The Structure of Investor Preferences and Asset Returns, and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds

DAVID CASS and JOSEPH E. STIGLITZ

Received September 3, 1969

1. INTRODUCTION

Over ten years ago, Tobin [5] showed that under certain conditions the investor's portfolio allocation decision could be considered as a two-stage process: The investor first decides in what proportions to purchase the available risky assets, and then he decides how to divide his total investment between risky and safe assets. This kind of "separation" is a special case of a more general property of the investor's optimal portfolio allocation, namely, that given a market in which there are available \( n \) different assets, nonetheless all the opportunities relevant to the investor's decision can be provided by a set of \( m \) \((m < n)\) "mutual funds," i.e., a set of \( m \) linear combinations (with weights adding to one) of the available assets.

Interest in this sort of property derives from at least four sources: (i) Keynesian macro-economic models conventionally assume that such a separation property obtains. (ii) When such a separation property does obtain, achieving a pure exchange Pareto optimum may not require a full complement of Arrow-Debreu securities. And, of course, there are good reasons—for example, transaction costs—for believing that a complete set of Arrow-Debreu markets will not exist. (iii) Many of the results in modern portfolio theory depend crucially on the existence of a safe and just one risky asset or, equivalently, a single mutual fund composed of risky assets.\(^1\) (iv) The separation property represents a particular extension

\(^*\)The research described in this paper was carried out under grants from the National Science Foundation and from the Ford Foundation. We have benefited from the comments of many people, but in particular want to thank W. M. Gorman and J. D. Sargent for pointing out an omission in our original argument.

\(^1\) See, for instance, Arrow [1] and Cass-Stiglitz [2] for analysis of the problem of determining the effects of changes in wealth on portfolio allocation with and without this assumption, respectively.
of an aggregation property which has long interested economists.\textsuperscript{3}

The purpose of this paper is two-fold: First, to find necessary and sufficient conditions (in terms of investor preferences) under which such mutual funds can be formed, and second, to find conditions under which the more restrictive separation into risky and safe assets can take place. Our results are illuminating not only because they show how very restrictive the conditions are, but also because they isolate the peculiar role played by "money" (the usual interpretation given the safe asset).

2. FORMAL STATEMENT OF THE PROBLEM

The individual investor has an initial wealth \( W_0 \). With it, he can purchase \( n \) different risky assets or securities. If his investment in the \( i^{th} \) security is \( Z_i \), then his budget constraint is simply

\[
\sum_{i=1}^{n} Z_i \leq W_0 .
\]

The investor believes that the \( i^{th} \) security will yield the positive gross return (per dollar invested in it) \( \rho_{i\theta} > 0 \) if state \( \theta \) occurs. Thus, for any given portfolio choice \( Z_i \), \( i = 1, ..., n \), \( W_\theta = \sum_{i=1}^{n} Z_i \rho_{i\theta} \) is his wealth next period if state \( \theta \) occurs. He chooses his portfolio to maximize the expected utility of his wealth next period

\[
EU(W_\theta) = EU \left( \sum_{i=1}^{n} Z_i \rho_{i\theta} \right)
\]

subject to his budget constraint (1). We will assume that \( U \) is twice differentiable and strictly concave, \( U''(\bar{W}) < 0 \).\textsuperscript{3} Necessary and sufficient condi-

\textsuperscript{3} Traditional aggregation theory is concerned with the question, when can a function \( F(x_{11}, ..., x_{1n}, ..., x_{mk}) \) be written \( f(X(x_{11}, ..., x_{1n}, ..., X_{m1}(x_{m1}, ..., x_{m\k})) \), where \( X_i \) for \( i = 1, ..., m \) are the aggregates? In our problem, we seek conditions enabling us to write the function as \( f(X(x_{11}, ..., x_{1n}, ..., X_{m2}(x_{m1}, ..., x_{m\k})) \), where \( X_i \) for \( i = 1, ..., k < \sum n_i \) are aggregates additive in the original variables.

\textsuperscript{4} Clearly, for the problem to make sense, we must also assume \( U'(0) > 0 \). However, we do not forestall the possibility of saturation, \( U'(\bar{W}) = 0 \) for \( \bar{W} < \infty \), which would be the case, for example, with the quadratic utility function often postulated in portfolio theory.
tions for such expected utility maximization, in addition to (1), can then be written\(^4\)

\[
EU'(W_0) \rho_{i\theta} = \lambda, \quad i = 1, \ldots, n,
\]

(3)

where \(\lambda\) is the marginal utility of initial wealth.

We will occasionally find it useful to refer to the proportion of initial wealth allocated to the \(i^{th}\) security

\[
a_i = \frac{Z_i}{W_0}
\]

(4)

rather than the amount of initial wealth allocated to the \(i^{th}\) security \(Z_i\). The budget constraint in terms of \(a_i\) becomes

\[
\sum_{i=1}^{n} a_i \leq 1.
\]

(1')

Let a mutual fund be a security composed of any fixed linear combination, with weights adding to one, of the original securities available to the investor, so that the returns on the mutual fund, say \(\mu_\theta\), are of the form

\[
\mu_\theta = \sum_{i=1}^{n} \delta_i \rho_{i\theta} \quad \text{with} \quad \sum_{i=1}^{n} \delta_i = 1.
\]

(5)

Then, we shall say that a class of utility functions\(^6\) exhibits the generalized separability property if and only if for any arbitrary set of \(n\) original securities there are \(m < n\) mutual funds

\[
\mu_{j\theta} = \sum_{i=1}^{n} \delta_{i\theta} \rho_{i\theta} , \quad j = 1, \ldots, m
\]

(6)

\(^4\) Under two further assumptions about the rules of the securities market and the structure of perceived returns, namely, that (i) short sales are permitted and (ii) the returns from any security cannot be dominated by the returns from a linear combination of the other securities using weights (possibly negative) adding to one. Given the first assumption the second is necessary for the portfolio allocation problem to be well-defined; for example, if the investor believes some security to be dominated by a second, and can sell the first short, then he will also believe himself capable of earning an arbitrarily large return on his portfolio in every state or contingency. Though permitting short sales clearly has substantive implications for portfolio behavior, for our purpose the assumption is basically harmless; the necessary modifications of our results were short sales not permitted are straightforward.

\(^6\) An example of a "class of utility functions" would be the concave, quadratic functions \(U(W) = \alpha + \beta W + \gamma W^2\) with \(\beta > 0, \gamma < 0\).
such that the optimal allocation of an individual with a utility function in that class can be represented simply by some allocation among the mutual funds

$$a_i = \sum_{j=1}^{m} a_j \delta_{ij} \quad \text{with} \quad \sum_{i=1}^{n} a_i = 1, \quad i = 1, \ldots, n. \quad (7)$$

An equivalent statement of this property is that we can find $m$ mutual funds such that for an individual with a utility function in the class and at any level of wealth, his maximum expected utility when faced with the opportunities afforded by the mutual funds will be the same as that achieved when faced with the opportunities afforded by the $n$ original securities.\(^6\) Notice that in general the weights defining the mutual funds (and \emph{a fortiori}, the number of mutual funds) will depend on both the class of utility functions and the returns associated with the original securities.

A geometric interpretation of generalized separability is the following: Suppose there is no saturation, so that we need only be concerned with equality in (1). (The general case is similar.) Then the set of feasible investments, in terms of the proportions of initial wealth invested in each security $a_j$, is just an $n - 1$ dimensional hyperplane including the unit simplex (which would be the set of feasible investments if short sales were ruled out). Generalized separability means that the set of optimal investments (as the level of initial wealth varies), again in terms of $a_j$, is an $n - 2$ (or smaller) dimensional hyperplane which is also included in the feasible set. The possibilities for the case $n = 3$ are illustrated in Figure 1.

There the feasible set is the plane $P$, while for a utility function from a class exhibiting generalized separability the optimal set might be a line in the plane, say $AB$, or a point in the plane, say $C$. In the first case, if two mutual funds are formed with proportions of the three assets given by any pair of points on $AB$, then the investor can achieve his optimal allocation by buying just these two mutual funds. Similarly, in the second case, the investor need only have available a mutual fund with proportions of the assets given by $C$.

For convenient exposition we will refer to the situation where a class of utility functions exhibits the generalized separability property by saying that \emph{generalized separation} obtains. When the number of mutual funds is two, $m = 2$, we shall say that the class of utility functions exhibits the \emph{separability property} or that there is \emph{separation}. Most of our analysis will concentrate on this more restrictive property and situation.

Both the properties of generalized separability and separability are

\(^6\) Thus $a_j$ in (7) represents the optimal proportion of the portfolio invested in the $j^{th}$ mutual fund.
statements about utility functions independent of the structure of returns on available assets. We will also be concerned with the special situation where the structure of returns is such that either there is money, i.e., an asset yielding the same return in every state of nature, or it is possible to form a mutual fund which is money. In this situation, if (i) separation
obtains and (ii) one of the two mutual funds formed may always be money, then we shall say *monetary separation* obtains.

The distinction between separation and monetary separation is illustrated in Figure 2. Suppose there are three states of nature, and also that there are just three securities. (As we shall show explicitly in Appendix I, for this example to make sense when short sales are permitted requires, among other things, the possibility of forming a mutual fund which is money.) Then the set of feasible investments, now in terms of the return to initial wealth in each state $W_0/W_0 = \sum_{i=1}^{n} a_{i}r_{i}$, is given by the plane $P$ passing through the three points $(r_{1}, r_{2}, r_{3})$ corresponding to the returns to initial wealth from plunging in each of the three securities available. If separation obtains, then the investor will choose an investment along a straight line in this plane, say $AB$, while if monetary separation obtains, this straight line will pass through the point where

$$W_j/W_0 = W_i/W_0 = W_0/W_0 = r,$$

say $A'B'$.

3. Separation with Arrow-Debreu Securities

We begin the analysis of the classes of utility functions exhibiting separability by considering a very special structure of returns from available assets. The rationale for this approach is the following: It should be obvious that the utility functions which, for short, work must do so when the structure of returns is specialized if they are to do so in general. Moreover, if by restricting the structure of returns we find that only a relatively few functions work, then it’s a simple matter to check which of them also work without restricting the structure of returns. But this is in fact what happens if we consider initially only securities which provide claims against each contingency (where contingencies are represented by states of nature).

