisfies our partial differential equation (26), with boundary conditions (27). Therefore, $W(t) = F[X(t),V(t),t]$, and in particular, $W(0) = F[X(0),V(0),0]$. Purchase of the derivative security by the investor is thus always feasible.

Having both derived the procedure for creating this type of security and established the feasibility of its purchase by the investor, we now examine the role of financial intermediation in making the continuous-time model more robust. Assume an institutional environment with a competitive financial-services industry, in which financial intermediaries pay no transactions costs and consider an investor who must pay significant transactions costs to trade in the asset markets. Let the investor solve for the optimal lifetime consumption-and-investment program that he would choose if he could trade continuously with no transactions costs. Using the technique of Cox and Huang (forthcoming), this optimal program can be expressed as a stream of contingent payments, $G(X(t),V(t),t)$ for $t < T$ and a final payment of $H(X(T),V(T))$ at time T. By assumption, the investor cannot implement this strategy by trading directly in the asset markets. The investor can, however, negotiate with an intermediary for the purchase of a custom-designed derivative security with payoffs

250

specified by G and H. By assumption, financial intermediation is a competitive industry,[57] and therefore, price equals the marginal cost, $F[X(0),V(0),0]$. Thus, if such derivative securities are generally available from intermediaries, all investors can achieve optimal consumption-bequest allocations that are identical to the ones they would have selected if they could trade continuously with no transactions costs.[58]

If the creation of these securities is to provide a meaningful extension to the robustness of the frictionless-market version of the continuous-time model, the assumptions about institutional structure and the partitioning of information sets among agents that are necessary for wide-spread distribution of these securities must be plausible. It is, for example, unlikely that trading in derivative securities of this type could be supported in organized secondary markets.[59] The wealth of any one investor is small and the particular security demanded by each investor will in general be unique in terms of its specified payout structure, G and H. The structure of these securities is however ideally suited for broad distribution by financial intermediaries. After the sale of such a security, an intermediary can use the proceeds to finance the hedging of its liability by trading continuously in the asset markets according to the rules (28a) and (28b). As in the preceding section, the nature of these trading rules are such that hedging of custom-designed derivative securities is no more difficult than hedging of standard ones. The intermediary simply aggregates its individual exposures, as calculated from (28), across all its outstanding securities and then establishes positions in a relatively small number of traded assets to hedge its net exposure.

Wide-spread availability of these derivative securities requires arms-length transactions, and these can only be implemented if the contracts create no important moral-hazard problem for either party. The contingent payments on each derivative security depend only on time and the prices of the growth-optimum mutual fund and the risky asset. All three are observable, and hence, verifiable, by the investor and the intermediary. Provided that both the investor and the intermediary are price-takers in the traded-asset markets, neither can affect the time path of either $X(t)$ or $V(t)$. Hence, there is no inherent moral hazard to either party in these contracts.

To establish and evaluate the terms of the derivative security, each party needs a specific body of information. Successful implementation of these contracts thus requires that the

---

[57] In the spirit of footnote 17, the purely competitive assumption for financial intermediaries can be somewhat relaxed without significantly affecting the results. That is, provided that the unit mark-up for non-competitive intermediary profits is a small fraction of each investor's wealth, the optimal allocation achieved through intermediation will be close to the competitive one because $C^*(t)$ is a continuously differentiable function of the investor's wealth.

[58] Cox and Huang (forthcoming) show that the investor's optimal-portfolio demand for the risky asset at time t is given by $\{F_1[X(t),V(t),t](\alpha - r)X(t)/\sigma^2 + F_2[X(t),V(t),t]V(t)\}$, with the balance of his wealth held in the riskless asset. If an intermediary uses hedging rules (28a) and (28b), it follows from footnote 56 that the intermediary's demands for traded assets to hedge the derivative security are identical to those the investor would select if he could trade continuously without cost.

