ANALYTICAL OPTIMAL CONTROL THEORY
AS APPLIED TO
STOCHASTIC AND NON-STOCHASTIC ECONOMICS
by
ROBERT COX MERTON
S., Columbia School of Engineering & Applied Science
(1966)
M.S., California Institute of Technology
(1967)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF
PHILOSOPHY
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September, 1970

Signature of Author ........................................

Department of Economics, August 28, 1970

Certified by ......

Thesis Supervisor

Accepted by ........................................

Chairman, Departmental Committee
on Graduate Students

Dewey

MASS. INST. TECH.

OCT 22 1970
LIBRARIES
ABSTRACT

ANALYTICAL OPTIMAL CONTROL THEORY AS APPLIED TO STOCHASTIC AND NON-STOCHASTIC ECONOMICS

Robert Merton

Submitted to the Department of Economics on August 28, 1970 in partial fulfillment of the requirement for the degree of Doctor of Philosophy.

The thesis consists of five self-contained essays that have in common that each is an economic's problem in intertemporal maximization. Chapter I examines the lifetime consumption-portfolio problem under uncertainty where the individual acts to maximize the expected value of an integral of utility of consumption over time. Itô's Lemma and the Fundamental Theorem of stochastic optimal control are introduced as the means for analyzing the problem. The advantages of working with the continuous-time version are discussed, and a number of theorems are proved. The emphasis of the paper is on efficient portfolios and the effects of uncertain time horizon and wage income. Non-stationary and non-Markov expectation mechanisms are introduced.

Chapter II studies the same problem as I, but with different emphasis. The stochastic Bellman equation is derived in a heuristic but informative fashion. Similarly, the use of limits of discrete formulations eliminates the need for the formal use of stochastic differential equations. The explicit dynamics of the optimal rules over the lifetime of the individual are examined in detail. The technique of comparative statics is used to study the effects of changes in risk, return, and risk-aversion on the consumption-saving decision. As a by-product, generalized intertemporal income and substitution effects are defined. The infinite horizon problem is studied in detail.

Chapter III considers the Ramsey-type social planning problem in a neoclassical growth model where the rate of population growth is endogeneous and is a function of per capita wealth or income. Under the Samuelson-Diamond criterion of maximizing social welfare, it is shown that the turnpike is no longer at the golden rule capital labor ratio. For the Bentham-Lerner criterion, it is shown that the origin of utility of the representative man is not arbitrary in the sense that the optimal program is not independent of this origin. It is further shown that the Schumpeter zero-interest level is never an optimal steady-state.

Chapter IV is a study to find the equilibrium price relationship between two perfectly correlated securities. The particular security examined here is a warrant. Based on expected utility maximization, explicit formulas are derived for the warrant price as a function of the common stock price and the warrant's maturity date.
In 1965, P.A. Samuelson presented a simple theory of warrant pricing. For a special class of warrants called perpetual warrants, the theory leads to an explicit solution for the warrant price. In chapter V, an econometric investigation of the theory is made. Because of the "tight" specification of the model, the emphasis was on statistical testing of the assumptions of the model and the model itself. The model was pitted against alternative theories in a forecasting "contest". The results were quite favorable for so closely a specified model.

Thesis supervisor: Paul A. Samuelson
Title: Institute Professor
ACKNOWLEDGEMENTS

I owe an irreducible debt to Paul A. Samuelson for introducing me to mathematical economics in his course where I learned the basic techniques used throughout this thesis; for introducing me to economic research as his assistant for two and one-half years which lead to my first published economics paper, a joint paper with him; for the countless hours of his time which he made available for discussion of further research which lead to this thesis and two more published papers; for financial assistance through his National Science Foundation grant which allowed me the maximum time to pursue this research. I gratefully dedicate this thesis to him.

I would like to thank Robert Solow and Franklin Fisher for helpful discussion and for serving on my thesis committee. Chapter V is based on work done for Fisher's Econometrics course. Peter Diamond read earlier drafts of Chapter I. David Scheffman was a faithful listener and helped in some of the mathematical proofs. Stanley Fischer provided useful suggestions for the work in Chapter II.

I thank M.I.T. for financial support under a NDEA fellowship and the National Science Foundation for both a fellowship and aid under grant GS-1812. Chapter II was originally published in the August, 1969, Review of Economics and Statistics. Chapter III was originally published in the December, 1969, Western Economics Journal. Chapter IV is a shortened version of a paper with the same title, written jointly with Paul A. Samuelson, and published in the Winter, 1969, Industrial Management Review.

Sue Friedman and Jacquelyn Tricomi did an excellent job of typing a hard manuscript under rush conditions.

I must thank my wife, June, for her patience and her support during the many evenings and weekends I spent doing this research over the last three years, and for so much more. Without her, none of this would have happened.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>OPTIMUM CONSUMPTION AND PORTFOLIO RULES IN A CONTINUOUS-TIME MODEL</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>LIFETIME PORTFOLIO SELECTION UNDER UNCERTAINTY: THE CONTINUOUS-TIME CASE</td>
<td>48</td>
</tr>
<tr>
<td>III</td>
<td>A GOLDEN RULE FOR WELFARE-MAXIMIZATION IN AN ECONOMY WITH A VARYING POPULATION GROWTH RATE</td>
<td>73</td>
</tr>
<tr>
<td>IV</td>
<td>A COMPLETE MODEL OF WARRANT PRICING THAT MAXIMIZES UTILITY</td>
<td>88</td>
</tr>
<tr>
<td>V</td>
<td>AN EMPIRICAL INVESTIGATION OF THE SAMUELSON RATIONAL WARRANT PRICING THEORY</td>
<td>112</td>
</tr>
</tbody>
</table>
Optimum Consumption and Portfolio Rules in a Continuous-time Model

Robert C. Merton
M.I.T. March, 1970

1. Introduction. A common hypothesis about the behavior of (limited liability) asset prices in perfect markets is the random walk of returns or (in its continuous-time form) the "geometric Brownian motion" hypothesis which implies that asset prices are stationary and log-normally distributed. A number of investigators of the behavior of stock and commodity prices have questioned the accuracy of the hypothesis. In particular, Cootner [2] and others have criticized the independent increments assumption, and Osborne [2] has examined the assumption of stationariness. Mandelbroit [2] and Fama [2] argue that stock and commodity price changes follow a stable-Paretian distribution with infinite second-moments. The non-academic literature on the stock market is also filled with theories of stock price patterns and trading rules to "beat the market", rules often called "technical analysis" or "charting", and that presupposes a departure from random price changes.

I would like to thank P.A. Samuelson, R.M. Solow, P.A. Diamond, J.A. Mirrlees, J.S. Flemming, and D.T. Scheffman for their helpful discussions. Of course, all errors are mine. Aid from the National Science Foundation is gratefully acknowledged.

In an earlier paper [12], I examined the continuous-time consumption-portfolio problem for an individual whose income is generated by capital gains on investments in assets with prices assumed to satisfy the "geometric Brownian motion" hypothesis. I.e. I studied Max E \int_0^T U(C(t))dt where U is the instantaneous utility function; C is consumption; E is the expectation operator. Under the additional assumption of a constant relative or constant absolute risk-aversion utility function, explicit solutions for the optimal consumption and portfolio rules were derived. The changes in these optimal rules with respect to shifts in various parameters such as expected return, interest rates, and risk were examined by the technique of comparative statics.

The present paper extends these results for more general utility functions, price behavior assumptions, and for income generated also from non-capital gains sources. It is shown that if the "geometric Brownian motion" hypothesis is accepted, then a general "Separation" or "mutual fund" theorem can be proved such that, in this model, the classical Tobin mean-variance rules hold without the objectionable assumptions of quadratic utility or of normality of distributions for prices. Hence, when asset prices are generated by a geometric Brownian motion, one can work with the two-asset case without loss of generality. If the further assumption is made that the utility function of the individual is a member of the family of utility functions called the "HARA" family, explicit solutions for the optimal consumption and portfolio rules are derived and a number of theorems proved. In the last parts of the paper, the effects on the consumption and portfolio rules of alternative asset price dynamics, in which changes are neither stationary nor independent, are examined along with the effects of introducing wage income, uncertainty of life expectancy, and the possibility of default on (formerly) "risk-free" assets.
2. A Digression on Itô Processes. To apply the dynamic programming technique in a continuous-time model, the state variable dynamics must be expressible as Markov stochastic processes defined over time intervals of length \( h \), no matter how small \( h \) is. Such processes are referred to as infinitely divisible in time. The two processes of this type \(^2\) are: functions of Gauss-Wiener Brownian motions which are continuous in the "space" variables and functions of Poisson processes which are discrete in the space variables. Because neither of these processes is differentiable in the usual sense, a more general type of differential equation must be developed to express the dynamics of such processes. A particular class of continuous-time Markov processes of the first type called Itô Processes are defined as the solution to the stochastic differential equation \(^3\)

\[
dP = f(P,t)dt + g(P,t)dz
\]

where \( P, f, \) and \( g \) are \( n \)-vectors and \( z(t) \) is a \( n \)-vector of

\(^2\) I ignore those infinitely divisible processes with infinite moments which include those members of the stable Pareto family other than the normal.

\(^3\) Itô Processes are a special case of a more general class of stochastic processes called Strong diffusion processes (see Kushner [9], p.22). (1) is a short-hand expression for the stochastic integral

\[
P(t) = P(0) + \int_0^t f(P,s)ds + \int_0^t g(P,s)dz
\]

where \( P(t) \) is the solution to (1) with probability one. A rigorous discussion of the meaning of a solution to equations like (1) is not presented here. Only those theorems needed for formal manipulation and solution of stochastic differential equations are in the text and these without proof. For a complete discussion of Itô Processes, see the seminal paper of Itô [7], Itô and McKean [8], and McKean [11]. For a short description and some proofs, see Kushner [9], p.12-18. For a heuristic discussion of continuous-time Markov processes in general, see Cox and Miller [3], chapter 5.
standard normal random variables. Then \(dz(t)\) is called a multi-dimensional Wiener process (or Brownian motion).

The fundamental tool for formal manipulation and solution of stochastic processes of the Itô type is Itô's Lemma stated as follows

**Lemma**: Let \(F(P_1, \ldots, P_n, t)\) be a \(C^2\) function defined on \(\mathbb{R}^n \times [0, \infty)\) and take the stochastic integrals

\[
P_i(t) = P_i(0) + \int_0^t f_i(P, s) ds + \int_0^t g_i(P, s) dz_i,
\]

\(i = 1, \ldots, n;\)

then the time-dependent random variable \(Y = F\) is a stochastic integral and its stochastic differential is

\[
dY = \sum_{i=1}^n \frac{∂F}{∂P_i} dP_i + \frac{∂F}{∂t} dt + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{∂^2 F}{∂P_i ∂P_j} dP_i dP_j
\]

where the product of the differentials \(dP_i dP_j\) are defined by the multiplication rule

\[
dz_i dz_j = ρ_{ij} dt \quad i, j = 1, \ldots, n
\]

\[
dz_i dt = 0 \quad i = 1, \ldots, n
\]

where \(ρ_{ij}\) is the instantaneous correlation coefficient between the Wiener processes \(dz_i\) and \(dz_j\).
Armed with Itô's Lemma, we are now able to formally differentiate most smooth functions of Brownian motions (and hence integrate stochastic differential equations of the Itô type).\(^\text{[7]}\)

Before proceeding to the discussion of asset price behavior, another concept useful for working with Itô Processes is the differential generator (or weak infinitesimal operator) of the stochastic process \(P(t)\). Define the function \(G(P,t)\) by

\[
\tag{2} \quad 0 \quad G(P,t) \equiv \lim_{h \to 0} E_t \left[ \frac{G(P(t+h),t+h) - G(P(t),t)}{h} \right]
\]

when the limit exists and where "\(E_t\)" is the conditional expectation operator, conditional on knowing \(P(t)\). If the \(P_i(t)\) are generated by Itô Processes, then the differential generator of \(P, \mathcal{L}_P\), is defined by

\[
\tag{3} \quad \mathcal{L}_P = \sum_{i=1}^{n} f_i \frac{\partial}{\partial P_i} + \frac{\partial}{\partial t} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \frac{\partial^2}{\partial P_i \partial P_j}
\]

where \(f = (f_1, \ldots, f_n)\), \(g = (g_1, \ldots, g_n)\), and \(a_{ij} = g_i g_j \rho_{ij}\). Further, it can be shown that

\[
\tag{4} \quad 0 \quad G(P,t) = \mathcal{L}_P[G(P,t)]
\]

\(G\) can be interpreted as the "average" or expected time rate of change of the function \(G(P,t)\) and as such is the natural generalization of the ordinary time derivative for deterministic functions.\(^\text{[8]}\)

\(^{[7]}\) Warning: derivatives (and integrals) of functions of Brownian motions are similar to, but different from, the rules for deterministic differentials and integrals. For example, if

\[
\begin{align*}
P(t) &= P(o) e^{\int_0^t dz - \frac{1}{2} t} = P(o) e^{z(t) - z(o) - \frac{1}{2} t},
\end{align*}
\]

then \(dP = Pdz\). Hence \(\int_0^t \frac{dP}{P} = \int_0^t dz = \log \left( \frac{P(t)}{P(o)} \right)\).

Stratonovich [15] has developed a symmetric definition of stochastic differential equations which formally follows the ordinary rules of differentiation and integration. However, this alternative to the Itô formalism will not be discussed here.

\(^{[8]}\) A heuristic method for finding the differential generator is to take the conditional expectation of \(dG\) (found by Itô's Lemma), and "divide" by \(dt\). The result of this operation will be \(\mathcal{L}_P[G]\), i.e. formally, \(\frac{1}{dt} E_t (dG) = \mathcal{G} = \mathcal{L}_P[G]\).

The "\(\mathcal{L}_P\)" operator is often called a Dynkin operator and is written as "\(DP\)."
3. **Asset price dynamics and the budget equation.**
Throughout the paper, it is assumed that all assets are of the limited liability type; that there exist continuously-trading, perfect markets with no transactions costs for all assets; that the prices per share, \( P_1(t) \), are generated by Itô Processes, i.e.

\[
\frac{dP_1}{P_1} = \alpha_1(P,t)dt + \sigma_1(P,t)dz_1
\]

where we define \( \alpha_1 \equiv f_1/P_1 \) and \( \sigma_1 \equiv g_1/P_1 \) and \( \alpha_1 \) is the instantaneous conditional expected percentage change in price per unit time and \( \sigma_1^2 \) is the instantaneous conditional variance per unit time. In the particular case where the "geometric Brownian motion hypothesis is assumed to hold for asset prices, \( \alpha_1 \) and \( \sigma_1 \) will be constants. For this case, prices will be stationarily and log-normally distributed and it will be shown that this assumption about asset prices simplifies the continuous-time model in the same way that the assumption of normality of prices simplifies the static one-period portfolio model.

To derive the correct budget equation, it is necessary to examine the discrete-time formulation of the model and then to take limits carefully to obtain the continuous-time form. Consider a period model with periods of length \( h \) where all income is generated by capital gains and wealth, \( W(t) \) and \( P_1(t) \) are known at the beginning of period \( t \). Let the decision variables be indexed such that the indices coincide with the period in which the decisions are implemented. Namely, let

\[
N_i(t) \equiv \text{number of shares of asset } i \text{ purchased during period } t, \text{ i.e. between } t \text{ and } t + h
\]

(6) and

\[
C(t) \equiv \text{amount of consumption per unit time during period } t
\]

The model assumes that the individual "comes into" period \( t \) with wealth invested in assets so that

\[
W(t) = \sum_{i=1}^{n} N_i(t-h)P_i(t)
\]

(7)
Notice that it is $N_1(t-h)$ because $N_1(t-h)$ is the number of shares purchased for the portfolio in period $(t-h)$ and it is $P_1(t)$ because $P_1(t)$ is the current value of a share of the $i^{th}$ asset. The amount of consumption for the period, $C(t)h$, and the new portfolio, $N_1(t)$, are simultaneously chosen, and if it is assumed that all trades are made at (known) current prices, then we have that

\[(8) \quad -C(t)h = \sum_{i=1}^{n} [N_1(t) - N_1(t-h)]P_1(t)\]

The "dice" are rolled and a new set of prices is determined, $P_1(t+h)$, and the value of portfolio is now $\sum_{i=1}^{n} N_1(t)P_1(t+h)$. So the individual "comes into" period $(t+h)$ with wealth $W(t+h) = \sum_{i=1}^{n} N_1(t)P_1(t+h)$ and the process continues.

Incrementing (7) and (8) by $h$ to eliminate backward differences, we have that

\[(9) \quad -C(t+h)h = \sum_{i=1}^{n} [N_1(t+h) - N_1(t)]P_1(t+h)\]

\[= \sum_{i=1}^{n} [N_1(t+h) - N_1(t)][P_1(t+h) - P_1(t)] \]

\[+ \sum_{i=1}^{n} [N_1(t+h) - N_1(t)]P_1(t)\]

and

\[(10) \quad W(t+h) = \sum_{i=1}^{n} N_1(t)P_1(t+h)\]

Taking the limits as $h \to 0$, (9) we arrive at the continuous version of (9) and (10),

\[(9') \quad -C(t)dt = \sum_{i=1}^{n} dN_1(t) \, dP_1(t) + \sum_{i=1}^{n} dN_1(t) \, P_1(t)\]

and

\[(10') \quad W(t) = \sum_{i=1}^{n} N_1(t)P_1(t)\]

\[\text{\textsuperscript{9}}\] We use here the result that Itô Processes are right-continuous (see [9], p.15) and hence $P_1(t)$ and $W(t)$ are right-continuous. It is assumed that $C(t)$ is a right-continuous function and throughout the paper, the choice of $C(t)$ is restricted to this class of functions.
Using Itô's Lemma, we differentiate (10') to get

\[(11) \quad dW = \sum_{i=1}^{n} N_i dP_i + \sum_{i=1}^{n} dN_i P_i + \sum_{i=1}^{n} dN_i dP_i\]

The last two terms, \(\sum_{i=1}^{n} dN_i P_i + \sum_{i=1}^{n} dN_i dP_i\), are the net value of additions to wealth from sources other than capital gains. Hence, if \(dy(t) = \) (possibly stochastic) instantaneous flow of non-capital gains (wage) income, then we have that

\[(12) \quad dy - C(t)dt = \sum_{i=1}^{n} dN_i P_i + \sum_{i=1}^{n} dN_i dP_i\]

From (11) and (12), the budget or accumulation equation is written as

\[(13) \quad dW = \sum_{i=1}^{n} N_i(t) dP_i + dy - C(t)dt\]

It is advantageous to eliminate \(N_i(t)\) from (13) by defining a new variable, \(w_i(t) = N_i(t)P_i(t)/W(t)\), the percentage of wealth invested in the \(i^{th}\) asset at time \(t\). Substituting for \(dP_i/P_i\) from (5), we can write (13) as

\[(14) \quad dW = \sum_{i=1}^{n} w_i \sigma_i dt - C dt + dy + \sum_{i=1}^{n} w_i \sigma_i dz_i\]

where, by definition, \(\sum_{i=1}^{n} w_i \equiv 1\).

Until section seven, it will be assumed that \(dy \equiv 0\), i.e. all income is derived from capital gains on assets.

If one of the \(n\)-assets is "risk-free" (by convention, the \(n^{th}\) asset), then \(\sigma_n = 0\), the instantaneous rate of return, \(\alpha_n\), will be called \(r\), and (14) is re-written as

\[(14') \quad dW = \sum_{i=1}^{n} w_i (\alpha_i - r) dt + (rW - C) dt + dy + \sum_{i=1}^{n} w_i \sigma_i dz_i\]

This result follows directly from the discrete-time argument used to derive (9') where \(-C(t)dt\) is replaced by a general \(dv(t)\) where \(dv(t)\) is the instantaneous flow of funds from all non-capital gains sources.

It was necessary to derive (12) by starting with the discrete-time formulation because it is not obvious from the continuous version directly whether \(dy - C(t)dt\) equals \(\sum_{i=1}^{n} dN_i P_i + \sum_{i=1}^{n} dN_i dP_i\) or just \(\sum_{i=1}^{n} dN_i P_i\).

There are no other restrictions on the individual \(w_i\) because borrowing and short-selling are allowed.
Using Itô's Lemma, we differentiate (10') to get

\[ dW = \sum_{1}^{n} N_1 dP_1 + \sum_{1}^{n} dN_1 P_1 + \sum_{1}^{n} dN_1 dP_1 \]

The last two terms, \( \sum_{1}^{n} dN_1 P_1 + \sum_{1}^{n} dN_1 dP_1 \), are the net value of additions to wealth from sources other than capital gains. \(^{10}\) Hence, if \( dy(t) = (\text{possibly stochastic}) \) instantaneous flow of non-capital gains (wage) income, then we have that

\[ dy - C(t)dt = \sum_{1}^{n} dN_1 P_1 + \sum_{1}^{n} dN_1 dP_1 \]

From (11) and (12), the budget or accumulation equation is written as

\[ dW = \sum_{1}^{n} N_1(t) dP_1 + dy - C(t)dt \]

It is advantageous to eliminate \( N_1(t) \) from (13) by defining a new variable, \( w_1(t) = N_1(t)P_1(t)/W(t) \), the percentage of wealth invested in the \( i^\text{th} \) asset at time \( t \). Substituting for \( dP_1/P_1 \) from (5), we can write (13) as

\[ dW = \sum_{1}^{n} w_1 W_1 dW_1 dt - C dt + dy + \sum_{1}^{n} w_1 W_1 \sigma_1 dz_1 \]

where, by definition, \( \sum_{1}^{n} w_1 = 1 \). \(^{11}\)

Until section seven, it will be assumed that \( dy \equiv 0 \), i.e., all income is derived from capital gains on assets. If one of the \( n \)-assets is "risk-free" (by convention, the \( n^\text{th} \) asset), then \( \sigma_n = 0 \), the instantaneous rate of return, \( \alpha_n \), will be called \( r \), and (14) is re-written as

\[ dW = \sum_{1}^{n} w_1(\alpha_1 - r)Wdt + (rW-C)dt + dy + \sum_{1}^{m} W_1 \sigma_1 dz_1 \]

\(^{10}\) This result follows directly from the discrete-time argument used to derive (9') where \(-C(t)dt\) is replaced by a general \( dv(t) \) where \( dv(t) \) is the instantaneous flow of funds from all non-capital gains sources.

\(^{11}\) It was necessary to derive (12) by starting with the discrete-time formulation because it is not obvious from the continuous version directly whether \( dy - C(t)dt \) equals \( \sum_{1}^{n} dN_1 P_1 + \sum_{1}^{n} dN_1 dP_1 \) or just \( \sum_{1}^{n} dN_1 P_1 \).

\(^{11}\) There are no other restrictions on the individual \( w_i \) because borrowing and short-selling are allowed.
where \( m = n - 1 \) and the \( w_1, \ldots, w_m \) are unconstrained by virtue of the fact that the relation \( w_n = 1 - \sum_{i=1}^{m} w_i \) will ensure that the identity constraint in (14) is satisfied.

4. Optimal portfolio and consumption rules: the equations of optimality. The problem of choosing optimal portfolio and consumption rules for an individual who lives \( T \) years is formulated as follows,

\[
\text{Max } E_0 \left[ \int_0^T U(C(t), t) \, dt + B(W(T), T) \right]
\]

subject to: \( W(0) = W_0 \); the budget constraint (14), which in the case of a "risk-free" asset becomes (14'); and where the utility function (during life), \( U \), is assumed to be strictly concave in \( C \) and the "bequest" function, \( B \), is assumed also to be concave in \( W \).

To derive the optimal rules, the technique of stochastic dynamic programming is used. Define

\[
J(W, P, t) = \text{Max } E_t \left[ \int_t^T U(C, s) \, ds + B(W(T), T) \right]
\]

where as before, "\( E_t \)" is the conditional expectation operator, conditional on \( W(t) = W \) and \( P_1(t) = P_1 \). Define

\[
\phi(W, C; W, P, t) = U(C, t) + \mathcal{L}[J],
\]

given \( w_1(t) = w_1 \), \( C(t) = C \), \( W(t) = W \), and \( P_1(t) = P_1 \). \(^{(13)}\)

\(^{(12)}\) Where there is no "risk-free" asset, it is assumed no asset can be expressed as a linear combination of the other assets, implying that the \( nxn \) variance-covariance matrix of returns, \( \Omega = [\sigma_{ij}] \) where \( \sigma_{ij} = \rho_{ij} \sigma_i \sigma_j \), is non-singular. In the case when there is a "risk-free" asset, the same assumption is made about the "reduced" \( mxm \) variance-covariance matrix.

\(^{(13)}\) "\( J \)" is short for the rigorous \( \mathcal{L}_{P, W}^W C \), the Dynkin operator over the variables \( P \) and \( W \) for a given set of controls \( W \) and \( C \):

\[
\mathcal{L} = \frac{\partial}{\partial t} + \left[ \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} w_i w_j \right] \frac{\partial}{\partial W} + \sum_{i=1}^n \alpha_i p_i \frac{\partial}{\partial P_i} \]

\[+ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} w_i w_j w_i \frac{\partial^2}{\partial W^2} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n p_i p_j \sigma_{ij} \frac{\partial^2}{\partial P_i \partial P_j} \]

\[+ \sum_{i=1}^n \sum_{j=1}^n p_i w_j \sigma_{ij, \gamma} \frac{\partial^2}{\partial W \partial W} \]
From the theory of stochastic dynamic programming, the following theorem provides the method for deriving the optimal rules, $C^*$ and $W^*$.

Theorem I. (14) If the $P_x(t)$ are generated by a strong diffusion process, $U$ is strictly concave in $C$, and $B$ is concave in $W$, then there exists a set of optimal rules (controls), $W^*$ and $C^*$, satisfying $\sum_1^n W_{i1}^* = 1$ and $J(W, P, T) = B(W, T)$ and these controls satisfy
$$o = \phi(C^*, W^*; W, P, t) \geq \phi(C, W; W, P, t)$$
for $t \in [0, T]$.

From theorem I, we have that

(18) $o = \text{Max} \{\phi(C, W; W, P, t)\}$

$(C, W)$

In the usual fashion of maximization under constraint, we define the Lagrangian, $L = \phi + \lambda[1 - \sum_1^n W_{i1}]$ where $\lambda$ is the multiplier and find the extreme points from the first-order conditions

(19) $O = L_C(C^*, W^*) = U_C(C^*, t) - J_W$

(20) $O = L_{W_k} (C^*, W^*) = -\lambda + J_W a_k W + J_{WW} \sum_1^n \sigma_{kj} W_{j*} W_{j}^2$
$$+ \sum_1^n J_{Wj} \sigma_{kj} W_{j*} P_{jW}, \ k=1, \ldots, n$$

(21) $O = L_{\lambda} (C^*, W^*) = 1 - \sum_1^n W_{i1}^*$

where the notation for partial derivatives is $J_W = \frac{\partial J}{\partial W}$, $J_t = \frac{\partial J}{\partial t}$,

$U_C = \frac{\partial U}{\partial C}$, $J_1 = \frac{\partial J}{\partial P_1}$, $J_{ij} = \frac{\partial^2 J}{\partial P_i \partial P_j}$, and $J_{WW} = \frac{\partial^2 J}{\partial W^2}$.

(14) For a heuristic proof of this theorem and the derivation of the stochastic Bellman equation, see Dreyfus (4) and Merton (12). For a rigorous proof and discussion of weaker conditions, see Kushner (9), chapter IV especially theorem 7.
From the theory of stochastic dynamic programming, the following theorem provides the method for deriving the optimal rules, \( C^* \) and \( w^* \).

**Theorem I.** (14) If the \( P_i(t) \) are generated by a strong diffusion process, \( U \) is strictly concave in \( C \), and \( B \) is concave in \( W \), then there exists a set of optimal rules (controls), \( w^* \) and \( C^* \), satisfying \( \sum_1^n w_{i1}^* = 1 \) and \( J(W,P,T) = B(W,T) \) and these controls satisfy

\[
o = \phi(C^*;w^*;W,P,t) \geq \phi(C,w;W,P,t) \quad \text{for } t \in [0,T].
\]

From theorem I, we have that

\[
(18) \quad o = \text{Max} \{\phi(C,w;W,P,t)\}
\]

\[
(C,w)
\]

In the usual fashion of maximization under constraint, we define the Lagrangian, \( L \equiv \phi + \lambda[1 - \sum_1^n w_{i1}] \) where \( \lambda \) is the multiplier and find the extreme points from the first-order conditions

\[
(19) \quad o = L_C(C^*;w^*) = U_C(C^*,t) - J_W
\]

\[
(20) \quad o = L_{w_{ik}}(C^*;w^*) = -\lambda + J_Wa_kW + \sum_1^n \sigma_{kj}w_{i1}^*J_{W^2} + \sum_1^n J_{W^2} \sigma_{kj}p_{j1}w_i^*, \quad k=1,\ldots,n
\]

\[
(21) \quad o = L_{\lambda}(C^*;w^*) = 1 - \sum_1^n w_{i1}^*
\]

where the notation for partial derivatives is \( J_{W} = \frac{\partial J}{\partial t} \), \( J_{t} = \frac{\partial J}{\partial t} \),

\[
U_C = \frac{\partial U}{\partial C}, \quad J_i = \frac{\partial J}{\partial P_i}, \quad J_{ij} = \frac{\partial^2 J}{\partial P_i \partial P_j}, \quad \text{and} \quad J_{W} = \frac{\partial^2 J}{\partial P_j \partial W}.
\]

(14) For a heuristic proof of this theorem and the derivation of the stochastic Bellman equation, see Dreyfus [4] and Merton [12]. For a rigorous proof and discussion of weaker conditions, see Kushner [9], chapter IV especially theorem 7.
Because \( \dot{L}_{CC} = \dot{U}_{CC} < 0, \quad \dot{L}_{C_{W_k}} = \dot{\phi}_{C_{W_k}} = 0; \)

\[
\dot{L}_{W_k \cdot W_k} = \sigma_k^2 W_k^2 \dot{J}_{WW}; \quad \dot{L}_{W_k \cdot W_j} = 0, \quad k \neq j, \text{ a sufficient condition for a unique interior maximum is that } \dot{J}_{WW} < 0 (\text{i.e. that } J \text{ be strictly concave in } W). \]

That assumed, as an immediate consequence of differentiating (19) totally with respect to \( W \), we have

\[
(22) \quad \frac{\partial C^*}{\partial W} > 0
\]

To solve explicitly for \( C^* \) and \( w^* \), we solve the \( n+2 \) non-dynamic implicit equations, (19) - (21), for \( C^* \) and \( w^* \), and \( \lambda \) as functions of \( J_W, J_{WW}, J_{jW}, W, P \) and \( t \). Then \( C^* \) and \( w^* \) are substituted in (18) which now becomes a second-order partial differential equation for \( J \), subject to the boundary condition \( J(W,P,T) = B(W,T) \). Having (in principle at least) solved this equation for \( J \), we then substitute back into (19) - (21) to derive the optimal rules as functions of \( W, P, \) and \( t \). Define the inverse function

\[
G \equiv [U_C]^{-1}. \quad \text{Then from (19),}
\]

\[
(23) \quad C^* = G(J_W, t)
\]

To solve for the \( w^*_1 \), note that (20) is a linear system in \( w^*_1 \) and hence can be solved explicitly. Define

\[
(24) \quad \Omega = [\sigma_{ij}]; \quad \text{the } n \times n \text{ variance-covariance matrix}
\]

\[
[v_{ij}] = \Omega^{-1}
\]

\[
\Gamma = \sum_1^n \sum_1^n v_{ij}.
\]

Eliminating \( \lambda \) from (20), the solution for \( w^*_k \) can be written as

\[
(25) \quad w^*_k = h_k(P, t) + m(P,W,t)g_k(P,t) + f_k(P,W,t), \quad k = 1, \ldots, n
\]

where \( \sum_1^n h_k = 1, \sum_1^n g_k = 0, \) and \( \sum_1^n f_k = 0. \)

\( ^{15} \Omega^{-1} \) exists by the assumption on \( \Omega \) in footnote 12.

\( ^{16} h_k(P, t) = \sum_1^n v_{kj} / \Gamma; \quad m(P,W,t) = - J_W / \Gamma W J_{WW}; \)

(continued)
Substituting for \( w^\ast \) and \( C^\ast \) in (18), we arrive at the fundamental partial differential equation for \( J \) as a function of \( W, P, \) and \( t, \)

\[
\begin{align*}
O &= U[G, t] + J_t + J_W \left[ \Sigma_1^n \Sigma_1^n v_{k1} \alpha_k W \right] - G \\
&+ \Sigma_1^n J_1 \alpha_1 P_1 + \frac{1}{2} \Sigma_1^n \Sigma_1^n j_{ij} \sigma_{ij} P_1 P_j + \frac{W}{T} \Sigma_1^n J_{PW} P_j \\
&- \frac{J_W}{T J_{PW}} \left( \Sigma_1^n j_{PW} P_k P_k - \Sigma_1^n j_{PW} P_j \Sigma_1^n v_{kl} \alpha_1 \right) \\
&+ \frac{J_{PW} W^2}{2T} - \frac{1}{2T J_{PW}} \left[ \Sigma_1^n \Sigma_1^n j_{PW} P_j P_m \sigma_{jm} \right] + \frac{1}{2T J_{PW}} \left( \Sigma_1^n J_{PW} P_j \right)^2 \\
&- \frac{J_W^2}{2T J_{PW}} \left[ \Sigma_1^n \Sigma_1^n v_{kl} \alpha_k \alpha_1 r - \left( \Sigma_1^n \Sigma_1^n v_{kl} \alpha_k \right)^2 \right]
\end{align*}
\]

subject to the boundary condition \( J(W, P, T) = B(W, T) \). If (26) were solved, the solution \( J \) could be substituted into (23) and (25) to obtain \( C^\ast \) and \( w^\ast \) as functions of \( W, P, \) and \( t. \)

For the case where one of the assets is "risk-free", the equations are somewhat simplified because the problem can be solved directly as an unconstrained maximum by eliminating \( w_n \) as was done in (14'). In this case, the optimal proportions in the risky assets are

\[
w^\ast_k = -\frac{J_{PW}}{J_{PW}^{\ast}} \Sigma_1^n v_{kj} (\alpha_j - r) - \frac{J_{PW} P_k}{J_{PW}} , \ k = 1, \ldots, m.
\]

The partial differential equation for \( J \) corresponding to (26) becomes

\[
\begin{align*}
0 &= U[G, T] + J_t + J_W [T W - C] \Sigma_1^n J_1 \alpha_1 P_1 \\
&+ \frac{1}{2} \Sigma_1^n \Sigma_1^n j_{ij} \sigma_{ij} P_1 P_j - \frac{J_W}{J_{PW}} \Sigma_1^n J_{PW} P_j (\alpha_j - r) \\
&+ \frac{J_{PW}^2}{2J_{PW}} \Sigma_1^n \Sigma_1^n v_{ij} (\alpha_j - r)(\alpha_j - r) - \frac{1}{2T J_{PW}} \Sigma_1^n \Sigma_1^n j_{PW} J_{PW} \sigma_{ij} P_1 P_j
\end{align*}
\]

(16) (continued)

\[
g_k(P, t) \equiv \Sigma_1^n v_{k1} \left( r \alpha_1 - \Sigma_1^n v_{ij} \alpha_j \right) ; \ f_k(P, W, t) \equiv \left[ r J_{PW} P_k \\
- \Sigma_1^n j_{PW} P_1 \right] \overline{T_{PW} W}
\]
subject to the boundary condition \( J(W,P,T) = B(W,T) \).

Although (28) is a simplified version of (26), neither (26) nor (28) lend themselves to easy solution. The complexities of (26) and (28) are caused by the basic non-linearity of the equations and the large number of state variables. Although there is little that can be done about the non-linearities, in some cases, it may be possible to reduce the number of state variables.

5. Log-normality of prices and the continuous-time analog to Tobin-Markowitz mean-variance analysis. When, for \( k = 1, \ldots, n \), \( \alpha_k \) and \( \sigma_k \) are constants, the asset prices have stationary, log-normal distributions. In this case, \( J \) will be a function of \( W \) and \( t \) only and not \( P \). Then (26) reduces to

\[
0 = U[G,t] + J_t + J_W \left[ \sum_{i=1}^{n} \sum_{j=1}^{n} v_{jk} \alpha_k \alpha_j W - G \right] + \frac{J_{WW} W^2}{2\Gamma} - \frac{J^2 W^2}{2\Gamma J_{WW} \left[ \sum_{i=1}^{n} \sum_{j=1}^{n} v_{kl} \alpha_k \alpha_l - \left( \sum_{i=1}^{n} \sum_{j=1}^{n} v_{kl} \alpha_k \right)^2 \right]}
\]

From (25), the optimal portfolio rule becomes

\[
w^*_k = h_k + m(W,t)g_k
\]

where \( \sum_{i=1}^{n} h_k = 1 \) and \( \sum_{i=1}^{n} g_k = 0 \) and \( h_k \) and \( g_k \) are constants.

From (30), the following "separation" or "mutual fund" theorem can be proved:

Theorem II. \(^{17}\) Given \( n \) assets with prices \( p_i \) whose changes are stationarily and log-normally distributed, then (1) there exist a unique pair (except for scale) of "mutual funds" constructed from linear combinations

\(^{17}\) See Cass and Stiglitz \([1]\) for a general discussion of Separation theorems. The only degenerate case is when all the assets are identically distributed (i.e. symmetry) in which case, only one mutual fund is needed.
of these assets such that, independent of preferences (i.e., the form of the utility function) wealth distribution, or time horizon, individuals will be indifferent between choosing from a linear combination of these two funds or a linear combination of the original n assets. (2) If \( P_f \) is the price per share of either fund, then \( P_f \) is log-normally distributed. Further (3) if \( \delta_k = \text{percentage of one mutual fund's value held in the } k^{th} \text{ asset} \), and if \( \lambda_k = \text{percentage of the other mutual fund's value held in the } k^{th} \text{ asset} \), then one can find that

\[
\delta_k = h_k + g_k, \quad k = 1, \ldots, n
\]

and

\[
\lambda_k = h_k, \quad k = 1, \ldots, n.
\]

Proof: (1) (30) is a parametric representation of a line in the hyperplane defined by \( \sum_n w^*_k = 1 \). Hence, there exist two linearly independent vectors (namely, the vectors of asset proportions held by the two mutual funds) which form a basis for all optimal portfolios chosen by the individuals. Therefore, each individual would be indifferent between choosing a linear combination of the mutual fund shares or a linear combination of the original n assets.