To be more precise about what we mean by “provide claims against each contingency”: Suppose that there are $n$ states of nature as well as $n$ securities, and also that the return (again, per dollar invested) from security $i$ is $r_{i}\theta = r_{i}$ if state $\theta = i$ occurs but $r_{i}\theta = 0$ otherwise. Then the investor’s wealth in state $i$ depends only on his investment in security $i$, or to put it another way, the investor can allocate his portfolio to provide against each contingency by itself (subject, of course, to his overall budget constraint). We shall refer to such securities as *Arrow-Debreu* securities. (Under suitable restrictions, set out in Appendix I, when there are as many securities as states of nature, corresponding to any given set of securities it is possible to find an equivalent set of Arrow-Debreu securities.)
The important implication of this special structure of returns is that the first order conditions for expected utility maximization take on the special form

\[ U'(W_i) \pi_i \rho_i = \lambda, \quad i = 1, \ldots, n \]  

(3')

where now \( W_i = Z \rho_i = \) wealth next period if state \( i \) occurs, while \( \pi_i = \) probability that state \( i \) will occur (and therefore satisfies \( \pi_i > 0 \) and \( \sum_{i=1}^{n} \pi_i = 1 \)). Analyzing the condition (3') under the assumption that separation obtains irrespective of the values of the parameters \( \rho_i \) and \( \pi_i \) results in the following fundamental theorem:

**Theorem 3.1.** Given Arrow-Debreu securities, a necessary and sufficient condition for separation is that marginal utility satisfy

\[ AU'(W) + BU'(W)^\alpha = W \]  

(8)

or

\[ U'(W)^\alpha (A + B \log U'(W)) = W. \]  

(9)

**Remark 1.** These functions include as special cases some which have been extensively in previous discussions of behavior under uncertainty, and to which we shall subsequently pay special attention:

(i) If, in (8), \( \alpha = 0 \) then we obtain an equation for marginal utility of the form

\[ U'(W) = (a + bW)^\varepsilon, \]  

(10)

where \( b = 1/B, \ a = -bA, \ c = 1/\beta. \) Two special subcases of (10) are well-known, that for which also \( A = a = 0 \) (or equivalently, in (9), \( B = 0 \)),

\[ U'(W) = bW^\varepsilon, \]  

the constant relative risk aversion utility function,\(^7\) (10') and that for which also \( \beta = c = 1, \)

\[ U'(W) = a + bW, \]  

the quadratic utility function. (10")

(ii) If, in (9), \( \alpha = 0 \) then we obtain an equation for marginal utility of the form

\[ U'(W) = ae^{bW}, \]  

the constant absolute risk aversion function\(^7\). (11)

where \( b = 1/B, \ a = e^{-b}. \)

\(^7\) (10') is so called because the Arrow-Pratt [1, 4] measure of relative risk aversion, \(-U''(W)W/U'(W),\) is constant (\(= -c\)), (11) because their measure of absolute risk aversion, \(-U''(W)/U'(W),\) is constant (\(= -b\)).
Remark 2. In general, neither (8) nor (9) can be manipulated so as to derive a closed form for the utility function \( U(W) \). However, for the special cases mentioned above, as well as for another class of special cases of (8) (namely, those for which \( \alpha \) and \( \beta \), and \( A \) and \( B \) are complex conjugates with zero real parts; see footnote 1 in Appendix II), such a derivation is possible. As the details of these calculations are straightforward, we omit them. It is worth mentioning too, that even when (8) or (9) cannot be inverted to solve explicitly for \( U' \) in terms of \( W \) (the critical step in deriving a closed form), it may be possible to obtain an implicit form for \( U \) by inverting and integrating each term in (8) or (9)

\[
U(W) = \max_{W_1, W_2 \geq 0} A' W_1^{\alpha'} + B' W_2^{\beta'} + C'
\]

or

\[
U(W) = \max_{W_1, W_2 \geq 0} A' W_1^{\alpha'} + B' \int_{c'} f(z) \, dz,
\]

where \( f \) is the inverse of the function \( z = x^a \log x \).

Remark 3. Clearly, not every function which satisfies (8) or (9) yields a concave utility function. For completeness we list the restrictions on the parameters, domains or ranges of the functions (8)-(11) which imply \( U'(W) > 0 \) and \( U''(W) < 0 \) for \( 0 \leq W < W < W < \infty \):

<table>
<thead>
<tr>
<th>if ( U' \equiv x ) satisfies</th>
<th>then for ( U' &gt; 0 )</th>
<th>while, in addition, for ( U'' &lt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Ax^a + Bx^b = W )</td>
<td>( x &gt; 0 ) and ( Ax^a + Bx^b &gt; 0 )</td>
<td>( \alpha Ax^a + \beta Bx^b &lt; 0 )</td>
</tr>
<tr>
<td>( x^a(A + B \log x) = W )</td>
<td>( B &gt; 0 ) and ( x &gt; e^{-A/B} )</td>
<td>( \alpha &lt; 0 ) and ( x &gt; e^{-(A/B+1/a)} )</td>
</tr>
<tr>
<td></td>
<td>( B &lt; 0 ) and ( 0 &lt; x &lt; e^{-A/B} )</td>
<td>( \alpha &lt; 0 ) or ( \alpha &gt; 0 ) and ( x &lt; e^{-(A/B+1/a)} )</td>
</tr>
<tr>
<td>( x = (a + bW)^c )</td>
<td>( b &gt; 0 ) and ( W &gt; \max[0, -(a/b)] )</td>
<td>( c &lt; 0 )</td>
</tr>
<tr>
<td>( x = aebW )</td>
<td>( a &gt; 0, b &lt; 0 ) and ( 0 &lt; W &lt; -(a/b) )</td>
<td>( c &gt; 0 )</td>
</tr>
</tbody>
</table>

\( a > 0 \) and \( 0 \leq W \), \( b < 0 \)

* The value of the solution to a problem of the form \( \max g_1(W_1) + g_2(W_2) \), subject to \( W_1 + W_2 = W \), \( W_1, W_2 \geq 0 \), say, \( U(W) \) (e.g., (12)), may be described by the single equation \( \int_{c_1}^{c_2} [U(W)] + \int_{c_1}^{c_2} [U(W)] = W \) (e.g., (8)) derived from the first order conditions \( W_1 + W_2 = W \)

\( \int_{c_1}^{c_2} [U(W)] < U'(W) \) w/equality for \( W > 0 \)

when, for example, the functions \( g_1 \) are increasing, concave and satisfy \( \lim_{W \to \infty} g_1(W) = \infty \). The conditions under which (8) and (12) or (9) and (13) are equivalent follow directly from this sort of consideration.
Proof of 3.1. The proof will be broken into several steps. For expository
convenience the discussion will focus on the case \( n = 3 \); the argu-
ments are, however, perfectly general.

1. The separability property is equivalent to the demand functions
for securities having the form

\[
Z_i = A_i W_0 + B_i h(W_0), \quad i = 1, 2, 3
\]  

(14)

with the coefficients \( A_i \) and \( B_i \) satisfying \( \sum_{i=1}^3 A_i = 1, \sum_{i=1}^3 B_i = 0. \)

This follows immediately from the definition of separability (as (14) is
basically just a parametric representation of a line in the plane described
by \( \sum_{i=1}^3 a_i = 1 \)). To see this, observe that the fact that there are
two mutual funds such that an individual's demand for securities can be
written as linear combinations of the mutual funds (independent of
initial wealth) means just that

\[
Z_i = [a'(W_0) \delta_{i1} + (1 - a'(W_0)) \delta_{i2}] W_0
\]

\[
= \delta_{i2} W_0 + (\delta_{i1} + \delta_{i2}) a'(W_0) W_0
\]

where \( a'(W_0) \) is the proportion of the portfolio allocated to the first
mutual fund (cf. equations (6) and (7)).

2. A necessary condition for the demand functions to have the
form (14) is that \( U'^{-1} = G \neq 0 \) satisfy the functional equation

\[
G(x) = f(x) \tilde{f}(y) + g(x) \tilde{g}(y)
\]  

(15)

for some \( f, \tilde{f}, g \) and \( \tilde{g} \) such that (a) \( G(x), \tilde{f}(x) \) and \( \tilde{g}(x) \) are defined and
differentiable for \( 0 \leq x = \max[0, U'(\infty)] < x < U'(0) = \bar{x} < \infty \), and
(b) \( f(x) \) and \( g(x) \) are defined and differentiable for \( x > 0 \).

Given a particular utility function exhibiting separability, consider the
situation where there are three Arrow-Debreu securities yielding returns \( \bar{p}_i \)
with probability \( \bar{p}_i \). For short, let us refer to such a configuration of
returns and probabilities as an Arrow-Debreu market (or in general, any
configuration of returns and probabilities as a market). For this particular
market (1) and (3') can be rewritten (for levels of initial wealth precluding
possible saturation)

\[
\sum_{i=1}^3 Z_i = \sum_{i=1}^3 \frac{W_i}{\bar{p}_i} = W_0.
\]

\[
U'(W_i) \bar{p}_i = \lambda, \quad i = 1, 2, 3
\]  

(16)

Clearly, the representation of separability by (14) is valid irrespective of the structure
of returns, e.g., of whether or not the securities are Arrow-Debreu.
The risk aversion assumption \( U' < 0 \) implies first, that we can think of these four equations as defining \( \lambda \) as a decreasing function of \( W_0 \), or more appropriately for our purpose, \( W_0 \) as a decreasing function of \( \lambda \)

\[
W_0 = \varphi(\lambda) \quad \text{with} \quad \varphi' < 0, \tag{17}
\]

and second, that we can invert each of the last three equations to obtain another form for the demand functions

\[
Z_i = \frac{1}{\hat{\rho}_i} \left( -\frac{\lambda}{\hat{\pi}_i \hat{\rho}_i} \right), \quad i = 1, 2, 3. \tag{18}
\]

Equating (14) and (18), and utilizing (17), we finally get

\[
U^{-1} \left( \frac{\lambda}{\hat{\pi}_i \hat{\rho}_i} \right) = A_i \hat{\rho}_i \varphi(\lambda) + B_i \hat{\rho}_i h(\varphi(\lambda)) \\
= \hat{A}_i \hat{F}(\lambda) + \hat{B}_i \hat{G}(\lambda), \quad i = 1, 2, 3. \tag{19}
\]

The equations (19) are very close to yielding the functional equation (15) with \( x = (1/\hat{\pi}_i \hat{\rho}_i) \) and \( y = \lambda \). The only remaining hurdle is that we haven’t yet established that, for the given utility function, \( \hat{F} \) and \( \hat{G} \) are independent of the particular Arrow-Debreu market while \( \hat{A}_i \) and \( \hat{B}_i \) depend only on \( \hat{\pi}_i \hat{\rho}_i \). That this is so can be easily demonstrated by considering another Arrow-Debreu market, denoted by double tildes, for which

\[
\tilde{\pi}_i \tilde{\rho}_i = \hat{\pi}_i \hat{\rho}_i, \quad i = 1, 2 \\
\neq \hat{\pi}_i \hat{\rho}_i, \quad i = 3. \tag{20}
\]

For this market the argument culminating in (19) can be repeated to obtain

\[
U^{-1} \left( \frac{\lambda}{\tilde{\pi}_i \tilde{\rho}_i} \right) = \tilde{A}_i \tilde{F}(\lambda) + \tilde{B}_i \tilde{G}(\lambda), \quad i = 1, 2, 3. \tag{21}
\]

Now, from the assumption (20) it follows from the first pair of equations in both (19) and (21) that

\[
U^{-1} \left( \frac{\lambda}{\hat{\pi}_i \hat{\rho}_i} \right) = \hat{F}(\lambda) + \hat{B}_i \hat{G}(\lambda) = \tilde{A}_i \hat{F}(\lambda) + \tilde{B}_i \tilde{G}(\lambda), \quad i = 1, 2 \tag{22}
\]

or, assuming that \( \hat{A}_i \hat{B}_i - \tilde{A}_i \tilde{B}_i \neq 0 \), that we can solve for \( \hat{F} \) and \( \hat{G} \) in terms of \( \hat{F} \) and \( \hat{G} \), say,

\[
\hat{F}(\lambda) = C_F \hat{F}(\lambda) + C_G \hat{G}(\lambda) \\
\hat{G}(\lambda) = C_F \hat{F}(\lambda) + C_G \hat{G}(\lambda). \tag{23}
\]

\(^{10}\) In fact, if we had assumed a continuum rather than a finite number of Arrow-Debreu securities, the equations analogous to (19) would yield (15) directly.
Hence, we can certainly write the third equation in (21) in terms of $\tilde{F}$ and $\tilde{G}$ too

$$U'^{-1}\left(\lambda \frac{\tilde{\pi}_2\tilde{\rho}_3}{\tilde{\pi}_2\tilde{\rho}_3}\right) = A_3\tilde{F}(\lambda) + B_3\tilde{G}(\lambda)$$

with

$$A_3 = C_1\tilde{A}_3 + C_3\tilde{A}_3, \quad B_3 = C_2\tilde{B}_3 + C_4\tilde{B}_3.$$  \hspace{1cm} (24)

But (24) means that, provided $\tilde{A}_3\tilde{B}_3 - \tilde{A}_4\tilde{B}_4 \neq 0$ for every market described by (20), we have established that $U'^{-1}$ must satisfy (15). Moreover, as the original parameters $\tilde{\rho}_i$ and $\tilde{\pi}_i$ are basically arbitrary positive numbers, if the proviso were not true for any configuration of these parameters, then $U'$ would have to have the form (10'), and $U'^{-1}$ would still have to satisfy (15).