[59] This does not imply that the security could not be redeemable. The duration of the derivative-security is the lifetime of the investor, T. The magnitude of the liability as assessed by the intermediary at any time $t(\leq T)$ is $F[X(t),V(t),t] = W(t)$. Hence, at any time t, a competitive intermediary would be willing to repurchase the security from the investor for $W(t)$. By such a sale, the investor can therefore recover the same wealth that he would have had if he had been able to trade continuously during the time interval $[0,t]$.

task of acquiring the needed information by the respective parties be a reasonable one. To create and price the derivative security, the intermediary need only know the schedules $G(X(t),V(t),t)$ and $H(X(T),V(T))$, which are specified by the investor at the time that the security is created. The intermediary need not know either the investor's preferences, or even that the purpose of the security is to provide an optimal lifetime allocation. Because $G \geq 0$ and $H \geq 0$, the only contractual payment by the investor to the intermediary is the initial one, $F[X(0),V(0),0]$. Thus, we have, as a *derived* result, and not as an assumption, that this type of derivative security provides limited liability for the investor. It follows that to price and hedge these securities, the intermediary does not require monitoring of the investor's endowment or credit-worthiness. We thus conclude that the informational requirements for intermediaries to produce these securities are reasonable.

The information set required by the investor to determine the functions $G(X,V,t)$ and $H(X,V)$ is surely reasonable, because it is no larger than the one he would use to solve for his optimal consumption and investment program in the absence of such intermediation opportunities. Moreover, this is the only information the investor needs, provided that there is no uncertainty about the ability of the issuing intermediary to make the promised contingent payments. The prospect of default on the customer-held liabilities of intermediaries raises two important issues. First, if there is significant default risk, then derivative securities with identical promised payments, but issued by different intermediaries, are no longer perfect substitutes. Hence, to evaluate these securities, the investor requires information about the credit-worthiness of the issuing intermediaries. Therefore, default risk imposes on the investor a burden of additional information acquisition, albeit perhaps not an unreasonable one.

A second and more serious issue is that default risk reduces the functional efficiency of the derivative securities sold to individual investors. These investors want securities with payments (contingent on X, V, and t) that match their optimal consumption-bequest allocations. Unless the risk of default on these contracts is negligible, their payments are, *de facto*, also contingent on the value of the issuing intermediary and its entire structure of outstanding liabilities. The suboptimality of the contracts induced by this unwanted dependency underscores the distinction drawn in Section 4 between the liabilities of the intermediary held by its customers and the liabilities held by its investors. In the model here, only zero-cost transactors would trade in the asset markets. Hence, these are the only investors to have direct holdings of the stock and other general liabilities of individual financial intermediaries. As shown in the proof of the Modigliani-Miller Theorem in Merton (1977), default risk on liabilities held by zero-cost transactors causes no loss in efficiency because such investors can, if they choose, use continuous-trading strategies in the asset markets to hedge this risk. In contrast, the customers of intermediaries are investors who buy the custom-designed financial products of the intermediaries precisely because they cannot transact costlessly in the asset markets. Unlike investors in intermediaries, customers cannot use trading strategies in the asset markets to shed their unwanted exposure to the fortunes of the intermediaries whose products they buy.

Customers can reduce the effects of bankruptcy risk by buying a security with promised payments of $G(X(t),V(t),t)/N$ and $H(X(T),V(T))/N$ at a price of $F[X(t),V(t),t]/N$ from each of N financial intermediaries. Such diversification efforts alone may not, however, be

sufficient to eliminate the adverse effects of default.[60] Because default risk reduces the usefulness of their financial products, intermediaries have an incentive to apply risk-management techniques, make institutional arrangements, and support regulations, all designed to minimize the probability of default on customer-held liabilities. This incentive would of course exist in any model of financial intermediation. What distinguishes the continuous-time model is that the theory provides a feasible set of trading rules for each intermediary to unilaterally reduce the risk of default on its own products to a negligible level.[61] Hence, there exists a feasible institutional structure and partitioning of agents' information sets to support the creation of optimal default-free contracts, such that each investor can buy his entire lifetime consumption-bequest allocation in a single transaction.