(2) Let \( V = N_f P_f \) = the total value of (either) fund where \( N_f \) = number of shares of the fund outstanding. Let \( N_k \) = number of shares of asset k held by the fund and \( \mu_k = N_k P_k / V \) = percentage of total value invested in the \( k^{th} \) asset. Then \( V = \sum_{k=1}^n N_k P_k \) and

\[
dV = \sum_{k=1}^n N_k dP_k + \sum_{k=1}^n P_k dN_k + \sum dP_f dN_f
\]

\[= N_f dP_f + P_f dN_f + dP_f dN_f\]

\(^{(18)}\) See [1], page 15.
But

\[ \sum_{k}^{n} P_k dN_k + \sum_{k}^{n} dP_k dN_k = \text{net inflow of funds from non-capital gain sources} \]

\[ = \text{net value of new shares issued} \]

\[ = P_f dN_f + dN_f dP_f \]

From (31) and (32), we have that

\[ N_f dP_f = \sum_{k}^{n} N_k dP_k \]

By the definition of \( V \) and \( \mu_k \), (33) can be re-written as

\[ \frac{dP_f}{P_f} = \sum_{k}^{n} \mu_k \frac{dP_k}{P_k} \]

\[ = \sum_{k}^{n} \mu_k \alpha_k dt + \sum_{k}^{n} \mu_k \sigma_k dz_k \]

By Itô's Lemma and (34), we have that

\[ P_f(t) = P_f(0) \exp \left[ \sum_{k}^{n} \mu_k \alpha_k t - \frac{1}{2} \sum_{k}^{n} \sum_{j}^{n} \mu_k \mu_j \sigma_{kj} \right] \]

\[ + \sum_{k}^{n} \mu_k \sigma_k \int_{0}^{t} dz_k \]

So, \( P_f(t) \) is log-normally distributed.

(3) Let \( a(W,t;U) \equiv \text{percentage of wealth invested in the first mutual fund by an individual with utility function} \)

\( U \) and wealth \( W \) at time \( t \). Then, \( (1-a) \) must equal the percentage of wealth invested in the second mutual fund. Because the individual is indifferent between these asset holdings or an optimal portfolio chosen from the original \( n \) assets, it must be that

\[ w_k^* = h_k + m(W,t)g_k = a \delta_k + (1-a) \lambda_k \]

Clearly, the only solution to the linear system (36) for all \( W, t, \) and \( U \) is

\[ \delta_k = h_k + g_k \quad , \quad k = 1, \ldots, n \]

\[ \lambda_k = h_k \quad , \quad k = 1, \ldots, n \]

\[ a = m(W,t) \].
Note that \( \sum_{1}^{n} \delta_k = \sum_{1}^{n} (h_k + g_k) = 1 \) and \( \sum_{1}^{n} \lambda_k = \sum_{1}^{n} h_k = 1 \).

Q.E.D.

For the case when one of the assets is "risk-free", there is a corollary to theorem II. Namely,

Corollary. If one of the assets is "risk-free", then one of the two mutual funds will contain only this asset. If \( \delta_k \) is percentage of the total value of the "risky" mutual fund invested in the \( k^{th} \) asset, then

\[
\delta_k = \frac{\sum_{1}^{m} v_{kj}(\alpha_j-r)}{\sum_{1}^{m} \sum_{1}^{m} v_{ij}(\alpha_j-r)}, \ k=1, \ldots, m.
\]

Proof. By the assumption of stationary log-normal prices, (27) reduces to

\[
(38) \quad w_k^* = \frac{-J_{W}}{J_{WW}} \sum_{1}^{m} v_{kj}(\alpha_j-r), \quad k=1, \ldots, m
\]

and

\[
(39) \quad w_n^* = 1 - \sum_{1}^{m} w_k^* = 1 + \frac{J_{W}}{J_{WW}} \sum_{1}^{m} \sum_{1}^{m} v_{ij}(\alpha_j-r)
\]

By the same argument as in the proof of theorem II, (38) and (39) define a line in the hyperplane defined by \( \sum_{1}^{n} w_k^* = 1 \) and

\[
\delta_k = \frac{\sum_{1}^{m} v_{kj}(\alpha_j-r)}{\sum_{1}^{m} \sum_{1}^{m} v_{ij}(\alpha_j-r)}; \quad \lambda_k=0, \ k=1, \ldots, m
\]

\[
\delta_n = 0 \quad \text{; } \lambda_k=1 \ k=n.
\]

Q.E.D.

Thus, if we have an economy where all asset prices are log-normally distributed, the investment decision can be divided into two parts by the establishment of two financial intermediaries (mutual funds) to hold all individual securities and to issue shares of their own for purchase by individual investors. The separation is complete because the "instructions" given the fund managers, namely to hold pro-
portions $\delta_k$ and $\lambda_k$ of the $k^{th}$ security, $k=1,...,n$, depend only on the price distribution parameters and are independent of individual preferences, wealth distribution, or age distribution.

The similarity of this result to that of the classical Tobin-Markowitz analysis is clearest when one examines closely the investment rule given to the "risky" fund's manager when there exists a "risk-free" asset (money) with zero return ($r=0$). It is easy to show that the $\delta_k$ proportions prescribed in the corollary are derived by finding the locus of points in the (instantaneous) mean-standard deviation space of composite returns which minimize variance for a given mean, and then by finding the point where a line drawn from the origin is tangent to the locus. This point determines the $\delta_k$ as is illustrated in figure 1.

![Diagram](image)

**Figure 1.**

Given the $\alpha^*$, the $\delta_k$ are determined. So the log-normal assumption in the continuous-time model is sufficient to allow the same analysis as in the static mean-variance model but without the objectionable assumptions of quadratic utility or normality of the distribution of absolute price changes. (Log-normality of price changes is much less objectionable, since this does invoke "limited liability" and by the central limit theorem is the only regular solution to
any continuous-space, infinitely-divisible process in time.)

An immediate advantage for the present analysis is that whenever log-normality of prices is assumed, we can work, without loss of generality, with just two assets, one "risk-free" and one risky with its price log-normally distributed. The risky asset can always be thought of as a composite asset with price \( P(t) \) defined by the process

\[
\frac{dP}{P} = \alpha dt + \sigma dz
\]

where

\[
\alpha = \sum_{i=1}^{m} \sum_{j=1}^{m} v_{ij}(\alpha_j - r) a_k / \sum_{i=1}^{m} \sum_{j=1}^{m} v_{ij}(\alpha_j - r)
\]

\[
\sigma^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \delta_{ij} \delta_{kj} c_{kj}
\]

\[
dz = \sum_{i=1}^{m} \delta_{ik} \sigma_k dz_k / \sigma.
\]

6. Explicit solutions for a particular class of utility functions. On the assumption of log-normality of prices, some characteristics of the asset demand functions were shown. If a further assumption about the preferences of the individual is made, then equation (28) can be solved in closed form, and the optimal consumption and portfolio rules derived explicitly. Assume that the utility function for the individual, \( U(C,t) \), can be written as \( U(C,t) = e^{-\rho t V(C)} \) where \( V \) is a member of the family of utility functions whose measure of absolute risk aversion is positive and hyperbolic in consumption, i.e. \( A(C) \equiv -V''/V' = 1/ \left( \frac{C}{1-\gamma} + \eta/\beta \right) > 0 \), subject to the restrictions

\[
\gamma \neq 1; \beta > 0; \left( \frac{\beta C}{1-\gamma} + \eta \right) > 0; \eta = 1 \text{ if } \gamma = -\infty.
\]

All members of the HARA (hyperbolic absolute risk-aversion) family can be expressed as

\[
V(C) = \frac{(1-\gamma)}{\gamma} \left( \frac{\beta C}{1-\gamma} + \eta \right)^{\gamma}
\]
This family is rich, in the sense that by suitable adjustment of the parameters, one can have a utility function with absolute or relative risk aversion increasing, decreasing, or constant.  

\[ A(C) = \frac{1}{\frac{C}{1-\gamma} + \frac{n}{\beta}} > 0 \quad \text{(implies } n > 0 \text{ for } \gamma > 1) \]

\[ A'(C) = \frac{-1}{(1-\gamma)\left(\frac{C}{1-\gamma} + \frac{n}{\beta}\right)^2} < 0 \text{ for } -\infty < \gamma < 1 \]
\[ > 0 \text{ for } 1 < \gamma < \infty \]
\[ = 0 \text{ for } \gamma = \pm \infty \]

Relative risk aversion \( R(C) \equiv -\frac{\text{V}''C}{\text{V}'} = A(C)C \)

\[ R'(C) = \frac{\frac{n/\beta}{\left(\frac{C}{1-\gamma} + \frac{n}{\beta}\right)^2}} > 0 \text{ for } n > 0 \ (\ -\infty \leq \gamma \leq \infty, \ \gamma \neq 1 \)
\[ = 0 \text{ for } n = 0 \]
\[ < 0 \text{ for } n < 0 \ (\ -\infty < \gamma < 1 \)

Note that included as members of the HARA family are the widely-used iso-elastic (constant relative risk aversion), exponential (constant absolute risk aversion), and quadratic utility functions. As is well known for the quadratic case, the members of the HARA family with \( \gamma > 1 \) are only defined for a restricted range of consumption, namely \( 0 < C < (\gamma - 1)n/\beta \).

[6], [10], [5], [13], [1], and [12] discuss the properties of various members of the HARA family in a portfolio context. Although this is not done here, the HARA definition can be generalized to include the cases when \( \gamma, \beta, \) and \( n \) are functions of time subject to the restrictions in (42).
Without loss of generality, assume that there are two assets, one "risk-free" asset with return \( r \) and the other, a "risky" asset whose price is log-normally distributed satisfying (40). From (28), the optimality equation for \( J \) is

\[
J_t = \frac{(1-\gamma)^2}{\gamma} e^{-\rho t} e^{\rho t J_{\text{W}}} \left[ \frac{\gamma}{\beta} \right] J_{\text{W}}^\gamma - J_{\text{W}}^2 \frac{(\alpha-r)^2}{\alpha \sigma^2} J_{\text{W}}^2 + J_t + (1-\gamma)n/\beta + r W
\]

subject to \( J(W, T) = 0 \). The equations for the optimal consumption and portfolio rules are

\[
C^*(t) = \frac{(1-\gamma)}{\beta} e^{\rho t} J_{\text{W}}^\gamma \left[ \frac{1}{\beta} \right] J_{\text{W}} J_W - \frac{(1-\gamma)n}{\beta}
\]

and

\[
w^*(t) = \frac{J_{\text{W}}}{J_{\text{W}} W} \frac{(\alpha-r)}{\sigma^2}
\]

where \( w^*(t) \) is the optimal proportion of wealth invested in the risky asset at time \( t \). A solution (21) to (44) is

\[
J(W, t) = \delta e^{-\rho t} \left( \frac{\delta}{\rho - \gamma v} \right)^\delta \left[ \frac{n}{\delta + \frac{n}{\beta r} (1-e^{-r(T-t)})} \right]^\gamma
\]

where \( \delta = 1 - \gamma \) and \( v = r + (\alpha-r)^2/2\sigma^2 \).

---

[20] It is assumed for simplicity that the individual has a zero bequest function, i.e. \( B = 0 \). If \( B(W, T) = H(T) (aW+b)^t \), the basic functional form for \( J \) in (47) will be the same. Otherwise, systematic effects of age will be involved in the solution.

[21] By theorem 1, there is no need to be concerned with uniqueness although, in this case, the solution is unique.
From (45), (46), and (47), the optimal consumption and portfolio rules can be written in explicit form as

$$C^*(t) = \frac{\delta_n}{\delta \left(1 - \exp \left[\frac{(\rho - \gamma \nu)}{\delta} (t-T)\right]\right)} + \delta_n \frac{\delta_n}{\delta \left(1 - \exp \left[\frac{(\rho - \gamma \nu)}{\delta} (t-T)\right]\right)} - \delta_n \frac{\delta_n}{\delta \left(1 - \exp \left[\frac{(\rho - \gamma \nu)}{\delta} (t-T)\right]\right)}$$

and

$$w^*(t)W(t) = \frac{\alpha - r}{\delta \sigma^2} W(t) + \frac{\eta}{b \sigma^2} \left(1 - e^{-(t-T)}\right)$$

The manifest characteristic of (48) and (49) is that the demand functions are linear in wealth. It will be shown that the HARA family is the only class of concave utility functions which imply linear solutions. For notation purposes, define \(I(X,t) \subseteq \text{HARA}(X)\) if \(-I_{XX}/I_X = 1/(\alpha X + \beta) > 0\), where \(\alpha\) and \(\beta\) are, at most, functions of time and \(I\) is a strictly concave function of \(X\).

Theorem III. Given the model specified in this section, then \(C^* = aW + b\) and \(w^*W = gW + h\) where \(a, b, g,\) and \(h\) are, at most, functions of time if and only if \(U(C,t) \subseteq \text{HARA}(C)\).

Proof: "If" part is proved directly by (48) and (49).
"Only if" part: Suppose \(w^*W = gW + h\) and \(C^* = aW + b\). From (19), we have that \(U_C(C^*,t) = J_W(W,t)\). Differentiating this expression totally with respect to \(W\), we have that \(U_C \frac{dC^*}{dW} = J_W(W,t)\) or \(aU_C = J_{WW}\) and hence

$$\frac{-U_{CC}a}{U_C} = -\frac{J_{WW}}{J_W}$$

From (46), \(w^*W = gW + h = -J_W(\alpha - r)/J_{WW}\sigma^2\) or

$$J_{WW}/J_W = 1/\left[\left(\frac{\alpha^2 g}{(\alpha - r)}\right) W + \frac{\alpha^2 h}{(\alpha - r)}\right].$$
So, from (50) and (51), we have that \( U \) must satisfy

\[
-\frac{U_{GG}}{U_C} = \frac{1}{(a'c + b')}
\]

where \( a' \equiv \sigma^2 g / (a-r) \) and \( b' \equiv (a\sigma h - b\sigma^2 g) / (a-r) \).

Hence \( U \subset \text{HARA}(C) \). Q.E.D.

As an immediate result of theorem III, a second theorem can be proved.

Theorem IV. Given the model specified in this section, \( J(W,t) \subset \text{HARA}(W) \) if and only if \( U \subset \text{HARA}(C) \).

Proof: "If" part is proved directly by (47).

"Only if" part: suppose \( J(W,t) \subset \text{HARA}(W) \). Then from (46), \( w^W \) is a linear function of \( W \). If (28) is differentiated totally with respect to wealth and given the specific price behavior assumptions of this section, we have that \( C^w \) must satisfy

\[
C^w = rW + \frac{J_t W}{J_{WW}} + \frac{rJ_W}{J_{WW}} - w^W \frac{d(w^W)}{dW} - \frac{J_{WW}}{2J_{WW}} \left( \frac{J_W}{J_{WW}} \right)^2 \frac{(a-r)^2}{\sigma^2}
\]

But if \( J \subset \text{HARA}(W) \), then (53) implies that \( C^w \) is linear in wealth. Hence, by theorem III, \( U \subset \text{HARA}(C) \). Q.E.D.

Given (48) and (49), the stochastic process which generates wealth when the optimal rules are applied, can be derived. From the budget equation (14'), we have that

\[
dW = \left[ \left( \frac{(\alpha-r)}{\sigma^2 \delta} - \frac{\mu}{1-e^{\mu(t-T)}} \right) \frac{d}{dt} + \frac{(\alpha-r) \sigma}{\sigma^2 \delta} \right] X(t) dt + \sigma w^W dz
\]

where

\[
X(t) = \frac{\phi \mu}{\phi \mu(t-T)} \text{ for } 0 \leq t \leq T \text{ and } \mu = (\rho-v)/\delta.
\]

By Itô's Lemma, \( X(t) \) is the solution to

\[
\frac{dX}{X} = \left[ \delta - \frac{\mu}{1-e^{\mu(t-T)}} \right] dt + \frac{(\alpha-r) \sigma}{\sigma^2 \delta} dz
\]
Again using Itô's Lemma, integrating (55) we have that

$$X(t) = X(0) \exp \left[ \left( \delta - \mu - \frac{(\alpha - r)^2}{2\sigma^2} \right) t + \frac{(\alpha - r)}{\sigma^2} \int_0^t dz \right] \frac{1 - e^{\mu(t-T)}}{1 - e^{-\mu T}}$$

and hence, $X(t)$ is log-normally distributed. Therefore,

$$W(t) = X(t) - \frac{\delta \epsilon}{\beta} (1 - e^{\epsilon(t-T)})$$

is a "displaced" or "three-parameter" log-normally distributed random variable. By Itô's Lemma, solution (56) to (55) holds with probability one and because $W(t)$ is a continuous process, we have with probability one that

$$\lim_{t \to T} W(t) = 0.$$  

From (48), with probability one,

$$\lim_{t \to T} C^*(t) = 0.$$  

Further, from (48), $C^* + \frac{\delta \epsilon}{\beta}$ is proportional to $X(t)$ and from the definition of $U(C^*, t)$, $U(C^*, t)$ is a log-normally distributed random variable. The following theorem shows that this result holds only if $U(C, t) \subseteq HARA(C)$.

**Theorem V.** Given the model specified in this section and the time-dependent random variable $Y(t) \equiv U(C^*, t)$, then $Y$ is log-normally distributed if and only if $U(C, t) \subseteq HARA(C)$.

**Proof:** "If" part: It was previously shown that if $U \subseteq HARA(C)$ then $Y$ is log-normally distributed.

"only if" part: Let $C^* \equiv \tilde{g}(W, t)$ and $W^* \equiv f(W, t)$. By Itô's Lemma,

$$dY = U_C dC^* + U_t dt + \frac{1}{2} U_{CC} (dC^*)^2$$

$$dC^* = \tilde{g}_W dW + \tilde{g}_t dt + \frac{1}{2} \tilde{g}_{WW} (dW)^2$$

$$dW = [f(\alpha - r) + rW - g] dt + \sigma fdz$$

$$(22) \quad U = \frac{(1 - \gamma)}{\gamma} e^{-\beta t} \left[ \frac{BC}{1 - \gamma} + \epsilon \right]^\gamma$$

and products and powers of log-normal variates are log-normal with one exception: the logarithmic utility function ($\gamma = 0$) is a singular case where $U(C^*, t) = \log C^*$ is normally distributed.
Because $(dW)^2 = \sigma^2 f^2 dt$, we have that

\begin{align}
\text{(60)} \quad dC &= \left[ g_W f(\alpha-r) + g_W r - g_g W + \frac{1}{2} \left( g_{WW} \sigma^2 f^2 + \varepsilon_t \right) \right] dt + \sigma f g W dz \\
\text{and} \\
\text{(61)} \quad dY &= \left\{ U_C \left[ g_W f(\alpha-r) + r g_W W - g_g W + \frac{1}{2} g_{WW} \sigma^2 f^2 + \varepsilon_t \right] + U_t \\
&\quad + \frac{1}{2} U_C \sigma^2 f^2 g_W^2 \right\} dt + \sigma f g W U_C dz
\end{align}

A necessary condition for $Y$ to be log-normal is that $Y$ satisfy

\begin{align}
\text{(62)} \quad \frac{dY}{Y} &= F(Y) dt + b dz
\end{align}

where $b$ is, at most, a function of time. If $Y$ is log-normal, from (61) and (62), we have that

\begin{align}
\text{(63)} \quad b(t) &= \sigma f g_W U_C / U \\
\text{From the first-order conditions,} \quad f \quad \text{and} \quad g \quad \text{must satisfy}
\end{align}

\begin{align}
\text{(64)} \quad U_C g_W &= J_W W \\
f &= - J_W (\alpha-r) / \sigma^2 J_W W
\end{align}

But (63) and (64) imply that

\begin{align}
\text{(65)} \quad bU/\sigma U_C &= f g_W = - (\alpha-r) U_C / \sigma^2 U_C \\
or
\end{align}

\begin{align}
\text{(66)} \quad - U_C / U_C &= \eta(t) U_C / U
\end{align}

where $\eta(t) = (\alpha-r) / \sigma b(t)$. Integrating (66), we have that

\begin{align}
\text{(67)} \quad U &= \left[ (\eta+1) (C+\mu) \xi(t) \right] \frac{1}{\eta+1}
\end{align}

where $\xi(t)$ and $\mu$ are, at most, functions of time and hence $U \in \text{HARA}(C)$. Q.E.D.

For the case when asset prices satisfy the "geometric" Brownian motion hypothesis and the individual's utility function is a member of the HARA family, the consumption-portfolio problem is completely solved. From (48) and (49), one could examine the effects of shifts in various parameters on the consumption and portfolio rules by the methods of comparative statics as was done for the iso-elastic case in [12].
7. Non-capital gains income: wages. In the previous sections, it was assumed that all income was generated by capital gains. If a (certain) wage income flow, \( dy = Y(t) dt \), is introduced, the optimality equation (18) becomes

\[
(68) \quad \mathcal{O} = \max_{\{C, w\}} \left[ U(C, t) + \mathcal{J}(J) \right]
\]

where the operator \( \mathcal{J} \) is defined by \( \mathcal{J} \equiv \mathcal{L} + Y(t) \frac{\partial}{\partial w} \).

This new complication causes no particular computational difficulties. If a new control variable, \( \tilde{C}(t) \), and new utility function, \( \tilde{V}(C, t) \) are defined by \( \tilde{C}(t) \equiv C(t) - Y(t) \) and \( \tilde{V}(C, t) \equiv U(C(t) + Y(t), t) \), then (68) can be re-written as

\[
(69) \quad \mathcal{O} = \max_{\{C, w\}} \left[ \tilde{V}(C, t) + \mathcal{J} [J] \right]
\]

which is the same equation as the optimality equation (18) when there is no wage income and where consumption has been re-defined as consumption in excess of wage income.

In particular, if \( Y(t) \equiv Y \), a constant, and \( U \in \text{HARA}(C) \), then the optimal consumption and portfolio rules corresponding to (48) and (49) are

\[
(70) \quad C^*(t) = \left[ \alpha - \gamma \nu \right] \left[ W + \frac{Y \left( 1 - e^{r(t-T)} \right)}{r} + \frac{\delta \eta}{\beta} \left( 1 - e^{r(t-T)} \right) \right] \frac{\delta \eta}{\beta}
\]

and

\[
(71) \quad w^W = \frac{(\alpha - r)}{\delta \sigma^2} \left( W + \frac{Y \left( 1 - e^{r(t-T)} \right)}{r} \right) + \frac{(\alpha - r) \eta}{\beta \sigma^2} \left( 1 - e^{r(t-T)} \right)
\]

Comparing (70) and (71) with (48) and (49), one finds that, in computing the optimal decision rules, the individual capitalizes the lifetime flow of wage income at the market (risk-free) rate of interest and then treats the capitalized value.
as an addition to the current stock of wealth.\(^{23}\)

The introduction of a stochastic wage income will cause increased computational difficulties although the basic analysis is the same as for the no-wage income case. For a solution to a particular example of a stochastic wage problem, see example two of section eight.

8. Poisson processes. The previous analyses always assumed that the underlying stochastic processes were smooth functions of Brownian motions and therefore, continuous in both the time and state spaces. Although such processes are reasonable models for price behavior of many types of liquid assets, they are rather poor models for the description of other types. The Poisson process is a continuous-time process which allows discrete (or dis-continuous) changes in the variables. The simplest independent Poisson process defines the probability of an event occurring during a time interval of length \(h\) (where \(h\) is as small as you like) as follows,

\[
\begin{align*}
\text{Prob } \{ \text{the event does not occur in the time interval } (t, t+h) \} &= 1 - \lambda h + O(h) \\
\text{Prob } \{ \text{the event occurs once in the time interval } (t, t+h) \} &= \lambda h + O(h) \\
\text{Prob } \{ \text{the event occurs more than once in the time interval } (t, t+h) \} &= O(h)
\end{align*}
\]

where \(O(h)\) is the asymptotic order symbol defined by

\[
\psi(h) \text{ is } O(h) \text{ if } \lim_{h \to 0} \left( \frac{\psi(h)}{h} \right) = 0
\]

and \(\lambda\) = the mean number of occurrences per unit time.

Given the Poisson process, the "event" can be defined in a number of interesting ways. To illustrate the degree of

\(^{23}\)As Hakansson [6] has pointed out, (70) and (71) are consistent with the Friedman Permanent Income and the Modigliani Life-cycle hypotheses. However, in general, this result will not hold.
latitude, three examples of applications of Poisson processes in the consumption-portfolio choice problem are presented below. Before examining these examples, it is first necessary to develop some of the mathematical properties of Poisson processes. There is a theory of stochastic differential equations for Poisson processes similar to the one for Brownian motion discussed in section two. Let \( q(t) \) be an independent Poisson process with probability structure as described in (72). Let the event be that a state variable \( x(t) \) has a jump in amplitude of size \( \delta \) where \( \delta \) is a random variable whose probability measure has compact support. Then a Poisson differential equation for \( x(t) \) can be written as

\[
(74) \quad dx = f(x,t)dt + g(x,t)dq 
\]

and the corresponding differential generator, \( \mathcal{L}_x \), is defined by

\[
(75) \quad \mathcal{L}_x[h(x,t)] = h_t + f(x,t)h_x + E_t[\lambda[h(x + \delta g,t) - h(x,t)]] 
\]

where "\( E_t \)" is the conditional expectation over the random variable \( \delta \), conditional on knowing \( x(t) = x \), and where \( h(x,t) \) is a \( C^1 \) function of \( x \) and \( t \).(24) Further, theorem I holds for Poisson processes.(25)

Returning to the consumption-portfolio problem, consider first the two-asset case. Assume that one asset is a common stock whose price is log-normally distributed and that the other asset is a "risky" bond which pays an instantaneous rate of interest \( r \) when not in default but in the event of default, the price of the bond becomes zero.(26)

---

\( ^{(24)} \) For a short discussion of Poisson differential equations and a proof of (75) as well as other references, see Kushner [9], pages 18-22.

\( ^{(25)} \) See Dreyfus [4], p. 225 and Kushner [9], chapter IV.

\( ^{(26)} \) That the price of the bond is zero in the event of default is an extreme assumption made only to illustrate how default can be treated in the analysis. One could make the more reasonable assumption that the price in the event of default is a random variable. The degree of computational difficulty caused by this more reasonable assumption will depend on the choice of distribution for the random variable as well as the utility function of the individual.
From (74), the process which generates the bond's price can be written as

\[ \text{d}P = rP \text{d}t - P \text{d}q \]  

where \( \text{d}q \) is as previously defined and \( \mathcal{F} \equiv 1 \) with probability one. Substituting the explicit price dynamics into (14'), the budget equation becomes

\[ \text{d}W = \{wW(a-r) + rW-C\} \text{d}t + w\sigma \text{d}z - (1-w)\text{d}q \]  

From (75), (77), and theorem I, we have that the optimality equation can be written as

\[ 0 = U(C^*, t) + J_t(W, t) + \lambda [J(W^*, W, t) - J(W, t)] \]

\[ + J_{WW}(W, t) \left[ (W^*(a-r) + r)W - C^* \right] + \frac{1}{2} J_{WW}(W, t) \sigma^2 W^* W \]

where \( C^* \) and \( w^* \) are determined by the implicit equations

\[ 0 = U_C(C^*, t) - J_W(W, t) \]

and

\[ 0 = \lambda J_W(w^*, W, t) + J_W(W, t)(a-r) + J_{WW}(W, t) \sigma^2 W^* W. \]

To see the effect of default on the portfolio and consumption decisions, consider the particular case when \( U(C, t) \equiv C^{\gamma}/\gamma \), for \( \gamma < 1 \). The solutions to (79) and (80) are

\[ C^*(t) = \frac{A W(t)}{(1-\gamma)} \left( 1-\exp[A(t-T)/1-\gamma] \right) \]

where

\[ A = -\gamma \left[ \frac{(a-r)^2}{2\sigma^2(1-\gamma)} + r \right] + \lambda \left[ 1 - \frac{(2-\gamma)w^\gamma - \gamma(a-r)}{\gamma 2\sigma^2(1-\gamma)} \right] \]

and

\[ w^* = \frac{(a-r)}{\sigma^2(1-\gamma) + \sigma^2(1-\gamma)} \]

\[ (w^*)^{\gamma-1} \]

As might be expected, the demand for the common stock is an increasing function of \( \lambda \) and, for \( \lambda > 0 \), \( w^* > 0 \) holds for all values of \( a, r \), or \( \sigma^2 \).

\[ ^{(27)} \text{Note that (79') and (80') with } \lambda = 0 \text{ reduce to the solutions (48) and (49) when } \eta = \rho = 0 \text{ and } \beta = 1 - \gamma. \]
For the second example, consider an individual who receives a wage, \( Y(t) \), which is incremented by a constant amount \( \epsilon \) at random points in time. Suppose that the event of a wage increase is distributed Poisson with parameter \( \lambda \). Then, the dynamics of the wage-rate state variable are described by

\[
dY = \epsilon dq, \quad \text{with} \quad \mathcal{J} = 1 \quad \text{with probability one.}
\]

Suppose further that the individual's utility function is of the form \( U(C,t) = e^{-\rho t}V(C) \) and that his time horizon is infinite (i.e., \( T = \infty \)). (28) Then, for the two-asset case of section six, the optimality equation can be written as

\[
0 = V(C^*) - \rho I(W,Y) + \lambda [I(W,Y + \epsilon) - I(W,Y)] \\
+ I_W(W,Y) \left[ (W^*(\alpha-r) + r)W + Y - C^* \right] + \frac{1}{2} I_{WW}(W,Y)\sigma^2 W^*^2 W^2
\]

where \( I(W,Y) = e^{\rho t}J(W,Y,t) \). If it is further assumed that \( V(C) = e^{-\eta C}/\eta \), then the optimal consumption and portfolio rules, derived from (83), are

\[
C^*(t) = r \left[ W(t) + \frac{Y(t)}{r} + \frac{\lambda}{\eta r} \left( 1 - e^{\eta \epsilon} \right) \right] + \frac{1}{\eta r} \left[ \rho - r + \frac{(\alpha-r)^2}{2\sigma^2} \right]
\]

and

\[
w^*(t)W(t) = \frac{(\alpha-r)}{\eta \sigma^2 r}
\]

In (84), \( \left[ W(t) + Y(t)/r + \lambda \left( 1 - e^{\eta \epsilon} \right)/\eta r^2 \right] \) is the general wealth term, equal to the sum of present wealth and capitalized future wage earnings. If \( \lambda = 0 \), then (84) reduces to (70) in

I have shown elsewhere [(12), p.252] that if \( U = e^{-\rho t}V(C) \) and \( U \) is bounded or \( \rho \) sufficiently large to ensure convergence of the integral and if the underlying stochastic processes are stationary, then the optimality equation (18) can be written, independent of explicit time, as

\[
0 = \max_{\{C,W\}} \left[ V(C) + \mathcal{J} [I] \right]
\]

where \( \mathcal{J} = \mathcal{J} - \rho - \frac{\partial}{\partial t} \) and \( I(W,P) = e^{\rho t}J(W,P,t) \).

A solution to (82) is called the "stationary" solution to the consumption-portfolio problem. Because the time state variable is eliminated, solutions to (82) are computationally easier to find than for the finite-horizon case.
section seven, where the wage rate was fixed and known with certainty. When \( \lambda > 0 \), \( \frac{\lambda (1-e^{-\eta \epsilon})}{\eta r^2} \) is the capitalized value of (expected) future increments to the wage rate, capitalized at a somewhat higher rate than the risk-free market rate reflecting the risk-aversion of the individual.\(^{29}\)

Let \( X(t) \) be the "Certainty-equivalent wage rate at time \( t \)" defined as the solution to

\[
U[X(t)] = E_0 U[Y(t)].
\]

For this example, \( X(t) \) is calculated as follows

\[
\frac{e^{-\eta X(t)}}{\eta} = \frac{1}{\eta} E_0 e^{-\eta Y(t)}
\]

\[
= \frac{1}{\eta} e^{-\eta Y(o)} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} e^{-\eta k \epsilon}
\]

\[
= \frac{1}{\eta} e^{-\eta Y(o)} - \lambda t + \lambda e^{-\eta \epsilon} t
\]

Solving for \( X(t) \) from (87), we have that

\[
X(t) = Y(o) + \lambda t \left(1-e^{-\eta \epsilon}\right)/\eta.
\]

The capitalized value of the (certainty equivalent wage income flow is

\[
\int_0^\infty e^{-rs} X(s) ds = \int_0^\infty Y(o)e^{-rs} ds + \int_0^\infty \frac{\lambda (1-e^{-\eta \epsilon})}{\eta} se^{-rs} ds
\]

\[
= \frac{Y(o)}{r} + \frac{\lambda (1-e^{-\eta \epsilon})}{\eta r^2}
\]

Thus, for this example,\(^{30}\) the individual, in computing the

\(^{29}\) The usual expected present discounted value of the increments to the wage flow is \( E_t \int_t^\infty e^{-r(s-t)} [Y(s) - Y(t)] ds \)

\[
= \int_t^\infty \lambda \epsilon e^{-r(s-t)} (s-t) ds = \lambda \epsilon / r^2, \text{ which is greater than}
\]

\[
\frac{\lambda (1-e^{-\eta \epsilon})}{\eta r^2} \text{ for } \epsilon > 0.
\]

\(^{30}\) The reader should not infer that this result holds in general. Although (86) is a common definition of Certainty-equivalent in one-period utility-of-wealth models, it is not satisfactory for dynamic consumption-portfolio models. The reason it works for this example is due to the particular relationship between the \( J \) and \( U \) functions when \( U \) is exponential.
present value of future earnings, determines the Certainty-equivalent flow and then capitalizes this flow at the (certain) market rate of interest.

The third example of a Poisson process differs from the first two because the occurrence of the event does not involve an explicit change in a state variable. Consider an individual whose age of death is a random variable. Further assume that the event of death at each instant of time is an independent Poisson process with parameter \( \lambda \). Then, the age of death, \( \tau \), is the first time that the event (of death) occurs and is an exponentially distributed random variable with parameter \( \lambda \). The optimality criterion is to

\[
\text{Max } E_O \{ \int_0^T U(C_t, t) dt + B(W(\tau), \tau) \}
\]

and the associated optimality equation is

\[
0 = U(C^*, t) + \lambda [B(W, t) - J(W, t)] + \mathcal{L}[J].
\]

To derive (91), an "artificial" state variable, \( x(t) \), is constructed with \( x(t) = 0 \) while the individual is alive and \( x(t) = 1 \) in the event of death. Therefore, the stochastic process which generates \( x \) is defined by

\[
dx = dq \text{ and } \mathcal{F} = 1 \text{ with probability one}
\]

and \( \tau \) is now defined by \( x \) as

\[
\tau = \min \{ t | t > 0 \text{ and } x(t) = 1 \}.
\]

The derived utility function, \( J \), can be considered a function of the state variables \( W, x, \) and \( t \) subject to the boundary condition

\[
J(W, x, t) = B(W, t) \text{ when } x = 1.
\]

In this form, example three is shown to be of the same type as examples one and two in that the occurrence of the Poisson event causes a state variable to be incremented, and (91) is of the same form as (78) and (83).

A comparison of (91) for the particular case when \( B = 0 \) (no bequests) with (82) suggested the following theorem. \(^{31}\)

\(^{31}\) I believe that a similar theorem has been proved by (continued)
Theorem VI. If \( \tau \) is as defined in (93) and \( U \) is such that the integral \( E_0 [\int_0^\tau U(C,t)dt] \) is absolutely convergent, then the maximization of \( E_0 [\int_0^\tau U(C,t)dt] \) is equivalent to the maximization of \( \mathcal{E}_0 [\int_0^\infty e^{-\lambda t}U(C,t)dt] \) where \( \mathcal{E}_0 \) is the conditional expectation operator over all random variables including \( \tau \) and \( \mathcal{E}_0 \) is the conditional expectation operator over all random variables excluding \( \tau \).

Proof: \( \tau \) is distributed exponentially and is independent of the other random variables in the problem. Hence, we have that

\[
E_0 [\int_0^\tau U(C,t)dt] = \int_0^\infty \lambda e^{-\lambda t} \mathcal{E}_0 \int_0^\tau U(C,t)dt
\]

\[
= \int_0^\infty \int_0^\tau \lambda g(t) e^{-\lambda t} dt d\tau
\]

where \( g(t) \equiv \mathcal{E}_0 [U(C,t)] \). Because the integral in (95) is absolutely convergent, the order of integration can be interchanged, i.e., \( \mathcal{E}_0 \int_0^\tau U(C,t)dt = \int_0^\tau \mathcal{E}_0 U(C,t)dt \). By integration by parts, (95) can be re-written as

\[
\int_0^\infty \int_0^\tau e^{-\lambda t} g(t) dt d\tau = \int_0^\infty e^{-\lambda s} g(s) ds
\]

\[
= \mathcal{E}_0 \int_0^\infty e^{-\lambda t} U(C,t) dt. \quad Q.E.D.
\]

Thus, an individual who faces an exponentially-distributed uncertain age of death acts as if he will live forever, but with a subjective rate of time preference equal to his "force of mortality", i.e. to the reciprocal of his life expectancy.

9. Alternative price expectations to the geometric Brownian motion. The assumption of the geometric Brownian motion hypothesis is a rich one because it is a reasonably good

---

(continued)

model of observed stock price behavior and it allows the proof of a number of strong theorems about the optimal consumption-portfolio rules, as was illustrated in the previous sections. However, as mentioned in the Introduction, there have been some disagreements with the underlying assumptions required to accept this hypothesis. The geometric Brownian motion hypothesis best describes a stationary equilibrium economy where expectations about future returns have settled down, and as such, really describes a "long-run" equilibrium model for asset prices. Therefore, to explain "short-run" consumption and portfolio selection behavior one must introduce alternative models of price behavior which reflect the dynamic adjustment of expectations.

In this section, alternative price behavior mechanisms are postulated which attempt to capture in a simple fashion the effects of changing expectations, and then comparisons are made between the optimal decision rules derived under these mechanisms with the ones derived in the previous sections. The choices of mechanisms are not exhaustive nor are they necessarily representative of observed asset price behavior. Rather they have been chosen as representative examples of price adjustment mechanisms commonly used in economic and financial models.