We demonstrate only the first part of the latter assertion here; the second part is established as a special case of a more general argument in Appendix II (case (i) second paragraph). Suppose for any choice of parameters (for simplicity we suppress the tildes) $\rho_i$ and $\pi_i$, $i = 1, 2$, satisfying $m < \rho_2/\rho_1 < M$ with $0 < m < 1 < M < \infty$ there is another pair of parameters such that in the demand functions (14) the coefficients have the property $A_3B_3 - A_4B_4 = 0$. This would mean that the incomes from securities 1 and 2 are necessarily in the same proportion $C$ irrespective of initial wealth $W_0$, or in terms of the first order conditions (3') that

$$U'(W_1)\pi_1\rho_1 = U'(CW_1)\pi_2\rho_2,$$  \hspace{1cm} (25)

where $C$ is clearly an increasing function of just $\pi_2\rho_2/\pi_1\rho_1$ over some range $n \leq C \leq N$ with $0 < n < 1 < N < \infty$ (the range, however, depending in general on all the parameters $\rho_i$ and $\pi_i$ as well as the form of $U'$). But equation (25) implies that

$$\frac{U'(W_1)W_1}{U'(W_1)} = \frac{U'(CW_1)CW_1}{U'(CW_1)},$$

which can be true for $n \leq C \leq N$ if and only if $U'$ has constant elasticity, i.e., has the form (10').

That the functions in (15) are differentiable follows immediately from the derivation of (24) given the assumption that $U$ is twice differentiable, while the intervals over which these functions are defined are simply determined by the first order conditions (16) in conjunction with the postulated structure of Arrow-Debreu markets.
3. In Appendix II it is shown that the only solutions to the equation (15) are

\[ f(x) = \frac{f(x)}{A} = x^a, \quad g(x) = \frac{g(x)}{B} = x^b \quad \text{and} \quad G(x) = Ax^a + Bx^b \quad (27) \]

and

\[ \bar{f}(x) = \bar{g}(x) = x^a \left( \frac{A}{2} + B \log x \right), \]

\[ \bar{f}(x) = \bar{g}(x) = x^a \quad \text{and} \quad G(x) = x^a(A + B \log x). \quad (28) \]

(27) and (28) yield just the asserted conditions on \( U' \) (8)–(9). We have thus established necessity. As the step remaining to prove sufficiency (i.e., the demonstration that (8)–(9) along with (3') imply (14)) is obvious, we leave it to the interested reader.

4. **Monetary Separation in Arrow-Debreu Markets**

In this section we investigate, still in the context of Arrow-Debreu markets, what further restrictions are imposed on the form of the utility function by the requirement that money can always be chosen as one of the mutual funds.\(^{11}\) The argument is pretty much self-contained,\(^ {12}\) and establishes

**Theorem 4.1.** Given Arrow-Debreu securities, a necessary and sufficient condition for monetary separation is that marginal utility have the form (10) or (11).

**Remark.** This result for Arrow-Debreu markets already tells us that the restrictions imposed by monetary separation are very severe. Indeed, the classes of utility functions (10) and (11) are precisely those which are already known to be sufficient for monetary separation in general markets with money (see, for instance, Hakansson [3]).

**Proof of 4.1.** (necessity) Consider any Arrow-Debreu market with three securities and assume that the utility function implies monetary

\(^{11}\) Note that it follows from the discussion in Appendix I that such an investigation really amounts to analyzing monetary separation in a market in which the available securities cover all contingencies and short sales are permitted (so that a mutual fund which is money is always possible).

\(^{12}\) In particular, the reader only interested in monetary separation may find all our results that are relevant to this problem either here or in the first part of section 6.
separation, i.e., separation such that one of the two mutual funds can be taken as the security with a return \( r = \frac{1}{\sum_{i=1}^{3} (1/\rho_i)} \) in every state (cf. equation I.6 in Appendix I). Now the other mutual fund is simply some linear combination (with weights adding to one) of the three Arrow-Debreu securities, say, \( \delta_1 \) of the first, \( \delta_2 \) of the second and \( \delta_3 = 1 - \delta_1 - \delta_2 \) of the third (e.g., the point \( B^* \) in Figure 2). Let the investment in money be \( x \), so that investment in the risky mutual fund is \( W_0 - x \). Then the investor's optimal allocation can be represented by

\[
\begin{align*}
W_1 &= \rho_1 Z_1 = rx + \delta_1 \rho_1 (W_0 - x) \\
W_2 &= \rho_2 Z_2 = rx + \delta_2 \rho_2 (W_0 - x) \\
W_3 &= \rho_3 Z_3 = rx + \delta_3 \rho_3 (W_0 - x).
\end{align*}
\]

(29)

By solving for \( W_1 - W_3 \) and \( W_2 - W_3 \) in terms of \( W_0 - x \), (29) yields a necessary condition for monetary separation

\[
\frac{W_1 - W_3}{W_2 - W_3} = \frac{\delta_1 \rho_1 - \delta_3 \rho_3}{\delta_2 \rho_2 - \delta_3 \rho_3}
\]

or (differentiating and simplifying)

\[
\frac{dW_1}{dW_0} (W_2 - W_3) + \frac{dW_2}{dW_0} (W_3 - W_1) + \frac{dW_3}{dW_0} (W_1 - W_2) = 0. \quad (30)
\]

But from the conditions describing the optimal allocation in this market ((16) without tildes) we have additional information about \( dW_i/dW_0 \), namely,

\[
\sum_{i=1}^{3} \frac{dW_i}{dW_0} \frac{1}{\rho_i} = 1
\]

\[
U'(W_i) \pi_i, \quad \frac{dW_i}{dW_0} = \frac{d\lambda}{dW_0}, \quad i = 1, 2, 3
\]

or (solving explicitly)

\[
\frac{dW_i}{dW_0} = \frac{V_i}{\sum_{i=1}^{3} V_i \rho_i}.
\]

(31)

where \( V_i = -U''(W_i)/U''(W_i) \), the reciprocal of Pratt's absolute risk aversion function [4]. (Notice, in passing, that (31) entails that Arrow-Debreu securities are superior goods for the risk-averse investor.) Thus,
substitution from (31) into (30) yields another necessary condition for monetary separation

\[ V_1(W_2 - W_3) + V_2(W_3 - W_4) + V_3(W_4 - W_5) = 0. \]  \( (32) \)

We now observe, again appealing to the condition (16), that in fact (32) must be identically true for \( \max[U'(\infty), 0] < U'(W_2) < U'(0) \) and given \( W_2, W_3, W_4 \), that is, must be true if we vary \( \rho_2 \) and \( W_6 \) so as to keep \( \lambda \) constant (an argument similar to that we used earlier in the derivation of the functional equation (15)). Hence, we can write (dropping the subscript 1)

\[ - \frac{U'(W)}{U''(W)} = \alpha + \beta W \]

or

\[ \frac{d \log U'(W)}{dW} = \frac{1}{\alpha + \beta W}. \]  \( (33) \)

The closed solution to (33) is just (10) if \( \beta \neq 0 \), (11) if \( \beta = 0 \).

(sufficiency) That (10) and (11) imply monetary separation for Arrow-Debreu markets follows directly from solving the conditions (16) (where 3 is replaced by \( n \geq 3 \)) after introducing these forms for \( U' \):

\[ U'(W) = (a + bW)^c \]  \( (i) \)

In this case the conditions (16) become

\[ \sum_{i=1}^{n} \frac{W_i}{\rho_i} = W_6. \]

\[ (a + bW)^c \pi_i \rho_i = \lambda, \quad i = 1, \ldots, n \]  \( (34) \)

Solving each of the last \( n \) equations in (34) for \( W_i \), substituting these results into the first and solving for \( \lambda \) and, finally, utilizing this result to express \( W_i \) solely in terms of initial wealth \( W_6 \) and the parameters \( a, b, c, \rho \), and \( \pi \), yields

\[ W_i = rx + \delta_i \rho_i (W_6 - x), \quad i = 1, \ldots, n \]  \( (35) \)

where

\[ r = \frac{1}{\sum_{i=1}^{n} \frac{1}{\rho_i}}, \quad x = -\frac{a}{\beta} \sum_{i=1}^{n} \frac{1}{\rho_i} \text{ and } \delta_i = \frac{(\pi_i \rho_i)^{1/c}}{\rho_i} \left/ \sum_{i=1}^{n} (\pi_i \rho_i)^{1/c} \right.. \]
(35) simply says that the optimal allocation consists of investing

\[ x = -\frac{a}{b} \sum_{i=1}^{n} \frac{1}{\rho_i} \]

in money and \( W_0 - x \) in a mutual fund combining

\[ \delta_i = \frac{(\pi_i \rho_i)^{-1/e}}{\rho_i} \left/ \sum_{i=1}^{n} \frac{(\pi_i \rho_i)^{-1/e}}{\rho_i} \right. \]

of each Arrow-Debreu security.

\[ U'(W) = ae^{bW} \tag{ii} \]

Precisely the same algebraic manipulations as in (i) now yield

\[ W_i = rx' + \delta_i \rho_i (W_0 - x'), \quad i = 1, \ldots, n \tag{35'} \]

where

\[ r = 1 \frac{1}{\sum_{i=1}^{n} \rho_i}, \quad x' = \frac{1}{b} \sum_{i=1}^{n} \frac{\log a \pi_i \rho_i}{\rho_i} + W_0 \]

and

\[ \delta_i = \frac{\log a \pi_i \rho_i}{\rho_i} \left/ \sum_{i=1}^{n} \frac{\log a \pi_i \rho_i}{\rho_i} \right. \]

Again, (35') has the straightforward interpretation of optimal allocation being accomplished by purchasing only money and a single risky mutual fund.

One interesting aspect of (35) and (35') is worth mentioning explicitly, namely, that with mutual funds formed in the way indicated (clearly there are many other ways to represent monetary separation than by these particular choices) when marginal utility has the form (10), investment in money is independent of initial wealth, while when it has the form (11), investment in the risky mutual fund is independent of initial wealth.