In Merton (1973b), an intertemporal equilibrium model of allocations and asset prices is developed for an economy with frictionless markets where all investors can trade continuously without cost. This model has a simple financial sector with no explicit derivative-security markets or financial intermediaries, and further assumes that all business firms are financed entirely by equity. However, as noted at the outset here, derivative securities, mutual funds, and corporate liabilities, with payoffs that can be replicated by continuous trading in existing assets, serve no important function in a frictionless environment in which all agents can trade continuously without cost. Because every investor can trade continuously, each investor can synthesize the payoff patterns of these financial instruments, using the existing set of traded assets alone. Hence, any allocation chosen by the investor after such instruments are added to the menu of investments could have been achieved before their introduction. Thus, the addition of these financial instruments into that model would leave unchanged the equilibrium time paths of consumption, production, and asset prices.

It follows that equilibrium in a frictionless-market model that explicitly includes all financial markets, intermediaries, and corporate-liability structures can be analyzed in two separate parts: in the first, the intertemporal model with a simplified financial sector is used to determine real-sector allocations and asset prices. In the second, the equilibrium dynamics of asset prices as determined in the first part, are used together with CCA pricing methods, to find the equilibrium prices of derivative securities, financial products, and corporate liabilities. Such a separation among sectors of the economy is generally called *dichotomy*. Dichotomy occurs here because many of the financial securities, markets, and institutions are redundant, and hence, much of the financial sector "doesn't matter".[62]

---

[60] The degree of success of this strategy in statistically eliminating the risk of default depends on the extent that the Law of Large Numbers applies. Valid application of this theorem requires not only that the number of intermediaries, N, is large, but also that the events of default among the selected intermediaries are sufficiently statistically independent. Because many economic events that affect the fortunes of one intermediary also affect the fortunes of others, the required independence assumption may not be satisfied.

[61] In the model here, these rules are given by (28a) and (28b). For the corresponding rules with n assets, see Merton (forthcoming, Chapter 16).

[62] The conditions under which the creation of a new financial instrument or market can change equilibrium allocations are discussed in Breeden (1979), Cox, Ingersoll and Ross (1985), and Merton (1973b; forthcoming). If a set of existing securities and markets leads to an equilibrium that is an unconstrained Pareto optimum, then any additional securities or markets will be redundant.

In the model of this section, the financial sector does matter, because not all agents can trade continuously without cost.[63] As we have seen, there exists an institutional structure of well-functioning and competitive financial intermediaries such that every investor can achieve the same consumption-bequest allocation that he or she would have selected in a frictionless-market environment. Although these investors need only transact once to buy their optimal allocation, the issuing intermediaries must transact continuously in the asset markets to hedge their customer-held liabilities. As discussed in footnote 58, the increments to a financial intermediary's hedging demands for traded assets that are induced by the sale of an optimal-allocation contract to an investor will, at every point in time after the sale, be identical to the optimal-portfolio demands that this investor would have chosen in a frictionless-market model. Hence, the individual consumption functions and aggregate demand functions for traded assets in this model can be computed *as if* all investors could trade continuously without cost. It follows that the equilibrium real-sector allocations and asset prices will be the same as the ones derived in the Breeden-Cox-Ingersoll-Ross-Merton-Solnik models.

It would seem therefore that the same dichotomy derived in the frictionless-market case applies also to our model with transactions costs. In a formal-computational sense, it does. We can still solve for the equilibrium asset-price dynamics in the frictionless-market model, and then compute the prices of derivative securities, mutual funds, and corporate liabilities using CCA. There is, however, one significant difference: the first-part calculation of equilibrium real-sector allocations is valid if and only if there exists a sufficiently rich set of securities and intermediaries in the financial sector to make these allocations feasible for agents who cannot trade without cost. Because of this conditioning, perhaps a more-appropriate term for this separation is "quasi-dichotomy".