Little can be said in general about the form of a solution to (28) when \( a_k \) and \( \sigma_k \) depend in an arbitrary manner on the price levels. If it is specified that the utility function is a member of the HARA family, i.e.

\[
U(C,t) = \frac{(1-\gamma)}{\gamma} F(t) \left( \frac{BC}{1-\gamma} + \eta \right) ^\gamma
\]

subject to the restrictions in (42), then (28) can be simplified because \( J(W,P,t) \) is separable into a product of functions, one depending on \( W \) and \( t \), and the other on \( P \) and \( t \).\(^{32}\)

\(^{32}\) This separability property was noted in [1], [5], [6], [10], [12], and [13]. It is assumed throughout this section that the bequest function satisfies the conditions of footnote 20.
In particular, if we take \( J(W,P,t) \) to be of the form

\[
J(W,P,t) = \frac{(1-\gamma)}{\gamma} H(P,t)P(t) \left( \frac{W}{1-\gamma} + \frac{n}{\delta r} \left[ 1-e^{r(t-T)} \right] \right)^{\gamma},
\]

substitute for \( J \) in (28), and divide out the common factor \( F(t) \left( \frac{W}{1-\gamma} + \left[ \frac{n}{\delta r} \left[ 1-e^{r(t-T)} \right] \right] \right)^{\gamma} \), then we derive a "reduced" equation for \( H \),

\[
0 = \frac{(1-\gamma)^2}{\gamma} \left( \frac{H}{\beta} \right)^{\gamma-1} + \frac{(1-\gamma)}{\gamma} \left( \frac{F}{F} + H_t \right) + (1-\gamma)rH
\]

\[+ \frac{(1-\gamma)}{\gamma} \sum_{i=1}^{m} \alpha_i P_i H_i + \frac{(1-\gamma)}{2\gamma} \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_{ij} P_i P_j H_i H_j \]

\[+ \sum_{i=1}^{m} (\alpha_i-r) P_i H_i + \frac{H}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} v_{ij} (\alpha_i-r)(\alpha_j-r) \]

\[+ \frac{1}{2H} \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_{ij} P_i P_j H_i H_j \]

and the associated optimal consumption and portfolio rules are

\[
C^*(t) = \frac{(1-\gamma)}{\beta} \left[ \left( \frac{H}{\beta} \right)^{\gamma-1} \left( \frac{W}{1-\gamma} + \frac{n}{\delta r} \left[ 1-e^{r(t-T)} \right] \right) \right] - \eta
\]

and

\[
w_k^*(t)W = \left[ \sum_{j=1}^{m} v_{jk} (\alpha_j-r) + \frac{H_k P_k}{H} \left( \frac{W}{1-\gamma} + \frac{n}{\delta r} \left[ 1-e^{r(t-T)} \right] \right) \right]
\]

\[k=1,\ldots,m.\]

Although (99) is still a formidable equation from a computational point of view, it is less complex than the general equation (28), and it is possible to obtain an explicit solution for particular assumptions about the dependence of \( \alpha_k \) and \( \sigma_k \) on the prices. Notice that both consumption and the asset demands are linear functions of wealth.

For a particular member of the HARA family, namely the Bernoulli logarithmic utility (\( \gamma=0=\eta \) and \( \beta=1-\gamma=1 \)) function, (28) can be solved in general. In this case, \( J \) will be of the form

\[
J(W,P,t) = a(t) \log W + H(P,t) \quad \text{with} \quad h(P,t) = a(T) = 0,
\]
with \( a(t) \) independent of the \( \alpha_k \) and \( \sigma_k \) (and hence, the \( P_k \)).

For the case when \( F(t) \equiv 1 \), we find \( a(t) = T-t \) and the optimal rules become

\[
C^* = \frac{W}{T-t}
\]

and

\[
w_k^* = \sum_{j=1}^{m} v_{kj} (\alpha_j - r), \quad k=1, \ldots, m.
\]

For the log case, the optimal rules are identical to those derived when \( \alpha_k \) and \( \sigma_k \) were constants, with the understanding that the \( \alpha_k \) and \( \sigma_k \) are evaluated at current prices. Hence, although we can solve this case for general price mechanisms, it is not an interesting one because different assumptions about price behavior have no effect on the decision rules.

The first of the alternative price mechanisms considered is called the "asymptotic 'normal' price-level" hypothesis which assumes that there exists a "normal" price function, \( F(t) \), such that

\[
\lim_{t \to \infty} E_T [P(t)/F(t)] = 1, \quad \text{for } 0 \leq T < t < \infty,
\]

i.e. independent of the current level of the asset price, the investor expects the "long-run" price to approach the normal price. A particular example which satisfies the hypothesis is that

\[
F(t) = F(o) e^{\nu t}
\]

and

\[
\frac{dP}{P} = \beta \left[ \phi + \nu t - \log \left( \frac{P(t)}{P(o)} \right) \right] dt + \sigma dz
\]

where \( \phi = k + \nu / \beta + \sigma^2 / 2 \beta \) and \( k = \log \left( \frac{F(o)}{P(o)} \right) \).

For the purpose of analysis, it is more convenient to work with the variable \( Y(t) = \log \left( \frac{P(t)}{P(o)} \right) \) rather than \( P(t) \). Substituting for \( P \) in (107) by using Itô's Lemma, we can write the dynamics for \( Y \) as

\[33\]

In the notation used in previous sections, (107) corresponds to (5) with \( a(P, t) = \beta \left[ \phi + \nu t - \log \left( \frac{P(t)}{P(o)} \right) \right] \). Note: "normal" does not mean "Gaussian" in the above use, but rather the normal long-run price of Alfred Marshall.
(108) \( \text{d}Y = \beta [\mu + \nu t - Y] \text{d}t + \sigma \text{d}z \)
where \( \mu = \phi - \sigma^2/2\beta \). Before examining the effects of this price mechanism on the optimal portfolio decisions, it is useful to investigate the price behavior implied by (106) and (107). (107) implies an exponentially-regressive price adjustment toward a normal price, adjusted for trend. By inspection of (108), \( Y \) is a normally-distributed random variable generated by a Markov process which is not stationary and does not have independent increments. \(^{34}\) Therefore, from the definition of \( Y \), \( P(t) \) is log-normal and Markov. Using Itô's Lemma, one can solve (108) for \( Y(t) \), conditional on knowing \( Y(T) \), as

(109) \( Y(t) - Y(T) = \left( k + \nu T - \frac{\sigma^2}{4\beta} - Y(T) \right) (1 - e^{-\beta T}) + \nu \text{t} + \sigma e^{-\beta t} \int_T^t e^{\beta z} \text{d}z \)
where \( \tau = t - T \geq 0 \). The instantaneous conditional variance of \( Y(t) \) is

(110) \( \text{Var} [Y(t) | Y(T)] = \frac{\sigma^2}{2\beta} \left(1 - e^{-2\beta T}\right) \).

Given the characteristics of \( Y(t) \), it is straightforward to derive the price behavior. For example, the conditional expected price can be derived from (110) and written as

(111) \( E_T(P(t)/P(T)) = E_T \exp [Y(t) - Y(T)] \)

\[
= \exp \left[ \left( k + \nu T - \frac{\sigma^2}{4\beta} - Y(T) \right) (1 - e^{-\beta T}) + \nu T + \frac{\sigma^2}{4\beta} \left(1 - e^{-2\beta T}\right) \right]
\]

It is easy to verify that (105) holds by applying the appropriate limit process to (111). Figure 2 illustrates the behavior of the conditional expectation mechanism over time.

\(^{34}\) Processes such as (108) are called Ornstein-Uhlenbeck processes and are discussed, for example, in [3], p. 225.
figure 2.

For computational simplicity in deriving the optimal consumption and portfolio rules, the two-asset model is used with the individual having an infinite time horizon and a constant absolute risk-aversion utility function, \(U(C,t) = -\frac{e^{-\gamma C}}{n}\).

The fundamental optimality equation then is written as

\[
0 = -e^{-\gamma C} + J_t + J_W \left[ w^*(\beta(\phi + \nu t - Y) - r)W + rW - C^* \right] + \frac{1}{2} J_{WW} w^* W^2 \sigma^2 + J_Y \beta(\mu + \nu t - Y) + \frac{1}{2} J_{YY} \sigma^2 + J_{YW} w^* W \sigma^2
\]

and the associated equations for the optimal rules are

\[
(113) \quad w^* W = -J_W \left[ \beta(\phi + \nu t - Y) - r \right]/J_{WW} \sigma^2 - J_{YW}/J_{WW}
\]

and

\[
(114) \quad C^* = -\log \left(\frac{J_W}{n}\right)
\]

Solving (112), (113), and (114), we write the optimal rules in explicit form as

\[
(115) \quad w^* W = \frac{1}{\eta r \sigma^2} \left[ \left(1 + \frac{\beta}{r}\right) (\alpha(P,t) - r) + \frac{\beta^2}{r^2} \left( \frac{\sigma^2}{2} + \nu - r \right) \right]
\]
and

\[(116) \quad C^* = rW + \frac{\beta^2}{2\sigma^2 \eta r^2} \gamma^2 - \frac{\beta}{\eta r^2} \left( \beta v t + \beta \phi - r + \beta \left( \nu + \frac{\sigma^2}{2} - r \right) \right) \gamma + a(t)\]

where \(a(P, t)\) is the instantaneous expected rate of return defined explicitly in footnote 33. To provide a basis for comparison, the solutions when the geometric Brownian motion hypothesis is assumed are presented as

\[(117) \quad w^* = \frac{(a-r)}{\eta r^2}\]

and

\[(118) \quad C^* = rW + \frac{1}{\eta r^2} \left[ \frac{(a-r)^2}{2\sigma^2} - r \right] \]

To examine the effects of the alternative "normal price" hypothesis on the consumption-portfolio decisions, the (constant) \(a\) of (117) and (118) is chosen equal to \(a(P, t)\) of (115) and (116) so that, in both cases, the instantaneous expected return and variance are the same at the point of time of comparison. Comparing (115) with (117), we find that the proportion of wealth invested in the risky asset is always larger under the "normal price" hypothesis than under the geometric Brownian motion hypothesis. In particular, notice that even if \(a < r\), unlike in the geometric Brownian motion case, a positive amount of the risky asset is held. Figures 3a and 3b illustrate the behavior of the optimal portfolio holdings.

\[a(t) = \frac{1}{n} \left\{ \frac{r}{2\sigma^2} - 1 + \frac{\beta}{\sigma^2} \left( \phi + 1 - \frac{\sigma^2}{2r} \right) + \frac{\beta^2}{r \sigma^2} \left[ (1 - \frac{\sigma^2}{2r}) \left( \phi + \frac{\nu}{r} + \frac{\sigma^2}{2r} - 1 \right) \right. \right. \]

\[\left. - \frac{\phi^2}{2} - \frac{\sigma^2}{2r} \right] + \frac{\beta \nu}{\sigma^2 r^2} \left( \frac{r}{2r} - \beta - \beta \phi \right) - \frac{\beta^2 \nu^2}{2 \sigma^2 r^2} \]

\[+ \frac{\beta v t}{r \sigma^2} \left[ \frac{r}{2r} - \beta - \beta \phi - \frac{\beta \sigma^2}{2r} \right] - \frac{\beta^2 \nu^2 t^2}{2 \sigma^2 r^2} \right\}.

\[\text{Footnote 35}\]

\[\text{Footnote 36}\]

\[\text{Footnote 37}\]

For a derivation of (117) and (118), see [12], p. 256.

It is assumed that \(\nu + \frac{\sigma^2}{2} > r\), i.e. the "long-run" rate of growth of the "normal" price is greater than the sure rate of interest so that something of the risky asset will be held in the short and long run.
The most striking feature of this analysis is that, despite the ability to make continuous portfolio adjustments, a person who believes that prices satisfy the "normal" price hypothesis will hold more of the risky asset than one who believes that prices satisfy the geometric Brownian motion hypothesis, even though they both have the same utility function and the same expectations about the instantaneous mean and variance.

The primary interest in examining these alternative price mechanism is to see the effects on portfolio behavior, and so, little will be said about the effects on consumption other than to present the optimal rule.

The second alternative price mechanism assumes the same type of price-dynamics equation as was assumed for the geometric Brownian motion: namely,

\[ \frac{dP}{P} = \alpha dt + \sigma dz. \]
However, instead of the instantaneous expected rate of return, \( \alpha \), being a constant, it is assumed that \( \alpha \) is itself generated by the stochastic differential equation

\[
(120) \quad d \alpha = \beta (\mu - \alpha) dt + \delta \left( \frac{dP}{P} - \alpha dt \right) \\
= \beta (\mu - \alpha) dt + \delta \sigma dz.
\]

The first term in (120) implies a long-run, regressive adjustment of the expected rate of return toward a "normal" rate of return, \( \mu \), where \( \beta \) is the speed of adjustment. The second term in (120) implies a short-run, extrapolative adjustment of the expected rate of return of the "error-learning" type, where \( \delta \) is the speed of adjustment. I will call the assumption of a price mechanism described by (119) and (120) the "De Leeuw" hypothesis for Frank De Leeuw who first introduced this type mechanism to explain interest rate behavior.

To examine the price behavior implied by (119) and (120), we first derive the behavior of \( \alpha \), and then \( P \). The equation for \( \alpha \), (120), is of the same type as (108) described previously. Hence, \( \alpha \) is normally distributed and is generated by a Markov process. The solution of (120), conditional on knowing \( \alpha(T) \) is

\[
(121) \quad \alpha(t) - \alpha(T) = \left( \mu - \alpha(T) \right) \left( 1 - e^{-\beta \tau} \right) + \delta \sigma e^{-\beta \tau} \int_T^t e^{\beta z} dz,
\]

where \( \tau \equiv t - T > 0 \). From (121), the conditional mean and variance of \( \alpha(t) - \alpha(T) \) are

\[
(122) \quad E_T \left( \alpha(t) - \alpha(T) \right) = \left( \mu - \alpha(T) \right) \left( 1 - e^{-\beta \tau} \right)
\]

and

\[
(123) \quad \text{Var} \left[ \alpha(t) - \alpha(T) \mid \alpha(T) \right] = \frac{\delta^2 \sigma^2}{2 \beta} \left( 1 - e^{-2 \beta \tau} \right).
\]

To derive the dynamics of \( P \), note that, unlike \( \alpha \), \( P \) is not Markov although the joint process \([P, \alpha]\) is. Combining the results derived for \( \alpha(t) \) with (119), we solve directly
for the price, conditional on knowing \( P(T) \) and \( \alpha(T) \),

\[
Y(t) - Y(T) = \left( \mu - \frac{1}{2}\sigma^2 \right) \tau - \frac{\mu - \alpha(T)}{\beta} (1 - e^{-\beta \tau}) \\
+ \sigma \delta \int_T^t \int_T^s e^{-\beta (s-s')} ds' dz' ds + \sigma \int_T^t dz,
\]

where \( Y(t) \equiv \log [P(t)] \). From (124), the conditional mean and variance of \( Y(t) - Y(T) \) are

\[
E_T [Y(t) - Y(T)] = \left( \mu - \frac{1}{2}\sigma^2 \right) \tau - \frac{\mu - \alpha(T)}{\beta} (1 - e^{-\beta \tau})
\]

and

\[
\text{Var} [Y(t) - Y(T) | Y(T)] = \sigma^2 \tau + \frac{\sigma^2 \delta^2}{2\beta^3} \left[ \beta \tau - 2(1 - e^{-\beta \tau}) \right] \\
+ \frac{\sigma^2}{2} \left[ (1 - e^{-\beta \tau}) \right] + \frac{2\delta \sigma^2}{\beta^2} \left[ \beta \tau - (1 - e^{-\beta \tau}) \right].
\]

Since \( P(t) \) is log-normal, it is straightforward to derive the moments for \( P(t) \) from (124), (125), and (126). Figure 4 illustrates the behavior of the expected price mechanism.

The equilibrium or "long-run" (i.e. \( \tau \rightarrow \infty \)) distribution for \( \alpha(t) \) is stationary gaussian with mean \( \mu \) and variance \( \frac{\delta^2 \sigma^2}{2\beta} \), and the equilibrium distribution for \( P(t)/P(T) \) is a stationary log-normal. Hence, the long-run behavior of prices under the De Leeuw hypothesis approaches the geometric Brownian motion.
Again, the two-asset model is used with the individual having an infinite time horizon and a constant absolute risk-aversion utility function, \( U(C,t) = -\frac{e^{-\eta C}}{\eta} \). The fundamental optimality equation is written as

\[
0 = -\frac{e^{-\eta C^*}}{\eta} + J_t + J_W \left[ w^*(\alpha - r)W + rW - C^* \right] \\
+ \frac{1}{2} J_{WW} w^{*2} W^2 \sigma^2 + J_{\alpha} \beta (\mu - \alpha) + \frac{1}{2} J_{\alpha\sigma^2} \sigma^2 \\
+ J_{W\sigma^2} w^* W
\]

Notice that the state variables of the problem are \( W \) and \( \sigma \), which are both Markov, as is required for the dynamic programming technique. The optimal portfolio rule derived from (127) is,

\[
w^*W = -\frac{J_W(\alpha - r)}{J_{WW} \sigma^2} - \frac{J_{W\sigma^2} \beta}{J_{WW}}
\]

The optimal consumption rule is the same as in (114). Solving (127) and (128), the explicit solution for the portfolio rule is

\[
w^*W = \frac{1}{\eta r \sigma^2 (r + 2\delta + 2\beta)} \left[ (r + \delta + 2\beta)(\alpha - r) - \frac{\delta \beta (\mu - r)}{r + \delta + \beta} \right]
\]

Comparing (129) with (117) and assuming that \( \mu > r \), we find that under the De Leeuw hypothesis, the individual will hold a smaller amount of the risky asset than under the geometric Brownian motion hypothesis. Note also that \( w^*W \) is a decreasing function of the long-run normal rate of return, \( \mu \). The interpretation of this result is that as \( \mu \) increases for a given \( \alpha \), the probability increases that future "\( \alpha \)'s" will be more favorable relative to the current \( \alpha \), and so there is a tendency to hold more of one's current wealth in the risk-free asset as a "reserve" for investment under more favorable conditions.

The last type of price mechanism examined differs from the previous two in that it is assumed that prices satisfy the geometric Brownian motion hypothesis. However, it is also
assumed that the investor does not know the true value of the parameter \( \alpha \), but must estimate it from past data. Suppose \( P \) is generated by equation (119) with \( \alpha \) and \( \sigma \) constants, and the investor has price data back to time \(-\tau\). Then, the best estimator for \( \alpha \), \( \hat{\alpha}(t) \), is

\[
(130) \quad \hat{\alpha}(t) = \frac{1}{t+\tau} \int_{-\tau}^{t} \frac{dP}{P}
\]

where we assume, arbitrarily, that \( \hat{\alpha}(-\tau) = 0 \). From (130), we have that \( \mathbb{E}(\hat{\alpha}(t)) = \alpha \), and so, if we define the error term \( \varepsilon_t \equiv \alpha - \hat{\alpha}(t) \), then (119) can be re-written as

\[
(131) \quad \frac{dP}{P} = \alpha dt + \sigma d\hat{\varepsilon}_t
\]

where \( d\hat{\varepsilon}_t \equiv dz + \varepsilon_t dt/\sigma \). Further, by differentiating (130)
we have the dynamics for \( \alpha \), namely

\[
(132) \quad d\alpha = \frac{\sigma}{t+\tau} d\hat{\varepsilon}_t
\]

Comparing (131) and (132) with (119) and (120), we see that this "learning" model is equivalent to the special case of the De Leeuw hypothesis of pure extrapolation (i.e. \( \beta = 0 \)) where the degree of extrapolation (\( \delta \)) is decreasing over time. If the two-asset model is assumed with an investor who lives to time \( T \) with a constant absolute risk-aversion utility function, and if (for computational simplicity) the risk-free asset is money (i.e. \( r = 0 \)), then the optimal portfolio rule is

\[
(133) \quad w^*W = \frac{(T+\tau)}{\eta \sigma^2} \log\left(\frac{T+\tau}{t+\tau}\right) \hat{\alpha}(t)
\]

and the optimal consumption rule is

\[
(134) \quad C^* = \frac{W}{T-t} - \frac{1}{\eta} \left[ \log(T+\tau) + \frac{2}{T-t} \left( T-t - (T+\tau) \log(T+\tau) \right) + (T+\tau) \log(T+\tau) + \frac{\delta^2}{2\sigma^2} \left( \frac{(T+\tau)^2}{T-t} \log\left[ \frac{T+\tau}{t+\tau}\right] - \frac{T-t}{T+\tau} \right) \right]
\]

By differentiating (133) with respect to \( t \), we find that \( w^*W \) is an increasing function of time for \( t < T \), reaches a
maximum at \( t = \bar{t} \), and then is a decreasing function of time for \( \bar{t} < t < T \), where \( \bar{t} \) is defined by

\[(135) \quad \bar{t} = [T + (1-e)\tau]/e. \]  
The reason for this behavior is that, early in life (i.e. for \( t < \bar{t} \)), the investor learns more about the price equation with each observation, and hence investment in the risky asset becomes more attractive. However, as he approaches the end of life (i.e. for \( t > \bar{t} \)), he is generally liquidating his portfolio to consume a larger fraction of his wealth, so that although investment in the risky asset is more favorable, the absolute dollar amount invested in the risky asset declines.

Consider the effect on (133) of increasing the number of available previous observations (i.e. increase \( \tau \)). As expected, the dollar amount invested in the risky asset increases monotonically. Taking the limit of (133) as \( \tau \to \infty \), we have that the optimal portfolio rule is

\[(136) \quad w*W = \left( \frac{T-t}{\eta \sigma^2} \right) \alpha \quad \text{as } \tau \to \infty \]

which is the optimal rule for the geometric Brownian motion case when \( \alpha \) is known with certainty. Figure 5. illustrates graphically how the optimal rule changes with \( \tau \).

![Figure 5](image-url)

Figure 5.
10. Conclusion. By the introduction of Itô's Lemma and the Fundamental Theorem of Stochastic Dynamic Programming (Theorem I), we have shown how to systematically construct and analyze optimal continuous-time dynamic models under uncertainty. The basic methods employed in studying the consumption-portfolio problem are applicable to a wide-class of economic models of decision making under uncertainty.

A major advantage of the continuous-time model over its discrete time analog is that one need only consider two types of stochastic processes: functions of Brownian motions and Poisson processes. This result limits the number of parameters in the problem and allows one to take full advantage of the enormous amount of literature written about these processes. Although I have not done so here, it is straightforward to show that the limits of the discrete-time model solutions as the period spacing goes to zero are the solutions of the continuous-time model.¹³⁸¹

A basic simplification gained by using the continuous-time model to analyze the consumption-portfolio problem is the justification of the Tobin-Markowitz portfolio efficiency conditions in the important case when asset price changes are stationarily and log-normally distributed. With earlier writers (Hakansson [6], Leland [10], Fischer [5], Samuelson [13], and Cass and Stiglitz [1]), we have shown that the assumption of the HARA utility function family simplifies the analysis and a number of strong theorems were proved about the optimal solutions. The introduction of stochastic wage income, risk of default, uncertainty about life expectancy, and alternative types of price dynamics serve to illustrate the power of the techniques as well as to provide insight into the effects of these complications on the optimal rules.

¹³⁸¹ For a general discussion of this result, see Samuelson [14].
References


II. LIFETIME PORTFOLIO SELECTION UNDER UNCERTAINTY: THE CONTINUOUS-TIME CASE*

I. Introduction. Most models of portfolio selection have been one-period models. I examine the concept of optimal portfolio selection and consumption by an individual in a continuous-time model where his income is generated by returns on assets and these returns have "growth rates" that are stochastic. P.A. Samuelson has a similar model in discrete-time for more general distributions in a companion paper. [8]

I derive the optimality equations for a model when the rate of returns are generated by a Wiener motion process. A particular case examined is a two-asset model with constant relative risk-aversion and elastic marginal utility. An explicit solution for the case of constant absolute risk-aversion technique employed can be used to examine a wide intertemporal economic problems under uncertainty.

In addition to the Samuelson paper [8], a period analysis of Tobin [9], Phelps [6] has determined the optimal consumption rule for a model where income is partly generated by an asset with a return. Mirrlees [5] has developed a continuous consumption model of the neoclassical type with a random variable.

II. Dynamics of the Model: The Budget Constraint

the usual continuous-time model under certainty equation is a differential equation. However, the budget equation is introduced by a random variable, the budget

* This work was done during the tenure of a National Education Act Fellowship. Aid from the National Science Foundation is gratefully acknowledged. I am grateful to A. Samuelson for many discussions and his help. I wish to thank Stanley Fischer, Massachusetts Institute of Technology, for his comments on section 7 and for his criticism of an earlier version.
be generalized to become a stochastic differential equation. To see the meaning of such an equation, it is easiest to work out the discrete-time version and then pass to the limit of continuous time.

Define

\[ W(t) = \text{total wealth at time } t \]
\[ P_i(t) = \text{price of the } i^{th} \text{ asset at time } t, \ (i=1,\ldots,m) \]
\[ C(t) = \text{consumption per unit time at time } t \]
\[ w_i(t) = \text{proportion of total wealth in the } i^{th} \text{ asset at time } t, \ (i=1,\ldots,m) \]

\[ \left( \sum_{i=1}^{m} w_i(t) = 1 \right) \]

The budget equation can be written as

\[ W(t) = \left[ \sum_{i=1}^{m} w_i(t_0) \frac{P_i(t)}{P_i(t_0)} \right] \]

\[ [W(t_0) - C(t_0)h] \quad (1) \]

where \( t \equiv t_0 + h \) and the time interval between periods is \( h \). By subtracting \( W(t_0) \) from both sides and using \( \sum_{i=1}^{m} w_i(t_0) = 1 \), we can rewrite (1) as,

\[ W(t) - W(t_0) = \left[ \sum_{i=1}^{m} w_i(t_0) \frac{P_i(t) - P_i(t_0)}{P_i(t_0)} \right] \]

\[ = \left[ \sum_{i=1}^{m} w_i(t_0) \right] \frac{g_i(t_0)h}{-1} \]

\[ = \left[ W(t_0) - C(t_0)h \right] - C(t_0)h \quad (2) \]

where

\[ g_i(t_0)h \equiv \log \left[ \frac{P_i(t)}{P_i(t_0)} \right], \]

the rate of return per unit time on the \( i^{th} \) asset. The \( g_i(t) \)
are assumed to be generated by a stochastic process.

In discrete time, I make the further assumption that \( g_i(t) \) is determined as follows,
\[ g_1(t)h = (a_1 - \sigma_1^2/2)h + \Delta Y_1 \]  

(3)

where \( a_1 \), the "expected" rate of return, is constant; and \( Y_1(t) \) is generated by a Gaussian random-walk as expressed by the stochastic difference equation,

\[ Y_1(t) - Y_1(t_0) = \Delta Y_1 = \sigma_1 Z_1(t) \sqrt{h} \]  

(4)

where each \( Z_1(t) \) is an independent variate with a standard normal distribution for every \( t \), \( \sigma_1^2 \) is the variance per unit time of the process \( Y_1 \), and the mean of the increment \( \Delta Y_1 \) is zero.

Substituting for \( g_1(t) \) from (3), we can rewrite (2) as,

\[ W(t) - W(t_0) = \sum_{l=1}^{m} w_1(t_0)(e^{(a_1 - \sigma_1^2/2)h + \Delta Y_1} - 1) \]

\[ (W(t_0) - C(t_0)h) \]

\[ - C(t_0)h. \]  

(5)

Before passing in the limit to continuous time, there are two implications of (5) which will be useful later in the paper.

\[ E(t_0)[W(t) - W(t_0)] = \sum_{l=1}^{m} w_1(t_0)a_1W(t_0) \]

\[ - C(t_0) \]

\[ h + O(h^2) \]  

(6)

and

\[ E(t_0)[(W(t) - W(t_0))^2] = \sum_{i=1}^{m} \sum_{j=1}^{m} w_1(t_0) w_j(t_0) \]

\[ E(t_0)(\Delta Y_1 \Delta Y_j). \]

\[ - - - \]

\[ W^2(t_0) + O(h^2) \]  

(7)

where \( E(t_0) \) is the conditional expectation operator (conditional on the knowledge of \( W(t_0) \)), and \( O(\cdot) \) is the usual asymptotic order symbol meaning "the same order as."

The limit of the process described in (4) as \( h \to 0 \) (continuous time) can be expressed by the formalism of the stochastic
differential equation, \(1\)

\[ dY_i = \sigma_i Z_i(t) \sqrt{dt} \]  

and \( Y_i(t) \) is said to be generated by a Wiener process.

By applying the same limit process to the discrete-time budget equation, we can write (5) as

\[ dW = \left[ \sum_{i=1}^{m} w_i(t) \alpha_i W(t) - C(t) \right] dt + \sum_{i=1}^{m} w_i(t) \sigma_i Z_i(t) W(t) \sqrt{dt}. \]  

The stochastic differential equation (5') is the generalization of the continuous-time budget equation under uncertainty.

A more familiar equation would be the averaged budget equation derived as follows: From (5), we have

\[ E(t_0) \left[ \frac{W(t) - W(t_0)}{h} \right] = \sum_{i=1}^{m} w_i(t_0) \alpha_i \left[ W(t_0) - C(t_0) \right] - C(t_0) + O(h). \]  

Now, take the limit as \( h \to 0 \), so that (8) becomes the following expression for the defined "mean rate of change of wealth":

\[ \dot{\bar{W}}(t_0) = \lim_{h \to 0} E(t_0) \left[ \frac{W(t) - W(t_0)}{h} \right] = \sum_{i=1}^{m} w_i(t_0) \alpha_i W(t_0) - C(t_0). \]  

III. The Two-Asset Model. For simplicity, I first derive the optimal equations and properties for the two-asset model and then, in section 8, display the general equations and results for the m-asset case.

Define

\[ \dot{\bar{W}}(t_0) \]

See K. Ito \[4\], for a rigorous discussion of stochastic differential equations.
\[ w_1(t) = w(t) = \text{proportion invested in the risky asset} \]
\[ w_2(t) = 1 - w(t) = \text{proportion invested in the sure asset} \]
\[ g_1(t) = g(t) = \text{return on the risky asset} \quad \text{(Var } g_1 > 0) \]
\[ g_2(t) = r = \text{return on the sure asset} \quad \text{(Var } g_2 = 0) \]

Then, for \( g(t)h = (\alpha - \sigma^2/2)h + \Delta Y \), equations (5), (6), (7), and (8') can be written as,
\[
W(t) - W(t_0) = [w(t_0)(e^{(\alpha - \sigma^2/2)h + \Delta Y} - 1) \\
+ (1 - w(t_0))(e^{rh} - 1)]. \\
(W(t_0) - C(t_0)h) - C(t_0)h. \quad (9)
\]

\[
E(t_0) [W(t) - W(t_0)] = \left\{ [w(t_0)(\alpha - r) + r]W(t_0) \\
- C(t_0) \right\} h + o(h^2). \quad (10)
\]

\[
E(t_0) [W(t) - W(t_0)]^2 = w^2(t_0)W^2(t_0)E(t_0)[(\Delta Y)^2] \\
+ o(h^2) = w^2(t_0)W^2(t_0)\sigma^2 h \\
+ o(h^2). \quad (11)
\]

\[
dW = [(w(t)(\alpha - r) + r)W(t) - C(t)]dt \\
+ w(t)\sigma Z(t)W(t) \sqrt{dt}. \quad (12)
\]

\[
\dot{W}(t) = [w(t)(\alpha - r) + r]W(t) - C(t). \quad (13)
\]

The problem of choosing optimal portfolio selection and consumption rules is formulated as follows,
\[
\text{Max } E \{ \int_0^T e^{-\rho t} \mathcal{U}[C(t)]dt + B[W(T), T] \} \quad (14)
\]
subject to: the budget constraint (12),
\[ C(t) \geq 0; \quad W(t) > 0; \quad W(0) = W_0 > 0 \]

and where \( U(C) \) is assumed to be a strictly concave utility function (i.e., \( U'(C) > 0; \quad U''(C) < 0 \)); where \( g(t) \) is a random variable generated by the previously described Wiener process. \( B[W(T), T] \) is to be a specified "bequest valuation function" (also referred to in production growth models as the "scrap function," and usually assumed to be concave in \( W(T) \)). "E" in (14) is short for \( E(0) \), the conditional expectation operator, given \( W(0) = W_0 \) as known.

To derive the optimality equations, I restate (14) in a dynamic programming form so that the Bellman principle of optimality \(^{2}\) can be applied. To do this, define,

\[
I[W(t), t] = \max_{\{C(s), W(s)\}} \left[ \int_t^T e^{-ps} U[C(s)] ds + B[W(T), T] \right]
\]

where (15) is subject to the same constraints as (14). Therefore,

\[
I[W(T), T] = B[W(T), T]. \quad (15')
\]

In general, from definition (15),

\[
I[W(t_0), t_0] = \max_{\{C(s), W(s)\}} \left[ \int_{t_0}^t e^{-ps} U[C(s)] ds + I[W(t), t] \right]
\]

and, in particular, (14) can be rewritten as

\[
I(W_0, 0) = \max_{\{C(s), W(s)\}} \left[ \int_0^T e^{-ps} U[C(s)] ds + I[W(T), T] \right]. \quad (14')
\]

\(^{2}\) The basic derivation of the optimality equations in this section follows that of S.E. Dreyfus [2], Chapter VII.
If \( t = t_0 + h \) and the third partial derivatives of \( I[W(t_0), t_0] \) are bounded, then by Taylor's theorem and the mean value theorem for integrals, (16) can be rewritten as

\[
I[W(t_0, t_0)] = \max_{(C, W)} \{ e^{-\rho t} U[C(t)] + h \\
+ I[W(t_0), t_0] + \frac{\partial I[W(t_0), t_0]}{\partial t} : h \\
+ \frac{\partial I[W(t_0), t_0]}{\partial W}[W(t) - W(t_0)] \\
+ \frac{1}{2} \frac{\partial^2 I[W(t_0), t_0]}{\partial W^2} (W(t) - W(t_0))^2 + o(h^2) \}
\]

where \( \varepsilon_{t_0, t} \).

In (17), take the \( E(t_0) \) operator onto each term and, noting that \( I[W(t_0), t_0] = E(t_0) I[W(t_0), t_0] \), subtract \( I[W(t_0), t_0] \) from both sides. Substitute from equations (10) and (11) for \( E(t_0)[W(t) - W(t_0)] \) and \( E(t_0)[[W(t) - W(t_0)]^2] \), and then divide the equation by \( h \). Take the limit of the resultant equation as \( h \to 0 \) and (17) becomes a continuous-time version of the Bellman-Dreyfus fundamental equation of optimality, (17')

\[
0 = \max_{\{C(t), W(t)\}} \left[ e^{-\rho t} U[C(t)] + \frac{\partial I_t}{\partial t} \\
+ \frac{\partial I_t}{\partial W} [W(t)(\alpha - r) + r] W(t) - C(t)] \\
+ 1/2 \frac{\partial^2 I_t}{\partial W^2} \sigma^2 w^2(t) W^2(t) \right]
\]

(17')

where \( I_t \) is short for \( I[W(t), t] \) and the subscript on \( t_0 \) has been dropped to reflect that (17') holds for any \( t \in [0, T] \).

If we define \( \phi(w, C; W; t) \equiv \{ e^{-\rho t} U(C) \\
+ \frac{\partial I_t}{\partial t} + \frac{\partial I_t}{\partial W} [W(t)(\alpha - r) + r] W(t) - C(t)] \\
+ 1/2 \frac{\partial^2 I_t}{\partial W^2} \sigma^2 w^2(t) W^2(t) \} \), then (17') can be written in

\[\text{footnote}^3\]

\( \phi(w, C; W; t) \) is short for the rigorous \( \phi[w, C; \partial I_t/\partial t; \partial I_t/\partial W; \partial^2 I_t/\partial W^2; I_t; W; t] \).
the more compact form,
\[ \max_{C,W} \phi(w,C;W,t) = 0 \]  \hspace{1cm} (17'')

The first-order conditions for a regular interior maximum to (17'') are,
\[ \phi_C[w^*,C^*;W,t] = 0 = e^{-\theta t} u'(c) - \frac{\partial}{\partial W} \]  \hspace{1cm} (18)

and
\[ \phi_W[w^*,C^*;W;t] = 0 = (\alpha - \gamma) \frac{\partial}{\partial W} \frac{\partial^2 I_t}{\partial W^2} + \frac{\partial^2 I_t}{\partial W^2} wW^2. \]  \hspace{1cm} (19)

A set of sufficient conditions for a regular interior maximum is
\[ \phi_{WW} < 0; \phi_{CC} < 0; \det \begin{bmatrix} \phi_{WW} & \phi_{WC} \\ \phi_{CW} & \phi_{CC} \end{bmatrix} > 0. \]  \hspace{1cm} (20)

\[ \phi_{WC} = \phi_{CW} = 0, \text{ and if } I[W(t),t] \text{ were strictly concave in } W, \text{ then} \]
\[ \phi_{CC} = u''(c) < 0, \text{ by the strict concavity of } U \]  \hspace{1cm} (20)

and
\[ \phi_{WW} = W(t) \sigma^2 \frac{\partial^2 I_t}{\partial W^2} < 0, \text{ by the strict concavity of } I_t \]  \hspace{1cm} (21)

and the sufficient conditions would be satisfied. Thus a candidate for an optimal solution which causes \( I[W(t),t] \) to be strictly concave will be any solution of the conditions (17') \( \sim \) (21).

The optimality conditions can be re-written as a set of two algebraic and one partial differential equation to be solved for \( w^*(t), C^*(t), \) and \( I[W(t),t] \).

\[ \phi[w^*,C^*;W;t] = 0 \]  \hspace{1cm} (17'')

(* \( \phi_C[w^*,C^*;W;t] = 0 \) \hspace{1cm} (18)

\[ \phi_W[w^*,C^*;W;t] = 0 \]  \hspace{1cm} (19)

\[ (4) \]  \hspace{1cm} (continued)

(4) By the substitution of the results of (18) into (19) at \( (C^*,w^*) \), we have the condition \( w^*(\alpha - r) > 0 \) if and only if \( \frac{\partial^2 I_t}{\partial W^2} < 0. \)
subject to the boundary condition
\[ I[W(T),T] = B[W(T),T] \] and the solution being a feasible solution to (14).

IV. Constant Relative Risk Aversion. The system (*) of a nonlinear partial differential equation coupled with two algebraic equations is difficult to solve in general. However, if the utility function is assumed to be of the form yielding constant relative risk-aversion (i.e., iso-elastic marginal utility), then (*) can be solved explicitly. Therefore, let \( U(C) = C^\gamma/\gamma, \gamma < 1 \) and \( \gamma \neq 0 \) or \( U(C) = \log C \) (the limiting form for \( \gamma = 0 \)) where \( -U''(C) \) \( C/U'(C) = 1 - \gamma \equiv \delta \) is Pratt's [7] measure of relative risk aversion. Then, system (*) can be written in this particular case as

\[
0 = \left( \frac{1-\gamma}{\gamma} \right) \frac{\partial I_t}{\partial W} \gamma/\gamma - 1 \cdot e^{-pT/1-\gamma} \\
+ \frac{\partial I_t}{\partial t} + \frac{\partial I_t}{\partial W} \cdot r \cdot W \\
- \frac{(a-r)^2}{2\sigma^2} \frac{\left[ \frac{\partial I_t}{\partial W} \right]^2}{\frac{\partial^2 I_t}{\partial W^2}} \\
(17')
\]

\[
C^*(t) = \left[ e^{pT} \frac{\partial I_t}{\partial W} \right]^{1/\gamma - 1} \\
w^*(t) = \frac{-\frac{(a-r)\partial I_t}{\partial W}}{\frac{\partial^2 I_t}{\partial W^2}} \\
(18)
\]

subject to \( I[W(T),T] = \epsilon^{1-\gamma} e^{-\rho T} \)

\([W(T)]^{\gamma/\gamma}, \text{ for } 0 < \epsilon < 1\)

where a strategically-simplifying assumption has been made as to the particular form of the bequest valuation function.