5. Separation in General Markets without Money

We now turn to the question of when separation will occur in markets in which securities are not available to cover all contingencies, i.e., the number of securities is less than the number of states, and there is no money, i.e., it is not possible to form a mutual fund with the same return in every state. Our method is again to consider a subset of all such markets, and to show that for this subset only certain classes of utility functions will
work. It is then straightforward to verify that the same classes will also work in general markets without money.\textsuperscript{13}

The particular subset of markets which is useful for our purpose here consists of those markets for which the first \( n - 1 \) are Arrow-Debreu securities, the last a security that has positive returns in the remaining \( \nu - n + 1 \) states (letting \( \nu \) be the number of states, and, as before, \( n \) the number of securities), so that the pattern of returns \( R \) has the structure

\[
R = \begin{bmatrix}
\rho_0 \\
0 & \rho_1 & 0 & 0 & \cdots & 0 \\
\vdots & & & & & \vdots \\
0 & 0 & \cdots & \rho_{n,n} & \rho_{n,n+1} & \cdots & \rho_{n,\nu}
\end{bmatrix}
\]

with \( 2 < n < \nu, \rho_{n,\theta} \neq \rho_{n,\theta'} \) for some \( \theta \neq \theta' \). \hspace{1cm} (36)

Notice especially that neither a mutual fund which dominates any original security nor a mutual fund which is money are possible in the markets described by (36). Thus by considering optimal allocation with short sales in such markets we can demonstrate that

\textbf{Theorem 5.1.} A necessary and sufficient condition for separation in general markets is that the utility function be quadratic (10') or constant relative risk aversion (10").

\textit{Proof of 5.1.} (necessity) From our discussion of separation in Arrow-Debreu markets, it follows that we need only consider the classes of utility functions represented by (8) and (9). It turns out useful to consider first, the subcases (10) and (11), and second, the general cases (excluding the subcases) (8) and (9). Our procedure will be to check and see whether, for each of these classes, separation obtains in the special markets (36). The case of such markets with \( n = 3, \nu = 4 \) is sufficiently general for our argument.

\[
U'(W) = (a + bW)^{\theta}
\]

For this particular class of functions the conditions (3) may be written

\[
(a + bZ_{\theta})^{\theta} \pi_\theta = \lambda, \quad i = 1, 2
\]

\[
\sum_{\theta=0}^{4} (a + bZ_{\theta})^{\theta} \pi_\theta = \lambda
\]

\textsuperscript{13} It is well-known that the utility functions which we find will entail separation in general markets will, when money is available, entail monetary separation. In the next section we probe the deeper question of whether there are additional utility functions (from the list of possible candidates (8) and (9)) which will entail either separation or monetary separation when money is available, i.e., the role of money per se in providing separation.
or, assuming separation and substituting from the demand functions in the form (14),

\[ (a + bA_i\rho_i W_0 + bB_i\rho_i h(W_\theta))^\epsilon \pi_i \rho_i = \lambda, \quad i = 1, 2. \] (37)

\[ \sum_{\theta = 3}^4 (a + bA_{i\theta}\rho_{i\theta} W_0 + bB_{i\theta} \rho_{i\theta} h(W_\theta))^\epsilon \pi_{i\theta} \rho_{i\theta} = \lambda \]

If \( h(W_\theta) \) is not identically constant, then from differentiating the first pair of equations in (37) twice with respect to \( W_\theta \) it follows that it must be the case that

\[ \pi_{1\theta} = \pi_{2\theta}. \]

Hence, without loss of generality we can assume \( h(W_\theta) \) is identically constant. But if, say, \( h(W_\theta) = K \), then from solving either of the first pair of equations in (37) it follows that

\[ \lambda = (\gamma + \delta W_\theta)^\epsilon \] (38)

where \( \gamma = (a + bB_i\rho_i K)(\pi_i \rho_i)^{1/c} \) and \( \delta = bA_i\rho_i(\pi_i \rho_i)^{1/c}, \quad i = 1, 2. \) Moreover, when \( h(W_\theta) = K \), the third equation in (37) can be written

\[ \sum_{\theta = 3}^4 (a + bB_{i\theta} \rho_{i\theta} K + bA_{i\theta} \rho_{i\theta} W_0)^\epsilon \pi_{i\theta} \rho_{i\theta} = \lambda \]

or, letting \( \gamma_{i\theta} = (a + bB_{i\theta} \rho_{i\theta} K)(\pi_{i\theta} \rho_{i\theta})^{1/c} \) and \( \delta_{i\theta} = bA_{i\theta} \rho_{i\theta}(\pi_{i\theta} \rho_{i\theta})^{1/c} \), and substituting from (38)

\[ \sum_{\theta = 3}^4 (\gamma_{i\theta} + \delta_{i\theta} W_\theta)^\epsilon = (\gamma + \delta W_\theta)^\epsilon. \] (39)

Finally, upon defining \( x = (\gamma + \delta W_\theta)^{-1}, \quad \gamma' = \gamma - (\delta/\delta) \gamma \) and \( \delta' = (\delta/\delta) \), (39) simplifies to

\[ \sum_{\theta = 3}^4 (\delta' + \gamma' x)^\epsilon = 1, \] (40)

which yields, after differentiating twice with respect to \( x, \)

\[ c(c - 1) \sum_{\theta = 3}^4 (\delta' + \gamma' x)^{c-2} \gamma'^2 = 0. \]

Thus (recalling that \( U'' < 0 \) requires \( c \neq 0 \), and noting that the derivation
of (40) requires $\delta' + \gamma' x > 0$ either $c = 1$, in which case the utility function is quadratic (10'), or

$$\gamma' = 0, \quad \text{i.e.,} \quad \frac{\gamma}{\delta} = \frac{\gamma}{\delta'}, \quad \text{i.e.,} \quad \frac{a + bB_2 \rho_{33} K}{bA_2 \rho_{33}} = \frac{a + bB_2 \rho_{34} K}{bA_2 \rho_{34}}. \quad (41)$$

But (41) means either $a = 0$, in which case the utility function is the constant relative risk aversion function (10'), or $\rho_{33} = \rho_{34}$, contrary to our assumption about the structure of the market (36).

Hence, we have established that, given (36), separation with (10) requires either the special case (10') or the special case (10').

$$U'(W) = ae^{bW} \quad \text{(ii)}$$

Essentially the same argument as the preceding also rules out separation for this class of utility functions. In particular, a step by step repetition of that argument yields, instead of (40), a condition

$$\sum_{\theta=0}^{t} \delta' e^{x' \theta} = 1 \quad (42)$$

for appropriately chosen variable $x$ and parameters $\gamma'$ and $\delta'$. We omit the additional details of this argument.

$$AU'(W)^x + BU'(W)^y = W \quad (iii)$$

For the pattern of returns (36) with $n = 3, \nu = 4$, the first order conditions (3) may be written in general

$$U'(Z_i \rho_i) \pi_i \rho_i = \lambda, \quad i = 1, 2. \quad (43)$$

$$\sum_{\theta=0}^{t} U'(Z \rho_{3i}) \pi_0 \rho_{30} = \lambda$$

Assuming both that marginal utility has the form (8) and that the demand functions have the form (14) it can be easily seen from the first pair of equations in (43), that we can solve for $W_0$ and $h(W_0)$ in terms of $\lambda^e$ and $\lambda^d$,

$$W_0 = C_1 \lambda^e + C_2 \lambda^d,$$

$$h(W_0) = C_3 \lambda^e + C_4 \lambda^d$$

so that the third equation in (43) can be represented by

$$U'(W_3) \pi_3 \rho_{33} + U' \left( \frac{1}{\gamma} W_3 \right) \pi_3 \rho_{34} = \lambda \quad \text{with} \quad W_3 = \delta \lambda^e + \epsilon \lambda^d \quad (44)$$
where $\gamma = \rho_{x2}/\rho_{x1}$, $\delta = \rho_{x2}(A_2C_1 + B_2C_2)$ and $\epsilon = \rho_{x2}(A_2C_2 + B_2C_2)$. We shall show that equation (44) cannot hold in general for an arbitrary utility function in the class (8) (with $A \neq 0$, $B \neq 0$ and $\alpha \neq \beta \neq 0$, that is, except perhaps for the special cases (10), (10') and (10''), which have already been considered separately).

Let $u = U'(W_3)$, $v = U''(1/\gamma)W_3$. Then (8) and (44) together yield the system of equations in $u$, $v$ and $\lambda$

$$Au^\alpha + Bu^\beta = \delta \lambda^\alpha + \epsilon \lambda^\beta$$

$$Av^\alpha + Bv^\beta = \frac{1}{\gamma} [\delta \lambda^\alpha + \epsilon \lambda^\beta]$$

$$u\pi_x \rho_{x3} + v\pi_x \rho_{x4} = \lambda$$

which must be satisfied in some nonnegative region. From solving each of the first two equations in (45) for $u^{\beta-\alpha}$ and equating the results we obtain

$$\frac{B}{A} u^{\beta-\alpha} = \frac{1 - \gamma \left(\frac{v}{u}\right)^\alpha}{1 - \gamma \left(\frac{v}{u}\right)^\beta} = \frac{1 - \frac{\delta}{A} \left(\frac{\lambda}{u}\right)^\alpha}{1 - \frac{\epsilon}{B} \left(\frac{\lambda}{u}\right)^\beta}$$

while by redefining $\lambda$ appropriately the last equation in (45) can be rewritten

$$\mu + (1 - \mu) \frac{v}{u} = \frac{\lambda}{u}.$$  \hfill (47)

Now let $x = v/u$. Then, finally, substitution from (47) into (46) yields the single equation in $x$

$$\frac{1 - \gamma x^\alpha}{1 - \gamma x^\beta} = \frac{1 - \frac{\delta}{A} (\mu + (1 - \mu) x)^\alpha}{1 - \frac{\epsilon}{B} (\mu + (1 - \mu) x)^\beta}$$

which must be satisfied in some nonnegative interval (as by (46) $x$ cannot be constant unless $u$ is constant). But this means that for the analytic function $f$ defined by

$$f(x) = (1 - \gamma x^\alpha) \left[1 - \frac{\epsilon}{B} (\mu + (1 - \mu) x)^\beta\right]$$

$$- (1 - \gamma x^\beta) \left[1 - \frac{\delta}{A} (\mu + (1 - \mu) x)^\alpha\right] \text{ for } x > 0$$ \hfill (49)
it must be true that

$$f(x) = 0 \quad \text{for all} \quad x > 0. \quad (50)$$

And by considering particular values of $x$ (namely, $x = 1$, $x = (1/\gamma)^{1/\alpha}$ and $x = (1/\gamma)^{1/\beta}$) it is easy to show that (50) cannot be true in general.\(^{14}\)

Hence, we have demonstrated that (8) does not exhibit separability in general markets without money.

$$U'(W)(A + B \log U'(W)) = W \quad \text{(iv)}$$

Again, essentially a repetition of the preceding argument rules out separation for this class of utility functions, and is therefore omitted.

(sufficiency) The argument simply consists of writing down the budget constraint (1) and first order conditions (3) for each of these functions, and then solving the resulting $n + 1$ equations for $\lambda$ and $z$, to verify (14). This exercise will be carried out in detail in section 8, where we analyze the exact form of the demand functions given the various utility functions which exhibit separability.