In summary, our manifest hypothesis is that the continuous-time model is a useful tool of analysis for both the theory and practice of financial intermediation. With the assumption that not all agents can trade continuously without cost, a significant economic role is established for derivative-security markets and financial intermediaries in the model. Somewhat paradoxically, we find that if intermediation is efficient, then derivative securities and the financial products of intermediaries will be priced *as if* they are redundant.[64] Under these conditions, quasi-dichotomy obtains, and equilibrium real-sector allocations and asset prices can be derived from the frictionless-market version of the model. Thus, we arrive at a second – perhaps more-latent – hypothesis: as real-world financial intermediation becomes increasingly more efficient, the continuous-time model's predictions about actual financial prices, products and institutions will become increasingly more accurate. In short, reality will eventually imitate theory.

---

[63] The term "transactions costs" can, of course, mean more than just the bid-ask spread or commissions paid for buying and selling assets. For example, if an investor literally traded his portfolio continuously, he would have no time to do anything else! Because investment management and financial intermediation involve primarily information processing and security trading, there are significant economies to scale. One would thus expect substantially lower costs for the individual investor to use intermediaries instead of transacting directly in the asset markets.

[64]Thus, our model provides a resolution to "The Catch 22 of Option Pricing" discussed in footnote 14. Although the *equilibrium* prices of options and other derivative securities satisfy the redundancy condition of CCA, they are nevertheless needed to support the equilibrium allocations, and hence, their elimination would reduce social welfare.

6. Afterword

The focus of my remarks has been on the theory of intermediation. Allusions to practice were concentrated on the micro, quantitative applications of the continuous-time model to financial-product technologies and risk management. As I said at the outset, this was indeed my intent. Still, I cannot wholly resist the temptation to call attention to the model's potential for addressing broader issues in the practice of intermediation. And so an afterword that shall ever so briefly touch upon implications of the model for intermediation policy and strategy in the hope that it will lead others, better qualified, to explore this matter elsewhere.

Derivative securities have long been an integral part of the financial markets. Indeed, from Joseph de la Vega's 1688 treatise detailing the workings of the Amsterdam stock exchange, it appears that options and securities quite similar to modern financial futures contracts dominated trading activities in this leading financial center of the seventeenth-century world. But, most would mark the creation of the Chicago Board Options Exchange in April 1973 as the start of the current wave of financial innovations involving derivative securities. The seminal Black-Scholes theory of option pricing coincidentally appeared in print just a month later. The succeeding fifteen years have witnessed an unprecedented proliferation of organized trading markets in both equity and fixed-income derivative instruments. In turn, these markets made possible the creation of a wide range of financial products, many custom-designed to meet selected needs of investors and issuers. Concurrently, mainstream financial institutions increasingly adopted quantitative techniques, including computerized trading strategies, to help manage their portfolios, often on a global scale. These changes have been accompanied by an explosion of trading volume in just about every sector of the financial markets. Of course, all this everyone knows.

There are some in the financial and regulatory communities – I assume, by self-selection, none here – who see all this alleged innovation as nothing more than a giant fad, driven by institutional investors and issuers with wholly unrealistic expectations of greater returns with less risk, and fueled by financial-services firms and organized exchanges that see huge profits from this vast activity. Perhaps. There have surely been instances of financial products and trading strategies that have not delivered in practice the performance promised in theory. But, notwithstanding such examples, there is another, quite different interpretation of the events of the preceding decade.

From the perspective of our theory, these same facts about change are seen as consistent with a real-world dynamic path evolving toward an idealized target of an efficient financial-market and intermediation system. On this premise, these changes can be interpreted as part of a "financial-innovation spiral". That is, the proliferation of new trading markets makes feasible the creation of new financial products; to hedge these products, producers trade in these new markets and volume expands; increased volume reduces marginal transactions costs and thereby, makes possible further implementation of new products and trading strategies, which in turn leads to still more volume. Success of these trading markets encourages investment in creating additional markets, and so on it goes... spiralling toward the theoretically limiting case of zero marginal transactions costs and dynamically-complete markets.

Consider now a small sampling of the implications for strategy and policy from this view of the process. In this scenario, aggregate volume expands secularly and trading is

increasingly dominated by institutions. As more institutions employ dynamic strategies to hedge their product liabilities, incentives rise for expansion to round-the-clock trading that permits more-effective implementation of these strategies. Supported by a potentially limitless technology for creating financial products, financial-services firms will increasingly focus on providing individually-tailored solutions to their client's investment and financing problems.