(4)

The paper considers only interior optimal solutions. The problem could have been formulated in the more general Kuhn-Tucker form in which case the equalities of (18) and (19) would be replaced with inequalities.
To solve (17") of (*'), take as a trial solution,

\[ \bar{I}_t[W(t), t] = \frac{b(t)}{\gamma} e^{-\rho t [W(t)]^\gamma}. \]  

(22)

By substitution of the trial solution into (17"), a necessary condition that \( \bar{I}_t[W(t), t] \) be a solution to (17") is found to be that \( b(t) \) must satisfy the following ordinary differential equation,

\[ \dot{b}(t) = \mu b(t) - (1-\gamma) [b(t)]^{-\gamma/(1-\gamma)} \]  

subject to \( b(T) = \epsilon^{1-\gamma} \), and where \( \mu = \rho - \gamma [(\alpha - r)^2/2\sigma^2(1-\gamma) + r] \). The resulting decision rules for consumption and portfolio selection, \( C^*(t) \) and \( w^*(t) \), are from equations (18) and (19) of (*'), then

\[ C^*(t) = [b(t)]^{1/(\gamma)} W(t) \]  

(24)

and

\[ w^*(t) = \frac{(\alpha - r)}{\sigma^2(1-\gamma)}. \]  

(25)

The solution to (23) is

\[ b(t) = \{(1 + (\nu \epsilon - 1) e^{\nu(t-T)})/\nu\}^{1-\gamma} \]  

(26)

where \( \nu = \mu/(1-\gamma) \).

A sufficient condition for \( \bar{I}[W(t), t] \) to be a solution to (*') is that \( \bar{I}[W(t), t] \) satisfy

A. \( \bar{I}[W(t), t] \) be real (feasibility)

B. \( \frac{\partial^2 \bar{I}_t}{\partial w^2} < 0 \) (concavity for a maximum)

C. \( C^*(t) \geq 0 \) (feasibility)

5 The form of the bequest valuation function (the boundary condition), as is usual for partial differential equations, can cause major changes in the solution to (*). The particular form of the function chosen in (*') is used as a proxy for the "no-bequest" condition (\( \epsilon = 0 \)). A slightly more general form which can be used without altering the resulting solution substantively is \( B[W(T), T] = e^{-\rho t G(T)} [W(T)]^{\gamma/\gamma} \) for arbitrary \( G(T) \). If \( B \) is not of the iso-elastic family, systematic effects of age will appear in the optimal decision-making.
The condition that $A$, $B$, and $C$ are satisfied in the isoelastic case is that

$$[1 + (\nu \varepsilon - 1)e^\nu(t-T)]/\nu>0, \quad 0 \leq t \leq T$$

which is satisfied for all values of $\nu$ when $T < \infty$.

Because (27) holds, the optimal consumption and portfolio selection rules are,

$$C^*(t) = {\nu/(1 - (\nu \varepsilon - 1)e^\nu(t-T))}W(t),$$

for $\nu \neq 0$

$$= [1/(T-t+\varepsilon)]W(t),$$

for $\nu = 0$  \hspace{1cm} (28)

and

$$W^*(t) = \frac{(\alpha - r)}{\sigma^2(1-\gamma)} \equiv w^*, \text{ a constant independent or}$$

$W$ or $t$. \hspace{1cm} (28)

V. Dynamic Behavior and the Bequest Valuation Function. The purpose behind the choice of the particular bequest valuation function in ($^*$) was primarily mathematical. The economic motive is that the "true" function for no bequests is $B[W(T),T] = 0$ (i.e., $\varepsilon = 0$). From (28), $C^*(t)$ will have a pole at $t = T$ when $\varepsilon = 0$. So, to examine the dynamic behavior of $C^*(t)$ and to determine whether the pole is a mathematical "error" or an implicit part of the economic requirements of the problem, the parameter $\varepsilon$ was introduced. From figure 1, $(C^*/W)_{t=T} \rightarrow \infty$ as $\varepsilon \rightarrow 0$. However, one must not interpret this as an infinite rate of consumption. Because there is zero utility associated with positive wealth for $t > T$, the mathematics reflects this by requiring the optimal solution to drive $W(t) \rightarrow 0$ as $t \rightarrow T$. Because $C^*$ is a flow and $W(t)$ is a stock and, from (28), $C^*$ is proportional to $W(t)$, $(C^*/W)$ must become

\[\text{[6]}\]

Although not derived explicitly here, the special case ($\gamma = 0$) of Bernoulli logarithmic utility has (29) with $\gamma = 0$ as a solution, and the limiting form of (28), namely

$$C^*(t) = \left[\frac{\rho}{1 + (\rho \varepsilon - 1)e^{\rho(t-T)}}\right]W(t).$$
Figure 1.

larger and larger as $t+T$ to make $W(T) = 0$.\(^7\) In fact, if $W(T -) > 0$, an "impulse" of consumption would be required to make $W(T) = 0$. Thus, equation (28) is valid for $\varepsilon=0$.

To examine some of the dynamic properties of $C^*(t)$, let $\varepsilon=0$, and define $V(t) = [C^*(t)/W(t)]$, the instantaneous marginal (in this case, also average) propensity to consume out of wealth. Then, from (28),

$$V(t) = [V(t)]^2 e^{v(t-T)}$$

and, as observed in figure 1 (for $\varepsilon=0$), $V(t)$ is an increasing function of time. In a generalization of the half-life calculation of radioactive decay, define $\tau$ as that $t \in [0,T]$ such that $V(\tau) = nV(0)$ (i.e., $\tau$ is the length of time required for $V(t)$ to grow to $n$ times its initial size). Then from (28),

$$\tau = \log \left[ e^{VT} \left(1- \frac{1}{n} \right) + \frac{1}{n} \right] /v; \quad \text{for} \quad v \neq 0$$

$$\tau = \frac{(n-1)}{(n)} T, \quad \text{for} \quad v = 0. \quad (31)$$

To examine the dynamic behavior of $W(t)$ under the optimal decision rules, it only makes sense to discuss the expected

\(^{7}\) The problem described is essentially one of exponential decay. If $W(t) = W_0 e^{-f(t)}$, $f(t) > 0$, finite for all $t$, and $W_0 > 0$, then it will take an infinite length of time for $W(t) = 0$. However, if $f(t) \to \infty$ as $t \to T$, then $W(t) \to 0$ as $t \to T$. 
or "averaged" behavior because $W(t)$ is a function of a random variable. To do this, we consider equation (13) the averaged budget equation, and evaluate it at the optimal $(w^*, c^*)$ to form

$$
\frac{\dot{W}(t)}{W(t)} = \alpha_\# - V(t),
$$

(13')

where $\alpha_\# = \left[ \frac{(\alpha-r)^2}{\sigma^2(1-\gamma)} + \frac{r}{\sigma^2(1-\gamma)} \right]$, and, in section VII, $\alpha_\#$ will be shown to be the expected return on the optimal portfolio.

By differentiating (13') and using (30), we get

$$
\frac{d}{dt} \left[ \frac{\dot{W}}{W} \right] = -V(t) < 0
$$

(32)

which implies that for all finite-horizon optimal paths, the expected rate of growth of wealth is a diminishing function of time. Therefore, if $\alpha_\# < V(0)$, the individual will dis-invest (i.e., he will plan to consume more than his expected income, $\alpha_\#W(t)$). If $\alpha_\# > V(0)$, he will plan to increase his wealth for $0 < t < \bar{t}$, and then, dis-invest at an expected rate $\alpha_\# - V(t)$ for $\bar{t} < t < T$ where $\bar{t}$ is defined as the solution to

$$
\bar{t} = T + \frac{1}{\nu} \log \left[ \frac{\alpha_\# - V}{\alpha_\#} \right].
$$

(33)

Further, $\partial \bar{t} / \partial \alpha_\# > 0$ which implies that the length of time for which the individual is a net saver increases with increasing expected returns on the portfolio. Thus, in the case $\alpha_\# > V(0)$, we find the familiar result of "hump saving." (8)

### VI. Infinite Time Horizon

Although the infinite time horizon case ($T=\infty$) yields essentially the same substantive results as in the finite time horizon case, it is worth examining

---

(8) "Hump saving" has been widely discussed in the literature. (See J. De V. Graaff [3] for such a discussion). Usually "hump saving" is discussed in the context of work and retirement periods. Clearly, such a phenomenon can occur without these assumptions as the example in this paper shows.
separately because the optimality equations are easier to solve than for finite time. Therefore, for solving more complicated problems of this type, the infinite time horizon problem should be examined first.

The equation of optimality is, from section III,

\[ 0 = \max_{\{C,w\}} \left[ e^{-\rho t} U(C) + \frac{\partial I_t}{\partial t} \right. \]

\[ + \frac{\partial I_t}{\partial w} \left[ (w(t) (\alpha-r) + r) W(t) - C(t) \right] \]

\[ + \frac{1}{2} \frac{\partial I_t}{\partial w} \sigma^2 w^2(t) W^2(t) \right] . \tag{17'} \]

However (17') can be greatly simplified by eliminating its explicit time-dependence. Define

\[ J[W(t), t] = e^{\rho t} I[W(t), t] \]

\[ = \max_{\{C,w\}} E(t) \int_t^\infty e^{-\rho (s-t)} U[C] ds \]

\[ = \max_{\{C,w\}} E \int_0^\infty e^{-\rho v} U[C] dv, \]

independent of explicit time. \tag{34} \]

Thus, write \( J[W(t), t] = J[W] \) to reflect this independence. Substituting \( J[W] \), dividing by \( e^{\rho t} \), and dropping all \( t \) subscripts, we can rewrite (17') as,

\[ 0 = \max_{\{C,w\}} \left[ u(C) - \rho J + J'(W) \right. \]

\[ \left. (w(t)(\alpha-r) + r) W-C \right] \]

\[ + \frac{1}{2} J''(W) \sigma^2 w^2 W^2 \right]. \tag{35} \]

Note: when (35) is evaluated at the optimum \( (C^*, w^*) \), it becomes an ordinary differential equation instead of the usual partial differential equation of (17'). For the iso-elastic case, (35) can be written as

\[ 0 = \left( \frac{1-\gamma}{\gamma} \right) [J'(W)]^{-\gamma/1-\gamma} - \rho J(W) \]

\[ - \frac{\gamma}{2\sigma^2} \frac{[J'(W)]}{J''(W)} + rWJ'(W). \tag{36} \]
where the functional equations for \( C^* \) and \( w^* \) have been substituted in equation (36).

The first-order conditions corresponding to (18) and (19) are

\[
0 = U'(C) - J'(W) \tag{37}
\]

and

\[
0 = (\alpha-r)J'(W) + J''WW\sigma^2 \tag{38}
\]

and assuming that \( \lim_{T \to \infty} B[W(T),T] = 0 \), the boundary condition becomes the transversality condition,

\[
\lim_{t \to \infty} E[I[W(t),t]] = 0 \tag{39}
\]

or

\[
\lim_{t \to \infty} E[e^{-\rho t}J[W(t)]] = 0
\]

which is a condition for convergence of the integral in (14). A solution to (14) must satisfy (39) plus conditions A, B, and C of section IV. Conditions A, B, and C will be satisfied in the iso-elastic case if

\[
V^* \equiv \nu = \frac{\rho}{1-\gamma} - \gamma \left[ \frac{(\alpha-r)^2}{2\sigma^2(1-\gamma)^2} + \frac{r}{1-\gamma} \right] > 0 \tag{40}
\]

holds where (40) is the limit of condition (27) in section IV, as \( T \to \infty \) and \( V^* = C^*(t)/W(t) \) when \( T = \infty \). Condition (39) will hold if \( \rho > \gamma \frac{\delta}{W} \) where, as defined in (13), \( \frac{\delta}{W} \) is the stochastic time derivative of \( W(t) \) and \( \frac{\delta}{W}W(t)/W(t) \) is the "expected" net growth of wealth after allowing for consumption. That (39) is satisfied can be rewritten as a condition on the subjective rate of time preference, \( \rho \), as follows:

- For \( \gamma < 0 \) (bounded utility), \( \rho > 0 \)
- \( \gamma = 0 \) (Bernoulli log case), \( \rho > 0 \)
- \( 0 < \gamma < 1 \) (unbounded utility), \( \rho > \gamma \)

\[
\left[ \frac{-(\alpha-r)^2(2-\gamma)}{2\sigma^2(1-\gamma)} + r \right]. \tag{41}
\]
Condition (41) is a generalization of the usual assumption required in deterministic optimal consumption growth models when the production function is linear: namely, that \( \rho > \max [0, \gamma \beta] \) where \( \beta = \text{yield on capital} \). If a "diminishing-returns," strictly-concave "production" function for wealth were introduced, then a positive \( \rho \) would suffice.

If condition (41) is satisfied, then condition (40) is satisfied. Therefore, if it is assumed that \( \rho \) satisfies (41), then the rest of the derivation is the same as for the finite horizon case and the optimal decision rules are,

\[
C_\infty^*(t) = \left\{ \frac{\rho}{1-\gamma} - \gamma \left[ \frac{(a-r)^2}{2\sigma^2(1-\gamma)^2} + \frac{r}{1-\gamma} \right] \right\} W(t)
\]

and

\[
w_\infty^*(t) = \frac{(a-r)}{\sigma(1-\gamma)}
\]

The ordinary differential equation (35), \( J'' = f(J,J') \), has "extraneous" solutions other than the one that generates (42) and (43). However, these solutions are ruled out by the transversality condition, (39), and conditions A, B, and C of section IV. As was expected, \( \lim_{T \to \infty} C^*(t) = C_\infty^*(t) \) and \( \lim_{T \to \infty} w^*(t) = w_\infty^*(t) \).

The main purpose of this section was to show that the partial differential equation (17') can be reduced in the case of infinite time horizon to an ordinary differential equation.

VII. Economic Interpretation of the Optimal Decision Rules for Portfolio Selection and Consumption. An important result is the confirmation of the theorem proved by Samuelson [8], for the discrete-time case, stating that, for iso-elastic

\[\text{If one takes the limit as } \sigma^2 \to 0\text{ (where } \sigma^2 \text{ is the variance of the composite portfolio) of condition (41), then (41) becomes the condition that } \rho > \max[0, \gamma \alpha^*] \text{ where } \alpha^* \text{ is the yield on the composite portfolio. Thus, the deterministic case is the limiting form of (41).}\]
marginal utility, the portfolio-selection decision is independent of the consumption decision. Further, for the special case of Bernoulli logarithmic utility (γ=0), the separation goes both ways, i.e., the consumption decision is independent of the financial parameters and is only dependent upon the level of wealth. This is a result of two assumptions: (1) constant relative risk-aversion that one's attitude toward financial risk is independent of one's wealth level, and (2) the stochastic process which generates the price changes (independent increments assumption of the Wiener process). With these two assumptions, the only feedbacks of the system, the price change and the resulting level of wealth, have zero relevance for the portfolio decision and hence, it is constant.

The optimal proportion in the risky asset, \( w^* \), can be rewritten in terms of Pratt's relative risk-aversion measure, \( \delta \), as

\[
  w^* = \frac{(\alpha - r)}{\sigma^2 \delta} .
\]

The qualitative results that \( \partial w^*/\partial \alpha > 0 \), \( \partial w^*/\partial r < 0 \), \( \partial w^*/\partial \sigma^2 < 0 \), and \( \partial w^*/\partial \delta < 0 \) are intuitively clear and need no discussion. However, because the optimal portfolio selection rule is constant, one can define the optimum composite portfolio and it will have a constant mean and variance. Namely,

\[
  \alpha^* = E[w^*(\alpha + \Delta Y) + (1 - w^*)r] = w^*\alpha
\]

\[
  + (1 - w^*)r = \frac{(\alpha - r)^2}{\sigma^2 \delta} + r
\]

\[
\sigma^2 = \text{Var} \left[ w^*(\alpha + \Delta Y) + (1 - w^*)r \right] = w^*\sigma^2\sigma^2 = \frac{(\alpha - r)^2}{\sigma^2 \delta^2} .
\]

\( ^{10} \) Note: no restriction on borrowing or going short was imposed on the problem, and therefore, \( w^* \) can be greater than one or less than zero. Thus, if \( \alpha < r \), the risk-averted will short some of the risky asset, and if \( \alpha > r + \sigma^2 \delta \), he will borrow funds to invest in the risky asset. If one wished to restrict \( w^* \in [0,1] \), then such a constraint could be introduced and handled by the usual Kuhn-Tucker methods with resulting inequalities.
After having determined the optimal $w^*$, one can now think of the original problem as having been reduced to a simple Phelps-Ramsey problem, in which we seek an optimal consumption rule given that income is generated by the uncertain yield of an (composite) asset.

Thus, the problem becomes a continuous-time analog of the one examined by Phelps [6] in discrete time. Therefore, for consistency, $C^*(t)$ should be expressible in terms of $\alpha^*, \sigma^*, \delta, \rho$, and $W(t)$ only. To show that this is, in fact, the result, (42) can be rewritten as, \[C^*(t) = \left[ \frac{\delta}{\delta^2} + (\delta-1) \frac{\sigma^*}{\delta^2} \right] W(t) = VW(t) \quad (46)\]

where $V$ is the marginal propensity to consume out of wealth.

The tools of comparative statics are used to examine the effect of shifts in the mean and variance on consumption behavior in this model. The comparison is between two economies with different investment opportunities, but with the individuals in both economies having the same utility function.

If $\theta$ is a financial parameter, then define $\left[ \frac{\partial C^*}{\partial \theta} \right]_{I^*_0}$ the partial derivative of consumption with respect to $\theta$, $I^*_0[W^*_0]$ being held fixed, as the intertemporal generalization of the Hicks-Slutsky "substitution" effect, $\left[ \frac{\partial C^*}{\partial \theta} \right]_U$ for static models. $[\partial C^*/\partial \theta - (\partial C^*/\partial \theta)_{I^*_0}]_{\bar{W}}$ will be defined as the intertemporal "income" or "wealth" effect. Then, from equation (22) with $I^*_0$ held fixed, one derives by total differentiation,

$$0 = -\frac{1}{\delta-1} \frac{\partial b(0)}{\partial \theta} W^*_0 + b(0) \left( \frac{\partial W^*_0}{\partial \theta} \right)_{I^*_0}.$$ \quad (47)

From equations (24) and (46), $b(0) = V^{-\delta}$, and so solving for $(\partial W^*_0/\partial \theta)_{I^*_0}$ in (47), we can write it as

--Because this section is concerned with the qualitative changes in the solution with respect to shifts in the parameters, the more simple form of the infinite-time horizon case is examined. The essential difference between $C_{\infty}^*(t)$ and $C^*(t)$ is the explicit time dependence of $C^*(t)$ which was discussed in section V. For simplicity, the "\infty" on subscript $C_{\infty}^*(t)$ will be deleted for the rest of this section.
\[
\left( \frac{\partial W_0}{\partial \theta} \right)_{I_0} = \frac{-\delta W_0}{(\delta-1)\nu} \frac{\partial V}{\partial \theta}. \tag{48}
\]

Consider the case where \( \theta = \alpha_\star \), then from (46),
\[
\frac{\partial V}{\partial \alpha_\star} = \frac{(\delta-1)}{\delta} \tag{49}
\]

and from (48)
\[
\left( \frac{\partial W_0}{\partial \alpha_\star} \right)_{I_0} = -\frac{W_0}{V}. \tag{50}
\]

Thus, we can derive the substitution effect of an increase in the mean of the composite portfolio as follows,
\[
\left( \frac{\partial C^\star}{\partial \alpha_\star} \right)_{I_0} = \left[ \frac{\partial V}{\partial \alpha_\star} W_0 + V \frac{\partial W_0}{\partial \alpha_\star} \right]_{I_0} = -\frac{W_0}{\delta} < 0. \tag{51}
\]

Because \( \frac{\partial C^\star}{\partial \alpha_\star} = \left( \frac{\partial V}{\partial \alpha_\star} \right) W_0 = \left[ \frac{(\delta-1)}{\delta} \right] W_0 \), then the income or wealth effect is
\[
\left[ \frac{\partial C^\star}{\partial \alpha_\star} - \left( \frac{\partial C^\star}{\partial \alpha_\star} \right)_{I_0} \right] = W_0 > 0. \tag{52}
\]

Therefore, by combining the effects of (51) and (52), one can see that individuals with low relative risk-aversion (0 < \( \delta < 1 \)) will choose to consume less now and save more to take advantage of the higher yield available (i.e., the substitution effect dominates the income effect). For high risk-aversions (\( \delta > 1 \)), the reverse is true and the income effect dominates the substitution effect. In the borderline case of Bernoulli logarithmic utility (\( \delta = 1 \)), the income and substitution effect just offset one another.\(^{12}\)

In a similar fashion, consider the case of \( \theta = -\sigma^2_\star \),\(^{13}\) then from (46) and (48), we derive
\[
\left( \frac{\partial W_0}{\delta \left( -\sigma^2_\star \right)} \right)_{I_0} = -\frac{\delta W_0}{2\nu} \tag{53}
\]

\[
\left( \frac{\partial C^\star}{\delta \left( -\sigma^2_\star \right)} \right)_{I_0} = -\frac{W_0}{2} < 0, \text{ the substitution effect.} \tag{54}
\]

\(^{12}\) Many writers have independently discovered that Bernoulli utility is a borderline case in various comparative-static situations. See, for example, Phelps [6] and Arrow [1].

\(^{13}\) Because increased variance for a fixed mean usually (continued)
Further, \( \frac{\partial C^*}{\partial (-\sigma^*_0)^2} = (\delta-1)W_0/2 \), and so

\[
\left[ \frac{\partial C^*}{\partial (-\sigma^*_0)^2} \right] - \left( \frac{\partial C^*}{\partial (-\sigma^*_0)^2} \right) \sigma_0^2 = \frac{\delta}{2} W_0 > 0, \text{ the income effect} \tag{55}
\]

To compare the relative effect on consumption behavior of an upward shift in the mean versus a downward shift in variance, we examine the elasticities. Define the elasticity of consumption with respect to the mean as

\[
E_1 \equiv \alpha^*_0 \frac{\partial C^*}{\partial \alpha^*_0} / C^* = \alpha^*_0 (\delta-1)/\delta V \tag{56}
\]

and similarly, the elasticity of consumption with respect to the variance as,

\[
E_2 \equiv \sigma^*_0 \frac{\partial C^*}{\partial \sigma^*_0} / C^* = -\sigma^*_0 (\delta-1)/2 V \tag{57}
\]

For graphical simplicity, we plot \( e_1 = [V E_1/\alpha^*_0] \) and \( e_2 = -[V E_2/\alpha^*_0] \) and define \( k \equiv \sigma^*_0^2/2\alpha^*_0 \). \( e_1 \) and \( e_2 \) are equal at \( \delta = 1, 1/k \). The particular case drawn is for \( k < 1 \).

\[\text{figure 2.}\]

\[\text{figure 2.}\]

\[13\text{\(1\) (continued)\}}

(always always for normal variates) decreases the desirability of investment for the risk-averter, it provides a more symmetric discussion to consider the effect of a decrease in variance.
For relatively high variance \((k>1)\), the high risk averter \((\delta>1)\) will always increase present consumption more with a decrease in variance than for the same percentage increase in mean. Because a high risk-averter prefers a steadier flow of consumption at a lower level than a more erratic flow at a higher level, it makes sense that a decrease in variance would have a greater effect than an increase in mean. On the other hand, for relatively low variance \((k<1)\), a low risk averter \((0<\delta<1)\) will always decrease his present consumption more with an increase in the mean than for the same percentage decrease in variance because such an individual (although a risk-averter) will prefer to accept a more erratic flow of consumption in return for a higher level of consumption. Of course, these qualitative results will vary depending upon the size of \(k\). If the riskiness of the returns is very small (i.e., \(k<<1\)), then the high risk-averter will increase his present consumption more with an upward shift in mean. Similarly, if the risk-level is very high (i.e., \(k>>1\)) the low risk averter will change his consumption more with decreases in variance.

The results of this analysis can be summed up as follows: Because all individuals in this model are risk-avers, when risk is a dominant factor \((k>>1)\), a decrease in risk will have the larger effect on their consumption decisions. When risk is unimportant (i.e., \(k<<1\)), they all react stronger to an increase in the mean yield. For all degrees of relative riskiness, the low risk-averter will give up some present consumption to attain an expected higher future consumption, while the high risk averter will always choose to increase the amount of present consumption.

VII. Extension to Many Assets. The model presented in section IV, can be extended to the \(m\)-asset case with little difficulty. For simplicity, the solution is derived in the infinite time horizon case, but the result is similar for finite time. Assume the \(m^{th}\) asset to be the only certain asset
with an instantaneous rate of return $\alpha_m = r^{(14)}$. Using the
general equations derived in section II, and substituting
for $w_m(t) = 1 - \sum_{i=1}^{n} w_i(t)$ where $n = m - 1$, equations (6)
and (7) can be written as,

\[
E(t_0) \left[ W(t) - W(t_0) \right] = \left[ w'(t_0)(\alpha - \hat{r}) + r \right] W(t_0)h - C(t_0)h + O(h^2) \tag{6}
\]

and

\[
E(t_0) \left[ W(t) - W(t_0) \right]^2 = w'(t_0)\Omega w(t_0)W^2(t_0)h + O(h^2) \tag{7}
\]

where

\[
w'(t_0) = [w_1(t_0), \ldots, w_n(t_0)], \text{ a } n\text{-vector}
\]

\[
\alpha' = [\alpha_1, \ldots, \alpha_n]
\]

\[
\hat{r}' = [r, \ldots, r], \text{ a } n\text{-vector}
\]

\[
\Omega = [\sigma_{ij}], \text{ the } n \times n \text{ variance-covariance}
\]

\[
\text{matrix of the risky assets}
\]

\[
\Omega \text{ is symmetric and positive definite.}
\]

Then, the general form of (35) for $m$-assets is, in matrix
notation,

\[
0 = \text{Max} \left\{ U(C) - \rho J(W) \right\}
\]

\[
+ J'(W) \left\{ \left[ w'(\alpha - \hat{r}) + r \right] W - C \right\}
\]

\[
+ \frac{1}{2} J''(W)w'\Omega wW^2 \right\} \tag{58}
\]

\[
^{(14)}\text{Clearly, if there were more than one certain asset, the one}
\]

\[
\text{with the highest rate of return would dominate the others.}
\]
and instead of two, there will be \( m \) first-order conditions corresponding to a maximization of (35) with respect to \( w_1, \ldots, w_n \) and \( C \). The optimal decision rules corresponding to (42) and (43) in the two-asset case, are

\[
C_\infty^*(t) = \left\{ \frac{\rho}{1-\gamma} - \gamma \left[ \frac{(\alpha-\hat{r})^\prime \Omega^{-1}(\alpha-\hat{r})}{2(1-\gamma)^2} \right] + \frac{r}{1-\gamma} \right\} W(t)
\]

(59)

and

\[
w_\infty^*(t) = \frac{1}{(1-\gamma)} \Omega^{-1}(\alpha-\hat{r})
\]

(60)

where \( w_\infty^*(t) = [w_1^*(t), \ldots, w_n^*(t)] \).

IX. Constant Absolute Risk Aversion. System (*) of section III, can be solved explicitly for a second special class of utility functions of the form yielding constant absolute risk-aversion. Let \( U(C) = e^{-\eta C/\eta}, \eta > 0 \), where \(-U''(C)/U'(C) = \eta \) is Pratt's [17] measure of absolute risk-aversion. For convenience, I return to the two-asset case and infinite-time horizon form of system (*) which can be written in this case as,

\[
0 = \frac{-J'(W)}{\eta} - \rho J(W) + J'(W)rW
+ \frac{J'(W)}{\eta} \log [J'(W)]
- \frac{(\alpha-r)^2}{2\sigma^2} \frac{[J'(W)]^2}{J''(W)}
\]

(*")

(17")

\[
C^*(t) = -\frac{1}{\eta} \log [J'(W)]
\]

(18)

\[
w^*(t) = -\frac{J'(W)(\alpha-r)/\sigma^2 WJ''(W)}{J'(W)}
\]

subject to limit \( t\to\infty \)

\[
E[e^{-\rho t}J(W(t))] = 0
\]

(19)

where \( J(W) = e^{\rho t}I[W(t),t] \) as defined in section VI.

To solve (17") of (*)", take as a trial solution,

\[
\bar{J}(W) = \frac{-\rho}{q} e^{-qW}
\]

(61)
By substitution of the trial solution into (17''), a necessary condition that \( \tilde{J}(\tilde{W}) \) be a solution to (17'') is found to be that \( p \) and \( q \) must satisfy the following two algebraic equations:

\[
q = \eta r
\]

and

\[
\left( \frac{r - \rho - (a-r)^2/2\sigma^2}{r} \right)
\]

\[p = e\]

The resulting optimal decision rules for portfolio selection and consumption are,

\[
C^*(t) = r\tilde{W}(t) + \left[ \frac{\rho - r + (a-r)^2/2\sigma^2}{\eta r} \right]
\]

and

\[
w^*(t) = \frac{(a-r)}{\eta r \sigma^2 \tilde{W}(t)}
\]

Comparing equations (64) and (65) with their counterparts for the constant relative risk-aversion case, (42) and (43), one finds that consumption is no longer a constant proportion of wealth (i.e., marginal propensity to consume does not equal the average propensity) although it is still linear in wealth. Instead of the proportion of wealth invested in the risky asset being constant (i.e., \( w^*(t) \) a constant), the total dollar value of wealth invested in the risky asset is kept constant (i.e., \( w^*(t)\tilde{W}(t) \) a constant). As one becomes wealthier, the proportion of his wealth invested in the risky asset falls, and asymptotically, as \( \tilde{W} \to \infty \), one invests all his wealth in the certain asset and consumers all his (certain) income. Although one can do the same type of comparative statics for this utility function as was done in section VII for the case of constant relative risk-aversion, it will not be done in this paper for the sake of brevity and because I find this special form of the utility function behaviorally less plausible than constant relative risk aversion. It is interesting to note that the substitution effect in this
case, \[ \frac{\partial C}{\partial \theta} \], is zero except when \( r = 0 \).

X. Other Extensions of the Model. The requirements for the general class of probability distributions which could be acceptable in this model are,

(1) the stochastic process must be Markovian.

(2) the first two moments of the distribution must be \( O(\Delta t) \) and the higher-order moments \( o(\Delta t) \) where \( o(\cdot) \) is the order symbol meaning "smaller order than."

So, for example, the simple Wiener process postulated in this model could be generalized to include \( \alpha_i = \alpha_i(X_1, \ldots, X_m, W, t) \) and \( \sigma_i = \sigma_i(X_1, \ldots, X_m, W, t) \), where \( X_i \) is the price of the \( i \)th asset. In this case, there will be \( (m + 1) \) state variables and \((17')\) will be generated from the general Taylor series expansion of \( I[X_1, \ldots, X_m, W, t] \) for many variables. A particular example would be if the \( i \)th asset is a bond which fluctuates in price for \( t < t_1 \), but will be called at a fixed price at time \( t = t_1 \). Then \( \alpha_i = \alpha_i(X_i, t) \) and \( \sigma_i = \sigma_i(X_i, t) > 0 \) when \( t < t_1 \) and \( \sigma_i = 0 \) for \( t > t_1 \).

A more general production function of a neo-classical type could be introduced to replace the simple linear one of this model. Mirrlees [5] has examined this case in the context of a growth model with Harrod-neutral technical progress a random variable. His equations (19) and (20) correspond to my equations (35) and (37) with the obvious proper substitutions for variables.

Thus, the technique employed for this model can be extended to a wide class of economic models. However, because the optimality equations involve a partial differential equation, computational solution of even a slightly generalized model may be quite difficult.
References


III. A GOLDEN RULE FOR WELFARE-MAXIMIZATION IN
AN ECONOMY WITH A VARYING POPULATION GROWTH RATE*

The planner seeking to maximize welfare in an economy with
a growing population is told that the most efficient golden-age
path results from following the Swan-Phelps Golden Rule [4].
This Golden Rule states that, for constant population growth,
the most efficient golden-age path equates the interest rate
to the rate of population growth [4]. This is true for an
economy with a constant rate of population growth. The question
arises, is it still true for an economy in which the rate of
population growth varies? The answer is no. The optimal golden-
age state will be at a lower capital-labor ratio than the Swan-
Phelps Golden-Rule State for the case where the rate of popu-
lation growth is an increasing function of the capital-labor
ratio, and it will be at a higher capital-labor ratio for the
rate of population growth a decreasing function.

I. The Planner's Economy. The planner is postulated
to choose maximizing the utility of the representative man over
the many years of his plan as the criterion for maximizing the
welfare of the economy.

Consider a planner faced with an economy whose representa-
tive man's fertility is affected only by his wealth or lack of
it. The wealthier he becomes, the more children he has. The
assumption of a representative man implies that all members of this
society have the same tastes, education, and wealth. If the
people in such a society enjoy having children, then it is
reasonable to assume that the wealthier people become, the more

* This work was done during the tenure of a National Defense
Education Act Fellowship. I am indebted to Paul A. Samuelson
for many helpful suggestions. I should like to thank Andrew
M. Lewis for his comments on the behavioral assumptions for
the rate of population growth.
children they have, and there is considerable empirical evidence for this when extraneous factors such as education and religion are allowed for by partial correlation.

It is possible that empirically the opposite could be true; namely, that fertility declines with per capita economic well-being (as e.g. when people can better afford to practice contraception effectively). Probably, a more realistic assumption would be a blend: for relatively low per capita wealth, increases in wealth would lead to increases in the rate of population growth until at some level of per capita wealth, the growth rate would reach a peak and then begin to decline with further increases. This decline would gradually slow down, and the growth rate would approach some constant level asymptotically. Actually, the derivation of the optimal golden-age path is the same whether the rate of population growth is an increasing or decreasing function of the capital-labor ratio. The main restriction on the function is that its second derivative have a particular lower-bound described in equation (8) later in the paper. However, except as otherwise noted, I suppose it to be strictly monotone-increasing.

II. The Planner's Problem. Mathematically, the planner's problem is described as follows:

\[
\text{Maximize } \int_0^T U(c(t)) \, dt \text{ for } k(0) \leq k^0; k(T) \geq k^T \leq 0 \quad (1)
\]

where

\[
U(c) = \text{utility of the representative man}
\]

\[
U \text{ is assumed monotone-increasing and strictly concave in } c, \text{ i.e. } U'(c) > 0; U''(c) < 0
\]

\[
c(t) = \text{consumption per person} = \text{consumption per worker}
\]

\(\text{Because this derivation parallels Samuelson's derivation of the Per Capita Consumption Turnpike Theorem [6], I have tried to keep the same notation. As is usual for continuous models, new people are assumed to work from the time they are born (into the labor force). There are various other alternative expressions one might care to maximize instead of the utility of per capita consumption. One such variate, the Bentham-Lerner criterion, is explored later.}\)
Constant-returns to scale, diminishing marginal returns (to varying proportions), and capital-saturation at a finite capital-labor ratio are assumed.

\[ k(t) = \text{capital-labor ratio} \]
\[ \dot{k}(t) = \text{net change in capital per worker} \]
\[ f(k) = \text{annual per capita output}, \quad (2) \]
\[ f'(k) > 0 \text{ for } k < k_s \]
\[ f'(k_s) = 0 \text{ where } k_s = \text{Schumpeter capital-saturation point} \]
\[ f''(k) < 0 \]

Using the behavioral assumption previously described,
\[ n(k) = \text{the rate of population growth}, \]
\[ n(0) = 0; \quad n'(k) > 0 \]

The technological budget constraint is:
\[ c(t) = f(k) - n(k)k - k. \]

Substituting the budget constraint for \( c(t) \), the problem becomes:
\[
\text{Maximize } \int_0^T V(k,\dot{k})dt \text{ for } k(0) \leq k^0; k(T) \geq k^T \geq 0 \quad (3)
\]

where
\[ V(k,\dot{k}) \equiv U(f(k) - n(k)k - k) \quad (4) \]

A necessary condition for a smooth interior maximum is the Euler equation:
\[
E_t(R,\dot{R},k) \equiv \frac{d}{dt} [V_2] - V_1 \equiv 0 \quad (5)
\]

where
\[ V_1 = \frac{\partial V(k,\dot{k})}{\partial k} ; \quad V_2 = \frac{\partial V(k,\dot{k})}{\partial \dot{k}} \]
\[ V_{12} = \frac{\partial^2 V(k,\dot{k})}{\partial k \partial \dot{k}} = V_{21} \]
\[ V_{11} = \frac{\partial^2 V(k,\dot{k})}{\partial k^2} ; \quad V_{22} = \frac{\partial^2 V(k,\dot{k})}{\partial \dot{k}^2} \]
The Euler equation is also sufficient provided $V(k, \dot{k})$ is strictly concave because then the strong Legendre condition is satisfied, i.e.,

$$V_{11} < 0; \ V_{22} < 0; \ \Delta \equiv \det \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} > 0 \quad (7)$$

where

$$V_{11} = U''(c)(f'(k)-n'(k)k-n(k))^2 + U'(c)(f''(k)-2n'(k)-n''(k))$$
$$V_{22} = U''(c)$$
$$V_{12} = V_{21} = -U''(c)(f'(k)-n'(k)k-n(k))$$
$$\Delta = U''(c)U'(c)(f''(k)-2n'(k)-n''(k))$$

Because it is assumed that $U'(c) > 0$, $U''(c) < 0$, and $f''(k) < 0$, the strict concavity conditions will be satisfied for:

$$n''(k) > (f''(k) - 2n'(k))/k. \quad (3)$$

Therefore, assume $n''(k)$ satisfies this condition which also assures no conjugate points and uniqueness of solution for a given set of end points.

Let $k^*$ be the unique stationary solution of the Euler equation, $E_t(0,0,k^*) = 0$. Because the turnpike is the optimal golden-age or stationary-state path, $k^*$ is the turnpike. Then,

$$E_t(0,0,k^*) = 0 - V_1(k^*,0) \quad (9)$$
$$= -U'(c^*)(f'(k^*)-n(k^*)-n'(k^*)k^*) = 0$$

where

$$c^* = f(k^*) - n(k^*)k^*.$$

Because it is assumed that $U'(c^*) > 0$, $k^*$ satisfies:

$$f'(k^*) = n(k^*) + n'(k^*)k^*. \quad (10)$$
This is our new Golden Golden-Rule. \(^{(2)}\)

The Swan-Phelps Golden-Rule level for constant population growth, call it \(\bar{K}\), satisfies \(f'(\bar{K}) = n(\bar{K})\). Because \(f'(k)\) is a decreasing function of \(k\), \(\bar{K}\) is greater (less) than \(k^*\) for \(n'(k)\) greater (less) than zero.

The essence fo the Turnpike theorem is the local catenary behavior of optimal dynamic paths near the turnpike. The Turnpike theorem states that, for large enough \(T\), the optimal path will spend most of the time arbitrarily near the turnpike.

Samuelson has shown that, for large enough \(T\), the optimal path will arch toward the turnpike like a catenary \([6]\). I suspect that the catenary behavior could be used to prove for consumption-turnpikes an analogous theorem to the one proved for production-turnpikes by Furuya and Inada \([2]\): namely, that any path, other than the one which spends most of the time near the turnpike, will move away from the turnpike and will eventually hit the axis, contradicting any claim to optimality.