6. Separation in General Markets with Money

In the last section we found that only quadratic or constant relative-risk-aversion utility functions exhibit separability in perfectly general securities markets. We now consider the question—whose importance stems from the fact that the availability of a riskless asset is a very commonly employed simplifying assumption—of what bearing the existence of money (either as an original security or as a mutual fund) has on separation. Such influence might be of two sorts: First, it might be that the existence of money allows additional classes of utility functions (obviously just from those described by (8) and (9)) to exhibit separability, irrespective of the possibility of monetary separation. Indeed, a careful reading of the argument for the necessity part of Theorem 5.1 reveals that for each case (i)-(iv) the assertion proved is of the form, for example, for (i), if the market has the special structure (36) without any restriction on $\rho_{ab}$, the utility function is of the class (10) with $a \neq 0$, $c \neq 0$ and separation obtains, then $\rho_{ab} = \rho_{ab}$, for all $a \leq \theta \leq b$. That is, the argument of the preceding section demonstrated that either the market (36) without restriction on the $n$th security had to reduce to one in which money is

\(^{14}\)Indeed, it will only be true if $\gamma = 1$ or $\rho_{ab} = \rho_{ab}$, which has been ruled out by assumption (36). We mention in passing that it is at this step that the argument fails for the cases $\alpha = 0$ or $\beta = 0$ and $\alpha = \beta$. 
before, all we will do is sketch the necessary changes in the earlier arguments.\textsuperscript{17}

\[ AU'(W)\delta + BU'(W)\theta = W \]  

(iii')

Now there is no simple relationship between \( W_2 \) and \( W_4 \) (whereas before \( W_3 = \gamma W_1 \)). This has the effect of requiring transformation of the system of equations (45) to the more general system of equations

\[ Au\alpha + Bu\theta = \delta u \lambda \alpha + \epsilon \omega \lambda \theta \]

\[ Au\nu + Bu\theta = \delta \nu \lambda \nu + \epsilon \nu \lambda \theta \]  

(45')

\[ \xi u + \eta v = \lambda \]

for appropriately defined variables and parameters, and the addition of a straightforward but tedious demonstration that in fact (45') must be equivalent to (45).\textsuperscript{18}

\[ U'(W)\nu (A + B \log U'(W)) = W \]  

(iv')

This is still fundamentally the same as case (iii').

We should emphasize, perhaps, that the attempt to elaborate the argument which ruled out separation given (10) or (11) in markets without money so that it encompasses markets with money fails; this is, of course, a central aspect of the theorem. It might be worthwhile noting further that such an attempt breaks down just at the last step, that is, for instance in the argument involving (10), in trying to go from the analogue of (41) to the contradiction \( \rho_{35} = \rho_{34} = \rho_{35} \).

7. Generalized Separability

Thus far we have been concerned exclusively with characterizing the utility functions which entail separation into just two mutual funds, and have found, roughly speaking, that very few functions exhibit such separability. The question arises naturally, then, as to how restrictive is the requirement that only two mutual funds be involved. The following

\textsuperscript{17} The smallest dimensionality of the market (36') which is sufficiently general for the present argument is clearly \( n = 4, \nu = 5 \).

\textsuperscript{18} The argument consists essentially of reducing (45') to an analogue of (49)

\[ f(x) = (\delta x - \epsilon x^2) - (\epsilon x - \epsilon x^2) = (\delta x - \epsilon x^2) \]  

(49')

and then showing that if (50) were true, it would also be true that \( W_5 = \gamma W_4 \).
Theorem concerning the possibilities for generalized separation\(^{19}\) in Arrow-Debreu markets strongly supports the conjecture that it is the requirement that there be any mutual funds, and not the limitation on the number of mutual funds which is the restrictive feature of the property of separability. The theorem also plainly suggests why we haven’t pursued the analysis of generalized separability (as opposed to separability) any further.

**Theorem 7.1.** Given Arrow-Debreu securities, a necessary and sufficient condition for generalized separation is that the inverse of marginal utility \(G = U'^{-1}\) may be represented by

\[
G(xy) = \sum_{i=1}^{m} \sum_{j=1}^{m} f_i(x) C_{ij} f_j(y)
\]

where

\[
f_i(x) = \sum_{k=1}^{K} x^{a_k} \sum_{l=1}^{L_k} D_{l} \log^{l-1} x \text{ with } 2 \leq m < n, 1 \leq L_k \text{ and } \sum_{k=1}^{K} L_k = m.
\]

**Remark.** What the theorem means is that even in Arrow-Debreu markets generalized separability turns out to be a property limited to a very, very few utility functions. Notice especially that implicit in (51) is the condition that the parameters \(C_{ij}\) and \(D_{l}\) imply \(G(xy) = \text{constant}\) when \(xy = \text{constant}\). Thus, for example, it is easily shown that for \(m = 3\), i.e., for all relevant market opportunities to be provided by at most three mutual funds, marginal utility must satisfy

\[
AU'(W)^x + BU'(W)^y + CU'(W)^z = W;
\]

\[
U'(W)^x (A + B \log U'(W)) + CU'(W)^z = W
\]

or

\[
U'(W)^x (A + B \log U'(W) + C \log^2 U'(W)) = W.
\]

**Proof of 7.1.** The proof is simply a generalization of the argument presented in section 3 and the appendix to the paper. As, except for the

\(^{19}\) That is, recalling our earlier definition, a situation where the utility function permits all relevant market opportunities to be provided by at most \(2 < m < n\) mutual funds.
Theorem concerning the possibilities for generalized separation\textsuperscript{19} in Arrow-Debreu markets strongly supports the conjecture that it is the requirement that there be any mutual funds, and not the limitation on the number of mutual funds which is the restrictive feature of the property of separability. The theorem also plainly suggests why we haven’t pursued the analysis of generalized separability (as opposed to separability) any further.

**Theorem 7.1.** Given Arrow-Debreu securities, a necessary and sufficient condition for generalized separation is that the inverse of marginal utility \( G \equiv U'^{-1} \) may be represented by

\[
G(xy) = \sum_{i=1}^{m} \sum_{j=1}^{m} f_i(x) C_{ij} f_j(y)
\]

where

\[
f_i(x) = \sum_{s=1}^{K} x_s^{L_s} \sum_{t=1}^{L_s} D_{stl} \log^{t-1} x \text{ with } 2 \leq m < n, 1 \leq L_s \text{ and } \sum_{s=1}^{K} L_s = m.
\]

**Remark.** What the theorem means is that even in Arrow-Debreu markets generalized separability turns out to be a property limited to a very, very few utility functions. Notice especially that implicit in (51) is the condition that the parameters \( C_{ij} \) and \( D_{stl} \) imply \( G(xy) = \) constant when \( xy = \) constant. Thus, for example, it is easily shown that for \( m = 3 \), i.e., for all relevant market opportunities to be provided by at most three mutual funds, marginal utility must satisfy

\[
AU'(W)^a + BU'(W)^b + CU'(W)^c = W,
\]

\[
U'(W)^a (A + B \log U'(W)) + CU'(W)^c = W
\]

or

\[
U'(W)^a (A + B \log U'(W)) + C \log^e U'(W)) = W.
\]

**Proof of 7.1.** The proof is simply a generalization of the argument presented in section 3 and the appendix to the paper. As, except for the

\textsuperscript{19} That is, recalling our earlier definition, a situation where the utility function permits all relevant market opportunities to be provided by at most \( 2 < m < n \) mutual funds.
last few steps, it is identical in spirit if not in detail—that is, there is nothing intrinsically interesting in the generalization not already contained in the argument previously given—we do not bother to include it.

8. Separation Formulas

The bulk of the foregoing argument has been devoted to demonstrating that only the very narrow classes of functions (10') and (10'') or (provided there is money) (10) and (11) exhibit separability with perfectly general securities markets. For completeness we now briefly list the precise formulas which describe the behavior of an investor having preferences represented by a utility function from one of these classes.

**Proposition 8.1.** Given (10') or (10''), or the existence of money and (10) or (11), the demand for each security is linear in initial wealth

\[ Z_i = A_i W_0 + B_i, \quad i = 1, \ldots, n. \]  

(52)

For (10') the parameters \( A_i \) and \( B_i \) are given by

\[ A_i = -\sum_{j=1}^{n-1} \sigma_{ij} E[\mu_i(\mu_j - \mu_{ij})], \quad i = 1, \ldots, n - 1, \]

\[ = 1 - \sum_{j=1}^{n-1} A_j, \quad i = n \]  

(53)

\[ B_i = -\frac{a}{b} \sum_{j=1}^{n-1} \sigma_{ij} E(\mu_i - \mu_{ij}), \quad i = 1, \ldots, n - 1 \]

\[ = -\sum_{j=1}^{n-1} B_j, \quad i = n \]

---

20 The structure of the general problem requires the analogue of rewriting the matrix \([\bar{E} \bar{E}]\) on the righthand side of II.15 in Jordan canonical form—rather than the analogue of differentiating both sides of II.15—and then solving the resulting system of differential equations.
where \( \sigma_{ij}^{-1} \) is the typical entry in the inverse of the matrix with elements \( E(\rho_{it} - \rho_{nt})(\rho_{it} - \rho_{nt}) \); for (10) by

\[
A_i, \ i = 1, \ldots, n, \text{ are the solution to }
\begin{align*}
E(\rho_{n0} + \sum_{i=1}^{n-1} A_i(\rho_{it} - \rho_{nt})^2 (\rho_{it} - \rho_{nt}) = 0, \ j = 1, \ldots, n - 1; \\
A_n = 1 - \sum_{i=1}^{n-1} A_i,
\end{align*}
\]

(54)

\[
B_i = 0, \ i = 1, \ldots, n
\]

for (10) by

\[
A_i = C_i b r, \quad i = 1, \ldots, n - 1
\]

\[
= 1 - \left( \sum_{i=1}^{n-1} C_i \right) b r, \quad i = n
\]

\[
B_i = C_i a, \quad i = 1, \ldots, n - 1
\]

\[
= - \left( \sum_{i=1}^{n-1} C_i \right) a, \quad i = n
\]

where

\[
C_i, \ i = 1, \ldots, n - 1 \text{ are the solution to }
\]

\[
E[1 + b] \left( \sum_{i=1}^{n-1} C_i(\rho_{it} - r)^2 (\rho_{it} - r) \right) = 0, \ j = 1, \ldots, n - 1; \quad (55)
\]

and for (11) by\(^{21}\)

\[
A_i = 0, \quad i = 1, \ldots, n - 1
\]

\[
= 1, \quad i = n
\]

\[
B_i, \ i = 1, \ldots, n, \text{ are the solution to }
\]

\[
\begin{align*}
E \exp \left( b \sum_{i=1}^{n-1} B_i(\rho_{it} - r) \right) (\rho_{it} - r) = 0, \ j = 1, \ldots, n - 1. \quad (56)
\end{align*}
\]

\(^{21}\) Note that the nth security is arbitrary in (53) and (54), but must be money (again with return denoted \( r \)) in (55) and (56). However, the nth security in all four can be taken as money when such a safe asset exists.
Proof of 8.1. The validity of these formulas is verified simply by direct substitution into the appropriate form of the first order conditions (3). Thus, for example, given a utility function from the class (10') the first order conditions are

$$E \left( a + b \sum_{i=1}^{n} Z_i \rho_{ia} \right) \rho_{ja} = \lambda, \quad j = 1, \ldots, n. \quad (57)$$

By utilizing the budget constraint (1) (to get rid of $Z_n$) and the last equation in (57) (to get rid of $\lambda$) the first $n - 1$ equations in (57) can be rewritten

$$E \left( a + b \rho_{ia} W_0 + b \sum_{i=1}^{n-1} Z_i (\rho_{ia} - \rho_{ja}) \right) (\rho_{ja} - \rho_{ia}) = 0, \quad j = 1, \ldots, n - 1. \quad (58)$$

It is easily seen that the $Z_i$ given by the definition (52) and (53) satisfy (1) and (58) and therefore also (1) and (3).