Whether the financial-intermediation industry becomes more concentrated or more diffuse in this scenario is not clear. The central functions of information and transactions processing would favor economies of scale. Similarly, from the analysis in Section 4, the greater opportunities for netting and diversifying risk exposures by an intermediary with a diverse set of products would suggest a decline in hedging-transactions-per-dollar-of-product-liability as size increases. On the other hand, expansion in the types of organized trading markets, reductions in transactions costs, and continued improvements in information-processing and telecommunications technologies will all make it easier for a greater variety of firms to serve the functions of a financial intermediary. Continuing the scenario, existing intermediaries will be capable of offering a broader range of financial products and servicing a wider geographic area. Traditional institutional identifications with specific types of products are likely to become increasingly blurred. Geo-political advantages currently enjoyed by some financial institutions will be reduced.

Along that hypothesized path of development, the need to distribute a larger-volume and more-diverse set of products promises continued relative growth of the sales activity within financial intermediaries. The trading activity is also likely to expand to meet the execution requirements for implementing more-complex product technologies. As in other innovating industries, competition to create new products and to find new ways to produce established ones at lower costs could make the research-and-development activity the lifeblood of the financial intermediary. As we know, many but not all of these changes have been underway and further, this represents one scenario, growing out of our model of the intermediation process.

The dramatic increases over the past decade in the size and complexity of transactions together with the global-linking of financial markets have, however, raised concerns about macro credit risk and the possibility of broad financial-market "breakdown." The crash in world stock markets last October surely heightened those concerns. The changes in practice projected by the theory imply, *ipso facto*, further increases in the interdependence among institutions and markets in the financial system. Moreover, as discussed in Sections 4 and 5, minimizing the default risk for customer-held liabilities of financial intermediaries is a key element in our theory of efficient intermediation. As also noted, much the same point can be made about markets for standardized instruments, such as options and futures. Thus, the theory selectively concurs with the belief that credit risk is a major macro issue for financial markets.

Time, not source material, is the scarce resource for discussion of the credit-risk issue on this occasion. I shall therefore not even try to summarize the various detailed lines of inquiry that show promise for finding policy solutions that are congruent with evolution to an

256

efficient financial system.[65] Instead, I use my time to comment on a single guideline for policy that is central from the perspective of the theory.

The overriding theme of the theory has financial innovation as the engine driving the financial system on its prospective journey to efficiency and complete risk-sharing markets. With its focus on product innovation, this theory largely abstracts from the concurrent changes in financial infrastructure (including institutional and regulatory practices, organization of trading facilities, and the communication and information processing systems for transactions) required to support realization of this journey. But, perhaps the single most important implication for policy is the explicit recognition of the interdependence between product and infrastructure innovations and of the inevitable conflicts that arise between the two.

As an analogy of supreme simplicity, consider the creation of a high-speed passenger train, surely a beneficial product innovation. Suppose however, that the tracks of the current rail system are inadequate to handle such high speeds. In the absence of policy rules, the innovator, either through ignorance or a willingness to take risk, could choose to fully implement his product and run the train at high speed. If the train subsequently crashes, it is, of course, true that the innovator and his passengers will pay a dear price. But, if in the process the track is also destroyed, then those, such as freight operators, who use the system for a different purpose will also be greatly damaged. Hence, the need for policy to safeguard the system. A simple policy that fulfills that objective is to permanently fix a safe, but low speed limit. But, of course, this narrowly-focused policy has as a rather unfortunate consequence that the benefits of innovation will never be realized. An obviously better, if more-complex, policy solution is to facilitate the needed upgrading of the track and at the same time, to set transient limits on speed, while there is a technological imbalance between the product and its infrastructure.