Because of the importance of the catenary behavior, it is necessary to determine the local behavior of the dynamic solution to the Euler equation \((5)\) near the \(k^*\)-turnpike \((10)\). In the one-sector model with no time preferences, the catenary behavior is implied by the characteristic roots of the linear stability equation, associated with the Euler equation, being real, equal-in-magnitude, and opposite-signed.

To examine stability in a neighborhood of the \(k^*\)-turnpike, linearize the Euler equation as follows:

---

\(^{(2)}\) Remark: The assumption that \(n'(k) > 0\) is used implicitly in \((10)\) to assure the existence of the \(k^*\) solution. Nowhere else in the derivation of the turnpike solution was it needed. Therefore, \(n'(k) < 0\) would give a similar result. However, from \((2)\) \(f'(k) > 0\) for \(k < k^*\), and so the additional restriction that \(n(k) > |n'(k)|k^*\) for \(k < k^*\) would be required for existence of a stationary solution. Similarly, for the case where: \(n'(k) > 0\) for \(0 < k < k_1; n'(k_1) = 0; n'(k) < 0\) for \(k_1 < k < k_2; n'(k) = 0\) for \(k_2 < k < k_3; n(k) > 0\) for \(0 < k < k_3\), uniqueness and existence of a solution are assured provided the condition \((8)\) on the second derivative is satisfied. Also it should be noted that \((8)\) is satisfied when \(n(k)\) is a constant.
Let \( y(t) = k(t) - k^* \); \( \dot{y}(t) = \dot{k}(t) \); \( \ddot{y}(t) = \ddot{k}(t) \).

Then, using equations (5) and (9), the linearized Euler equation becomes:

\[
L_{11}\ddot{y}(t) - L_{22}y(t) = 0
\]  
(11)

where

\[
L_{11} = U''(c^*) < 0
\]

\[
L_{22} = U'(c^*)(f'(k^*) - 2n'(k^*) - n''(k^*)k^*) < 0
\]

by the condition (8) on \( n''(k) \).

The solution,

\[
y(t) = A_1e^{\lambda t} + A_2e^{-\lambda t}, \quad \lambda = \sqrt{\frac{L_{22}}{L_{11}}},
\]  
(12)

is Samuelson's catenary in the Per Capita Consumption Turnpike Theorem [6], but now applied to the case of endogenously variable population growth.

### III. The Planner's Dilemma

Should the planner follow the Golden Rule for constant population growth and have society spend most of the time near the capital-labor ratio which equates the interest rate to the population growth rate, or should he follow the rule of the new Turnpike theorem and spend most of the time near the \( k^* \)-turnpike (10)?

To help solve the dilemma, consider the special case: for boundary conditions in (3), let \( k^0 = k^T = \bar{k} \) where:

\[
f'(\bar{k}) = n(\bar{k}), \quad \bar{k} \text{ is the Golden-Rule capital-labor ratio.}
\]  
(13)

In this case, a planner who follows the Golden Rule would tell society to remain at the initial capital-labor ratio and to consume what is left after widening of capital is provided for. The planner who follows the rule of the new Turnpike theorem will tell society to consume more in the beginning, which will lower per capita wealth and hence the rate of population growth until the economy is operating near the new \( k^* \) level. \(^3\)

\[^3\] By a similar argument for \( n'(k) < 0 \) case, society should consume less in the beginning and raise per capita wealth which would lower the rate of population growth.
Then, society should continue to consume at this level until in the last moments of time period \((0, T)\), it would then move away from the turnpike back toward the pre-assigned old \(k\) level so that at time \(T\), it would end-up satisfying the boundary condition, \(k(T) = k^T = \bar{k}\) (see Figure 1).

Clearly, by the general Turnpike theorem, path ABCD gives a larger \(\int_0^T U(c(t))dt\) than the Swan-Phelps path AED. Further, because path AED is in the steady-state (\(\dot{k} = 0\)) for the whole period \((0, T)\) and path ABCD spends most of the period arbitrarily near the steady-state (\(\dot{k}^* = 0\)), it is of interest to compare per capita consumption between the two steady-states.

Mathematically, it is clear that per capita consumption among steady-states (\(\dot{k} = 0\)) is highest at the new \(k = k^*\).

\[
\begin{align*}
\text{Figure 1} \\
\begin{array}{c}
\text{capital/worker} \\
\text{\(k\)}
\end{array}
\begin{array}{c}
\text{A} \\
\text{B} \\
\text{C} \\
\text{D}
\end{array}
\begin{array}{c}
\text{\(f'(k) = n(k)\)} \\
\text{\(f'(k^*) = n(k^*) + n'(k^*)k^* + n''(k^*)k^* k \geq 0\)}
\end{array}
\begin{array}{c}
\text{k^*} \\
\text{0} \\
\text{T}
\end{array}
\text{time}
\end{align*}
\]

Figure 1.

Maximize \(c = \max_k (f(k) - n(k)k)\) \hspace{1cm} (14)
\[
\begin{align*}
dc/dk &= f'(k) - n'(k)k - n(k) = 0 \text{ for } k = k^* \\
d^2c/dk^2 &= f''(k) - 2n'(k) - n''(k)k < 0 \text{ for } k = k^*, \\
&\text{by condition (8).}
\end{align*}
\]
So per capita consumption is maximized among golden ages at
k = k*.

The planner's problem is solved. With the rate of population
growth variable, society should spend most of its time away from
the Swan-Phelps Golden Rule for constant population
growth and close to the k*-turnpike, the Golden Golden-Rule.

The results of this section, which were formerly part of the
main text, are essentially covered in a recent note by E. Davis [1].

GENERALIZATIONS FOR THE PLANNER OF AN
ECONOMY WITH DIFFERENT BEHAVIOR:
1. Population Growth Rate as a Function of Per Capita Con-
sumption. If the planner had faced an economy whose rate of
population growth was a function of per capita consumption,
n = n(c), the resulting stationary solution to (9), the Euler
equation, would be:

\[
\frac{f'(k^*) - n(c^*)}{1 + n'(c^*)k^*} = 0,
\]

which implies f'(k^*) = n(c^*), the Swan-Phelps Golden Rule
(i.e., k^* = \bar{k}), if \(1 + n'(c^*)k^* \neq 0\). Although k = \bar{k}
clearly maximizes per capita consumption among stationary states
(k = 0), this case is qualitatively different from n = n(k)
because c = c(k, \bar{k}) which implies n = n(k, \bar{k}). Therefore, the
dynamic behavior of the optimal path (k \neq 0) is different.
The required second-order conditions on V(k, \bar{k}), (7), which
satisfy the strong Legendre conditions assuring uniqueness
and a smooth interior maximum are complex to derive.

2. Population Growth Rate as a Function of Per Capita Output.
   If the planner had faced an economy whose rate of population
growth was a function of per capita output, n = n(f(k)), the
resulting stationary solution to the Euler equation would be:

\[
f'(k^*) = \frac{n(f(k^*))}{1 - n'(f(k^*))k^*}
\]

where n'(f(k^*)) = \frac{dn}{df} and 1 - n'(f(k^*))k^* > 0.

This is a particular case of the Golden Golden-Rule (10).
n = n(f(k)) is a particular form of n = n(k), and because
f(k) is a strictly monotone function, all order relations are
preserved. The apparent new restriction that n'(f(k^*))k < 1
is implicit in the original equation:

\[
f'(k^*) = \left(\frac{dn}{dk}\right)k^* + n(k^*)
\]

By assumption (2), f'(k) > 0 for k < \bar{k}, and n(k) > 0.
Therefore, to assure a solution to (10), f'(k^*) > (dn/dk)k^*.
In this particular case, \frac{dn}{dk} = n'(f(k))f'(k). Therefore,
the condition becomes f'(k^*) > n'(f(k^*))f'(k^*)k^* and because
f'(k^*) > 0, this implies 1 > n'(f(k^*))k^*.

Notice for n'(f(k)) > 0 (< 0), the k^*-turnpike level is lower
(higher) than \bar{k}, the Swan-Phelps Golden Rule (13), as was
derived in the general case of the Golden Golden-Rule.

(continued)

Consider the plight of the planner who rejects the Samuelson-Diamond criterion for social welfare and instead prefers the Bentham-Lerner criterion of maximizing the total utility of all people who will ever live. For a constant rate of population growth, Samuelson has shown that the turnpike which optimizes Bentham-Lerner social welfare is the Schumpeter zero-interest level where capital is saturated [7]. Therefore, the planner already knows that society should spend most of its time away from the Golden-Rule state. However, is the Schumpeter point still the optimal capital-labor ratio for an economy in which the rate of population growth is a decision variable? The answer is no.

When the rate of population growth is treated as a rising function of the capital-labor ratio rather than as an exogenous constant, the Schumpeter zero-interest level cannot be an optimal solution, and the optimal path will always lie below it. In general, the optimal path for the Bentham-Lerner criterion will not be uniquely determined for the same sufficient conditions (8) specified for the Samuelson-Diamond criterion. However, for somewhat stronger sufficient conditions, a unique turnpike will exist.

Mathematically, the Bentham-Lerner criterion can be written as a constrained calculus of variations problem as follows:

\[
\text{(4)} \text{ (continued)}
\]

In the two-sector model where \( f = f(k, \dot{k}) \), a problem arises in deriving the required second-order conditions which satisfy the strong Legendre conditions, similar to the \( n = n(c) \) case previously described.

3. Time Preference. If the planner has a systematic time preference, his problem would be to

Maximize \( \int_0^T e^{-\rho t} U(c(t)) dt, \quad (\rho < 0) \)  \hspace{1cm} (17)

The new stationary turnpike solution, the Golden Golden-Rule with time preference, is

\[
f'(k^*) = n(k^*) + n'(k^*)k^* + \rho \quad \hspace{1cm} (18)
\]

which is a lower (higher) turnpike for \( \rho > 0(< 0) \) than without time preference. The solution to the resulting linearized Euler equation corresponding to equations (11) and (12) is an unbalanced catenary [6].

(5)
\[
\text{Max } \int_0^T L(t)U(f(k) - n(k)k - \dot{k})dt \quad k(0) \leq k^0; k(T) \geq k^T \quad (18)
\]

subject to \( L(t) = n(k)L(t) \)

For \( U(\cdot) \) and \( f(\cdot) \) concave functions, finding a solution to (18) is equivalent to finding the saddlepoint to the problem

\[
\text{Maximize } \text{Minimize } \int_0^T W(k, k, L, \dot{L}; \lambda)dt \quad k(0) \leq k^0; K(T) \geq K^T = 0 \quad (18')
\]

where

\[
W \equiv L(t)(V(k, \dot{k}) - M) + \lambda(t)(\dot{L}(t) - n(k)L(t))
\]

\( L(t) \) = labor force \( = \) population size

\( \lambda(t) \) = Lagrange multiplier

\( M \) = constant chosen by the planner (see the next paragraph)

\( V(k, \dot{k}) = U(c) = U(f(k) - n(k)k - \dot{k}) \)

\( U, k, \dot{k}, f(k), n(k), c \) are as previously defined in (1) and (2).

Before deriving the optimality conditions, it is important to explain the meaning of the constant \( M \). Because \( U(c) \) is a measure of cardinal utility, it is uniquely determined only up to an affine transformation. Therefore, \( U(c) \) must be written, in general, as \( \alpha U(c) - b \), where \( \alpha \) and \( b \) are constants and \( \alpha > 0 \). Clearly, only the origin of utility, \( b \), could have any effect on the optimal solution (as e.g., maximize \( \int_0^T L(t)(\alpha U(c) - b)dt \) = a maximize \( \int_0^T L(t)(U(c) - M)dt \) where \( M = b/\alpha, \alpha > 0 \)). This shift of the utility origin did not affect the results of the Samuelson-Diamond criterion because the addition of the term \( -\int_0^T Mdt \) is not a function of any decision variables and so does not affect the optimal path. Similarly, for the Bentham-Lerner criterion when \( L(t) \) is not a decision variable, Ramsey [5] and Samuelson [7] could choose

\[\text{5}\] In addition to Lerner and Samuelson, the Bentham-Lerner criterion is discussed by Meade in his chapter on Optimum Population [3, Ch. 6].
M = U(c_B), where c_B is bliss-point consumption, for convenience because the resulting optimal path was independent of the origin for U(c). However, when the rate of population growth is endogenous and L(t) is a decision variable, the value of M chosen by the planner is crucial in determining the turnpike level of the capital-labor ratio.

One interpretation of the economic meaning of the constant M is as follows: Define M = U(c_M), where c_M is that level of per capita consumption below which the planner would not want to bring any new people into the world. Thus, a planner, given a c_M so high that is is not possible for the economy to produce for a long period of time at a level where per capita consumption is greater than or equal to c_M, would prefer, as an optimal solution, to "shut-down" the economy and the human race (i.e., set L(t) = 0). If one maximizes \( \int_0^T L(t)(V(k,k) - M)dt \), L(t) \( \geq 0 \), and the optimal k is such that \( (V - M) < 0 \) for most of the long time-period, (0,T), then \( \int_0^T L(t)(V(k,k) - M)dt \) will be maximized by setting L(t) identically equal to zero. Because U(c) = V(k,k) is a strictly monotone-increasing function of c, the condition \( (V - M) \) less than zero implies that c is less than c_M. For the rest of the paper, I shall refer to optimal paths k for (18) as being admissible optimal paths if the associated optimal L is positive rather than zero (e.g., if k is a steady-state path and if \( (V(k,0) - M) \) is non-negative then k is an admissible steady-state path).

Return now to finding the extremal paths to (18'). The necessary conditions for a saddlepoint are a set of Euler equations:

\[
\frac{d}{dt} \left[ \frac{\partial W}{\partial k} \right] - \frac{\partial W}{\partial k} = 0
\]  
(19)

\[
\frac{d}{dt} \left[ \frac{\partial W}{\partial L} \right] - \frac{\partial W}{\partial L} = 0
\]  
(20)

\[
\frac{d}{dt} \left[ \frac{\partial W}{\partial \lambda} \right] - \frac{\partial W}{\partial \lambda} = 0
\]  
(21)
Equation (21) returns the constraint, \( \dot{L} = n(k)L \). Substituting this constraint and dividing out \( L(t) \) in equation (19), the resulting equations are:

\[
V_{22}k + V_{12}k + n(k)V_2 - V_1 + \lambda n'(k) = 0 \quad (19')
\]

\[
\dot{\lambda} + n(k)\lambda = \lambda V(k,k) - M \quad (20')
\]

Let \( k^{**} \) and \( \lambda^{**} \) be the stationary turnpike solution to equations (19') and (20'). Then, \( k^{**} = \dot{k}^{**} = \lambda^{**} = 0 \)

\[
f'(k^{**}) = n'(k^{**})k^{**} + n'(k^{**})\lambda^{**}/U'(c^{**}) \quad (22)
\]

\[
\lambda^{**} = (U(c^{**}) - M)/n(k^{**}) \quad (23)
\]

where

\[
c^{**} = f(k^{**}) - n(k^{**})k^{**} \quad (24)
\]

By eliminating \( \lambda^{**} \) from (22), the \( k^{**} \)-turnpike solution for the Bentham-Lerner criterion is defined by

\[
f'(k^{**}) = n'(k^{**})k^{**} + n'(k^{**})(U(c^{**}) - M)/n(k^{**})U'(c^{**}). (25)
\]

Query: is \( k^{**} \) unique for a given value of the parameter \( M \) (i.e., is \( k^{**} \) (M) single-valued)? A sufficient condition for the single-valuedness of \( k^{**} \) is that \( dM/dk^{**} \neq 0 \). Solve (25) for \( M \) and calculate \( dM/dk^{**} \):

\[
M = U(c^{**}) - n(k^{**})U'(c^{**})h(k^{**})/n'(k^{**}) \quad (26)
\]

\[
dM/dk^{**} = U'(c^{**})n(k^{**})(n''(k^{**})f'(k^{**}) - n'(k^{**})f''(k^{**}))/
\]

\[
(n'(k^{**}))^2 - U''(c^{**})n(k^{**})(h(k^{**}) - n(k^{**}))h(k^{**})/n'(k^{**}) \]

where

\[
h(k^{**}) \equiv f'(k^{**}) - n'(k^{**})k^{**}.
\]

\( ^6 \) Remark: for the special case, \( n \) constant (i.e., \( n'(k) = 0 \)), the \( \lambda^{**} \) disappears from (22) and, as expected, \( M \) has no effect on the solution, which is then the Schumpeter zero-interest level. In general, it is clear by inspection of (22) and (23) that \( k^{**} = k^{**}(M) \) and \( \lambda^{**} = \lambda^{**}(M) \).
To examine the behavior of $dM/dk^{**}$, divide the domain of $k^{**}$ into two regions:

I. All $k^{**}$ such that $h(k^{**}) \geq n(k^{**})$ (28)

II. All $k^{**}$ such that $n(k^{**}) > h(k^{**}) > 0$

By inspection of (27), for $k^{**}$ contained in I., a new sufficient condition,

$$n''(k) > n'(k)f''(k)/f'(k),$$

replacing (8), will assure $dM/dk^{**}$ is positive and $k^{**}(M)$ is a single-valued function monotone-increasing in region I. In region II, condition (29a) is not sufficient. A stronger condition on $n''(k)$ which assures $dM/dk^{**}$ is positive in II is:

$$n''(k) > [n'(k)/U'(c)]^2U''(c)h(k)(h(k)-n(k))/f'(k) + n'(k)f''(k)/f'(k)$$

(29b)

Therefore, for the remainder of the paper, assume $n''(k)$ satisfies conditions (29), and $k^{**} = k^{**}(M)$ is a single-valued, monotone-increasing function of $M$.

Now consider the effect of various choices for the parameter $M$ upon the optimal stationary-state solution. Because $k^{**}$ is an increasing function of $M$, $L^{**}/L^{**} = n(k^{**}(M))$ is an increasing function of $M$ which implies that the optimal population size, $L^{**}(t)$, will be larger for larger values of $M$ at each instant of time. Therefore, a planner who favors a large population will be one with a large $M$. Examination of three critical values of $M$ will serve to illustrate the effect of $M$ on the turnpike level.

1. Consider a planner with $M = m$ such that $k^{**}(M_1) = k^*$, the Golden Golden-Rule level of maximum per capita

Clearly, if $h(k^{**}) < 0$, from (25), $U(c^{**}) < M$, which implies that $L^{**}(t) = 0$ is the optimal solution.
consumption (10). Then, \( f'(k^{**}(M_1)) = n'(k^{**})k^{**} + n(k^{**}) \) and by substitution into (26),

\[
M_1 = U(c^*) - U'(c^*)n^2(k^*)/n'(k^*),
\]

where \( c^* = f(k^*) - n(k^*)k^* = c^{**}(M_1) \), as defined in (24). From (30), \( M_1 \) is less than \( U(c^{**}) \) which implies that \( k^{**}(M_1) \) is an admissible optimal solution and \( L^{**}(M_1) \) is positive.

2. Consider a planner with \( M = M_2 \), where \( M_2 \) is the largest \( M \) such that \( k^{**}(M) \) is an admissible optimal solution. By the definition of admissible optimal and because \( dM/dk^{**} \) is positive, \( M_2 = U(c^{**}(M_2)) \). Substituting into (25), the equation for the \( k^{**}(M_2) \)-turnpike is:

\[
f'(k^{**}(M_2)) = n'(k^{**})k^{**}
\]

This turnpike level is the generalization of the Samuelson-Lerner-Ramsey turnpike when \( n \) is a constant (i.e., \( n'(k) = 0 \)). Essentially, when \( M = M_2 = U(c^{**}) \), the original extremal integral (18) is minimizing the divergence from bliss (Ramsey, [5]). Also, to follow the technique used by Samuelson [7] to determine the general turnpike solution for the Bentham-Lerner criterion, one could apply a time preference, \( \rho = -n(k^{**}) \), to derive the above turnpike level.

3. Consider a planner with \( M = M_3 \) such that \( k^{**}(M_3) = k_s \), the Schumpeter capital-saturation level. Then, \( f'(k^{**}(M_3)) = f'(k_s) = 0 \), and by substitution into (26),

\[
M_3 = U(c_s) + n(k_s)k_sU'(c_s),
\]

(8) Because the population size in the turnpike state at a given time, \( L^{**}(M) \), is a strictly increasing functional of \( M \), the planner with \( M = M_2 \) will favor the largest population size among all planners. A planner with \( M \) greater than \( M_2 \) will prefer to "shut-down" the economy, i.e., \( L^{**}(M) = 0 \). One could consider \( M_2 \) as a "Malthusian-type" lowest standard of living at which the society will continue to exist. Thus if that standard is higher than the economy can sustain (e.g., a society of aristocrats), then the optimal solution is for that society to cease to exist.
where \( c_s = f(k_s) - n(k_s)k_s = c^{**}(M_3) \) as defined in (24). From (32), \( M_3 \) is larger than \( U(c^{**}) \) which implies that \( k^{**}(M_3) \) is not an admissible optimal solution (i.e., \( L^{**}(M_3) \equiv 0 \)), and the planner would prefer to "shut-down" the economy. Therefore, the Schumpeter point can never be an optimal solution for \( n'(k) \) positive.

To complete the description of the effect of \( M \) on the turnpike solution, consider the two cases: (1) \( n'(k) \equiv 0 \) and (2) \( n'(k) < 0 \). For \( n'(k) \equiv 0 \), from (30), \( M_1 = -\infty \) which confirms earlier findings that the Golden Golden-Rule (or in this case, the Golden-Rule) level is never an optimal solution for \( n \) positive under the Bentham-Lerner criterion. By taking limits on (26), \( M_2 = M_3 \), and the solution from (31) is \( k^{**}(M_2) = k^{**}(M_3) = k_s \), the Schumpeter point. These results were expected because, for \( n \) fixed, the turnpike solution is unique, independent of \( M \). For \( n'(k) < 0 \), there exists no stationary-state solution for the Bentham-Lerner criterion other than the trivial one, \( L^{**}(M) \equiv 0 \). By substituting into (25) the condition for all admissible optimal solutions, \( U(c^{**}) \) greater or equal to \( M \), the result is \( f'(k^{**}) \) less than zero which has no solution for \( k \) less than \( k_s \) by assumption (2). For \( U(c^{**}) \) less than \( M \), \( f'(k^{**}) \) could possibly be non-negative, but the solution is not admissible and so, \( L^{**}(M) \equiv 0 \).

In summary, for \( n'(k) \) positive, the turnpike solution under the Bentham-Lerner criterion is a function of the origin of utility. There exists an upper-bound on the optimal capital-labor ratio which is always below the Schumpeter capital-saturation level. In general, when the optimal capital-labor ratio is an increasing function of the utility origin, the higher the utility origin specified for the planner, the larger is the population size favored by the planner at each moment of time. However, if the utility origin is specified too high, the planner will favor, as an optimal solution, a zero population size.
V. Conclusion. Throughout the paper, I have considered plans only for large finite time, and therefore, I have not been concerned with the convergence of the integrals of utility as the length of the period goes to infinity. For the Bentham-Lerner criterion, conditions for convergence of the integral may not exist because, for \( n = n(k) \), it is no longer possible to pick the origin of utility arbitrarily as Ramsey and Samuelson did for \( n \) constant.

Although some attempt was made to justify, on empirical grounds, the behavioral assumptions about the rate of population growth, clearly these are gross simplifications of the behavior of any actual society.

With these reservations, it was found that for an economy with a variable rate of population growth, the Samuelson-Diamond criterion for social welfare is not optimized at the Swan-Phelps Golden-Rule level for constant population growth, but instead at a lower level, the Golden Golden-Rule. The introduction of a subjective rate of time preference shifts the turnpike level in a straightforward manner.

If the Bentham-Lerner criterion is preferred, the turnpike solution depends upon the utility origin. However, there is an upper-bound on the optimal capital-labor ratio which is always below the Schumpeter capital-saturation level.

In general, for welfare maximization among golden ages, there is nothing sacred about the Golden Rule, the Schumpeter capital-saturation level, or even the Golden Golden-Rule. One would want now to consider the effect of changing the \( n(k) \) function. Such a change would "invalidate" the Golden Golden-Rule.

\(^{(9)}\) I have analyzed, but not here, the cases
Maximize \( \int_{0}^{T} U(f(k) - k\dot{L}/L - \dot{k})dt \) and Maximize \( \int_{0}^{T} U(f(k) - k\dot{L}/L - \dot{k})L(t)e^{-\rho t}dt \).
References


IV. A COMPLETE MODEL OF WARRANT PRICING THAT MAXIMIZES UTILITY

Introduction. †

In a paper written in 1965, one of us developed a theory of rational warrant pricing. 1 Although the model is quite complex mathematically, it is open to the charge of over-simplification on the grounds that it is only a "first-moment" theory. 2 We now propose to sketch a simple model that overcomes such deficiencies. In addition to its relevance to warrant pricing, the indicated general theory is of interest for the analysis of other securities since it constitutes a full supply-and-demand determination of the outstanding amounts of securities.

Cash-Stock Portfolio Analysis.

Consider a common stock whose current price $X_t$ will give rise $n$ periods later to a finite-variance, multiplicative probability distribution of subsequent prices, $X_{t+n}$, of the form:

$$\text{Prob } \{X_{t+n} \leq X \mid X_t = Y\} P(X_t,Y;n) = P(X/Y;n) \quad (1)$$

where the price ratios, $X_{t+n}/X_t = Z = Z_1 Z_2 \ldots Z_n$, are assumed to be products of uniformly and independently distributed distributions, of the form,

$$\text{Prob } Z_1 \leq Z = P(Z;1), \text{ and where, for all integral } n \text{ and } m, \text{ the Chapman-Kolmogorov relation } P(Z;n+m) = \int_0^\infty P(Z/z;n)dP(z;m)$$

is satisfied. This is the "geometric-Brownian motion," which at least asymptotically approached the familiar log-normal.

Ignoring for simplicity any dividends, we know that a risk averter, one with concave utility and diminishing marginal utility, will hold such a security in preference to zero-yielding safe cash only if the stock has an expected positive gain:

$$0 < E[Z] - 1 = \int_0^\infty ZdP(Z;n) - 1 = e^{\alpha n} - 1, \text{ that is, } \alpha > 0 \quad (2)$$

† Footnotes for this article appear at the end of the article.

‡ Aid from the National Science Foundation is gratefully acknowledged.
where the integral is the usual Stieltjes integral that handles discrete probabilities and densities, and $\alpha$ is the mean expected rate of return on the stock per unit time. (We have ensured that $\alpha$ is constant, independent of $n$.)

A special case would be, for $n=1$, the following discrete distribution, where $\lambda > 1$:

$$X_{t+1} = \lambda X_t \quad \text{with probability } p \geq 0$$

$$X_{t+1} = \lambda^{-1} X_t, \quad \text{with probability } 1 - p \geq 0$$

(3)

This simple geometric Brownian motion leads asymptotically to the log-normal distribution. Condition (2) becomes, in this special case, $0 < E[Z] - 1 = p\lambda + (1-p)\lambda^{-1} - 1$. If, for example, $\lambda = 1.1$ and $p = 1 - p = 0.5$, then $E[Z] - 1 = 1/2(1.1 + 10/11) = .004545$. If our time units are measured in months, this represents a mean gain of almost one-half a per cent per month, or about 5 1/2 per cent per year, a fair approximation to the recent performance of a typical common stock.

To deduce what proportion cash holding will bear to the holding of such a stock, we must make some definite assumption about risk aversion. A fairly realistic postulate is that everyone acts now to maximize his expected utility at the end of $n$ periods and that his utility function is strictly concave. Then by portfolio analysis 3 in the spirit of the classical papers of Domar-Musgrave and Markowitz (but free of their approximations), the expected utility is maximized when $w = w^*$, where $w$ is the fraction of wealth in the stock:

$$\max_{w} U(w) = \max_{w} \int_{0}^{\infty} U[(1 - w) + wZ]dP(Z;n)$$

(4)

where $w = w^*$ is the root of the regular condition for an interior maximum

$$0 = U'(w^*) = \int_{0}^{\infty} [U'[(1 - w^*) + w^*Z] - U'[(1-w^*)+ w^*Z]] dP(Z;n)$$

or

$$1 = \frac{\int_{0}^{\infty} ZU'[(1 - w^*) + w^*Z]dP(Z;n)}{\int_{0}^{\infty} U'[(1 - w^*) + w^*Z]dP(Z;n)}$$

(5)
Since $U$ is a concave function, $U''$ is everywhere negative, and the critical point does correspond to a definite maximum of expected utility. (Warning: Equations like (4) posit that no portfolio changes can be made before the $n$ periods are up, an assumption modified later.)

If zero-yielding cash were dominated by a safe asset yielding an instantaneous force of interest $r$, and hence $e^{rn}$ in $n$ periods, terms like $(1-w^*)$ would be multiplied by $e^{rn}$ and (5) would become

$$e^{rn} = \frac{\int_0^\infty U'[(1 - w^*)e^{rn} + w^*Z]dP(Z;n)}{\int_0^\infty U'[(1 - w^*)e^{rn} + w^*Z]dP(Z;n)} < e^{rn}, \text{ if } w^* > 0 \quad (5a)$$

This relationship might well be called the Fundamental Equation of Optimizing Portfolio theory. Its content is worth commenting on. But first we can free it from any dependence on the existence of a perfectly safe asset. Re-writing (4) to involve any number $m$ of alternative investment outlets, subject to any joint probability distribution, gives the multiple integral:

$$\max_{\{w_j\}} U[w_1, \ldots, w_m] = \max_{\{w_j\}} \int_0^\infty U[\sum_j w_j Z_j]dP(Z_1, \ldots, Z_m;n) \quad (4a)$$

Introducing the constraint, $\sum_{j=1}^m w_j = 1$, into the Lagrangian expression, $L = U + \gamma[1 - \sum_{j=1}^m w_j]$, we derive as necessary conditions for a regular interior maximum:

$$\frac{\delta L}{\delta w_k} = 0 = \int_0^\infty Z_k U'[\sum_{j=1}^m w_j Z_j]dP(Z_1, \ldots, Z_m;n) - \gamma, \text{ for } k=1, \ldots, m$$

Dividing through by a normalizing factor, we get the fundamental equation:

$$\int_0^\infty Z_1 dQ(Z_1, \ldots, Z_m;n) = \int_0^\infty Z_2 dQ(Z_1, \ldots, Z_m;n) = \ldots$$

$$\int_0^\infty Z_m dQ(Z_1, \ldots, Z_m;n) \quad (5b)$$

where
\[ dQ(Z_1, \ldots, Z_m; n) = \frac{U' \left[ \sum_{i=1}^{m} w_j Z_i \right] dB(Z_1, \ldots, Z_m; n)}{\int_{0}^{\infty} U' \left[ \sum_{j=1}^{m} w_j Z_j \right] dB(Z_1, \ldots, Z_m; n)} \]

The probability-cum-utility function \( Q(Z; n) \) has all the properties of a probability distribution, but it weights the probability of each outcome so to speak by the marginal utility of wealth in that outcome.

Figure 1 illustrates the probability density of good and bad outcomes; Figure 2 shows the diminishing marginal utility of money; and Figure 3 plots the "effective-probability" density whose integral \( \int_{0}^{Z} dB(Z; n) \) defines \( Q \). Conditions (5), (5a), and (5b) say, in words, that the "effective probability" mean of every asset must be equal in every use, and, of course, be equal to the yield of a safe asset if such an asset is held. Note that \( U'(0) = E[Z] - e^{rn} = e^{\alpha n} - e^{rn} \), and this must be positive if \( w^* \) is to be positive. Also \( U'(1) = \int_{0}^{\infty} ZdQ(Z; n) - e^{rn} \), and this cannot be positive if the safe asset is to be held in positive amount. By Kuhn-Tucker methods, interior conditions of (5) could be generalized to the inequalities needed if borrowing or short-selling are ruled out.

For the special probability process in (3) with \( p = 1/2 \) and Bernoulli logarithmic utility, we can show that expected utility turns out to be maximized when wealth is always divided equally between cash and the stock, i.e. \( w^* = 1/2 \) for all \( \lambda \):

\[
\begin{align*}
\text{Max } U(w) &= \text{Max } \left\{ \frac{1}{2} \log (1-w+w\lambda) + \frac{1}{2} \log (1-w+w\lambda^{-1}) \right\} \\
&= \frac{1}{2} \log \left( \frac{1}{2} + \frac{1}{2} \lambda \right) + \frac{1}{2} \log \left( \frac{1}{2} + \frac{1}{2} \lambda^{-1} \right), \text{ for all } \lambda.
\end{align*}
\]

The maximum condition corresponding to (5) is

\[
0 = U'(w^*) = \frac{1}{2 + \frac{1}{2} \lambda} (-1 + \lambda) + \frac{1}{2 + \frac{1}{2} \lambda^{-1}} (-1 + \lambda^{-1}), \text{ and } \]

\[
w^* = \frac{1}{2} \text{ for all } \lambda. \text{ Q.E.D.}
\]
(The portfolio division is here so definitely simple because we have postulated the special case of an "unbiased" logarithmic price change coinciding with a Cernoulli logarithmic utility function; otherwise changing the probability distribution and the typical person's wealth level would generally change the portfolio proportions.)

Recapitulation of the 1965 Model.

Under what conditions will everyone be willing to hold a warrant (giving the right to buy a share of the common stock for an exercise price of $1 per share at any time in the next n periods), and at the same time be willing to hold the stock and cash? Since the warrant's price will certainly move with the common rather than provide an opposing hedge against its price movements, if its expected rate of return were not in excess of the safe asset's yield, the warrant would not get held. In the 1965 paper, it was arbitrarily postulated that the warrant must have a specified gain per dollar which was as great or greater than the expected return per dollar invested in the common stock. Thus, if we write \( Y_t(n) \) for the price at time \( t \) of a warrant with \( n \) periods still to run, the 1965 paper assumed for stock and warrant,

\[
E[X_{t+T}/X_t] = e^{\alpha T} > e^{rT} \tag{8a}
\]

\[
E[Y_{t+T(n-T)}/Y_t(n)] = e^{\beta T} > e^{\alpha T}, \text{ if the warrant is to be held} \tag{8b}
\]

In (8b), we recognize that after the passage of \( T \) periods of time, the warrant has \( n-T \) rather than \( n \) periods left to run until its exercise privilege expires. It should be stressed that the warrant can be exercised any time (being of "American" rather than "European" option type) and hence in (8b) the warrant prices can never fall below their arbitrage exercise value, which in appropriate units (i.e., defining the units of common so that the exercise price of the warrant is unity) is given by \( \text{Max} (0, X_t - 1) \). Thus, we can always convert the
warrant into the common stock and sell off the stock (commi-
missions are here neglected).

In the 1965 model, the expected percentage gain of $\beta$ of
a warrant and the expected percentage gain $\alpha$ of a common were
arbitrarily postulated as exogenously given data, instead of
being deduced from knowledge of the risk aversion properties of
$U$. Postulating a priori knowledge of $\alpha$ and $\beta$, the model was
derived by beginning with the known arbitrage value of a warrant
about to expire, namely:

$$Y_t(0) = \text{Max}(0, X_t - 1) = F_0(X_t)$$

(9)

Then if the warrant is to be held, we can solve (8b) for
$Y_t(1) = F_1(X_t)$ from the equation:

$$e^\beta = \frac{E[F_0(XZ)/F_1(X)|X]}{F_1(X)}$$

(10)

In this integral and elsewhere we can write $X$ for $X_t$. If
(10) is not achievable, the warrant will be converted, and
will now be priced at its $F_0(X)$ value. Hence, in every case:

$$F_1(X) = e^{-\beta} \int_0^\infty F_0(XZ)dP(Z;1), \text{ if held}$$

$$= X-1 \geq e^{-\beta} \int_0^\infty F_0(XZ)dP(Z;1), \text{ if now converted}$$

$$= \text{Max}[0, X-1, e^{-\beta} \int_0^\infty F_0(XZ)dP(Z;1)], \text{ in all cases}$$

(10a)

Successively putting in these expression $F_2$ and $F_1$ for $F_1$
and $F_0, \ldots, F_{n+1}$ and $F_n$ for $F_1$ and $F_0$, the 1965 model deduced
rational warrant price formulas, $F_n(X) = F_n(X_t) = Y_t(n)$, for
any length of life; and the important perpetual warrant case,
$F_\infty(X) = F(X)$, can be deduced by letting $n \to \infty$:

$$F(X) = e^{-\beta} \int_0^\infty F(XZ)dP(Z;1), \text{ if } X \leq C(\alpha, \beta)$$

$$= X-1 \geq e^{-\beta} \int_0^\infty F(XZ)dP(Z;1), \text{ if } X > C(\alpha, \beta)$$

(11)
where \( C(\alpha, \beta) \) is the critical level at which the warrant will be worth more dead than alive. This critical level will be defined by the above relations and will be finite if \( \beta > \alpha \).

The special case of the 1965 theory in which \( \alpha = \beta \) is particularly simple, and its mathematics turns out to be relevant to the new utility theory presented here. In this case, where conversion is never profitable (for reasons which will be spelled out even more clearly in the present paper), the value of the warrants of any duration can be evaluated by mere quadrature, as the following linear integrals show:

\[
F_n(X) = e^{-\alpha n} \int_0^\infty \int_0^\infty F_{n-T}(XZ)dP(Z;T)
\]

\[
= e^{-\alpha n} \int_0^\infty F_0(XZ)dP(Z;n)
\]

\[
= e^{-\alpha n} \int 1/x (XZ-1)dP(Z;n)
\]  

(12)

In concluding this recapitulation, let us note that the use of short, discrete periods here gives a good approximation to the mathematically difficult, limiting case of continuous time in the 1965 paper and its appendix.

Determining Average Stock Yield.

To see how we can deduce rather than postulate in the 1965 manner the mean return that a security must provide, let us first assume away the existence of a warrant and try to deduce the mean return of a common stock. The answer must depend on supply and demand: supply as dependent upon risk-averters willingness to part with safe cash, and demand as determined by the opportunities nature affords to invest in real, risky processes along a schedule of diminishing returns.

To be specific, suppose one can invest today's stock of real output (chocolates or dollars when chocolate always sell for $1 each) either (a) in a safe (storage-type) process - cash, so to speak - that yields in the next period exactly one chocolate; or (b) in a common stock, which in the special case (3) gives for each chocolate "invested" today \( \lambda \) chocolates tomorrow
with probability \( p \), or \( \lambda^{-1} \) chocolates with probability \( 1-p \). If we allocate today's stock of chocolates so as to maximize the expected utility, we shall shun the risk process unless its expected yield exceeds unity. For the special case, \( p = 1-p = 1/2 \), this will certainly be realized, and as seen in the earlier discussion of (7), for all \( \lambda \), a Bernoulli-utility maximizer will choose to invest half of present resources in the safe (cash) process and half in the risky (common-stock) process.