The other three cases are exactly parallel, and are left as exercises to the interested reader.

Several additional comments are worth making. Firstly, utilizing (52) and (53) it can easily be shown that for any quadratic utility function, the mutual funds can be chosen as those for which the variance of return and coefficient of variation are minimum, respectively.\(^{28}\) Secondly, from (52) and (54) it follows immediately that for any given constant relative risk aversion function only a single mutual fund is necessary, namely, the one in which securities are available in proportions $A_i$. Thirdly, (52) and (56) imply that for any given constant absolute risk aversion function, investment in any particular risky security is constant and, as a proportion of total investment in all risky securities, independent of the parameters of the utility function.

Finally, we note expressly that proposition 8.1 together with the general representation of separation (14) immediately entail

**Proposition 8.2.** (i) A necessary and sufficient condition for separation in general markets (markets with money) is that the demand function for any security be linear in initial wealth. (ii) A necessary and sufficient condition that the demand function for any security be proportional to initial wealth is that the utility function be constant relative risk aversion (10'), that the demand function for any security be linear in initial wealth is that the utility function be quadratic (10') (or satisfy (10) or (11)).

\(^{28}\) This is simply a reflection of the fact that the investor with a quadratic utility function is really only interested in the mean and variance of return on his portfolio; see the following section 9.
9. **Classes of Probability Distributions of Returns**

The investor with a quadratic utility function is in fact behaving as if he evaluates his portfolio in terms of mean and variance of return.\(^{20}\) This means that for this special case one can interpret separation as simply reflecting the fact that all efficient (i.e., minimum variance for given mean return) portfolios can be generated from a pair of mutual funds.

One can—as Tobin pointed out in his seminal piece—also arrive at a mean-variance analysis of portfolio allocation by assuming, rather than a quadratic utility function and an arbitrary pattern of returns, a normal distribution of returns and an arbitrary utility function. This naturally suggests the question, are there other classes of distributions which entail mean-variance analysis and therefore portfolio separation?

It is easily seen that this is not the case by considering the simplest example where \( n = 2 \) and \( \rho_{ij}^\alpha, \ i = 1, 2, \) are independent, identically distributed random variables. For the mean-variance analysis to be applicable here, it must be true that the distribution of all linear combinations of \( \rho_{1\alpha} \) and \( \rho_{2\alpha} \) differ only in scale and location parameters from the original distribution of \( \rho_{1\alpha} \) (or \( \rho_{2\alpha} \)). But the class of distributions for which this is true is just the stable or Pareto-Levy distributions, of which the only member with finite variance is the normal.

Moreover, it is straightforward to show that if (i) the returns from risky securities each have a stable distribution with the same characteristic exponent\(^{23}\) and (ii) there is money, then the efficient (i.e., now minimum dispersion given mean return) portfolios can be generated by money and a single mutual fund, while without money there is such separation only for the normal distribution.

Thus, even extending the notion of “risk” does not appreciably alter the conclusion that it is essentially only normal distributions of returns which entail portfolio separation.

10. **Summary**

In this paper we have established that the following sets of assertions are equivalent.

\(^{20}\) Such evaluation clearly depends on initial wealth when the investor is concerned with total (and not just rate of) return, as we have assumed all along.

\(^{23}\) Notice that the latter must be greater than one for the mean return to be finite, which itself has the undesirable implication (given limited liability institutions) that the return may be any number (rather than any nonnegative number).
In general:

A. (i) The utility function is constant relative risk aversion.
   (ii) All relevant market opportunities may be provided by a single mutual fund (i.e., a fixed bundle of securities).
   (iii) The demand for any security is proportional to initial wealth.

B. (i) The utility function is quadratic or constant relative risk aversion.
   (ii) All relevant market opportunities may be provided by a pair of mutual funds.
   (iii) The demand for any security is linear in initial wealth.

If there is money (i.e., an asset yielding the same return in every contingency):

C. (i) The utility function satisfies (10) or (11).
   (ii) All relevant market opportunities may be provided by a pair of mutual funds, one of which may be chosen as money.
   (iii) The demand for any security is linear in initial wealth.

If there are Arrow-Debreu securities (i.e., a security for every contingency):

D. (i) The utility function satisfies

\[ AU'(W)^a + BU'(W)^b = W \quad \text{or} \quad U'(W)^a[A + B \log U'(W)] = W. \]

   (ii) All relevant market opportunities may be provided by a pair of mutual funds.

Two fairly general conclusions emerging from our study are worthwhile emphasizing. First, the conditions under which separation is possible are very restrictive indeed (and encompass the utility functions which have usually been employed in the analysis of portfolio allocation and, more generally, behavior towards risk). Second, the reference to “separation” is perhaps misleading; while the possibility of representing all the market opportunities relevant to optimal portfolio allocation by means of two mutual funds is enhanced by the existence of money—and therefore being able to separate the decision about allocation among risky assets from that about allocation between safe and risky assets—it is only marginally so.

**APPENDIX I**

In this appendix we analyze the relation between any collection of securities and Arrow-Debreu securities. The key to this relation is the number of securities vis-a-vis states of nature. Assuming for the purposes
of the comparison that the number of states is finite\(^1\) and letting the number of states be denoted \(\nu\), there are three possible cases (i) \(\nu < n\), (ii) \(\nu = n\) and (iii) \(\nu > n\). Consider the first case, \(\nu < n\). In this case there can be at most \(\nu\) securities with returns which are linearly independent. What this means is that at least \(n - \nu\) securities have returns which are dominated by or are equivalent to the returns from certain mutual funds formed from the remaining securities. To see this, assume a subset of securities \(i = 1, \ldots, n'(\leq \nu + 1)\) for which there are weights \(\alpha_i \neq 0\) such that

\[
\sum_{i=1}^{n'} \alpha_i \rho_{i\theta} = 0. \tag{I.1}
\]

Then either \(\sum_{i=1}^{n'} \alpha_i > 0\) or \(\sum_{i=1}^{n'} \alpha_i = 0\) (as the weights are only unique up to a multiple). In both cases we may choose \(\alpha_1 > 0\) and write

\[
\rho_{1\theta} = \sum_{i=2}^{n'} -\frac{\alpha_i}{\alpha_1} \rho_{i\theta} \quad \text{with} \quad \sum_{i=2}^{n'} -\frac{\alpha_i}{\alpha_1} = 1 - \frac{\sum_{i=1}^{n'} \alpha_i}{\alpha_1} \leq 1. \tag{I.2}
\]

Thus, if \(\sum_{i=1}^{n'} \alpha_i = 0\), we can form a mutual fund with returns identical to those expected from the first security, while if \(\sum_{i=1}^{n'} \alpha_i > 0\), we can form one with returns which dominate those expected from the first security, simply by choosing in (5), the weights

\[
\delta_i = -\frac{\alpha_i}{\alpha_1}, \quad i = 2, \ldots, n' - 1
\]

\[
= 1 + \sum_{i=2}^{n'-1} \frac{\alpha_i}{\alpha_1}, \quad i = n'
\]

\[
= 0, \quad \text{otherwise}
\]  

Hence we can conclude that the case \(\nu < n\) is either inconsistent with permitting short sales or results in an inessential indeterminacy in the optimal portfolio allocation. More broadly, this argument establishes that the assumption of short sales all but requires that the returns from the securities be linearly independent; for convenience we therefore just assume linear independence of returns hereafter.\(^2\)

\(^1\) In this context the implications of a continuum of states are the same as those of a large but finite number of states.

\(^2\) Strictly speaking the assumption of short sales doesn’t require us to rule out the sort of linear dependence which merely allows formation of equivalent mutual funds. But our argument loses no generality if we rule out this possibility; the investor who can sell short is indifferent between a security and an equivalent mutual fund (given that there are no transaction costs).
The second case, \( v = n \), is slightly more complicated. It offers the possibility that the original securities may be—from the viewpoint of the investor—equivalent to Arrow-Debreu securities. That is, because the number of securities and states are equal, it may be possible to form \( n \) mutual funds with the property that the returns from the \( i \)th are \( \mu_{i\theta} = \rho_i > 0 \) if state \( \theta = i \) occurs and \( \mu_{i\theta} = 0 \) otherwise. To check this possibility, we begin by observing that if the formation of such funds is feasible, then as the reciprocal of the return per dollar from the \( i \)th security, \((1/\rho_i)\), represents the price of a dollar’s wealth in state \( i \), the value of the returns from each of the original securities at these prices must be a dollar

\[
1 = \sum_{\theta=1}^{v} \frac{1}{\rho_{i\theta}}, \quad i = 1, \ldots, n
\]

or, in matrix notation,

\[
e = RM^{-1}e
\]

(1.4)

where

\[
e = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad R = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1v} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2v} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_n \end{bmatrix}.
\]

Given that the returns from the original securities are linearly independent, it follows that \((1/\rho_i) > 0, i = 1, \ldots, n, \) if and only if

\[
M^{-1}e = R^{-1}e > 0.
\]

(1.5)

It is easily shown that this further restriction on \( R \) (i) is equivalent to the existence of a unique set of \( n \) mutual funds which are Arrow-Debreu securities and (ii) implies the existence of a unique mutual fund which is money.

Let the weights defining the mutual funds which are Arrow-Debreu securities be denoted by

\[
D' = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \cdots & \delta_{nn} \end{bmatrix}
\]

the weights defining the mutual fund which is money by

\[
d' = [\delta_1 \delta_2 \cdots \delta_n]
\]

and money itself by \( \mu_{\theta} = r > 0 \) for \( \theta = 1, \ldots, v \). Then these assertions can be restated formally as
Proposition 1.1. There is a unique \( D' \) such that \( D'R = M \) with \( m_{ii} > 0 \), \( i = j, m_{ij} = 0, i \neq j \) and \( D'e = e \) if and only if \( R^{-1}e > 0 \).

Proposition 1.2. If \( R^{-1}e > 0 \) then there are unique \( d \) and \( r > 0 \) such that \( d'R = re' > 0 \) and \( d'e = 1 \).

Proof of I.1. (necessity) \( D'R = M \) with \( m_{ii} > 0 \), \( i = j, m_{ij} = 0, i \neq j \), implies \( M^{-1} \) exists with \( m^{ii} > 0, i = j, m_{ij} = 0, i \neq j \), and \( M^{-1}D' = R^{-1} \), which together with \( D'e = e \) implies \( M^{-1}e = M^{-1}D'e = R^{-1}e > 0 \). (sufficiency) \( R^{-1}e > 0 \) implies a unique \( M \) with \( m_{ii} > 0, i = j, m_{ij} = 0, i \neq j \), exists such that \( M^{-1}e = R^{-1}e \), implies a unique \( D' \) exists such that \( D'R = M \) exists, implies \( D'e = MR^{-1}e = MM^{-1}e = e \).