As in this hypothetical rail system, the financial system is used by many for a variety of purposes. When treated atomistically, financial innovations in products and services can be implemented unilaterally and rather quickly. Hence, these innovations take place in an entrepreneurial and opportunistic manner. In contrast, innovations in financial infrastructure must be more coordinated and therefore, take longer to implement. It is thus wholly unrealistic to expect financial innovation to proceed along a balanced path of development for all elements in the system. It is indeed possible that at times, the imbalance between product innovation and infrastructure could become large enough to jeopardize the functioning of the system. Hence, the need for policy to protect against such breakdown. But, as we have seen, a single-minded policy focused exclusively on this concern could derail the engine of innovation and bring to a halt the financial system's trip to greater efficiency.

From abstract mathematics to concrete policy evaluations, this has surely been a song of unbridled praise for the continuous-time theory of finance. But, having opened discussion of that theory's relevance to the real world, due diligence requires that I not close it without some qualifying caveats.

---

[65] These range from finding feasible mechanisms to centralize (or at least coordinate) the clearing systems and global standards of collateral for all major markets, a line more rooted in the disciplines of political science and systems analysis than finance, to the risk management of macro "jumps" in asset prices, a subject very much in the mainstream of current finance research.

It is a dictum that financial markets will not long let stand prices that violate arbitrage. However, unlike with true arbitrage, the derived price relations in the continuous-time model are delicately conditional on the assumed stochastic processes for the underlying securities and on the opportunity to trade continuously. Hence, with only conditional arbitrage, there are no assurances in the real world of an inexorable and swift convergence of actual prices to their theoretical model values. Thus, the simple, but powerful argument of no violation of true arbitrage cannot be invoked to, *a priori*, validate the model.

Validation must therefore rely on traditional *a posteriori* assessments. As we shall hear much about empirical study of the model in the conference sessions to come, I simply note here that the cumulative statistical evidence appears encouraging, but also contains many anomalies. The clinical evaluations from more than a decade of widespread and ongoing application of the continuous-time model by practitioners seem to provide inferentially favorable evidence of a different sort. But, those who see the last decade of innovations as a fad are also likely to interpret this evidence as merely an instance of a self-fulfilling prophecy, serving only to confirm their hypothesis.[66] Even without accepting such a polar view, we must surely concede that practitioners, like academics, are not immune from following the paths of error. Thus, although the prognosis may be good, the continuous-time model as an empirical hypothesis, strictly speaking, remains unproved. In a word, as our Chinese friends might put it, these are interesting times for finance theory and for finance practice.

---

[66] That is, the proclamation that the model is widely used induces each practitioner to use it only because he believes that all others are and therefore, that prices will conform to the model, independently of its economic validity. As formulated in R. K. Merton (1948), the concept of the self-fulfilling prophecy applies only if the prophesized event would not have occurred in the absence of its public proclamation. Hence, even if widespread public knowledge of the model's adoption leads others to use it, there is no self-fulfilling prophecy if as our theory predicts, the model is economically valid without the proclamation.

BIBLIOGRAPHY

ADLER, M. and DUMAS, B. (1983), "International Portfolio Choice and Corporation Finance: A Synthesis", *Journal of Finance*, 38 (June): 925-984.

ARROW, K.J. (1953), "Le Rôle des Valeurs Boursières pour la Répartition la meilleure des Risques", *Econometrie*, Colloques Internationaux du Centre National de la Recherche Scientifique, Vol. XI, Paris, 41-47.

ARROW, K.J. (1964), "The Role of Securities in the Optimal Allocation of Risk Bearing", *Review of Economic Studies*, 31 (April): 91-96.

BACHELIER, L. (1900), "Théorie de la Spéculation", Annales de l'Ecole Normale Supérieure, 3, Paris: Gauthier-Villars.

BANZ, R.W. and MILLER, M.H. (1978), "Prices for State-Contingent Claims: Some Estimates and Applications", *Journal of Business*, 51 (October): 653-672.

BLACK, F. (1987), "This Week's Citation Classic", *Current Contents / Social and Behavioral Sciences*, August 17, 1987, Vol. 19, No. 33, Philadelphia: Institute for Scientific Information: p. 16.