Now suppose that the risky process—say growing chocolate on the shady side of hills where the crop has a .5 chance of being large or small—is subject to diminishing returns. With the supply of hill land scarce, the larger the number of chocolates planted, rather than merely stored, the lower the mean return per chocolate (net of any competitive land rents for which the limited supply of such land will bid to at each level of total investment in risk chocolates). Although it is admittedly a special-case assumption, suppose that \( \lambda \) in (3) drops toward unity as the absolute number of chocolates invested in the risky process rises but that \( p = 1-p = 1/2 \) throughout. Then the expected yield, \( a = e^\alpha - 1 \), drops toward zero as \( \lambda \) drops toward one.

Given the initial supply of chocolates available for safe or risk allocations, the expected yield of the common stock, \( a \), will be determined at the equilibrium intersection of total supply and demand—in our simple case at the level determined by the \( \lambda \) and \( a \) yields on the diminishing returns curve where exactly half of the available chocolates go into the risk process.

Determining Warrant Holdings and Prices.

Using the general method outlined above, we can now deduce what warrants must yield if a prescribed amount of them is to be held alongside of cash and the common stock by a maximizer of expected utility.

Specifically, assume that cash in an insured bank account, or a safe process, has a sure yield of \( e^r - 1 \) per unit time.
Assume that each dollar invested in the common stock has a mean ex-ante yield \( \int_0^\infty ZdP(Z;1) - 1 = e^\alpha - 1 \) per period. It will be desirable now to specialize slightly our assumption of concave total utility so that the behavior of a group of investors can be treated as if it resulted from the deliberation of a single mind. In order that asset totals should behave in proportions independent of the detailed allocations of wealth among individuals, we shall assume that every person has a constant elasticity of marginal utility at every level of wealth and that the value of this is the same for all individuals. \(^9\) Just as assuming uniform homothetic indifference curves frees demand curve analysis in non-stochastic situations from problems of disaggregation, a similar trick comes in handy here.

Finally, we must specify how many of the warrants are to be outstanding and in need of being voluntarily held. There is a presumption that to induce people to hold a larger quantity of warrants, their relative yields will have to be sweetened. Let the amounts of total wealth, \( W \), to be invested in cash, common stock, and warrants be respectively \( w_1, w_2, \) and \( w_3 \). As already seen, there is no loss of generality in setting \( W = 1 \). Then subject to the constraint, \(^ {10} \) \( w_1 + w_2 + w_3 = W = 1 \), we consider the following special case of (4a) and generalization of (4):

\[
\max \{ w_1, w_2, w_3 \} = \max \int_0^\infty \left[ w_1 e^{rT} + w_2 Z + w_3 \frac{F_n(XZ)}{F_{n+T}(X)} \right] dP(Z;T)
\]

(13)

where, as before, we assume that the decision is made for a period of length \( T \). (Setting \( T = 1 \), a small period, would be typical.) To explain (13), note that \( e^{rt} \) is the sure return to a dollar invested in the common stock. Since we can with \$1\ buy \( 1/P_{n+T}(X) \) units of a warrant with \( n + T \) periods to go, and since these turn out after \( T \) periods to have the random-variable price \( F_n(XZ) \), clearly \( w_3 \) is to be multiplied by the per-dollar return \( F_n(XZ)/F_{n+T}(X) \) as indicated. \(^{11}\) As in (4a), we seek a critical point for the Langrangian expression
\[ L = U + \lambda [1 - \sum_{j=1}^{3} w_j] \] to get the counterpart of (5b), namely:

\[ \frac{\partial L}{\partial w_1} = 0 = -\lambda + \alpha \int_0^\infty e^{-T} U'[w_1 e^{rT} + w_2 Z + w_3 \frac{F_n(XZ)}{F_{n+T}(X)}] dP(Z;T) \]

\[ \frac{\partial L}{\partial w_2} = 0 = -\lambda + \int_0^\infty ZU'[w_1 e^{rT} + w_2 Z + w_3 \frac{F_n(XZ)}{F_{n+T}(X)}] dP(Z;T) \]

\[ \frac{\partial L}{\partial w_3} = 0 = -\lambda + \int_0^\infty \frac{F_n(XZ)}{F_{n+T}(X)} U'[w_1 e^{rT} + w_2 Z + w_3 \frac{F_n(XZ)}{F_{n+T}(X)}] dP(Z;T) \]

\[ \frac{\partial L}{\partial \lambda} = 0 = 1 - w_1 - w_2 - w_3 \quad (14) \]

Eliminating \( \lambda \), we end up with special cases of (5b), namely:

\[ e^{rT} = \frac{\int_0^\infty ZU'[\lambda - w_2^* - w_3^*] e^{rT} + w_2^* Z + w_3^* \frac{F_n(XZ)}{F_{n+T}(X)}] dP(Z;T)}{C} \quad (15) \]

\[ e^{rT} = \frac{\int_0^\infty \frac{F_n(XZ)}{F_{n+T}(X)} U'[\lambda - w_2^* - w_3^*] e^{rT} + w_2^* Z + w_3^* \frac{F_n(XZ)}{F_{n+T}(X)}] dP(Z;T)}{C} \quad (16) \]

where we have the normalizing factor

\[ C = \int_0^\infty U'[\lambda - w_2^* - w_3^*] e^{rT} + w_2^* Z + w_3^* \frac{F_n(XZ)}{F_{n+1}(X)}] dP(Z;T) \]

so that, as in (5b),

\[ dQ(Z;T) = \frac{U'[\lambda - w_2^* - w_3^*] e^{rT} + w_2^* Z + w_3^* \frac{F_n(XZ)}{F_{n+1}(X)}] dP(Z;T)}{C} \]
If the \( w_j^* \) were prescribed—e.g. as the solution to a simultaneous-equation supply and demand process that auctions off the exogenously given supplies of common stock and warrants at the prices that will just get them voluntarily—then, for \( T=1 \), (16) would become an implicit equation enabling us to solve for the unknown function \( F_{n+1}(X) \) recursively in terms of assumed known function \( F_n(X) \). Since \( F_0(X) \) is known from arbitrage-conversion considerations, (16) does provide an alternative theory to the 1965 first-moment theory.

Let us now call attention to the fact that the implicit equation in (16) for \( F_{n+1}(X) \) can be enormously simplified in the special case where the number of warrants held is "small". Thus, for \( w_j^*=0 \), or nearly so, the dependence of \( U'[. \) on \( F_{n+1}(X) \) becomes zero, or negligible; and (16) becomes a simple, linear relationship for determining \( F_{n+1}(.) \) recursively from \( F_n(.) \). If \( w_j^*=0 \), (15) and (16) become:

\[
e^{rT} = \int_0^\infty \frac{U'[1-w_2^*]e^{rT+w_2^*Z}}{C} dP(Z;T) \tag{15a}
\]

\[
e^{rT} = \int_0^\infty \frac{F_n(XZ)/F_{n+1}(X)U'[1-w_2^*]e^{rT+w_2^*Z}}{C} dP(Z;T) \tag{16a}
\]

Our task will thus be simplified when we specify that the number of warrants to be held is "small"; that is, warrant pricing is to be determined at the critical level just necessary to induce an incipient amount of them to be voluntarily held. This is an interesting case because it is also the critical level at which hedging transactions, involving buying the common and selling a bit of the warrant short, just become desirable. Most of our paper will be concerned with this interesting "incipient warrant" case based on (15a) and (16a), but on pages 25-26 of [7], it is shown how one might deduce the quantitative level of all \( w_j^* \) in terms of given supplies of the various securities.

Utility-Maximizing Warrant Pricing: The Important "Incipient" Case

The factors in (16a) can be rearranged to get, for \( T=1 \),

\[
F_{n+1}(X) = e^{-r} \int_0^\infty F_n(XZ) dQ(Z;1) \text{ where } dQ(Z;1) \text{ is short for: } \tag{20}
\]

\[
dQ(Z;1;r,w_2^*) = \frac{U'[1-w_2^*]e^{r+w_2^*Z}dP(Z;1)}{\int_0^\infty U'[1-w_2^*]e^{r+w_2^*Z}dP(Z;1)}
\]
Here \( w_2^* \) is a parameter already determined from solving (15a), and indeed is precisely the same as the \( w^* \) determined earlier from solving equation (5a).

It will be recalled that \( Q(Z;1) \) is a kind of util-prob distribution. Precisely because of (15a), we know that the expected value of \( Z \) calculated, not in terms of true objective probability distribution \( dP(Z;1) \), but rather in terms of the util-prob distribution \( dQ(Z;1) \), has a "yield" per unit time exactly equal to that of the safe asset. Rearranging (14a), we have:

\[
\int_0^\infty zdQ(Z;1) = e^x < e = \int_0^\infty zdP(Z;1).
\]

(21)

Taken together with the initial condition from (9): \( F_0(X) = \text{Max}(0,X-1) \), equations (20) and (21) give us linear recursion relationships to solve our problems completely, provided we can be sure that they always yield \( F_n(X) \) values that definitely exceed the conversion value of \( F_0(X) \). Because of (21), we are here in a mathematical situation similar to the 1965 special case in which \( \chi = \beta \), and indeed no premature conversion is ever possible. But of course there is this significant difference: in the 1965 case, \( dP \) rather than \( dQ \) is used to compute \( \chi \) and \( \beta \), and to emphasize this, we write \( \chi = \chi_p = \beta_p \) for that case; in the present case, where \( dQ \) is used in computation, we write \( \chi_Q \) and \( \beta_Q \), recognizing from (21) that \( \chi_Q = r \) and from (20) that \( \beta_Q = \alpha_Q \). The \( \chi_Q \) and \( \beta_Q \) "yields" are purely "hypothetical" or subjective; they should not be identified with the higher "objective" \( \chi_p \) and \( \beta_p \) yields computed with actual probability \( dP \). These are the true ex ante expected percentage yields calculated from actual dollar gains and losses; they are objective in the sense that Monte Carlo experiments replicated a large number of times will, within this probability model characterized by \( P(Z;1) \), actually average out ex post with mean yields of \( \chi_p \) and \( \beta_p \) on the common stock and warrants respectively. (15)

The mathematics does not care about this \( dP \) and \( dQ \) distinction. The same kind of step-by-step algorithm is yielded whatever the interpretation of the probability distribution used. But this new approach does raise an awkward question. In the 1965 paper, it could be taken as almost self-evident that conversion can never be mandatory if both warrant and stock have the same ex ante yield. In this case, where the yields calculated with \( dQ(Z;1) \) are of a hypothetical kind, it is desirable to provide a rigorous proof that our new theory of warrant pricing never impinges on the inequalities set by arbitrage.
as discussed above and in the 1965 paper.

If we are assured of non-conversion, the value of a perpetual warrant can be determined from the linear integral equation (20). For \( n \) so large that it and \( n+1 \) are indistinguishable, we can write:

\[
F_n(X) = F_{n+1}(X) = F_\infty(X) = F(X),
\]

and (20) becomes:

\[
F(X) = e^{-X} \int_{0}^{\infty} F(XZ) dQ(Z;1)
\]  
(22)

Substituting \( F(X)=X \) into (22) does turn out to provide a solution. So too would \( cX \), but only for \( c=1 \) can we satisfy the two-sided arbitrage conditions \( X > F(X) > X-1 \).

Actually, the homogeneous integral equation (22) has other solutions of the form \( cX^m \), where substitution entails

\[
cX^m = e^{-X} cX^m \int_{0}^{\infty} Z^m dQ(Z;1)
\]

\[
1 = e^{-X} \int_{0}^{\infty} Z^m dQ(Z;1) = \phi(m).
\]

This last equation will usually be a transcendental equation for \( m \), with an infinite number of complex roots of which only \( m=1 \) is relevant in view of our boundary conditions. (16)

That our new theory leads to the perpetual warrant being priced equal to the common stock may seem paradoxical, just as in the 1965 special case where \( \phi_0 = \phi_p \). In [7], pages 30–31, this paradox is discussed in detail. In appendix B of that paper, it is shown that the introduction of dividends essentially resolves the paradox by "forcing" \( F(X) < X \).

**Explicit Solutions**

In a sense, our new theory is completed by the step-by-step solution of (20). In the 1965 theory, however, it was possible to display explicit formulas for non-converted warrants by quadrature or direct integration over the original \( F_0(X) \) function. The same procedure is possible here by introducing some further generalizations of our util-prob distribution \( Q(Z;1) \).

There are some by-no-means obvious complications in our new theory. Given the quadrature formula:

\[
F_1(X) = e^{-X} \int_{0}^{\infty} F_0(XZ) dQ(Z;1)
\]

(24)
one is tempted at first to write, as would be possible in the 1965 case, where
\( dP \) replaced \( dQ \):

\[
F_2(X) = e^{-2r} \int_0^\infty F_0(XZ) \, dQ(Z;2)
\]

................................................

................................................

(25)

or, in general:

\[
F_n(X) = e^{-nr} \int_0^\infty F_0(XZ) \, dQ(Z;n)
\]

(26)

where, as in (5b), we define

\[
dQ(Z;n) = \frac{U'[w_1 e^{I_u + w_2 \lambda Z}]dP(Z;n)}{\int_0^\infty U'[w_1 e^{I_u + w_2 \lambda Z}]dP(Z;n)}
\]

But these relations are not valid. They would be valid only if, say in the case \( n = 2 \), we locked ourselves in at the beginning to a choice of portfolio that is frozen for both periods, regardless of the fact that after one period has elapsed, we have learned the outcomes \( X_{t+1} \) and by (20) would want to act anew to create the proper \( w_j^* \) proportions for the final period. (For example, suppose as in (7), we have \( U = \log W \) and there is an equal chance of the stock's doubling or halving, with \( \lambda = 2 \), \( p = 1/2 = 1-p \). Suppose we put half our wealth into cash at the beginning and freeze our portfolio for two periods. Then we are violating the step-by-step solution of (20) if, after we have learned that the stock has doubled, we do not sell-out half our gain and put it into cash for the second period.)

(17) In summary, (25) is not consistent with (24) and

\[
F_2(X) = e^{-r} \int_0^\infty F_1(XZ) dQ(Z;1)
\]

(27)

If direct quadrature with \( Q(Z;n) \) is not valid, what is? What we need are new iterated integrals, \( Q_2(Z), \ldots, Q_n(Z) \), which reflect the compound probabilities for \( 2, \ldots, n \) periods ahead when the proper non-frozen portfolio changes have been made. Rather than derive these by tortuous economic intuition, let us give the mathematics its head and merely make successive substitutions. Thus, from (20) applied twice, we get:
\[
F_{n+2}(X) = e^{-r} \int F_{n+1}(XZ) \, dQ(Z;1) \\
= e^{-r} \int [e^{-r} \int F_n(XZ \mid V) \, dQ(V;1)] \, dQ(Z;1) \\
= e^{-2r} \int F_n(XZ) \, d[\int Q(Z \mid V;1) \, dQ(Z;1)] \\
= e^{-2r} \int F_n(XR) \, dQ_2(R) 
\]

where

\[
Q_2(R) = \int Q(R \mid Z;1) \, dQ(Z;1)
\]

and where the indicated interchange in the order of integration of the double integral can be straightforwardly justified.

This suggests defining the iterated integrals (18) by a process which becomes quite like that of convolution when we replace our variables by their logarithms, namely, relations like those of Chapman-Kolmogorov:

\[
Q_1(Z) = Q(Z;1), \text{ by definition} \\
Q_2(Z) = \int Q_1(Z \mid V) \, dQ_1(V) \neq Q(Z;2) \\
\vdots \\
Q_{n+1}(Z) = \int Q_n(Z \mid V) \, dQ_n(V) 
\]

Then, by repeated use of (28)'s substitutions, the results of the step-by-step solution of (20) can be written in terms of mere quadratures, namely,

\[
F_1(X) = e^{-r} \int F_0(XZ) \, dQ_1(Z) \\
F_2(X) = e^{-2r} \int F_0(XZ) \, dQ_2(Z) \\
\vdots \\
F_n(X) = e^{-nr} \int F_0(XZ) \, dQ_n(Z) 
\]

Fortunately, the "subjective yields," \( \gamma_Q \) and \( \gamma_{Q_i} \), calculated for the new generalized util-prob functions \( Q_\ell(Z) \), do all equal \( r \) per unit time. That is, we can prove by induction:
\[ \int_0^\infty ZdQ_1(z) = e^r < e^\alpha \]
\[ \int_0^\infty ZdQ_2(z) = e^{2r} \]

\[ \int_0^\infty ZdQ_n(z) = e^{nr} \]

This is an important fact, needed to ensure that the solutions to our new theory never fall below the arbitrage levels at which conversion would be mandatory.

**Warrants Never to be Converted**

It was shown in the 1965 paper that for \( \beta > \alpha \) and \( \beta \), a constant, the warrants would always be converted at a finite stock price level. We will show that in the present model with its explicit assumption of no dividends, the warrants are never converted (i.e., \( F_n(X) > F_0(X) \)).

**Theorem:** If \( \int_0^\infty ZdQ_n(z) = e^{-rn} \) and \( F_n(X) = e^{-rn} \int_0^\infty F_0(XZ) dQ_n(z) \), then \( F_n(X) \geq F_0(X) = \text{Max}(0, X-1) \) and we are in the case where the warrants need never be converted prior to expiration.

**Proof:** Since \( F_0(X) \geq X-1 \), it is sufficient to show that

\[ X-1 \leq e^{-rn} \int_0^\infty F_0(XZ) dQ_n(z) \equiv \Phi_n(X;r) \]

holds for all \( r > 0 \), \( n > 0 \), and \( X > 0 \). We show this as follows:

\[ \Phi_n(X;r) \equiv e^{-rn} \int_0^\infty (XZ-1)dQ_n(z) \text{ because } F_0(XZ) \geq XZ-1 \text{ and } dQ_n(z) \geq 0 \]

\[ \geq Xe^{-rn} \int_0^\infty ZdQ_n(z) - e^{-rn} \]

\[ \geq X - e^{-rn} \geq X - 1 \text{ from (31) for all } r \geq 0, \ n > 0, \text{ and } X \geq 0. \]

Therefore, (32) holds and the theorem is proved.

Thus, we have validated the step-by-step relations of (20) or the one-step quadrature formula of (30).

As an easy corollary of this theorem, we do verify that longer life of a warrant can at most enhance its value, i.e., \( F_{n+1}(X) \geq F_n(X) \).
For, from the theorem itself, \( F_1(X) \geq F_0(X) \), and hence
\[
F_2(X) = e^{-\int_0^\infty F_1(XZ) \, dQ(Z;1)} \geq e^{-\int_0^\infty F_0(XZ) \, dQ(Z;1)} = F_1(X).
\]
And, inductively, if \( F_t(X) \geq F_{t-1}(X) \) for all \( t \leq n \), it follows that
\[
F_{n+1}(X) = e^{-\int_0^\infty F_n(XZ) \, dQ(Z;1)} \geq e^{-\int_0^\infty F_{n-1}(XZ) \, dQ(Z;1)} = F_n(X).
\]
If \( Q(Z;1) > 0 \) for all \( Z > 0 \) and \( Q(Z;1) < 1 \) for all \( Z < \infty \), we can write strong
inequalities \( F_{n+1}(X) > F_n(X) > F_{n-1}(X) > \ldots > F_1(X) > F_0(X) \). The log-normal case.
belongs to this class. If, however, as in example three, \( Q(Z;1) = 0 \) for
\( Z < \lambda \) and for \( Z > \lambda \), \( F_1(X) \) will vanish for some of the same \( X \) values
where \( F_0(X) \) vanishes; \( F_1(X) \) will equal \( (X - 1) = F_0(X) \) for large enough \( X \) values.
Hence, our weak inequalities are needed in general. However, for \( n \) large enough
and \( X \) fixed, we can still write the strong inequality, namely \( F_{n+1}(X) > F_n(X) \)
for \( n < n^a(X) \). The crucial test is this: if for a given \( X \), one can in \( T \) steps
end up both above or below the conversion price of 1, then \( F_T(X) > F_0(X) \) and
\( F_{n+T}(X) > F_n(X) \). Also, if \( F_n(X) > F_0(X) \) for a particular \( X \), \( F_{n+T}(X) > F_n(X) \)
for that \( X \).

Proof of the Superiority of Yield of Warrants Over Yield of Common Stock

First we wish to state an important lemma upon which this proof and other
results rest. Proof of this lemma and indeed of a wider lemma of which this is
a special case is found in appendix A of [7]. Broadly speaking, what we wish
to show is that if two perfectly positively correlated securities are to be
held in the same portfolio, with the outcome of one being a monotone-increasing
function of the other, but with its possessing greater "volatility" in the
sense of its elasticity with respect to the other exceeding one, the mean
yield of the volatile security must exceed the mean yield of the less volatile
one.

We define the elasticity of the function \( \eta(Y) \) with respect to \( Y \), \( E_\eta \),
in the usual fashion as,
\[
E_\eta \equiv d(\log \eta)/d(\log Y) = Y \frac{\eta'(Y)}{\eta(Y)}.
\]
Although we work here with functions possessing a derivative, this could be
dispensed with and be replaced by working with finite-difference arc elasticities.
Lemma: (a) Let \( \psi(Y) \) be a differentiable, non-negative function whose elasticity, \( E \psi \), is strictly greater than one for all \( Y \in (0, \omega) \). (b) Let \( v(Y) \) be a positive, monotone-decreasing, differentiable weighting function (i.e., \( v(Y) > 0, v'(Y) < 0 \)) and \( dp(Y) \) be a probability distribution function over non-negative \( Y \) such that its cumulative distribution function must grow at more than one positive point (so that \( P(Y) \) takes on at least three positive values for positive \( Y \)'s). If \( \int_{c}^{\omega} \psi(Y)v(Y) dp(Y) = \int_{c}^{\omega} Yv(Y)dp(Y) \), then

\[
\int_{c}^{\omega} \psi(Y)dp(Y) > \int_{c}^{\omega} Ydp(Y).
\]

With this Lemma, we can then proceed to state and prove the following theorem.

Theorem: If \( F_n(X) \) is generated by the process described in equations (20) and (21), or in (29), (30), and (31) and if the actual \( \beta(X,n) \) is defined by

\[
e^{\beta}(X,n) = \int_{c}^{\omega} F_n(XZ)/F_{n+1}(X) dp(Z;1),
\]

then for all finite \( n \), \( \beta(X,n) > \alpha \).

Proof: Now, writing \( F_n(XZ)/F_{n+1}(X) = \psi(Z) \), we must show that \( \psi \) has the properties hypothesized by part (a) of the Lemma, i.e., \( \psi \geq 0 \) and \( E \psi > 1 \). Clearly, \( \psi(Z) \geq 0 \) and, even more, because \( F_n \) is an increasing function of its argument, \( \psi'(Z) > 0 \) for all \( Z > 0 \). From equation (30) and the definition of \( F_0(X) \), for all \( X > 0 \) such that \( F_n(X) > 0 \), we have:

\[
0 < F_n'(X) = \frac{\int_{\omega}^{\omega} z d\nu_n(Z)}{\int_{c}^{\omega} (Xz-1) d\nu_n(Z)} > \frac{1}{X}
\]

So, for \( X > 0 \) such that \( F_n(X) > 0 \),

\[
XF_n'(X)/F_n(X) > 1
\]

Therefore, from (42), \( E \psi > 1 \). If we write \( v(Z) = U'[(1-w_2*\varepsilon^r + v_2*Z] \), we must show that \( U' \) satisfies condition (b) of the Lemma. Clearly, by the definition of \( U, U' > 0 \) and \( U' < 0 \), condition (b) is satisfied. From (29), (30), and (31), with \( n = 1 \), all the conditions for the hypothesis of the Lemma are satisfied:

\[
\int_{c}^{\omega} F_n(XZ)/F_{n+1}(X) dq(Z;1) = e^r = \int_{c}^{\omega} Z dq(Z;1).
\]
Therefore, by the Lemma,
\[
\int_{c}^{\infty} \frac{F_n(XZ)}{F_{n+1}(X)} \, dP(Z;1) > \int_{c}^{\infty} Z \, dP(Z;1)
\]
or,
\[e^{\mathcal{Q}(X,n)} > e^{\alpha}\]
therefore,
\[\mathcal{Q}(X,n) > \alpha\].

So the theorem is proved. Using the Lemma, as generalized in appendix A of [7], one could give a second proof that the common itself, being more "volatile" than the safe asset, must have a greater expected yield: namely \(\mathcal{Q} > r\) as expressed earlier in equation (21).

**Conclusion**

This completes the theory of utility-warranted warrant pricing. We have proved a number of strong theorems about warrant pricing and have answered the basic criticisms made of the 1965 "first-moment" theory.

**Footnotes**

1. See Samuelson, [6].

2. See Kassouf, [2].

3. See Samuelson, [5], where theorems like this one are proved without making the mean-variance approximations of the now classical Markowitz-Tobin type. Since units are arbitrary, we can take any prescribed wealth level and by dimensional convention make it unity in all of our formulas. This enables expressions like \(w\mathcal{W}\) to be written simply as \(w\) where \(\mathcal{W} = \) total wealth. As will be specified later, working with iso-elastic marginal utility functions that are uniform for all investors will make the scale of prescribed wealth of no importance.

4. The concavity of \(U\) is sufficient to achieve the negative semi-definiteness of the constrained quadratic forms and bordered Hessian minors of \(L\) needed to insure that any solution to the first-order conditions does provide a global as well as local maximum. Although the maximum is unique, the port-
folio proportions could take on more than one set of optimizing values in singular cases where the quadratic forms were semi-definite rather than definite, e.g., where a perpetual warrant and its common stock are perfectly linearly correlated, making the choice between them indifferent and not unique. This example will be presented later.

5. At a Washington conference in 1953, P.A. Samuelson once shocked the late J.M. Clark by saying, "Although the probability of a serious 1954 recession is only one-third, that probability should be treated as though it were two-thirds." This was a crude and non-marginal use of a util-prob notion akin to dQ.

6. In the 1965 paper [6], pp. 30-31, it was mentioned that the possibility of hedges, in which the common stock is sold short in some proportion and the warrant is bought long, would be likely to set limits on the discrepancies that, in the absence of dividend payments, could prevail between \( \rho \) and \( \gamma \).

7. If the probability of good and bad crops were not equal of if the safe investment process had a non-zero yield, the proportion of the risk asset held would be a function of the \( \lambda \) yield factor; and for utility functions other than the Bernoulli log-form and a probability distribution different from the simple binomial, \( w^* \) would be a more complicated calculable function.

8. Strictly speaking, a will probably be a function of time, \( a_{t-1} \) being high in the period following a generally poor crop when the \( \lambda_t^{-1} \) yield factor, rather than \( \lambda_t^1 \), has just occurred and the investable surplus is small.

We have here a stationary time series in which total output vibrates around an equilibrium level. Spelling all this out would be another story: here a will be taken as constant.

9. For the family, \( U(X) = a + b X^e/e, e < 1, -UX''(X)/U'(X) = 1 - e \). The singular case where \( e = 0 \) can be found by L'Hopital's evaluation of an indeterminate form to correspond to the Bernoulli case, \( U(X) = a + b \log X \). As Arrow[1], Pratt[4], and others have shown, optimal portfolio proportions are independent of the absolute size of wealth for any function that is a member of this utility family.
Actually, we can free our analysis from the assumption of iso-elastic marginal utility if we are willing to apply it to any single individual and determine from it the critical warrant price patterns at which he would be neither a buyer nor seller, or would hold some specified proportion of his wealth in the form of warrants. By pitting the algebraic excess demands of one set of individuals against the other, we could determine the market clearing pattern.

10. U being concave assures a maximum. The problem could be formulated with Kuhn-Tucker inequalities to cover the no-borrowing restriction, \( w_1 \leq 1 \), and the no-short-selling restriction, \( w_1 \geq 0 \).

11. The \( F_n \) function in (13) is the "utility-warranted" price of the warrant, which is not the same as the "rational" warrant price of the 1965 theory discussed above, even though we use the same symbol for both.

12. This would be a generalization of the analysis above to three rather than only two assets.

13. Thorp and Kassouf[8] advocate hedged short sales of overpriced warrants about to expire. The analysis here defines the levels at which one who holds the stock long can just benefit in the maximizing expected utility sense from short-sale hedges in the warrant.

15. We will show later that \( \hat{p}_n > \check{p}_n \) for finite-duration warrants, falling toward equality as the duration time becomes perpetual.

16. The Hertz-Herglotz-Lotka methods of renewal theory are closely related, once we replace X and Z by their logarithms. However, the fact that our dQ involves Z's on both sides of unity with positive weights introduces some new complications. Later, without regard to formal expansions of this type, we prove that \( F_n(X) \rightarrow F_\infty(X) = F(X) \). For references to this literature, including work by Fellner, see Lopez [3].
17. There is a further complication. If decisions are frozen for \( n \) periods, then (26) is valid, superseding (24) and (20). Or put differently, \( n \) of the old time periods are now equivalent to one new time period; and in terms of this new time period, (20) would be rewritten to have exactly the same content as (26). Now (24) and (25) would simply be irrelevant. One must not suppose that this change in time units is merely a representational shift to new dimensional units, as from seconds to minutes. If our portfolio is to be frozen for six months, that differs substantively from its being frozen for six weeks, even though we may choose to write six months as twenty-six weeks. But now for complication: one would not expect the \( U(W) \) function relevant for a six-week frozen-decision period to be relevant for a six-month period as well. Strictly speaking then, in using (26) for a long-frozen-period analysis, we should require that the \( U'(W) \) function which enters into \( dq(Z; n) \) be written as dependent on \( n \), or as \( \mathcal{F} U(W; n) / \mathcal{F} W \). One further remark. Consider the "incipient-cash" case where \( w_1^* = 0 \) because the common stock dominates the safe asset, with \( \alpha \gg r \). Combining this case with our incipient-warrant case, \( w_2^* \) remains at unity in every period, no matter what we learn about the outcomes within any larger period. In this case, the results of (20) and those of (26) are compatible and the latter does give us by mere quadrature a one-step solution to the problem. The 1965 proof that \( F_n(X) \rightarrow X \) as \( n \rightarrow \infty \) can then be applied directly.

18. If, as mentioned in footnote 9, we free the analysis from the assumption of iso-elastic marginal utility, the definitions of (29) must be generalized to take account of the changing \( v_j^* \) optimizing decisions, which will now be different depending on changing wealth levels that are passed through.

19. The results of this section also holds for calls. See appendix B of [7] for the results for dividend-paying stocks.
References


V. AN EMPIRICAL INVESTIGATION OF THE SAMUELSON RATIONAL WARRANT PRICING THEORY

Introduction. A warrant is a piece of paper, usually issued by a corporation, giving the bearer the right to purchase a specified number of shares of stock of the corporation at a given price (the "exercise" price) per share on or before a set date (the "expiration" date).¹ In a paper written in 1965, P.A. Samuelson [3] developed a theory of rational warrant pricing. Based on certain axioms of behavior postulated about stock price movements, he derives a set of equations which describe how the warrant should be priced as a function of the stock price and expiration date. The present paper is an empirical examination of his rational warrant pricing theory.

It will be useful to define several variables which will be used throughout the paper. Define:

\[ S_t = \text{the price of the common stock at time } t. \]

\[ W_t(n) = \text{price of a warrant at time } t \text{ which gives the bearer the right to purchase one share of common stock at $1 per share on or before the date } (n+t), \text{ i.e., the warrant will expire } n \text{ periods from time } t. \]² Such a warrant will be called a "n-period warrant".

¹There are often other provisions such as a changing exercise price as a function of time, protection against dividend payments to the common stock holders, etc. Similarly, stock options, rights, call options, etc. would be classified as warrants under the above definition. Although the theory could be modified to take into account these complications, it is not necessary for the purpose of this paper, and the part of the theory examined describes the price movements of simple "American-type" warrants which are widely held by the public and actively traded in organized markets.

²As defined, \( S_t \) and \( W_t(n) \) would be the "normalized" prices of the stock and warrant, respectively. Suppose the actual market prices (as read in the newspaper) are called \( S_t^* \) and \( W_t(n)^* \) and suppose \( W_t(n)^* \) is a warrant giving the bearer the right to purchase \( m \) shares at a price $P per share. Then,
\[
X_t = \log_e (S_t)
\]
\[
Y_t = \log_e (W_t(n)) = Y_t(n)
\]
\[
\mu = \mathbb{E}[X_{t+1} - X_t | X_t] ; \sigma^2 = \text{Var}[X_{t+1} - X_t | X_t] ; \nu = \mathbb{E}[Y_{t+1} - Y_t | Y_t] ;
\]
\[
\delta^2 = \text{Var}[Y_{t+1} - Y_t | Y_t]
\]
\[
e^\alpha = \mathbb{E}[S_{t+1} / S_t | S_t] ; e^\beta = \mathbb{E}[W_{t+1}(n-1) / W_t(n) | W_t(n)]
\]

where \( \mathbb{E} \) is the conditional expectation operator, conditional on the knowledge of the value of the variable to the right of the vertical line within the brackets.

Samuelson's theory of warrant pricing. The axioms of the Samuelson theory are as follows:

I. The stock price, \( S_t \), is generated by a geometric random walk ("Brownian" motion) defined by:

A. \( \text{Prob} \left[ S_{t+T} \leq s | S_t \right] = P(s,S_t;T) = P(Z_t;T) \)

where \( Z_t = S_{t+T} / S_t \) and the random variables \( Z_t \) are independent and identically distributed random variables whose distributions have finite moments.

B. \( \mathbb{E}[S_{t+T} / S_t] = S_t e^{\alpha T} = S_t \int_0^\infty Z_t dP(Z_t;T) \)

and \( \alpha \) is a non-negative constant.

II. The holders of warrants demand a minimum expected return, \( \beta \), and \( \beta \) is defined by

\[
\mathbb{E}[W_{t+T}(n-T) / W_t(n)] = W_t(n)e^{\beta T}
\]

where \( (n-T) \) reflects the fact that a \( n \)-period warrant bought at time \( t \) will be a \( (n-T) \)-period warrant at time \( (t+T) \).

Because the warrant price is perfectly positively correlated with the common stock price and because it is more volatile, it is assumed that \( \beta \alpha \). Finally, if it is not possible to price the warrant so that its minimum expected return is \( \beta \),

---

2 (Continued) the "normalized" prices of the stock and warrant \( S_t \) and \( W_t(n) \), would be: \( S_t = S_t^* / P \) and \( W_t(n) = W(n)^* / mP \). Thus the normalized prices use the price of one share of stock priced at the exercise price of the warrant as numerator. Not only is this normalization of theoretical convenience, but it makes empirical comparisons between warrants with different terms an easy task.
then the theory assumes that the warrant is converted into common stock, (its value is Max[0, S_t-1]).

If we now define $F(S_t; n) = W_t(n)$, the "rational" warrant price as a function of the stock price and number of periods before expiration, the price equation is derived, as follows: From axiom II.,

$$E[F(S_{t+1}, n)/F(S_t, n+1)] = e^\beta,$$ if the warrant is held

Therefore, from axiom I.,

$$F(S, n+1) = e^{-\beta \int_0^\infty F(ZS, n) dP(Z; 1)}, \text{ if held}$$

$$= S - 1 - e^{-\beta \int_0^\infty F(ZS, n) dP(Z; 1)}, \text{ if converted}$$

But, $F(S, 0) = \text{Max}[0, S - 1]$. So (1) can be solved recursively since $F(S, 0)$ is known. Thus, the "rational" warrant prices are

$$F(S, n+1) = \text{Max}[0, S - 1, e^{-\beta \int_0^\infty F(ZS, n) dP(Z; 1)}], \text{ } n > 0$$

and $F(S, 0) = \text{Max}[0, S - 1]$

Samuelson goes on to show that for $\beta < \alpha$, there exists a $S^\# = C_n < \infty$ such that for all $S \geq C_n$, $\text{Max}[0, S - 1, e^{-\beta \int_0^\infty F(ZS, n) dP(Z; 1)}]$, and hence $F(S, n) = S - 1$ for $S \geq C_n$. Further, $\exists F(C_n; n) = 1$, which is called the $S$ "high-contact" condition because it means that $F(S, n)$ and its first derivative are continuous for all $S$. $C_n$ is called the "conversion price" of a $n$-period warrant and it is that stock price above which the warrant can no longer yield an expected return $\beta$ and hence is converted. A graph of a typical solution of (2), $F(S, n)$ is shown in figure 1.
Note: the "high-contact" condition at \( S=C_n \), i.e.

\[
\frac{\partial F(C_n,n)}{\partial S} = \frac{\partial (S-1)}{\partial S} = 1.
\]

If we define \( \psi(X,n) = F(e^X,n) \) (\( X=\log(S) \)), then for the special case when the \( Z \) has a log-normal distribution, \( \psi \) satisfies the partial differential equation

\[
\frac{\sigma^2}{2} \frac{\partial^2 \psi}{\partial X^2} + \mu \frac{\partial \psi}{\partial X} - \frac{\partial \psi}{\partial n} - \beta \psi = 0
\]

with boundary conditions, \( \psi(X,0) = \text{Max}[0,e^X-1] \); \( \psi(-\infty,n) = 0 \)

\[
\psi(\log C_n,n) = \log C_n-1; \frac{\partial \psi}{\partial X} (\log C_n,n) = 0; \text{"high contact"}
\]

Equation (3) is a difficult equation to solve because \( \psi(X,n) \)
and \( C_n \) must be solved simultaneously. However, there are a
class of warrants for which equation (3) can be solved analytically.
This class of warrants is called the perpetual warrant class and
correspond to the case \( n=\infty \) (i.e., the warrants never expire.).
In this case, \( \psi(X,n) = \psi(X,\infty) = \psi(X) \), and (3) reduces to an
ordinary differential equation,

\[
\frac{\sigma^2}{2} \psi''(X) + \mu \psi'(X) - \beta \psi(X) = 0
\]

with boundary conditions,

\[
\psi(\log C) = C-1; \psi(-\infty) = 0
\]

\[
\psi'(\log C) = C \text{ where } C = C_\infty
\]

The solution to (4) is

\[
F(S) = F(S,\infty) = (C-1) (S/C)^\beta, \quad S \leq C
\]

\[
= S-1, \quad S > C
\]
Figure 1.
where \( \gamma = c/(c-1) \)

and \( \gamma = -\frac{\mu}{\sigma^2} + \sqrt{\left(\frac{\mu}{\sigma^2}\right)^2 + 2 \left(\frac{\beta}{\sigma^2} + \frac{1}{2}\right)} \)

It is this very special case of perpetual warrants whose associated stock price is distributed log-normally which will be examined empirically to test the theory. To summarize:

The assumption that \( S \) is distributed log-normally is:

\[ X_{t+1} = X_t + \mu + \epsilon_t \quad \text{where } \epsilon_t \text{ is distributed } N(0, \sigma^2). \quad (6) \]

If (6) holds, then from the theory for perpetual warrants,

\[ W_t = (c-1) \left(\frac{S_t}{c}\right)^\gamma \quad , \quad S_t \leq c \]

\[ = S_{t-1} \quad , \quad S_t \geq c \quad (7) \]

where

\[ \gamma = \frac{-\mu}{\sigma^2} + \sqrt{\left(\frac{\mu}{\sigma^2}\right)^2 + 2 \left(\frac{\beta}{\sigma^2} + \frac{1}{2}\right)} \]

\[ = \left(\frac{1}{2} - \alpha_{sigma}^2\right) + \sqrt{\left(\frac{1}{2} + \frac{\alpha}{\sigma^2}\right)^2 + \frac{2(\beta-\alpha)}{\sigma^2}} \quad (8) \]

the second equality holding because, from the log-normality of \( S \), \( \alpha = \mu + \frac{1}{2} \sigma^2 \). and

\[ C = \frac{\gamma}{\gamma-1} \quad (9) \]

Further, from (6) and (7),

\[ Y_{t+1} = Y_t + \nu + v_t \quad \text{where } v_t \text{ is distributed } N(0, \delta^2) \]
and \( v = \gamma \mu \) and \( \delta^2 = \gamma^2 \sigma^2 \).