Proof of I.2. \( R^{-1}e > 0 \) implies \( r = (1/e'R^{-1}e) > 0 \), \( d' = re'R^{-1} \) are well defined and \( d'R = re'R^{-1}R = re' > 0 \), \( d'e = re'R^{-1} = 1 \). Also, the existence of \( d \) and \( r > 0 \) such that \( d'R = re' \) and \( d'e = 1 \) requires \( d' = re'R^{-1} \) and \( 1/r = e'R^{-1}e \).

We note explicitly that when \( n = v \) (and condition (I.5) is satisfied) the return on money, or more precisely, the mutual fund which is money is given by the formula

\[
r = \frac{1}{e'R^{-1}e} = \frac{1}{e'M^{-1}e}. \tag{I.6}
\]

The foregoing tells us that for the case \( v = n \), if the original securities satisfy (I.5), then they are equivalent to both Arrow-Debreu securities and a set of securities composed of, for instance, \( n - 1 \) of the original securities plus money. The only question remaining is, what would it mean if condition (I.5) were not fulfilled? The answer is clear; as, roughly speaking, a nonpositive element in \( R^{-1}e \) reflects the fact that wealth in some state is at worst a free good (and at best a negatively priced good), such a circumstance must mean that at least one security is dominated. Thus, condition (I.5) follows naturally from our assumption that short sales are permitted.

These results and their interpretation are illustrated in Figure 3 for \( v = n = 2 \).

There is not much to be said about the last case, \( v > n \), except perhaps to emphasize the fact that when there are fewer securities than states, it is not possible to provide against every contingency. This makes this case in a sense the most interesting for the general purpose of describing portfolio behavior, as well as the most important for our particular purpose of analyzing portfolio separability. In particular, we show in the text that the inability to provide against every contingency imposes a very severe limitation on the classes of utility functions which exhibit separability.
APPENDIX II

In this appendix we prove the assertion made in section 3 of the text, that the only solutions to the functional equation

\[ G(xy) = f(x)f(y) + \bar{g}(x)\bar{g}(y), \tag{II.1} \]

when the functions \(G, f\) and \(\bar{g}(f\) and \(\bar{g})\) are defined and differentiable in some interval \((\bar{x}, \tilde{x})\) with \(0 \leq \bar{x} < \tilde{x} \leq \infty ((0, \infty))\), are given by

\[ f(x) = \frac{\bar{g}(x)}{A} = x^\alpha, \quad \bar{g}(x) = \frac{\bar{g}(x)}{B} = x^\beta \]

and

\[ G(x) = A x^\alpha + B x^\beta \tag{II.2} \]

and

\[ \bar{f}(x) = \bar{g}(x) = x^\alpha \left( \frac{A}{2} + B \log x \right), \]

\[ \bar{f}(x) = \bar{g}(x) = x^\alpha \quad \text{and} \quad G(x) = x^\alpha(A + B \log x). \tag{II.3} \]

1 There are several ways to write down (as well as to derive) the solutions to (II.1). For example, if in (II.2) we explicitly account for the fact that \(A, B, \alpha\) and \(\beta\) may be complex, \(A = A_1 + iA_2\) and so on, we obtain the representation \(G(x) = x^{\alpha}(A_1 \cos \alpha_2 \log x - A_3 \sin \beta_2 \log x) + x^{\beta}(B \cos \beta_2 \log x - A_3 \sin \beta_2 \log x)\). But, in order for \(G(x)\) to be real, the parameters must satisfy (a) \(\alpha_1 = \beta_1\), and (b) either \(A_1 = -A_2\), \(A_3 = -A_2\) or \(A_1 = A_2\), \(A_3 = -A_2\), and (c) either \(A_3 = -A_2\), \(\alpha_2 = \beta_2\) or \(A_3 = A_2\), \(\alpha_2 = \beta_2\). And when \(\alpha_2 = \beta_2\), \(G(x) = 0\), while when \(\alpha_2 = -\beta_2\) (i.e., when \(\alpha\) and \(\beta\), and \(A\) and \(B\) are complex conjugates)

\[ G(x) = 2x^{\alpha}(A_1 \cos \alpha_2 \log x - A_3 \sin \alpha_2 \log x). \]

Another example of a seemingly different representation of (II.2) can be obtained by substituting \(a + b\) and \(a - b\) for the exponents \(\alpha\) and \(\beta\), \(c + d\) and \(c - d\) for the
We begin by noting that it is clear that (II.2) and (II.3) satisfy (II.1). To see how these solutions may be derived, consider the following special cases:

\( \bar{f}(x) = \tilde{f}(x) = f(x), \quad \bar{g}(x) = \tilde{g}(x) = 0 \) \hspace{1cm} (i)

The solution to \( G(xy) = f(x)f(y) \) when the functions are differentiable is well-known to be \( f(x) = ax^b, \quad G(x) = a^x x^y \). This is shown simply by differentiating \( G(xy) \) with respect to \( x \), then \( y \)

\[
G'(xy) = f'(x) \frac{f(y)}{y} = \frac{f(x)}{x} f'(y) = G'(xy) \quad \text{for} \quad x, y \neq 0
\]

which implies

\[
\frac{f'(x) x}{f(x)} = \frac{f'(y) y}{f(y)} \quad \text{for} \quad x, y \neq 0
\]

which implies, for fixed \( y \),

\[
\frac{d \log f(x)}{d \log x} = a \neq 0 \quad \text{for} \quad x \neq 0
\]

which implies

\( f(x) = ax^a \).

\( \bar{f}(x) = \tilde{g}(x) = g(x), \quad \bar{f}(x) = \tilde{g}(x) = 1 \) \hspace{1cm} (ii)

The solution to \( G(xy) = g(x) + g(y) \) when the functions are differentiable is also well-known, and is just \( g(x) = a + b \log x, \quad G(x) = 2a + b \log x \). Again differentiating \( G(xy) \) with respect to \( x \), then \( y \) we have

\[
G'(xy) = \frac{g'(x)}{y} = \frac{g'(y)}{x} = G'(xy) \quad \text{for} \quad x, y \neq 0
\]

which implies

\[
g(x) = a + b \log x.
\]

\[
\bar{f} = \frac{\tilde{f}}{A} = f \quad \text{with} \quad \varphi(xy) = f(x)f(y), \quad \text{(iii)}
\]

\[
\bar{g} = \frac{\tilde{g}}{B} = g \quad \text{with} \quad \psi(xy) = g(x)g(y)
\]

Multiplicative constants \( A \) and \( B \), and hyperbolic functions for the power function

\[ x^y = e^{b \log x} = \sinh b \log x + \cosh b \log x, \]

\[ G(x) = 2e^{d \sinh b \log x + c \cosh b \log x} \].

It turns out useful later on to have these particular representations in hand. However, by and large, in this appendix we attempt to present an argument which balances avoiding unnecessary parameterization against illustrating why additional parameterization is in fact superfluous.
Clearly \((a)\) \(f(\psi)\) and \(g(\phi)\) have the forms derived in (i), say, \(f(x) = ax^a\) \((\psi(x) = a^2x^a)\) and \(g(x) = bx^b\) \((\phi(x) = b^2x^b)\), while \((b)\) the sum of \(A\psi(xy) = \int f(y)\) and \(B\phi(xy) = \int g(y)\) is only a function of the product \(xy\), say, \(G(xy) = A\psi(xy) + B\phi(xy) = Aa^2(xy)^a + Bb^2(xy)^b\). As there are only two degrees of freedom in picking the multiplicative constants in the latter function, by setting \(A\) and \(B\) arbitrary, \(a\) and \(b\) equal one, we arrive at (II.1).

\[
\begin{align*}
\int f &= \int g = f & \text{with } \quad \psi(xy) = f(x)f(y), \\
\int f &= \int g = fg & \text{with } \quad \phi(xy) = g(x) + g(y)
\end{align*}
\]

Now \(f\) must have the form derived in (i), say, \(f(x) = \beta x^a\), while \(g\) must have the form derived in (ii), say, \(g(x) = (A/2) + B \log x\). Moreover, though now neither \(\int f(y)\) nor \(\int g(y)\) is only a function of \(xy\), their sum is, say, \(G(xy) = \int f(x)f(y) + \int g(x)g(y) = \beta^2(xy)^a (A + B \log xy)\).

Again, as there are only two degrees of freedom in picking the multiplicative constants in the latter function, by letting \(A\) and \(B\) be arbitrary, \(\beta\) equal one, we arrive at (II.2).

Getting solutions to (II.1) is easy; the difficulty comes in showing that (II.2) and (II.3) are the only solutions. As in the argument in the text where (II.1) first appears, we shall break up this part of the proof into several steps. Also, for convenience we shall assume that all the functions involved are defined and differentiable for positive arguments; the necessary elaboration of the argument to encompass the possibility that \(0 < U'(\infty) < U'(0) < \infty\) is straightforward.

1. If \(G(xy) = \int f(x)f(y) + \int g(x)g(y)\) for \(x, y > 0\) has a solution with \(G \neq 0\), then either (a) there exists a function \(f\) such that

\[
G(xy) = f(x)f(y) \tag{II.4}
\]

or (b) there exist functions \(f\) and \(g\) such that

\[
G(xy) = f(x)g(y) + g(x)f(y) \tag{II.5}
\]

or (c) there exist functions \(f\) and \(g\) such that

\[
G(xy) = f(x)f(y) + g(x)g(y) \tag{II.6}
\]

By hypothesis we have for all \(x, y > 0\)

\[
G(xy) = \int f(x)f(y) + \int g(x)g(y) = \int f(y)f(x) + \int g(y)g(x) = G(yx). \tag{II.7}
\]
For any two values of \( y \), say, \( y_1 \) and \( y_2 \), (II.7) yields the equations
\[
\begin{align*}
\bar{f}(y_1) \bar{f}(x) + \bar{g}(y_1) \bar{g}(x) &= \bar{f}(y_2) \bar{f}(x) + \bar{g}(y_2) \bar{g}(x), \\
\bar{f}(y_2) \bar{f}(x) + \bar{g}(y_2) \bar{g}(x) &= \bar{f}(y_1) \bar{f}(x) + \bar{g}(y_1) \bar{g}(x).
\end{align*}
\] (II.8)

There are two cases to consider:
\[
\bar{f}(y_1) \bar{g}(y_2) - \bar{f}(y_2) \bar{g}(y_1) = 0 \quad \text{for all} \quad y_1, y_2 > 0
\] (i)

For this case, if \( \bar{f}(y) = 0 \) for all \( y > 0 \), then, since \( G \neq 0 \), there must be at least one value \( y_1 \) such that \( \bar{g}(y_1) \neq 0 \). Hence, by utilizing the first of the equations in (II.8) we have
\[
\bar{g}(x) = \frac{\bar{g}(y_1)}{\bar{f}(y_1)} \bar{g}(x) \quad \text{for} \quad x > 0,
\]
which, when substituted into the original functional equation (II.1) yields (II.4)
\[
G(xy) = \frac{\bar{g}(y_1)}{\bar{f}(y_1)} \bar{g}(x) \bar{g}(y) = f(x)f(y),
\]
where \( f = (\bar{g}(y_1)/\bar{g}(y_1))^{1/2} \bar{g} \). On the other hand, if \( \bar{f}(y_1) \neq 0 \) for some \( y_1 > 0 \), then the hypothesis \( \bar{f}(y_1) \bar{g}(y_2) - \bar{f}(y_2) \bar{g}(y_1) = 0 \) yields
\[
\bar{g}(y) = \frac{\bar{g}(y_1)}{\bar{f}(y_1)} \bar{f}(y) = C \bar{f}(y) \quad \text{for} \quad y > 0.
\]
Again, substitution into (II.1) yields (II.4)
\[
G(xy) = (\bar{f}(x) + C\bar{g}(x)) \bar{f}(y) = h(x) \bar{f}(y) = f(x)f(y),
\]
where now \( f = (h(y_1)/\bar{f}(y_1))^{1/2} \).
\[
\bar{f}(y_1) \bar{g}(y_2) - \bar{f}(y_2) \bar{g}(y_1) \neq 0 \quad \text{for some} \quad y_1, y_2 > 0
\] (ii)