BLACK, F. and SCHOLES, M. (1973), "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81 (May/June): 637-654.

BREEDEN, D.T. (1979), "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities", *Journal of Financial Economics*, 7 (September): 265-296.

BREEDEN, D.T. and LITZENBERGER, R. (1978), "Prices of State-Contingent Claims Implicit in Option Prices", *Journal of Business*, 51 (October): 621-651.

BRENNAN, M. and SCHWARTZ, E. (1976), "The Pricing of Equity-Linked Life Insurance Policies with an Asset Value Guarantee", *Journal of Financial Economics*, 3, (June): 195-214.

BRENNAN, M. and SOLANKI, R. (1981), "Optimal Portfolio Insurance", *Journal of Financial and Quantitative Analysis*, 16 (September): 279-300.

CONSTANTINIDES, G. (1986), "Capital Market Equilibrium With Transactions Costs", *Journal of Political Economy*, 94 (August): 842-862.

CONSTANTINIDES, G. and SCHOLES, M. (1980), "Optimal Liquidation of Assets in the Presence of Personal Taxes: Implications for Asset Pricing", *Journal of Finance*, 35 (May): 439-443.

COX, D.A. and MILLER, H.D. (1968), *The Theory of Stochastic Processes*, New York: John Wiley & Sons, Inc.

COX J.C. and HUANG, C. (forthcoming), "Optimum Consumption and Portfolio Policies When Asset Prices Follow a Diffusion Process", *Journal of Economic Theory*.

COX, J.C., INGERSOLL, J.E. and ROSS, S.A. (1985), "An Intertemporal General Equilibrium Model of Asset Prices", *Econometrica*, 53 (March): 363-384.

COX, J.C. and ROSS, S.A. (1976), "The Valuation of Options for Alternative Stochastic Processes", *Journal of Financial Economics*, 3 (January/March): 145-166.

COX, J.C., ROSS, S.A. and RUBINSTEIN, M (1979), "Option Pricing: A Simplified Approach", *Journal of Financial Economics*, 7 (September): 229-263.

COX, J.C. and RUBINSTEIN, M. (1985), *Options Markets*, Englewood Cliffs, New Jersey: Prentice Hall.

DEBREU, G. (1959), *Theory of Value*, New York: John Wiley & Sons, Inc.

DETTMAN, J.W. (1969), *Mathematical Methods in Physics and Engineering*, 2nd Edition, New York: McGraw-Hill Publishing Company.

DUFFIE, D. and HUANG, C. (1985), "Implementing Arrow-Debreu Equilibria by Continuous Trading of Few Long-Lived Securities", *Econometrica*, 53 (November): 1337-1356.

HAKANSSON, N.H. (1976), "The Purchasing Power Fund: A New Kind of Financial Intermediary", *Financial Analysts Journal*, 32 (November/December): 49-59.

HAKANSSON, N.H. (1979), "The Fantastic World of Finance: Progress and the Free Lunch", *Journal of Financial and Quantitative Analysis*, 14 (Proceedings Issue): 717-734.

KARLIN, S. and TAYLOR, H.M. (1981), *A Second Course in Stochastic Processes*, New York: Academic Press.

KRAUS, A. and ROSS, S.A. (1982), "The Determination of Fair Profits for the Property-Liability Insurance Firm", *Journal of Finance*, 37 (September): 1015-1028.

LELAND, H. (1980), "Who Should Buy Portfolio Insurance?", *Journal of Finance*, 35 (May): 581-594.

LELAND, H. (1985), "Option Pricing and Replication with Transactions Costs", *Journal of Finance*, 40 (December): 1283-1301.

MARKOWITZ, H. (1952), "Portfolio Selection", *Journal of Finance*, 7 (March): 77-91.

MARKOWITZ, H. (1959), *Portfolio Selection: Efficient Diversification of Investment*, New York: John Wiley & Sons, Inc.

MASON, S. and MERTON, R.C. (1985), "The Role of Contingent Claims Analysis in Corporate Finance", in E. Altmann and M. Subrahmanyan, eds., *Recent Advances in Corporate Finance*, Homewood, Illinois: Richard D. Irwin.