Before going on to the formal testing of the theory, it is valuable to discuss the good points and the short-comings of the theory. By doing this, one may be able to anticipate where some of the problems may arise in the testing.

The good points of the theory are that the basic axioms are simple and believable in the sense that it is reasonable to suppose that most investors would compare yields and relative risks and that one would expect that because the warrant is more volatile and stocks usually pay dividends, \( \beta > \alpha \). Although there have been many theories of warrant pricing, the Samuelson one is the first to satisfactorily build into the model an explanation of conversion of the warrant prior to expiration, and hence, to assign a value to the privilege of "American-type" warrants to convert the warrant prior to the expiration date. \(^{3}\)

The most important short-comings of the theory are:

1. it is a "first-moment" theory, i.e., investors only really compare expected returns without consideration of higher-order moments of the distribution (e.g., variance, skewness, and kurtosis) either explicitly, or implicitly, by modulating the return through some type of utility function.
2. the analysis is both partial equilibrium and micro-economic. Alternative assets and yields other than for the stock and its warrant are not considered although one could argue that these are implicit in the estimates and requirements of \( \alpha \) and \( \beta \). Differences among individuals in their estimates and demands for \( \beta \) are not considered. This is quite important because the theory states that when the warrant can no longer be priced to yield \( \beta \), it

\(^{3}\)"American-type" warrants give the warrant holder the right to exercise his warrant any time prior to the expiration date. "European-type" warrants can only be exercised on the expiration date.
is converted into common stock yielding $\alpha$. If there is a spread (i.e., $\beta > \alpha$) why wouldn't some individual accept a $\beta'$ yield such that $\beta > \beta' > \alpha$ rather than convert? Hence, unless everyone in the economy is identical in both tastes and expectations, one would expect the conversion price $C$, which is a function of $(\beta - \alpha)$, to be at most a "rubber" $C$ in the sense that one would expect to start observing some conversions as $S_t$ exceeds $C$, but not complete conversion. Thus, one should not expect the theory to be very accurate in predicting warrant prices when the stock prices are in a neighborhood of the conversion price. \(^4\)

**Empirical test of the Samuelson theory.** As was mentioned in the previous section, the only case ($\beta > \alpha$) which can be solved analytically is for perpetual warrants when the log of the stock price is generated by a simple Wiener process. Although perpetual warrants are relatively rare (there are only three actively traded on listed exchanges) and therefore not as important as finite-lived warrants, their analysis is worthwhile as a first test of the theory.

An outline of the analysis is as follows:

1. Examine equation (6), $X_{t+1} = X_t + \mu + \varepsilon_t$, and test whether the assumption $\varepsilon_t$ distributed $N(0, \sigma^2)$ is a reasonable one. As Part of the test of the theory, examine equation (10) $Y_{t+1} = Y_t + \nu + v_t$, and test whether $v_t$ is distributed $N(0, \delta^2)$ and if $v = \gamma \mu$ and $\delta^2 = \gamma^2 \sigma^2$. \(^5\) In the process of testing (6) and (10), compute estimates of $\alpha, \beta, \mu, \nu, \sigma^2$, and $\delta^2$. Then using (8) and (9), compute estimates of $\gamma$ and $C$.

\(^4\) Most of the short-comings of the 1965 theory have been answered in a later and more complicated version, P.A. Samuelson and R.C. Merton \(^4\). However, these short-comings were resolved at the expense of a more complicated theory and probably a far less practical one.

\(^5\) Continued
2. If (6) and (10) are judged sufficiently close to the distributional assumptions, then the main part of the testing of the theory is done by estimating equation (7). By taking the log of both sides of (7), we get

\[ Y_t = \gamma X_t + \log(C-1) - \gamma \log(C) \]  

(11)

From (9), \( C = C(\gamma) \), and now define

\[ f(\gamma) = \log(C-1) - \gamma \log(C) \]  

(12)

Then, the actual equation to be estimated is,

\[ Y_t = \hat{\gamma} X_t + \hat{\eta} \]  

(11')

To test the theory, we examine whether \((\hat{\gamma}, \hat{\eta}) = (\gamma, f(\gamma))\) as predicted in part 1. A different test (although clearly not independent) is to see whether the \( \beta \) predicted from equation (8) by putting \( \hat{\gamma} \) in for \( \gamma \) is statistically different from the \( \beta \) measured from equation (10). A final test which is essentially testing whether the estimated polynomial has high-contact with the line \( S-1 \) is to compare \( f(\gamma) \) with \( \hat{\eta} \).

3. The third part of the analysis will be to see how well the theory predicts in comparison with alternative theories of warrant pricing.

Data: In choosing the data, we wanted securities which were actively traded and on which there would be systematically accurate quotations. This requirement limited the choice to securities listed on either the New York or American Stock Exchange. Although commonly done in similar type analyses, it was decided that long finite-lived warrants would not be

\[ ^{5} \text{Continued.} \]

Actually, \( Y \), according to the theory from (7), is a composite of two random processes: one for \( S<C \) and the other for \( S>C \). For simplicity, and because in either case the process is normal, this distinction both here and in the empirical estimation is neglected. As previously mentioned, the theory was not expected to do well in the region \( S>C \). However, this simplification should be kept in mind when examining the results.
acceptable as an approximation to perpetual warrants. Most finite-lived warrants have lives of less than fifteen years and because a reasonably long observation period was desirable (six years was actually used), it was decided not to introduce further error due to the warrant becoming shorter-lived. Also, in a later study, it would be interesting to test whether the assumption is valid that long-but-finite life warrants can be approximated by perpetual warrant equations.

There are only three perpetual warrants listed on national exchanges: Tri-Continental, Allegheny, and Atlas. The common stocks are all listed on the New York Stock Exchange and the warrants on the American Exchange. All three warrants fit the conditions of simple perpetual warrants in that there are no variable conversion rates or call provisions (after completion of the study, it was discovered that the Atlas warrant was not quite a simple warrant in that there exists a preferred security which can be used at face value as "script" in conjunction with the warrant to purchase the common stock. The primary effect of neglecting this feature in the analysis is to change the constant term in (11').) Weekly data using the closing price each Friday (except for holidays, in which case, opening prices on the following Monday were used.) was used to provide a large number of observations so that we could take advantage of the various asymptotic statistical properties. Also, with these many observations, it would be possible to check some results for longer intervals without affecting the sample size significantly. The period of observation chosen was January, 1963 through December, 1968. This choice gives a long, but fairly homogeneous, period as far as economic and stock market conditions were concerned.

The data was transformed in the following fashion: both the stock and warrant price were normalized by dividing the stock price by the exercise price of the warrant and the warrant price by the exercise price times the number of shares into which each warrant was convertible. Adjustments were made for all stock splits and dividends because the warrants are not protected against cash
dividends paid to shareholders. There were 296 observations for Tri-Continental and Allegheny and 297 observations for Atlas.

Part 1: testing the distribution of $S_t$. The first set of regressions, R1 and R2, were run to get estimates of the basic parameters of the stock and warrant price series.

(R1) $X_{t+1} - X_t = \mu_0 + \epsilon_t$

<table>
<thead>
<tr>
<th></th>
<th>$\mu_0 \times 10^3$</th>
<th>$\sigma_0^2 \times 10^3$</th>
<th>$\alpha_0 \times 10^3$</th>
<th>Durbin-Watson</th>
<th>Number Of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri Continental</td>
<td>1.28 (.935)</td>
<td>0.53</td>
<td>1.53</td>
<td>2.1157</td>
<td>295</td>
</tr>
<tr>
<td>Allegheny</td>
<td>2.61 (.926)</td>
<td>2.34</td>
<td>3.77</td>
<td>1.8801</td>
<td>295</td>
</tr>
<tr>
<td>Atlas</td>
<td>3.77 (.074)</td>
<td>4.44</td>
<td>6.02</td>
<td>2.0700</td>
<td>296</td>
</tr>
</tbody>
</table>

Estimator for $\alpha_0$: $e^{\alpha_0} = \frac{1}{n} \sum_{i=1}^{n} \frac{S_{t+1}}{S_t}$

(R2) $Y_{t+1} - Y_t = v + V_t$
<table>
<thead>
<tr>
<th></th>
<th>$v_0 \times 10^3$</th>
<th>$\delta_0^2 \times 10^3$</th>
<th>$\beta_0 \times 10^3$</th>
<th>$\sigma_\beta^2 \times 10^3$</th>
<th>Durbin-Watson</th>
<th>Number of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>2.37 (1.40)</td>
<td>0.85</td>
<td>2.78</td>
<td>2.37</td>
<td>1.944</td>
<td>295</td>
</tr>
<tr>
<td>Allegheny</td>
<td>3.34 (.947)</td>
<td>3.67</td>
<td>5.19</td>
<td>3.35</td>
<td>1.900</td>
<td>295</td>
</tr>
<tr>
<td>Atlas</td>
<td>5.26 (1.03)</td>
<td>7.20</td>
<td>8.89</td>
<td>5.26</td>
<td>2.3239</td>
<td>296</td>
</tr>
</tbody>
</table>

Estimator for $\beta_0$: $e^{\beta_0} = \frac{1}{n} \sum_{t=1}^{n} \left( \frac{W_{t+1}}{W_t} \right)$

$\sigma_\beta^2 = \frac{1}{n-1} \sum_{t=1}^{n} \left( \frac{W_{t+1}}{W_t} - e^{\beta_0} \right)^2$

Because the Durbin-Watson statistic suggested some autocorrelation in the Atlas warrant time series, it was decided to re-run all the regressions in R1 and R2 using the Cochrane-Orcutt routine to correct for first-order autocorrelation. The results are shown in R5 and R6 later in the paper. As anticipated, all the estimates of the correlation coefficient, $\rho$, were near to zero except for the Atlas warrant case, and in every case, the estimates of the regression coefficients were essentially the same as in R1 and R2.

The next step is to use the estimates from R1 and R2 to calculate the "predicted" $\gamma, \gamma_0$, from equation (8), $C(\gamma_0)$ from (9), and $f(\gamma_0)$ from (12). The predicted values shown in T1., will be used throughout the paper for testing both the distributional assumptions and the validity of the theory.
\[ \begin{array}{|c|c|c|} \hline & \gamma_0 & f(\gamma_0) & C(\gamma_0) \\ \hline \text{Tri-Continental} & 1.63 & -1.09 & 2.58 \\ \text{Allegheny} & 1.27 & -0.656 & 4.72 \\ \text{Atlas} & 1.32 & -0.731 & 4.13 \\ \hline \end{array} \]

The hypothesis that \( (X_{t+1} - X_t) \) is generated by a simple gaussian random walk requires not only that the stochastic generator be normal, but that it be stationary and independent as well.

A. Tests for Normality (Log-normality of \( S_t \)). The first set of tests for normality were the standard skewness and kurtosis tests computing \( \sqrt{b_1} = \)

\[ \frac{1}{n} \sum_{t=1}^{n} \left( \frac{X_{t+1} - X_t - \mu_0}{\sigma_0} \right)^3 / \sigma_0^2 \text{ and } \beta_2 = \frac{1}{n} \sum_{t=1}^{n} \left( \frac{X_{t+1} - X_t - \mu_0}{\sigma_0} \right)^4 / \sigma_0^2, \]

These were the proper estimators in view of the large number of observations. The results of these tests are presented in T2, and in every case, the distributions were too skewed to the right and too peaked to be accepted as having come from a normal population. The negative outcomes of these tests were anticipated because they are quite sensitive, particularly for large samples to any deviation from normality.
The second test is based on the result that if \( S_t \) is log-normal, then \( \alpha_0 + \mu + \frac{1}{2} \sigma_0^2 \) and similarly, \( \beta_0 = \nu_0 + \frac{1}{2} \delta_0^2 \) if \( W_t \) is log-normal. The results presented in T3 were very close to the predicted relationship. However, after further examination, it was found that for \( S_{t+1}/S_t \) and \( W_{t+1}/W_t \) values close to one (note: \( E[S_{t+1}/S_t] \approx 1.005 \)), this relationship was practically an arithmetic identity, and hence, the results are completely vacuous.

\[(T3)\]

<table>
<thead>
<tr>
<th></th>
<th>( \alpha_0 \times 10^3 )</th>
<th>( (\mu_0 + \frac{1}{2} \sigma_0^2) \times 10^3 )</th>
<th>( \beta_0 \times 10^3 )</th>
<th>( (\nu_0 + \frac{1}{2} \delta_0^2) \times 10^3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>1.53</td>
<td>1.54</td>
<td>2.78</td>
<td>2.79</td>
</tr>
<tr>
<td>Allegheny</td>
<td>3.77</td>
<td>3.78</td>
<td>5.19</td>
<td>5.17</td>
</tr>
<tr>
<td>Atlas</td>
<td>6.02</td>
<td>5.99</td>
<td>8.89</td>
<td>8.86</td>
</tr>
</tbody>
</table>

Since it is clear that the distributions are not normal, it was decided to see how much they deviated from Normality. Cumulative distribution function tables (in 2% intervals) were constructed from the data and the scatter plotted on normal
probability graph paper where cumulative distribution functions of normal variates plot as straight lines (see figures 2-7). Then, by converting the cumulative probabilities into measures of standard deviations, it was possible to run regressions to find the "best-fitted" normal distribution to each scatter. The results of these regressions are summarized in R3 and R4 and the regression lines plotted in figures 2-7.

(R3)  \( \Delta X_t = 1_1 P + \mu_1 \)

<table>
<thead>
<tr>
<th></th>
<th>( \sigma_1^2 \times 10^3 )</th>
<th>( \mu_1 \times 10^3 )</th>
<th>( R^2 )</th>
<th>Standard Error ( \times 10^3 )</th>
<th>Durbin-Watson</th>
<th>Number of Observations</th>
<th>Figure Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>0.475 (28.9)</td>
<td>0.948 (1.16)</td>
<td>.948</td>
<td>5.1</td>
<td>0.5764</td>
<td>39</td>
<td>2</td>
</tr>
<tr>
<td>Allegheny</td>
<td>2.12 (35.4)</td>
<td>2.57 (1.94)</td>
<td>.9698</td>
<td>8.5</td>
<td>0.1886</td>
<td>41</td>
<td>4</td>
</tr>
<tr>
<td>Atlas</td>
<td>4.06 (23)</td>
<td>5.55 (1.82)</td>
<td>.9484</td>
<td>8.3</td>
<td>0.3706</td>
<td>31</td>
<td>6</td>
</tr>
</tbody>
</table>

(R4)  \( \Delta X_t = 1_1 P + v_1 \)

<table>
<thead>
<tr>
<th></th>
<th>( \delta_1^2 \times 10^3 )</th>
<th>( v_1 \times 10^3 )</th>
<th>( R^2 )</th>
<th>Standard Error ( \times 10^3 )</th>
<th>Durbin-Watson</th>
<th>Number of Observations</th>
<th>Figure Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>0.75 (45.3)</td>
<td>1.88 (3.04)</td>
<td>.9814</td>
<td>4.0</td>
<td>.4173</td>
<td>41</td>
<td>3</td>
</tr>
<tr>
<td>Allegheny</td>
<td>3.35 (47)</td>
<td>3.62 (2.83)</td>
<td>.9831</td>
<td>8.1</td>
<td>.2716</td>
<td>40</td>
<td>5</td>
</tr>
<tr>
<td>Atlas</td>
<td>7.20 (33.1)</td>
<td>4.24 (1.37)</td>
<td>.9786</td>
<td>15.8</td>
<td>.9979</td>
<td>26</td>
<td>7</td>
</tr>
</tbody>
</table>
TRI-CONTINENTAL STOCK: Cumulative Distribution Function

\[ \log(S_x/S_0) \times 10^3 \]

\[ \log(S_x) = (21.79E - 0.943) \times 10^{-1} \]

\[ R^2 = 0.9576 \quad 39 \text{ observations} \]
ALLEGHENY STOCK: Cumulative Distribution function

\[ \Delta \log \Phi = (45.95P + 2.57) \times 10^{-3} \]

\[ R^2 = 0.9698 \quad \text{41 observations} \]
ALLEGHENY WARRANT, Cumulative Distribution Function

\[ A \log\left(1 + \frac{N}{N_c}\right) \times 10^3 \]

\[ A \log\left(1 + \frac{N}{N_c}\right) \times 10^3 = \frac{5.621}{10^3} \]

\[ \chi^2 = 0.9271 \quad 40 \text{ observations} \]
The fit is rather good in terms of $R^2$. However, visual examination of figures 2-7 and a simple runs test on the residuals shown in T4, shows systematic deviation from normality.

(T4)

<table>
<thead>
<tr>
<th></th>
<th>Stock</th>
<th>Warrant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Runs</td>
<td>Number of Observations</td>
</tr>
<tr>
<td>Tri-Continental</td>
<td>6</td>
<td>39</td>
</tr>
<tr>
<td>Allegheny</td>
<td>3</td>
<td>41</td>
</tr>
<tr>
<td>Atlas</td>
<td>3</td>
<td>31</td>
</tr>
</tbody>
</table>

Further, as can be seen in figures 6 and 7, both the Atlas stock and warrant series were significantly different from normal (between 30 and 40% of the observations were within a small interval.).

This systematic deviation from normality agrees with the earlier results of the skewness and kurtosis tests. The distributions appear to be more peaked with fatter tails. Although the Durbin-Watson statistics were poor, I chose to stay with the ordinary least squares estimates because the cumulative distribution function is certainly not a time series and because there is no reason to believe that the systematic deviation from normality is generated by a first-order Markov process.
The fit of normality is sufficiently good for Tri-Continental and Allegheny to continue with the testing of the theory. However, this is only done with the reservation that the deviations could be due to either lack of stationarity of the process or that the distribution is a different member of the stable Pareto-Levy family. Furthermore, it is to be expected that the deviations will "spoil" the results sufficiently (for the large number of observations involved) to cause the usual statistical tests to reject the theory. Thus, as is often true for studies using a large number of observations, although formal statistical tests will be employed when applicable, the weighting will be toward more informal "eye-balling" of the results before reaching any conclusions.

In this spirit, new predictions, \( \gamma_1 \), \( f(\gamma_1) \), and \( C(\gamma_1) \), were computed using the \((\mu_1, \sigma_1^2)\) and \((v_1, \delta_1^2)\) implied by the best normal approximation from R3 and R4, and these were compared with \( \gamma_0 \), \( f(\gamma_0) \), and \( C(\gamma_0) \) predicted from the sample. Such comparisons give a rough idea of the effects on these crucial parameters of the measured deviations from normality.

\[
\begin{array}{|c|c|c|}
\hline
& \gamma_1 & f(\gamma_1) & C(\gamma_1) \\
\hline
\text{Tri-Continental} & 1.16 & -0.4654 & 7.25 \\
\hline
\text{Allegheny} & 1.28 & -0.6942 & 4.57 \\
\hline
\text{Atlas} & 1.0^+ & 0 & \infty \\
\hline
\end{array}
\]

As can be seen in T5, the deviations from normality caused large changes in the predicted \( \gamma \) for the Tri-Continental and Atlas series while little change took place for the Allegheny series.
B. Tests for Serial Correlation. As previously mentioned, in testing the hypothesis of a simple random walk, the series (6) and (10) must be examined for serial correlation. The standard test for autocorrelation is the Von-Neumann-ratic and, in what amounts to the same test, the Durbin-Watson statistics from R1 and R2.

\[(T6)\]

<table>
<thead>
<tr>
<th></th>
<th>Von Neumann Ratio</th>
<th>Confidence Level</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock</td>
<td>2.214</td>
<td>97%</td>
<td>2.1157</td>
</tr>
<tr>
<td>Warrant</td>
<td>1.948</td>
<td>67%</td>
<td>1.944</td>
</tr>
<tr>
<td>Allegheny</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock</td>
<td>1.880</td>
<td>85%</td>
<td>1.880</td>
</tr>
<tr>
<td>Warrant</td>
<td>1.902</td>
<td>80%</td>
<td>1.900</td>
</tr>
<tr>
<td>Atlas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock</td>
<td>2.07</td>
<td>73%</td>
<td>2.3239</td>
</tr>
<tr>
<td>Warrant</td>
<td>2.47</td>
<td>99+%</td>
<td>2.6605</td>
</tr>
</tbody>
</table>

Although all the series statistics but the ones for the Atlas warrant series did not reject the hypothesis of independence, regressions R1 and R2 were re-run using the Cochrane-Orcutt technique to see how large the estimate of the correlation coefficient, \(\rho\), is and to see if the coefficients or other statistics are significantly changed.

\[(R5)\]

\[
\left( x_{t+1} - x_t \right) - \rho \left( x_t - x_{t-1} \right) = \left( 1 - \rho \right) \mu + \epsilon_t - \rho \epsilon_{t-1}
\]
(R5) Continued

<table>
<thead>
<tr>
<th></th>
<th>(\mu X10^3)</th>
<th>(\sigma^2 X10^3)</th>
<th>(\rho)</th>
<th>Durbin-Watson</th>
<th>Number Of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>1.37 (1.064)</td>
<td>0.55</td>
<td>-.061</td>
<td>1.9987</td>
<td>294</td>
</tr>
<tr>
<td>Allegheny</td>
<td>2.54 (.8475)</td>
<td>2.30</td>
<td>.0595</td>
<td>1.9934</td>
<td>294</td>
</tr>
<tr>
<td>Atlas</td>
<td>3.79 (1.01)</td>
<td>4.44</td>
<td>-.356</td>
<td>2.006</td>
<td>295</td>
</tr>
</tbody>
</table>

(R6)

\[
(Y_{t+1} - Y_t) - \rho (Y_t - Y_{t-1}) = (1-\rho) \nu + (V_t - \rho V_{t-1})
\]

<table>
<thead>
<tr>
<th></th>
<th>(\nu X10^3)</th>
<th>(\delta^2 X10^3)</th>
<th>(\rho)</th>
<th>Durbin-Watson</th>
<th>Number Of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>2.56 (1.484)</td>
<td>0.84</td>
<td>.0214</td>
<td>1.9828</td>
<td>294</td>
</tr>
<tr>
<td>Allegheny</td>
<td>3.20 (.86)</td>
<td>3.68</td>
<td>.0487</td>
<td>1.9988</td>
<td>294</td>
</tr>
<tr>
<td>Atlas</td>
<td>5.44 (1.26)</td>
<td>7.40</td>
<td>-.1637</td>
<td>2.0388</td>
<td>295</td>
</tr>
</tbody>
</table>
With the exception of the Atlas warrant series, the estimated $\rho$'s were small and in all cases, the estimates of the coefficients did not change radically from the estimates of R1 and R2.

Finally, although not a very good test for autocorrelation, a simple runs test on the number of times $\Delta^2 S_{t+1}$ $\Delta^2 S_t < 0$ was made and the results displayed in T7.

![Table](image)

The results of the test were in agreement with previous tests although positive serial correlation was suggested (97% confidence) for the Allegheny stock series. Since this correlation did not show up in the other tests, a 97% confidence interval may not be significant.

The conclusion of this section is that there is a definite serial correlation in the Atlas warrant series and although the tests are not really set up for comparisons among series which do not reject the null hypothesis, it appears that again the Allegheny series fits best the hypothesis of a simple random walk.
C. Tests for Trend. Various tests for trend were made as part of the test for stationarity and also to determine whether part of the measurement of autocorrelation in R5 and R6 is due to a non-specified simple time trend in the mean of the series. The first test was to calculate the Spearman rank correlation coefficient. The Spearman coefficient, $s$, is distributed normally, $N(0,1/N-1)$, when the number of observations, $N$, is large.

\[ S = 1 - \frac{6}{N(N^2-1)} \sum_{t=1}^{n} (\theta_t - t)^2 \]

$\theta_t = \text{rank of } t^{th} \text{ observation}$

<table>
<thead>
<tr>
<th></th>
<th>$S$</th>
<th>Number Of Observations</th>
<th>Confidence Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental Stock Warrant</td>
<td>.041</td>
<td>295</td>
<td>52%</td>
</tr>
<tr>
<td></td>
<td>.045</td>
<td>295</td>
<td>56%</td>
</tr>
<tr>
<td>Allegheny Stock Warrant</td>
<td>.014</td>
<td>295</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>.017</td>
<td>295</td>
<td>23%</td>
</tr>
<tr>
<td>Atlas Stock Warrant</td>
<td>-.042</td>
<td>296</td>
<td>52%</td>
</tr>
<tr>
<td></td>
<td>-.090</td>
<td>296</td>
<td>88%</td>
</tr>
</tbody>
</table>

From the results of T8 the hypothesis of no trend could not be rejected although the Atlas warrant series was close to rejection. Both the Allegheny-series appear to satisfy the null hypothesis better than Tri Continental or Atlas.
Two further tests were made: (1) a runs test for monotone trend on the number of times the first differences of equations (6) and (10) were positive. Like T7, it is not a strong test although it is non-parametric. The results T9 show significant trend in the Atlas warrant series and no significance in the other series.

(T9)

\[ \Delta^2 X_t > 0 \]

<table>
<thead>
<tr>
<th></th>
<th>Number Of Times &gt; 0</th>
<th>Number Of Observations</th>
<th>Confidence Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>Stock 141</td>
<td>294</td>
<td>37%</td>
</tr>
<tr>
<td></td>
<td>Warrant 142</td>
<td>294</td>
<td>32%</td>
</tr>
<tr>
<td>Allegheny</td>
<td>Stock 139</td>
<td>294</td>
<td>49%</td>
</tr>
<tr>
<td></td>
<td>Warrant 137</td>
<td>294</td>
<td>59%</td>
</tr>
<tr>
<td>Atlas</td>
<td>Stock 129</td>
<td>295</td>
<td>87%</td>
</tr>
<tr>
<td></td>
<td>Warrant 117</td>
<td>295</td>
<td>98%</td>
</tr>
</tbody>
</table>

(2) two sets of regressions were run for each series. The first was a fit of a simple time trend and the second with six individual time trends (one for each year). The results of the first set are presented in R7 and R8 on the following page.
(R7)

\[ \Delta X_t = \mu + bt \]

<table>
<thead>
<tr>
<th></th>
<th>( \mu \times 10^3 )</th>
<th>( b \times 10^4 )</th>
<th>( R^2 )</th>
<th>Standard Error</th>
<th>Durbin-Watson</th>
<th>Number Of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>-.6725 (245)</td>
<td>1.32 (.82)</td>
<td>.0023</td>
<td>.02355</td>
<td>2.1206</td>
<td>295</td>
</tr>
<tr>
<td>Allegheny</td>
<td>-.413 (732)</td>
<td>.455 (1.38)</td>
<td>.0064</td>
<td>.04826</td>
<td>1.8923</td>
<td>295</td>
</tr>
<tr>
<td>Atlas</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>296</td>
</tr>
</tbody>
</table>

* This regression was lost, but because of the negative results of the others and of the second set which was of the same type, the regression was not re-run.

(R8)

\[ \Delta Y_t = \nu + bt \]

<table>
<thead>
<tr>
<th></th>
<th>( \nu \times 10^3 )</th>
<th>( b \times 10^4 )</th>
<th>( R^2 )</th>
<th>Standard Error</th>
<th>Durbin-Watson</th>
<th>Number Of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>-1.28 (-.378)</td>
<td>.2466 (1.24)</td>
<td>.0052</td>
<td>.0291</td>
<td>1.0546</td>
<td>295</td>
</tr>
<tr>
<td>Allegheny</td>
<td>-4.89 (-.692)</td>
<td>.5600 (1.34)</td>
<td>.0061</td>
<td>.0606</td>
<td>1.9119</td>
<td>295</td>
</tr>
<tr>
<td>Atlas</td>
<td>-1.70 (-.167)</td>
<td>.470 (788)</td>
<td>.0021</td>
<td>.08758</td>
<td>2.3288</td>
<td>296</td>
</tr>
</tbody>
</table>
In none of the series were the coefficients on the trend significant. In the cases where the Durbin-Watson improved, it did so only slightly and in the important Atlas warrant case, the Durbin-Watson became slightly worse. The results of the second set of regressions were equally negative and are not presented in the paper.

The final test for stationarity and independence is to compute \( n\delta_n^2/\sigma_n^2 \), \( n\mu/\mu_n \), \( n\sigma_n^2/\sigma^2 \), and \( n\nu/\nu_n \) where \( n \) is the number of weeks per interval and the subscripted statistics are the mean and variance associated with \( n \)-week-interval observations. For a stationary and independent time series (where these statistics are well-defined), these ratios should all equal one.
The statistics were computed for \( n = 2, 40 \) weeks in jumps of two. The results are displayed in T10 below.

(T10)

<table>
<thead>
<tr>
<th>Stock</th>
<th>Warrant</th>
<th>Stock</th>
<th>Warrant</th>
<th>Stock</th>
<th>Warrant</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>( \frac{n\sigma^2}{\delta^2n} )</td>
<td>( \frac{n\sigma^2}{\nu n} )</td>
<td>( \frac{n\delta^2}{\delta^2n} )</td>
<td>( \frac{n\delta^2}{\nu n} )</td>
<td>( \frac{n\sigma^2}{\delta^2n} )</td>
</tr>
<tr>
<td>2</td>
<td>1.02</td>
<td>0.97</td>
<td>0.98</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>4</td>
<td>1.05</td>
<td>0.94</td>
<td>0.98</td>
<td>0.95</td>
<td>0.93</td>
</tr>
<tr>
<td>6</td>
<td>1.07</td>
<td>0.91</td>
<td>1.00</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>8</td>
<td>1.15</td>
<td>0.91</td>
<td>1.04</td>
<td>0.92</td>
<td>0.95</td>
</tr>
<tr>
<td>10</td>
<td>1.23</td>
<td>0.94</td>
<td>1.05</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>12</td>
<td>1.32</td>
<td>0.96</td>
<td>1.09</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>14</td>
<td>1.35</td>
<td>0.98</td>
<td>1.09</td>
<td>0.99</td>
<td>0.93</td>
</tr>
<tr>
<td>16</td>
<td>1.33</td>
<td>1.01</td>
<td>1.06</td>
<td>1.01</td>
<td>0.93</td>
</tr>
<tr>
<td>18</td>
<td>1.34</td>
<td>1.04</td>
<td>1.08</td>
<td>1.04</td>
<td>0.93</td>
</tr>
<tr>
<td>20</td>
<td>1.40</td>
<td>1.07</td>
<td>1.11</td>
<td>1.06</td>
<td>0.94</td>
</tr>
<tr>
<td>22</td>
<td>1.50</td>
<td>1.09</td>
<td>1.13</td>
<td>1.08</td>
<td>0.96</td>
</tr>
<tr>
<td>24</td>
<td>1.59</td>
<td>1.10</td>
<td>1.15</td>
<td>1.08</td>
<td>0.99</td>
</tr>
<tr>
<td>26</td>
<td>1.68</td>
<td>1.10</td>
<td>1.16</td>
<td>1.08</td>
<td>1.00</td>
</tr>
<tr>
<td>28</td>
<td>1.73</td>
<td>1.11</td>
<td>1.16</td>
<td>1.08</td>
<td>1.01</td>
</tr>
<tr>
<td>30</td>
<td>1.72</td>
<td>1.12</td>
<td>1.15</td>
<td>1.08</td>
<td>1.02</td>
</tr>
<tr>
<td>32</td>
<td>1.70</td>
<td>1.15</td>
<td>1.12</td>
<td>1.10</td>
<td>1.04</td>
</tr>
<tr>
<td>34</td>
<td>1.67</td>
<td>1.16</td>
<td>1.09</td>
<td>1.11</td>
<td>1.08</td>
</tr>
<tr>
<td>36</td>
<td>1.70</td>
<td>1.16</td>
<td>1.07</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>38</td>
<td>1.69</td>
<td>1.17</td>
<td>1.06</td>
<td>1.13</td>
<td>1.16</td>
</tr>
<tr>
<td>40</td>
<td>1.73</td>
<td>1.18</td>
<td>1.05</td>
<td>1.14</td>
<td>1.19</td>
</tr>
</tbody>
</table>
Although $\frac{n\alpha^2}{\sigma_n^2}$ ratio suggests a F-test (or possibly a $X^2$ test), the two variances are computed from the same sample which violates the independence assumption, and so such tests are not valid. From T10, both the mean and variance ratios show significant differences from one and deviation increases as n increases. However, both the Allegheny common and warrant and the Tri-Continental warrant are reasonable trendless and near one for n less than 20. Because of the outcome of these ratios, a second "corrected" calculation of the variance ratios was run to see if the movement in this ratio observed in T10 was due primarily to the movement in the mean ratio series. The correction made was to use $n\mu$ instead of $\mu_n$ and $n\nu$ instead of $\nu_n$ as the mean around which the $\sigma_n^2$ and $\sigma_n^2$ variances were computed, so that the resulting variance ratios would be conditional ratios given $n\mu = \mu_n$ and $n\nu = \nu_n$. This correction is reasonable because for all values of n, the sample size is large ($\geq 256$). The corrected series is presented in T11, and as can be seen, the correction had no effect on the variance ratio series. This result suggests that the variation in the sample was large enough to be little affected by the relatively small shifts in the mean.

In an attempt to better understand the behavior of these ratios, another table, T11a, was constructed using the new ratios $(\mu_n/\sigma_n^2)/(\mu/\sigma^2)$ and $(\nu_n/\delta_n^2)/(\nu/\delta^2)$. In a crude sense, this ratio, which should equal one for a stationary series, is a measure of the relative variation of the series for one-week intervals versus n-week intervals. Although the exact statistical meaning of this ratio is not clear, heuristically, the relative variation of the Allegheny stock and warrant series are about the same and near one for all n. The Atlas series jumps up immediately and then stabilizes at numbers significantly larger than one implying smaller relative variation for larger n. The Tri-Continental stock series is closer to Atlas in behavior while the Tri-Continental warrants behaves more like the Allegheny (see T11a). The point of these analyses was to get some idea of
\[ \frac{n \sigma^2}{\sigma^2} \]

<table>
<thead>
<tr>
<th></th>
<th>Tri-Continental</th>
<th>Allegheny</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stock</td>
<td>Warrant</td>
<td>Stock</td>
</tr>
<tr>
<td>2</td>
<td>1.02</td>
<td>0.98</td>
<td>0.94</td>
</tr>
<tr>
<td>4</td>
<td>1.05</td>
<td>0.98</td>
<td>0.93</td>
</tr>
<tr>
<td>6</td>
<td>1.07</td>
<td>1.00</td>
<td>0.94</td>
</tr>
<tr>
<td>8</td>
<td>1.15</td>
<td>1.04</td>
<td>0.95</td>
</tr>
<tr>
<td>10</td>
<td>1.23</td>
<td>1.05</td>
<td>0.96</td>
</tr>
<tr>
<td>12</td>
<td>1.32</td>
<td>1.09</td>
<td>0.95</td>
</tr>
<tr>
<td>14</td>
<td>1.35</td>
<td>1.09</td>
<td>0.93</td>
</tr>
<tr>
<td>16</td>
<td>1.33</td>
<td>1.06</td>
<td>0.93</td>
</tr>
<tr>
<td>18</td>
<td>1.34</td>
<td>1.08</td>
<td>0.93</td>
</tr>
<tr>
<td>20</td>
<td>1.40</td>
<td>1.11</td>
<td>0.94</td>
</tr>
<tr>
<td>22</td>
<td>1.49</td>
<td>1.13</td>
<td>0.96</td>
</tr>
<tr>
<td>24</td>
<td>1.59</td>
<td>1.15</td>
<td>0.99</td>
</tr>
<tr>
<td>26</td>
<td>1.68</td>
<td>1.16</td>
<td>1.00</td>
</tr>
<tr>
<td>28</td>
<td>1.73</td>
<td>1.16</td>
<td>1.01</td>
</tr>
<tr>
<td>30</td>
<td>1.72</td>
<td>1.14</td>
<td>1.02</td>
</tr>
<tr>
<td>32</td>
<td>1.70</td>
<td>1.12</td>
<td>1.04</td>
</tr>
<tr>
<td>34</td>
<td>1.67</td>
<td>1.09</td>
<td>1.08</td>
</tr>
<tr>
<td>36</td>
<td>1.69</td>
<td>1.07</td>
<td>1.12</td>
</tr>
<tr>
<td>38</td>
<td>1.68</td>
<td>1.06</td>
<td>1.16</td>
</tr>
<tr>
<td>40</td>
<td>1.69</td>
<td>1.04</td>
<td>1.19</td>
</tr>
</tbody>
</table>
\[
\left( \frac{\mu_n}{\sigma_n^z} \right) / \left( \frac{\mu}{\sigma^z} \right)
\]

<table>
<thead>
<tr>
<th>n</th>
<th>Tri-Continental</th>
<th></th>
<th>Allegheny</th>
<th></th>
<th>Atlas</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stock</td>
<td>Warrant</td>
<td>Stock</td>
<td>Warrant</td>
<td>Stock</td>
<td>Warrant</td>
</tr>
<tr>
<td>2</td>
<td>1.05</td>
<td>1.02</td>
<td>0.92</td>
<td>0.92</td>
<td>1.05</td>
<td>1.13</td>
</tr>
<tr>
<td>4</td>
<td>1.12</td>
<td>1.03</td>
<td>0.96</td>
<td>0.94</td>
<td>1.20</td>
<td>1.31</td>
</tr>
<tr>
<td>6</td>
<td>1.16</td>
<td>1.08</td>
<td>1.04</td>
<td>1.04</td>
<td>1.20</td>
<td>1.46</td>
</tr>
<tr>
<td>8</td>
<td>1.26</td>
<td>1.12</td>
<td>1.05</td>
<td>1.09</td>
<td>1.25</td>
<td>1.58</td>
</tr>
<tr>
<td>10</td>
<td>1.31</td>
<td>1.11</td>
<td>1.04</td>
<td>1.10</td>
<td>1.28</td>
<td>1.74</td>
</tr>
<tr>
<td>12</td>
<td>1.37</td>
<td>1.13</td>
<td>1.03</td>
<td>1.07</td>
<td>1.27</td>
<td>1.86</td>
</tr>
<tr>
<td>14</td>
<td>1.38</td>
<td>1.10</td>
<td>1.01</td>
<td>1.06</td>
<td>1.31</td>
<td>1.94</td>
</tr>
<tr>
<td>16</td>
<td>1.31</td>
<td>1.05</td>
<td>1.01</td>
<td>1.06</td>
<td>1.32</td>
<td>1.93</td>
</tr>
<tr>
<td>18</td>
<td>1.29</td>
<td>1.04</td>
<td>0.98</td>
<td>1.04</td>
<td>1.30</td>
<td>1.93</td>
</tr>
<tr>
<td>20</td>
<td>1.31</td>
<td>1.04</td>
<td>0.95</td>
<td>1.00</td>
<td>1.29</td>
<td>1.91</td>
</tr>
<tr>
<td>22</td>
<td>1.37</td>
<td>1.05</td>
<td>0.94</td>
<td>0.99</td>
<td>1.27</td>
<td>1.94</td>
</tr>
<tr>
<td>24</td>
<td>1.44</td>
<td>1.06</td>
<td>0.94</td>
<td>0.98</td>
<td>1.27</td>
<td>1.99</td>
</tr>
<tr>
<td>26</td>
<td>1.52</td>
<td>1.08</td>
<td>0.94</td>
<td>0.99</td>
<td>1.29</td>
<td>1.92</td>
</tr>
<tr>
<td>28</td>
<td>1.56</td>
<td>1.08</td>
<td>0.94</td>
<td>1.00</td>
<td>1.28</td>
<td>1.90</td>
</tr>
<tr>
<td>30</td>
<td>1.53</td>
<td>1.06</td>
<td>0.94</td>
<td>1.00</td>
<td>1.31</td>
<td>1.92</td>
</tr>
<tr>
<td>32</td>
<td>1.48</td>
<td>1.02</td>
<td>0.96</td>
<td>1.03</td>
<td>1.33</td>
<td>1.91</td>
</tr>
<tr>
<td>34</td>
<td>1.44</td>
<td>0.98</td>
<td>0.98</td>
<td>1.06</td>
<td>1.34</td>
<td>1.97</td>
</tr>
<tr>
<td>36</td>
<td>1.46</td>
<td>0.96</td>
<td>1.00</td>
<td>1.08</td>
<td>1.33</td>
<td>1.98</td>
</tr>
<tr>
<td>38</td>
<td>1.44</td>
<td>0.94</td>
<td>1.03</td>
<td>1.11</td>
<td>1.32</td>
<td>2.01</td>
</tr>
<tr>
<td>40</td>
<td>1.44</td>
<td>0.92</td>
<td>1.04</td>
<td>1.14</td>
<td>1.32</td>
<td>2.07</td>
</tr>
</tbody>
</table>
how much the lack of stationarity would affect the estimates of the crucial parameters of the theory. From equation (9), $\gamma$ can be seen to depend only upon the ratios $\mu/\sigma^2$ and $\beta/\sigma^2$, and hence one might expect from the results in T1a that the estimates of $\gamma$ for the Allegheny series would be robust for different interval lengths. For the same reasons, the estimates for Atlas and Tri-Continental would be expected to vary considerably.