For this case we can solve the pair of equations (II.8) for \( \bar{f} \) and \( \bar{g} \) in terms of \( \bar{f} \) and \( \bar{g} \)
\[
\begin{align*}
\bar{f}(x) &= C_1 \bar{f}(x) + C_2 \bar{g}(x) \\
\bar{g}(x) &= C_3 \bar{f}(x) + C_4 \bar{g}(x),
\end{align*}
\] (II.9)

where
\[
\begin{bmatrix}
C_1 & C_2 \\
C_3 & C_4
\end{bmatrix} = \frac{1}{\bar{f}(y_1) \bar{g}(y_2) - \bar{f}(y_2) \bar{g}(y_1)} \begin{bmatrix}
\bar{f}(y_1) \bar{g}(y_2) - \bar{f}(y_2) \bar{g}(y_1) & \bar{g}(y_1) \bar{g}(y_2) - \bar{g}(y_2) \bar{g}(y_1) \\
\bar{f}(y_2) \bar{f}(y_1) - \bar{f}(y_1) \bar{f}(y_2) & \bar{g}(y_2) \bar{f}(y_1) - \bar{g}(y_1) \bar{f}(y_2)
\end{bmatrix}
\]
Notice that (a) \( G(y_1, y_2) = \tilde{f}(y_1) \tilde{g}(y_2) + \hat{g}(y_1) \hat{f}(y_2) = \hat{f}(y_2) \hat{g}(y_1) + \hat{g}(y_2) \hat{f}(y_1) \)
if and only if \( \hat{g}(y_1) \hat{g}(y_2) - \hat{g}(y_2) \hat{g}(y_1) = \hat{f}(y_2) \hat{f}(y_1) - \hat{f}(y_1) \hat{f}(y_2) \)
if and only if \( C_3 = C_4 \); and (b) \( C_1 = 0 \) if \( C_4 = 0 \) for all \( y_1, y_2 > 0 \) if and only if \( \hat{f}(y_1) \hat{g}(y_2) - \hat{f}(y_2) \hat{g}(y_1) = 0 \) \( (\hat{g}(y_2) \hat{f}(y_1) - \hat{g}(y_1) \hat{f}(y_2) = 0) \)
for all \( y_1, y_2 > 0 \). Because of this last consideration, we again have two cases to consider:

\[
C_1 = C_4 = 0
\]  

(i)

In this case (II.9) becomes

\[
\hat{f}(x) = C_2 \hat{g}(x) \]

\[
\hat{g}(x) = C_2 \tilde{f}(x)
\]

(II.10)

Substituting from (II.10) into (II.1), and using the fact that \( C_2 = C_3 \), yields (II.5)

\[
G(xy) = C_2 \tilde{g}(x) \tilde{f}(y) + C_2 \hat{f}(x) \hat{g}(y)
\]

\[
= f(x) g(y) + g(x) f(y)
\]

where \( f = C_2 \hat{g} = C_2 \hat{g}, \ g = \hat{f} \).

\[
C_1 \neq 0 \quad \text{or} \quad C_4 \neq 0
\]  

(ii)

Without loss of generality we can assume \( C_1 \neq 0 \). Then, substituting from (II.9) into (II.1), we find that this case yields (II.6)

\[
G(xy) = [C_1 \tilde{f}(x) + C_2 \hat{g}(x)] \tilde{f}(y) + [C_2 \hat{f}(x) + C_4 \tilde{g}(x)] \hat{g}(y)
\]

\[
= C_1 \tilde{f}(x) \tilde{f}(y) + C_2 \hat{f}(x) \hat{g}(y) + \tilde{f}(y) \hat{g}(x) + C_4 \hat{g}(x) \tilde{g}(y)
\]

\[
= C_1 \left\{ \left[ \tilde{f}(x) + \frac{C_2}{C_1} \hat{g}(x) \right] \tilde{f}(y) + \frac{C_2}{C_1} \hat{g}(y) \right\} + \left[ C_4 - C_2 \left( \frac{C_2}{C_1} \right) \right] \hat{g}(x) \tilde{g}(y)
\]

\[
= f(x) f(y) + g(x) g(y)
\]

where \( f = C_1^{1/2} \tilde{f} + (C_4/C_1) \hat{g} \) and \( g = (C_4 - (C_2 \hat{g}/C_1))^{1/2} \hat{g} \).

We have already characterized the solution to (II.4); the balance of the discussion will be concerned with the solutions to (II.5) and (II.6). Moreover, because the analysis for (II.6) is all but identical to that for (II.5), we will only detail the argument for the latter.
2. The functions $f$ and $g$ appearing on the RHS of (II.5) must either have the similar forms
\[ f(x) = ax^n \quad \text{and} \quad g(x) = bx^n, \]  
(II.11)
in which case $G$ clearly has the form (II.2), or satisfy the same differential equation
\[ Af(x) + 2Bf'(x) x + f''(x) x^2 = 0 \quad \text{and} \quad Ag(x) + 2Bg'(x) x + g''(x) x^2 = 0, \]  
(II.12)
in which case $G$ satisfies the differential equation
\[ AG(x) + BG'(x) x + G''(x) x^2 = 0. \]  
(II.13)

Differentiating (II.5) with respect to $x$ and multiplying the result by $x$ we obtain
\[ G'(xy) xy = f'(x) xg(y) + g'(x) xf(y). \]  
(II.14)

(II.14) is precisely the functional equation (II.1) (with $G'xy = G,f'x = f$, etc.). Hence, by what is essentially a repetition of the argument of the preceding section (and whose details we therefore omit) either $f$ and $g$ have the forms (II.11) or for some $C_3 = C_3'$ and $C_1 \neq 0$ or $C_4 \neq 0$
\[ f'(x) x = C_1' g(x) + C_2' f(x) \]  
(II.15)
\[ g'(x) x = C_3' g(x) + C_4' f(x) \]

By differentiating these equations we have
\[ f''(x) x = C_1' g'(x) + (C_2' - 1) f'(x) \]  
(II.16)
\[ g''(x) x = (C_3' - 1) g'(x) + C_4' f'(x) \]

Now, solving (II.15) for $f'$ in terms of $g$ and $g'$, and $g'$ in terms of $f$ and $f'$, substituting into (II.16), and rearranging the resulting expressions we find
\[ Af(x) + 2Bf'(x) x + f''(x) x^2 = 0 \quad \text{and} \quad Ag(x) + 2Bg'(x) x + g''(x) x^2 = 0, \]  
(II.12)

where $A = C_3 C_3' - C_1 C_4'$, $B = 1 - C_1 - C_4$. Finally, from differentiating (II.14) with respect to $x$ and multiplying the result by $x$
\[ G'(xy) xy = (f''(x) x^2 + f'(x) x) g(y) + (g''(x) x^2 + g'(x) x) f(y), \]  
(II.17)
it is easily seen by adding together $A \cdot (II.5)$, $B \cdot (II.14)$ and (II.17) that (II.12) implies (II.13).

\[ * \text{Note that if } f \text{ and } g \text{ are once differentiable, then (II.15) implies they are analytic.} \]
3. The only solutions to the differential equations (II.12) and (II.13) satisfying the functional equation (II.5) (and, a fortiori, (II.6)) yield the forms described by (II.2) and (II.3).

It is sufficient for our purpose to characterize the solutions to (II.13). Let \( z = \frac{G'(x)x}{G(x)} \). Then

\[
\frac{dz}{dx} x = \frac{G'(x)x}{G(x)} - \frac{G'(x)^2 x^2}{G(x)^2} + \frac{G'(x)x^2}{G(x)}
\]

or

\[
\frac{G'(x)x^2}{G(x)} = -z + z^2 + \frac{dz}{dx} x.
\]

Substituting from (II.18) into (II.13) and simplifying yields

\[ A + Cz + z^2 + \frac{dz}{dx} x = 0 \]

where \( C = B - 1 \), or

\[ -\int \frac{dx}{x} = \int \frac{dz}{A + Cz + z^2}. \]

The closed form solution to (II.19) (and thus to (II.13)) depends on the sign of \( q = C^2 - 4A \):

\[
\log x = -\frac{2}{\sqrt{-q}} \tan^{-1} \frac{2z + C}{\sqrt{-q}} + \text{constant of integration}, \quad q < 0
\]

\[
= \frac{2}{2z + C} + \text{constant of integration}, \quad q = 0
\]

\[
= -\frac{2}{\sqrt{q}} \tanh^{-1} \frac{2z + C}{\sqrt{q}} + \text{constant of integration}, \quad q > 0.
\]

We need only consider the first two cases, as the argument concerning the third is a repetition of that for the first with trigonometric replaced by hyperbolic functions:

\[
\log x = -\frac{2}{\sqrt{-q}} \tan^{-1} \frac{2z + C}{\sqrt{-q}} + \text{constant of integration with } q < 0 \quad (i)
\]

\(^a\) Indeed, the dependence of the form of the solution to (II.1) on the sign of \( q \) is much like the dependence of the form of the solution to a quadratic equation on the sign of its discriminant. In particular, it turns out that \( q < 0 \) corresponds to complex conjugate exponents (as well as multiplicative constants) in (II.2), while \( q > 0 \) corresponds to real, unequal exponents (as well as multiplicative constants) in (II.2). However, the relation between \( q = 0 \) and (II.3) as a special case of (II.2) appears not to be such a clear-cut analogy.
First inverting to get \( z \) as a function of \( x \)

\[
\frac{d \log G(x)}{d \log x} = z = \alpha_1 - \alpha_2 \tan(\alpha_2 \log x + b_1)
\]

and then integrating, we obtain the solution

\[
\log G(x) = \log b_2 + \alpha_1 \log x + \log \cos(\alpha_2 \log x + b_1)
\]
or

\[
G(x) = b_2 x^{\alpha_1} \cos(\alpha_2 \log x + b_1) = 2x^{\alpha_1} (A_1 \cos \alpha_2 \log x - A_2 \sin \alpha_2 \log x)
\] (II.21)

where \( \alpha_1 = -(C/2), \alpha_2 = -\sqrt(-q/2) \) and the \( A \), are essentially arbitrary constants of integration. But (II.21) is nothing but the particular representation of (II.2) when \( \alpha \) and \( \beta \), and \( A \) and \( B \) are complex conjugates (cf. footnote 1 at the beginning of this appendix).

\[
\log x = \frac{2}{2x + C} + \text{constant of integration} \quad (ii)
\]

Again inverting and integrating, we now obtain the solution

\[
G(x) = b_2 x^{\alpha} \log b_1 x = x^{\alpha}(A + B \log x)
\]

where \( \alpha = -(C/2) \) and \( A, B \) are essentially arbitrary constants of integration. Hence, this case yields (II.3) as a solution to (II.1).

BIBLIOGRAPHY