MERTON, R.C. (1969), "Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case", *Review of Economics and Statistics*, 51 (August): 247-257. Reprinted as Chapter 4 in Merton (forthcoming).

MERTON, R.C. (1971), "Optimum Consumption and Portfolio Rules in a Continuous-Time Model", *Journal of Economic Theory*, 3 (December): 373-413. Reprinted as Chapter 5 in Merton (forthcoming).

MERTON, R.C. (1973a), "Theory of Rational Option Pricing", *Bell Journal of Economics and Management Science*, 4 (Spring): 141-183. Reprinted as Chapter 8 in Merton (forthcoming).

MERTON, R.C. (1973b), "An Intertemporal Capital Asset Pricing Model", *Econometrica*, 41 (September): 867-887. Reprinted as Chapter 15 in Merton (forthcoming).

MERTON, R.C. (1976), "Option Pricing When Underlying Stock Returns are Discontinuous", *Journal of Financial Economics*, 3 (January/March): 125-144. Reprinted as Chapter 9 in Merton (forthcoming).

MERTON, R.C. (1977), "On the Pricing of Contingent Claims and the Modigliani-Miller Theorem", *Journal of Financial Economics*, 5 (November): 241-249. Reprinted as Chapter 13 in Merton (forthcoming).

MERTON, R.C. (1978), "On the Cost of Deposit Insurance When There are Surveillance Costs" *Journal of Business*, 51 (July): 439-452. Reprinted as Chapter 20 in Merton (forthcoming).

MERTON, R.C. (1982), "On the Microeconomic Theory of Investment Under Uncertainty", in K.J. Arrow and M. Intriligator, eds., *Handbook of Mathematical Economics*, Volume II, Amsterdam: North-Holland.

MERTON, R.C. (1983), "Financial Economics" in E.C. Brown and R.M. Solow, eds., *Paul Samuelson and Modern Economic Theory*, New York: McGraw-Hill Publishing Company.

MERTON, R.C. (forthcoming), *Continuous-Time Finance*, Oxford: Basil Blackwell, Inc.

MERTON, R.K. (1948), "The Self-Fulfilling Prophecy", *Antioch Review* (Summer): 193-210.

RADNER, R. (1972), "Existence of Plans, Prices, and Price Expectations in a Sequence of Markets", *Econometrica*, 40 (March): 289-303.

ROSS, S.A. (1976a),"Arbitrage Theory of Capital Asset Pricing", *Journal of Economic Theory*, 13 (December): 341-360.

ROSS, S.A. (1976b), "Options and Efficiency", *Quarterly Journal of Economics*, 90, (February): 75-89.

RUBINSTEIN, M. (1985), "Alternative Paths to Portfolio Insurance", *Financial Analysts Journal*, 41 (July/August): 45-52.

RUBINSTEIN, M. (1988), "Portfolio Insurance and the Market Crash of October 19, 1987", *Financial Analysts Journal*, 44 (January/February): 38-47.

SCHOLES, M. (1976), "Taxes and the Pricing of Options", *Journal of Finance*, 31 (May): 319-332.

SOLNIK, B.H. (1974), "An Equilibrium Model of the International Capital Market", *Journal of Economic Theory*, 8 (August): 500-524.

SUN, T. (1987), "Transactions Costs and Intervals in a Discrete-Continuous Time Setting for Consumption and Portfolio Choice", Chapter 1 in *Connections Between Discrete-Time and Continuous-Time Financial Models*, Ph.D. dissertation, Graduate School of Business, Stanford University, Stanford, California.

TOBIN, J. (1958), "Liquidity Preference as Behavior Towards Risk", *Review of Economic Studies*, 25 (February): 68-85.

VEGA, de la, J.P. (1688), *Confusion de Confusiones.* English translation by H. Kellenbenz (1957), Publication No. 13, The Kress Library Series of Publications, The Kress Library of Business and Economics, Harvard University, Cambridge, Massachusetts.