There are three reasonable alternative hypotheses to the simple gaussian random walk which might explain the lack of stationarity and independence suggested by the above results: (1) The time series are essentially stationary and independent, but the distribution, although stable, is not gaussian, i.e., the process is generated by some other member of the Pareto-Le\'vy family. (2) The time series is stable with finite moments but is generated by a complicated random walk such as an Ornstein-Uhlenbeck process. (3) The time series is being generated by a gaussian random walk, but with reflecting barriers (Cootner's hypothesis).

A complete analysis to determine which, if any, of the alternative hypotheses is preferable to the gaussian assumption, is beyond the scope of this paper. However, it is possible to look at the particular series used here to see which hypotheses are supported.

It is well known that for all members of the Pareto-Le\'vy family of distributions (with the exception of the gaussian) the second moment is infinite, and for a subset of these distributions, the first moment is also infinite (e.g. Cauchy; Arcsine). However, Mandelbrot [2] has shown that if one calculates the mean and variance ratios of the sample, then $n \sigma^2$ should tend toward zero as $n$ increases, and either $n\mu/\mu_n$ is constant (for the finite-mean part of the family) or else it also tends to zero with increasing $n$ (for the infinite-mean part). From T10 and T11, the exact opposite is occurring, i.e., the variance and mean are increasing more slowly than linearly in $n$. Thus, one can rule out the hypothesis of an infinite-moment process for these processes.
how much the lack of stationarity would affect the estimates of the crucial parameters of the theory. From equation (9), \( \gamma \) can be seen to depend only upon the ratios \( \mu/\sigma^2 \) and \( \beta/\sigma^2 \), and hence one might expect from the results in Tll a that the estimates of \( \gamma \) for the Allegheny series would be robust for different interval lengths. For the same reasons, the estimates for Atlas and Tri-Continental would be expected to vary considerably.

There are three reasonable alternative hypotheses to the simple gaussian random walk which might explain the lack of stationarity and independence suggested by the above results: (1) The time series are essentially stationary and independent, but the distribution, although stable, is not gaussian, i.e., the process is generated by some other member of the Pareto-Le\(^{\text{ey}}\) family. (2) The time series is stable with finite moments but is generated by a complicated random walk such as an Ornstein-Uhlenbeck process. (3) The time series is being generated by a gaussian random walk, but with reflecting barriers (Cootner's hypothesis).

A complete analysis to determine which, if any, of the alternative hypotheses is preferable to the gaussian assumption, is beyond the scope of this paper. However, it is possible to look at the particular series used here to see which hypotheses are supported.

It is well known that for all members of the Pareto-Le\(^{\text{ey}}\) family of distributions (with the exception of the gaussian) the second moment is infinite, and for a subset of these distributions, the first moment is also infinite (e.g. Cauchy; Arcsine). However, Mandelbrot [2] has shown that if one calculates the mean and variance ratios of the sample, then \( n \sigma^2 \_n \sigma^2 \) should tend toward zero as \( n \) increases, and either \( n\mu/\mu_n \) is constant (for the finite-mean part of the family) or else it also tends to zero with increasing \( n \) (for the infinite-mean part). From T10 and T11, the exact opposite is occurring, i.e., the variance and mean are increasing more slowly than linearly in \( n \). Thus, one can rule out the hypothesis of an infinite-moment process for these processes.
The Ornstein-Uhlenbeck (O.U.) process, approximated for discrete time, postulates that the rate of change of the increments is generated by a Wiener process with negative serial correlation, i.e.,

\[ \Delta^2 X_t = -\eta n \Delta X_t + \mu n + \eta_t \]

where \( \Delta \) is the usual forward difference operator for increments of size \( n \), \( \eta \) is a positive constant and \( [\eta_t] \) are independent, identical gaussian distributions with zero means. Then, it can be shown that, if \( U_t = \Delta X_t \), then

\[ E [U_t] = U_0 e^{-\eta n} + \frac{\mu}{\eta} [n(n-1)e^{-\eta n}] \]

and

\[ \text{Var} [U_t] = \frac{\sigma^2}{n} (1-e^{-2\eta n}) \]

By Taylor series expansion, the ratio \( \eta_0 / \mu_0 \) for such a process will be proportional to \( n \) for \( n \) small and will gradually become constant, asymptotically for large \( n \). This result does not agree with observed behavior. However, the same process would suggest that the variance ratio \( n \sigma^2 / \sigma_0^2 \) would be approximately constant for \( n \) small, and for \( n \) large, proportional to \( n \). This result does agree with the observations. Although the mean-ratio behavior seems to rule out the O-U hypothesis, more detailed analysis would be required before rejecting it.

The third alternative hypothesis (which is called Cootner's hypothesis because he was the first to formalize it in a stock market context [1]) suggests that the behavior of stock prices is a simple geometric random walk with reflecting barriers. These barriers are caused by the differentiation of knowledge between professional and non-professional investors. The behavior of the mean and variance ratios for a process of this type fit the results of Tl0 and Tl1. However, because his theory is rather broad and not well-defined, the most that one can draw from this result without further analysis is that the data does not reject his theory.
The Ornstein-Uhlenbeck (O.U.) process, approximated for discrete time, postulates that the rate of change of the increments is generated by a Wiener process with negative serial correlation, i.e.,

$$\Delta^2 x_t = -\eta n \Delta x_t + \mu n + \eta_t$$

where $\Delta$ is the usual forward difference operator for increments of size $n$, $\eta$ is a positive constant and $[\eta_t]$ are independent, identical gaussian distributions with zero means. Then, it can be shown that, if $U_t = \Delta x_t$, then

$$E[U_t] = U_0 e^{-\eta n} + \frac{\mu}{n} \left[ n n - 1 + e^{-\eta n} \right]$$

and

$$\text{Var}[U_t] = \sigma^2 \left( 1 - e^{-2\eta n} \right)$$

By Taylor series expansion, the ratio $\eta \mu / \mu_n$ for such a process will be proportional to $n$ for $n$ small and will gradually become constant, asymptotically for large $n$. This result does not agree with observed behavior. However, the same process would suggest that the variance ratio $n \sigma^2 / \sigma_n^2$ would be approximately constant for $n$ small, and for $n$ large, proportional to $n$. This result does agree with the observations. Although the mean-ratio behavior seems to rule out the O-U hypothesis, more detailed analysis would be required before rejecting it.

The third alternative hypothesis (which is called Cootner's hypothesis because he was the first to formalize it in a stock market context [1]) suggests that the behavior of stock prices is a simple geometric random walk with reflecting barriers. These barriers are caused by the differentiation of knowledge between professional and non-professional investors. The behavior of the mean and variance ratios for a process of this type fit the results of T10 and T11. However, because his theory is rather broad and not well-defined, the most that one can draw from this result without further analysis is that the data does not reject his theory.
Actually, there is a fourth explanation which applies primarily to the Atlas series. Atlas warrants were trading near $1 (market price, not normalized price) for a large portion of the series. Because the minimum increment in price above a/$1 is $1/8(12.5¢) and below $1, $1/16(6.25¢), any move at all in the Atlas price would imply a minimum of from 6-12% change. The approximation of stock prices as being continuous is no longer a good one, i.e., problems of indivisibilities. This bad approximation could give the appearance of a larger deviation for short-intervals than would be justified over longer intervals of time when Atlas' price moved to higher levels where indivisibility effects would be less pronounced. Although this may be the reason for the particularly poor behavior of the Atlas warrant series in T10, all three sets of series exhibited the same behavior to some degree which suggests that although indivisibility may have an effect, it does not explain the observed mean and variance-ratio behavior.

Before going on to the section which test the warrant pricing theory, there is a set of tests which depend on both the assumption of normality and the accuracy of the warrant pricing theory. These tests examine the relationship between the $\Delta X_t$ and $\Delta Y_t$ time series. As shown previously in (10) if the distributional assumptions and the theory hold, then $\delta^2 = \gamma^2 \sigma^2$ and $\nu = \gamma \mu$.

\[(t12)\]

<table>
<thead>
<tr>
<th></th>
<th>$\frac{\gamma^2 \sigma^2}{\delta^2}$</th>
<th>$\frac{\nu}{\gamma \mu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>1.650</td>
<td>1.140</td>
</tr>
<tr>
<td>Allegheny</td>
<td>1.025</td>
<td>1.006</td>
</tr>
<tr>
<td>Atlas</td>
<td>1.075</td>
<td>1.057</td>
</tr>
</tbody>
</table>
From T12, these two relationships hold well for Atlas and in the case of Allegheny, exceptionally well.

In summary of this section, the Atlas series appears to have enough severe distributional discrepancies from the assumed behavior, that it is considered unlikely to provide any serious test of the theory. Tri-Continental is a mixed result. Allegheny appears to satisfy most of the distributional assumption quite well, and although its series clearly deviates from the stationarity-normality assumptions, these deviations do not appear to be large. It will be Allegheny that will provide the best test of the warrant pricing theory. For ease of reference, T13 summarizes the results of this section's tests.

(T13)

<table>
<thead>
<tr>
<th>Normality (T2, R3, R4, T4, T12)</th>
<th>Tri-Continental</th>
<th>Allegheny</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock</td>
<td>Fair</td>
<td>Good</td>
<td>Poor</td>
</tr>
<tr>
<td>Warrant</td>
<td>Best</td>
<td>Good</td>
<td>Poor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Autocorrelation (T6, R5, R6, T7)</th>
<th>Tri-Continental</th>
<th>Allegheny</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock</td>
<td>Fair</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Warrant</td>
<td>Best</td>
<td>Good</td>
<td>Good</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trend (T8, T9, R7, R8)</th>
<th>Tri-Continental</th>
<th>Allegheny</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock</td>
<td>Good</td>
<td>Good</td>
<td>Poor</td>
</tr>
<tr>
<td>Warrant</td>
<td>Good</td>
<td>Good</td>
<td>Poor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independence (T10, T11, T11a)</th>
<th>Tri-Continental</th>
<th>Allegheny</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock</td>
<td>Poor</td>
<td>Good</td>
<td>Poor</td>
</tr>
<tr>
<td>Warrant</td>
<td>Good</td>
<td>Good</td>
<td>Poor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( Y_0 ) (T1)</th>
<th>Tri-Continental</th>
<th>Allegheny</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.63</td>
<td>1.27</td>
<td>1.32</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( Y_1 ) (T5)</th>
<th>Tri-Continental</th>
<th>Allegheny</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.16</td>
<td>1.28</td>
<td>1.0+</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \nu_0 ) ( \gamma_2 ) (T12)</th>
<th>Tri-Continental</th>
<th>Allegheny</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.86</td>
<td>1.28</td>
<td>1.39</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \delta_0 ) ( \gamma_3 ) (T12)</th>
<th>Tri-Continental</th>
<th>Allegheny</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.27</td>
<td>1.25</td>
<td>1.27</td>
<td></td>
</tr>
</tbody>
</table>
Testing the theory of "rational" warrant pricing. The strategy of the tests of this section is to estimate the parameters \((\gamma, f(\gamma))\) of the basic price relations, (11) and (12), between the warrant and the stock and ask the two questions:
1. How well do the observations fit the specified structure?
2. Are the estimated parameters equal to their theoretically-predicted values?

The first set of regressions are on the basic equation (11).

\[(R10)\]

\[Y_t - \hat{Y}_t X_t + \hat{\eta}_t\]

<table>
<thead>
<tr>
<th></th>
<th>(\hat{\gamma}_1)</th>
<th>(\hat{\eta}_1)</th>
<th>(R^2)</th>
<th>Standard Error</th>
<th>Durbin-Watson</th>
<th>Number Of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>1.800</td>
<td>-1.238</td>
<td>.9423</td>
<td>.0476</td>
<td>.3647</td>
<td>296</td>
</tr>
<tr>
<td></td>
<td>(69.3)</td>
<td>(-45.75)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allgheny</td>
<td>1.229</td>
<td>-.5716</td>
<td>.9840</td>
<td>.03725</td>
<td>.9113</td>
<td>296</td>
</tr>
<tr>
<td></td>
<td>(134)</td>
<td>(-53.54)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlas</td>
<td>1.284</td>
<td>-.5518</td>
<td>.9125</td>
<td>.1480</td>
<td>.3092</td>
<td>297</td>
</tr>
<tr>
<td></td>
<td>(55.46)</td>
<td>(-31.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because the Durbin-Watson Statistic suggests significant autocorrelation, a second version of equation (11) is estimated using the Cochrane-Orcutt\(^6\) technique to eliminate first-order autocorrelation.

\(^6\) In each regression, Hildreth-Lu estimates were also made to ensure that the Cochrane-Orcutt routine had not estimated \(\rho\)'s at a local minimum.
(R11)

\[ Y_{t+1} - \rho \ Y_t = \hat{\gamma}_2 \left( X_{t+1} - \rho \ X_t \right) + (1-\rho) \ \hat{\eta}_2 \]

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\gamma}_2 )</th>
<th>( \hat{\eta}_2 )</th>
<th>( \rho_1 )</th>
<th>( R^2 )</th>
<th>Standard Error</th>
<th>Durbin-Watson</th>
<th>Number Of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>0.919 (18.8)</td>
<td>-0.2257 (-2.02)</td>
<td>0.9902</td>
<td>0.9901</td>
<td>0.0197</td>
<td>2.3871</td>
<td>295</td>
</tr>
<tr>
<td>Allegheny</td>
<td>1.210 (75.5)</td>
<td>-0.5498 (-29.2)</td>
<td>0.5561</td>
<td>0.9888</td>
<td>0.03122</td>
<td>2.1861</td>
<td>295</td>
</tr>
<tr>
<td>Atlas</td>
<td>0.760 (12.1)</td>
<td>-0.8395 (-7.8)</td>
<td>0.9607</td>
<td>0.9789</td>
<td>0.0727</td>
<td>2.613</td>
<td>296</td>
</tr>
</tbody>
</table>

Because the Durbin-Watson statistic estimated in R11 suggests significant first-order autocorrelation in the transformed equation (second-order in the original R10) for Atlas and Tri-Continental and because there is no problem with losing degrees of freedom, the Cochran-Orcutt technique is applied to the transformed variables.  

(R12)

\[ Y_{t+2} - (\rho_1 + \rho_2) Y_{t+1} + \rho_1 \rho_2 Y_t = \]

\[ \hat{\gamma}_3 \left( X_{t+2} - (\rho_1 + \rho_2) X_{t+1} + \rho_1 \rho_2 X_t \right) + (1-\rho_1)(1-\rho_2) \ \hat{\eta}_3 \]

7 If the original error in R10 is \( Z_t \), then the results of R11 and R12 imply— (Continued on next page after R12)
(R12)

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\gamma}_3 )</th>
<th>( \hat{\eta}_3 )</th>
<th>( \rho_1 )</th>
<th>( \rho_2 )</th>
<th>( R^2 )</th>
<th>Standard Error</th>
<th>Durbin-Watson</th>
<th>Number Of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>0.966 (-20.4)</td>
<td>-.2455</td>
<td>.9902</td>
<td>-.2146</td>
<td>.5626</td>
<td>.01924</td>
<td>2.049</td>
<td>294</td>
</tr>
<tr>
<td>Allegheny</td>
<td>1.214 (82)</td>
<td>-.555</td>
<td>.5561</td>
<td>-.0967</td>
<td>.9513</td>
<td>.03112</td>
<td>1.9905</td>
<td>294</td>
</tr>
<tr>
<td>Atlas</td>
<td>0.829 (14.7)</td>
<td>-.804</td>
<td>.9607</td>
<td>-.3261</td>
<td>.3993</td>
<td>.0689</td>
<td>2.0853</td>
<td>295</td>
</tr>
</tbody>
</table>

\( Z_{t+2} - (\rho_1 + \rho_2) Z_{t+1} + \rho_1 \rho_2 Z_t = \zeta_t \)

Where \( E[\zeta_t \zeta_{t-\theta}] = 0 \) for \( \theta \neq 0 \).

Define:

\[
\lambda_1 \equiv (\rho_1 + \rho_2) \pm \sqrt{(\rho_1 + \rho_2)^2 - 4\rho_1 \rho_2} \quad ;
\]

\[
\lambda_2 \equiv (\rho_1 + \rho_2) - \sqrt{(\rho_1 + \rho_2)^2 - 4\rho_1 \rho_2} \quad ;
\]

<table>
<thead>
<tr>
<th></th>
<th>( \rho_1 )</th>
<th>( \rho_2 )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>.9902</td>
<td>-.2146</td>
<td>.9878</td>
<td>-.2122</td>
</tr>
<tr>
<td>Allegheny</td>
<td>.5561</td>
<td>-.0967</td>
<td>.5742</td>
<td>-.0803</td>
</tr>
<tr>
<td>Atlas</td>
<td>.9607</td>
<td>-.3261</td>
<td>.9623</td>
<td>-.3277</td>
</tr>
</tbody>
</table>

The behavior of \( Z_t \) is that any disturbance will be damped (although slightly oscillatory, this behavior is clearly dominated.). This result would support a more complicated theory of adaptive expectations where the regression line is the expected price of the warrant. However, this idea goes beyond any simple theory and will not be pursued further in the paper.
Summary of Theoretical and Estimated Parameters

(Tl4)

<table>
<thead>
<tr>
<th>Predictions</th>
<th>Estimators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tl: ( \gamma_0 ), ( f(\gamma_0) )</td>
<td>T5: ( \gamma_0 ), ( f(\gamma_1) )</td>
</tr>
<tr>
<td>Tri-Continental: ( \gamma )</td>
<td>1.63</td>
</tr>
<tr>
<td>( f(\gamma) )</td>
<td>-1.09</td>
</tr>
<tr>
<td>( \hat{\eta} )</td>
<td>-1.238</td>
</tr>
<tr>
<td>Allegheny: ( \gamma )</td>
<td>1.27</td>
</tr>
<tr>
<td>( f(\gamma) )</td>
<td>-0.656</td>
</tr>
<tr>
<td>( \hat{\eta} )</td>
<td>-0.5716</td>
</tr>
<tr>
<td>Atlas: ( \gamma )</td>
<td>1.32</td>
</tr>
<tr>
<td>( f(\gamma) )</td>
<td>-0.731</td>
</tr>
<tr>
<td>( \hat{\eta} )</td>
<td>-0.5518</td>
</tr>
</tbody>
</table>

(**not defined for \( \gamma < 1 \))

From summary table Tl4, both the Atlas and Tri-Continental estimates are not close to the values predicted by the theory. The estimates of \( \gamma \) were dispersed between \( (\gamma_0, \hat{\gamma}_1) \) and \( (\gamma_1, \hat{\gamma}_2, \hat{\gamma}_3) \). Further, the estimates of \( \gamma < 1 \) suggests that the warrant prices are concave functions of the stock price \( (\beta < \alpha) \) which completely contradicts one of the basic axioms of the theory. The large change in the estimate of the \( \gamma \) coefficient between the original and transformed equations raise doubt about how good the estimators are anyway. The estimates for Allegheny are reasonably close together and suggest at first look agreement with the theory.

Before making the formal tests for rejection of the theory, consider two assumptions implicit in equations R10-R12. The first is on the distribution of the \( X_t \) examined in the previous section. As was noted at the end of that section, Atlas' distribution deviated sufficiently from the Wiener process.
assumption to make it a poor test case for the theory. Tri-Continental did not deviate enough from the distributional assumptions to explain its erratic performance as shown in T14. The second implicit assumption is that \( S_t \leq C(\gamma) \), i.e. the equations R10-R12 are only valid for \( S_t \) less than the conversion price. To see how good this assumption is, the data is checked to see how many sample observations are equal to or exceed \( C(\gamma_0) \). The results appear in T15.

(T15)

<table>
<thead>
<tr>
<th></th>
<th>Tri-Continental</th>
<th>Allegheny</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total number of Observations of</strong> ( S_t ) ( C(\gamma_0) )</td>
<td>296</td>
<td>296</td>
<td>297</td>
</tr>
<tr>
<td><strong>Number of Observations</strong> ( S_t &gt; C(\gamma_0) ) ( C(\gamma_1) )</td>
<td>2.58</td>
<td>4.72</td>
<td>4.125</td>
</tr>
<tr>
<td><strong>Number of Observations</strong> ( S_t &gt; C(\gamma_1) ) ( C(\gamma_2) )</td>
<td>248</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td><strong>Number of Observations</strong> ( S_t &gt; C(\gamma_2) ) ( C(\gamma_3) )</td>
<td>2.25</td>
<td>5.367</td>
<td>4.521</td>
</tr>
<tr>
<td></td>
<td>296</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td><strong>Number of Observations</strong> ( S_t &gt; C(\gamma_3) ) ( C(\gamma_4) )</td>
<td>not defined</td>
<td>5.759</td>
<td>not defined</td>
</tr>
</tbody>
</table>

Most of the observations on Tri-Continental are for \( S_t > C(\gamma_0) \), the region where, as mentioned before, the theory was expected to do poorly. This is also the region where the warrant price changes from a power function of the stock price to a simple linear one, \( W_t = S_t - 1 \). Further, after looking into the corporate records, it was discovered that large scale conversions of the warrant had been taking place during the entire period of observation. To examine the effect on the \( \hat{\gamma} \) of this mis-specification, consider figure 1 re-produced with most of the observations above the predicted conversion price.
Thus, to fit the data to the specification, a higher curve is required than the theory would predict. To show this, note \( W_t = (C-1)(S_t/C)^Y \) so

\[
\frac{\partial \log W_t}{\partial Y}_{S_t} = \log \left( \frac{S_t}{C} \right) < 0 \quad \text{for } S_t < C
\]

and therefore, the fitted \( Y \) would be expected to be lower than the theoretically predicted \( Y \). Most of the data is for \( 2 \leq S_t \leq 4 \). Table T16 shows the values predicted by \( W_t = S_t - 1 \) and \( W_t = .78(S_t)^{.97} \), the fitted equation for \( 2 \leq S_t \leq 4 \).

(T16)

<table>
<thead>
<tr>
<th>S</th>
<th>( W = S - 1 )</th>
<th>( W = .78(S)^{.97} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>1.0</td>
<td>1.55</td>
</tr>
<tr>
<td>2.5</td>
<td>1.5</td>
<td>1.93</td>
</tr>
<tr>
<td>3.0</td>
<td>2.0</td>
<td>2.32</td>
</tr>
<tr>
<td>3.5</td>
<td>2.5</td>
<td>2.65</td>
</tr>
<tr>
<td>4.0</td>
<td>3.0</td>
<td>3.10</td>
</tr>
<tr>
<td>S</td>
<td>( W = S - 1 )</td>
<td>( W = .78(S)^{.97} )</td>
</tr>
</tbody>
</table>
Whether this mis-specification explains the full reason for a significantly lower \( \gamma \) for Tri-Continental is difficult to decide. In any case, one could certainly not reject the theory on the basis of the Tri-Continental results.

Because twenty of the \( S_t \) observations for Allegheny were for \( S_t \geq C(\gamma_0) \), regression R10 was re-run using only \( S_t < C(\gamma_0) \).

\[
(R13) \quad Y_t = \hat{\gamma}_4 X_t + \hat{\eta}_4 \quad \text{for} \quad S_t < C(\gamma_0)
\]

<table>
<thead>
<tr>
<th></th>
<th>( \hat{\gamma}_4 )</th>
<th>( \hat{\eta}_4 )</th>
<th>( R^2 )</th>
<th>Standard Error</th>
<th>Durbin-Watson</th>
<th>Number Of Observations</th>
<th>F(( \hat{\gamma} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allegheny</td>
<td>1.222</td>
<td>-.5644</td>
<td>.9676</td>
<td>.0379</td>
<td>.8535</td>
<td>276</td>
<td>-.576</td>
</tr>
<tr>
<td>((90.5))</td>
<td>(-37.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison:</td>
<td>( \hat{\gamma}_1 )</td>
<td>( \hat{\eta}_1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allegheny</td>
<td>1.229</td>
<td>-.5716</td>
<td>.9840</td>
<td>.0373</td>
<td>.9113</td>
<td>296</td>
<td>-.591</td>
</tr>
<tr>
<td>((134))</td>
<td>(-53.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In agreement with the explanation of how \( \gamma \) should change given in figure 8, \( \hat{\gamma}_4 < \hat{\gamma}_1 \). However, the difference is not significant. Also the measure of "high-contact" (\( f(\hat{\gamma})=\hat{\eta} \)) in R13 is better as would be expected.

To formally test the theory, four tests were performed:
1. Does \( (\gamma_0, f(\gamma_0)) = (\hat{\gamma}, \hat{\eta}_1) \)?
2. Does \( \gamma_0 = \hat{\gamma}_1 \)?
3. Does \( f(\gamma_0) = \hat{\eta}_1 \)?
4. Does \( f(\hat{\gamma}_1) = \hat{\eta}_1 \) (the test for high-contact).

The results of these tests are summarized in T17.
<table>
<thead>
<tr>
<th>(T17)</th>
<th>Tri-Continental</th>
<th>Allegheny</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>Confidence Level</td>
<td>Statistic</td>
</tr>
<tr>
<td>1. ( (\gamma_0, f(\gamma_0)) = (\hat{\gamma}_1, \hat{n}_1) )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test: R10 ( (\hat{\gamma}_1, \hat{n}_2) )</td>
<td>1.42</td>
<td>99%</td>
<td>2.12</td>
</tr>
<tr>
<td>F-test: R11 ( (\hat{\gamma}_2, \hat{n}_2) )</td>
<td>1.73</td>
<td>99%</td>
<td>1.34</td>
</tr>
<tr>
<td>F-test: R12 ( (\hat{\gamma}_3, \hat{n}_3) )</td>
<td>3.72</td>
<td>99%</td>
<td>2.04</td>
</tr>
<tr>
<td>2. ( \gamma_0 = \hat{\gamma}_1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-test: R10 ( \hat{\gamma}_1 )</td>
<td>6.54</td>
<td>99+%</td>
<td>-4.49</td>
</tr>
<tr>
<td>t-test: R11 ( \hat{\gamma}_2 )</td>
<td>-14.57</td>
<td>99+%</td>
<td>-3.73</td>
</tr>
<tr>
<td>t-test: R12 ( \hat{\gamma}_3 )</td>
<td>-14.0</td>
<td>99+%</td>
<td>-3.78</td>
</tr>
<tr>
<td>3. ( f(\gamma_0) = \hat{n}_1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-test: R10 ( \hat{n}_1 )</td>
<td>5.47</td>
<td>99+%</td>
<td>7.90</td>
</tr>
<tr>
<td>t-test: R11: ( \hat{n}_2 )</td>
<td>6.80</td>
<td>99+%</td>
<td>5.64</td>
</tr>
<tr>
<td>t-test: R12: ( \hat{n}_3 )</td>
<td>7.90</td>
<td>99+%</td>
<td>13.0</td>
</tr>
<tr>
<td>4. ( f(\hat{\gamma}_1) = \hat{n}_1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-test: R10 ( \hat{n}_1 )</td>
<td>.037</td>
<td>2.5%</td>
<td>1.81</td>
</tr>
<tr>
<td>t-test: R11 ( \hat{n}_2 )</td>
<td>*</td>
<td>*</td>
<td>0.47</td>
</tr>
<tr>
<td>t-test: R12 ( \hat{n}_3 )</td>
<td>*</td>
<td>*</td>
<td>1.47</td>
</tr>
</tbody>
</table>

(\#not defined for \( \gamma < 1 \))
Nearly all the tests rejected at the 99+\% level. The only test which did not totally reject was the one for "high contact". However, these tests should be taken for what they say: namely, they reject the hypothesis that the theory holds exactly, conditional on the estimate $\gamma_0$ and the assumption of normality of the errors in the time series being true. Because of the large number of observations, the exactly remark must be taken seriously. Just as the skewness and kurtosis tests in the previous section were quite sensitive to any deviations, so also are these tests. A more important point is the "conditional on the estimate $\gamma_0$...being true". $\gamma_0$ was computed on the basis of estimates of $\mu$, $\sigma^2$, and $\beta$. From R1 and R2, the t-statistics on $\mu$ and $\beta$ are small which suggests that the variation around these estimates is relatively large. To make the point, consider the following alternative test of the theory: take $\hat{\gamma}_1$ as estimated in the various regressions; using the estimates of $\mu$, $\sigma^2$ from R1 and formula (8) for $\gamma$, solve for $\hat{\beta}_1 = \hat{\beta}_1(\hat{\gamma}_1; \mu, \sigma^2)$. Now, take $\beta_0$ and $\delta^2$ as estimated in R2 and form the t-statistic $(\beta_0 - \hat{\beta}_1)/\delta$ and ask the question does $\beta_0 = \hat{\beta}_1$? (i.e., could $\hat{\beta}$ implied by the estimate $\hat{\gamma}$ be the $\beta$ estimate of the time series?). The results of this test are displayed in T18. The results were quite encouraging.

The test is discriminating enough to reject the theory for Tri-Continental and Atlas where it was known that the theory should not hold and at the same time, it did not reject the theory for Allegheny where all the evidence suggested rather strong agreement.
Nearly all the tests rejected at the 99+% level. The only test which did not totally reject was the one for "high contact". However, these tests should be taken for what they say: namely, they reject the hypothesis that the theory holds exactly, conditional on the estimate $\gamma_0$ and the assumption of normality of the errors in the time series being true. Because of the large number of observations, the exactly remark must be taken seriously. Just as the skewness and kurtosis tests in the previous section were quite sensitive to any deviations, so also are these tests. A more important point is the "conditional on the estimate $\gamma_0$...being true". $\gamma_0$ was computed on the basis of estimates of $\mu$, $\sigma^2$, and $\beta$. From R1 and R2, the $t$-statistics on $\mu$ and $\beta$ are small which suggests that the variation around these estimates is relatively large. To make the point, consider the following alternative test of the theory: take $\hat{\gamma}_1$ as estimated in the various regressions; using the estimates of $\mu$, $\sigma^2$ from R1 and formula (8) for $\gamma$, solve for $\hat{\beta}_1 = \hat{\beta}_1(\hat{\gamma}_1; \mu, \sigma^2)$. Now, take $\beta_0$ and $\sigma^2$ as estimated in R2 and form the $t$-statistic $(\beta_0 - \hat{\beta}_1)/\delta$ and ask the question does $\beta_0 = \hat{\beta}_1$? (i.e., could $\hat{\beta}$ implied by the estimate $\hat{\gamma}$ be the $\beta$ estimate of the time series?). The results of this test are displayed in Table 3. The results were quite encouraging. The test is discriminating enough to reject the theory for Tri-Continental and Atlas where it was known that the theory should not hold and at the same time, it did not reject the theory for Allegheny where all the evidence suggested rather strong agreement.
Testing the Theory's Ability to Predict. In this last section of the paper, the rational warrant pricing theory is pitted against four alternative theories. The period chosen is January, 1962 to December, 1962. This was a period when the stock market declined significantly. The prediction period was chosen near to the test period so that the basic structure of the economy and the individual companies could be taken as unchanged. However, because the period 1963-68 was primarily one of "bull" markets, by choosing this erratic period, it was thought that any merely extrapolative techniques would fail. Because Atlas and Tri-Continental did not fit the theory in the previous sections, the only stock tested in this period is Allegheny.

The four alternative theories are: (1) J.P. Shelton's formula [5]: \( W_t = 0.81875 S_t - 0.275 \). (2) A "rule of thumb" used by S.W. Fried, editor of the widely-read investment service on warrants, R.H.M. Warrant Service: \( W_t = 0.5 S_t - 0.1 \). (3) S. A.
Kassouf's formula (see [5]): $W_t = (1 + S_t^2) \cdot 0.5 - 1$. (4) A "naive theory": regress $W_t$ on $S_t$ and a constant for the observation period of 1963-68. The regressions are shown in R14 and R15 below.

Originally it was planned to test the theory over a variety of periods and to perform several tests on the prediction results. However, due to a time constraint and the fact that the author plans to make a more extensive study in the future, it was decided to compute the coefficient of variation, the mean and variance of the difference between predicted and observed warrant prices and of the percentage error between the predicted and observed prices. The percentage error used was the logarithmic percentage, $2(\hat{W}_t - W_t)/(\hat{W}_t + W_t)$, so that consistent under or over estimation would not bias the results.

(R14) $W_t = a_1 S_t + b_1$

<table>
<thead>
<tr>
<th></th>
<th>$a_1$</th>
<th>$b_1$</th>
<th>$R^2$</th>
<th>Durbin-Watson</th>
<th>Standard Error</th>
<th>Number Of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>1.289</td>
<td>-1.744</td>
<td>.9346</td>
<td>.2907</td>
<td>.1086</td>
<td>296</td>
</tr>
<tr>
<td>(64.8)</td>
<td>(-30.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allegheny</td>
<td>0.962</td>
<td>-0.696</td>
<td>.9879</td>
<td>.8714</td>
<td>.09718</td>
<td>296</td>
</tr>
<tr>
<td>(155)</td>
<td>(-33.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlas</td>
<td>.6430</td>
<td>-0.0765</td>
<td>.9416</td>
<td>.3498</td>
<td>.03691</td>
<td>297</td>
</tr>
<tr>
<td>(69)</td>
<td>(-13.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(R15) $W_{t+1} - \rho W_t = a_2 (S_{t+1} - \rho S_t) + (1-\rho) b_2$
<table>
<thead>
<tr>
<th></th>
<th>$a_2$</th>
<th>$b_2$</th>
<th>$\rho$</th>
<th>$R^2$</th>
<th>Durbin-Watson</th>
<th>Standard Error</th>
<th>Number Of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tri-Continental</td>
<td>.608</td>
<td>.855</td>
<td>.9963</td>
<td>.9926</td>
<td>2.4343</td>
<td>.03657</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td>(19.1)</td>
<td>(1.73)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allegheny</td>
<td>.9439</td>
<td>-.636</td>
<td>.588</td>
<td>.9918</td>
<td>2.3091</td>
<td>.08016</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td>(83.5)</td>
<td>(-16.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlas</td>
<td>.5213</td>
<td>-.0016</td>
<td>.982</td>
<td>.982</td>
<td>2.4342</td>
<td>.0205</td>
<td>296</td>
</tr>
<tr>
<td></td>
<td>(19.6)</td>
<td>(-.077)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results of the test are displayed on the following page in T19. The number of observations in all cases was 52.

**Conclusion.** Although the rational warrant theory did significantly better than either the Fried or Shelton formula, it was only slightly more accurate than the Kassouf formula or the linear "naive theory". Part of the reason that the linear regression did as well as it did is that most of the observations were between 1 and 2.5 so the convexity of the true relations would not show. Further, a linear fit (which by arbitrage) is forced through the origin would have done considerably worse.

This study should be treated as preliminary. Although one should not infer from these results that the Samuelson theory is the correct one, there appears to be significant enough agreement to warrant further study.

Further study planned by the author will include:
<table>
<thead>
<tr>
<th></th>
<th>Standard Deviation Percentage Error</th>
<th>Variance Percentage Error</th>
<th>Variance Percentage Error</th>
<th>Variance Percentage Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-5.3%</td>
<td>3.48x10^-3</td>
<td>2.84x10^-3</td>
<td>2.21x10^-3</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\bar{W}_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-5.29</td>
<td>3.46x10^-3</td>
<td>2.10x10^-3</td>
<td>2.25x10^-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.091</td>
<td>2.09</td>
<td>2.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.55</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.95x10^-3</td>
<td>1.95x10^-3</td>
<td>1.95x10^-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Derivation and testing of alternative hypotheses on the distributional theory of the stock and warrant price series. For those distributions for which no statistical tests exist, monte carlo simulations will be attempted.

2. A better theory of what determines \( \beta \), the yield on warrants. As has already been shown [4], \( \beta \), should be a function of the length of time until expiration, the stock price, and the yield on the stock. Other possibilities are yields on other warrants, leverage, and dividends paid on the common.

3. A study of the behavior of finite-lived warrants which will involve (numerical) solution of the partial differential equation(3).
References:


Biographical Note

Robert Merton was born in New York, New York, in 1944. He received a B.S. in Engineering Mathematics from Columbia University School of Engineering and Applied Science in 1966, and a M.S. in Applied Mathematics from California Institute of Technology in 1967. He is a member of Tau Beta Pi and Sigma Xi. He was a research assistant to Paul A. Samuelson from 1967 to 1970 and is the editor of the third volume of The Collected Scientific Papers of Paul A. Samuelson. In September, 1969, he was appointed Instructor in the department of Economics, M.I.T., and in July, 1970, appointed Assistant Professor of Finance, Sloan School of Management, M.I.T..
