Transaction Costs and the Pricing of Assets
Author(s): Joram Mayshar
Source: *The Journal of Finance*, Vol. 36, No. 3 (Jun., 1981), pp. 583-597
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2327520
Accessed: 27/11/2009 08:31

THE JOURNAL OF FINANCE • VOL. XXXVI, NO. 3 • JUNE 1981

Transaction Costs and the Pricing of Assets

JORAM MAYSHAR*

ABSTRACT

The existence of transaction costs explains why investors do not fully diversify their portfolios. This paper examines the implications of such limited diversification on equilibrium asset prices in the framework of the capital asset pricing model. In the pricing equation obtained here an asset's risk premium depends on a weighted average of its covariance with the market and its own variance.

THE ABSENCE OF TRANSACTION costs in the Sharpe-Lintner capital-asset pricing model (CAPM) is often claimed to be only a minor idealization which hardly affects the substantive implications of the model. This paper aims to demonstrate that this is not the case: incorporating transaction costs into the CAPM leads to a major reformulation of its basic asset pricing equation. In contrast with the CAPM formula

$$ER_j = r + \lambda \operatorname{Cov}(R_j, \hat{R}), \tag{1}$$

the pricing equation derived here is

$$ER_j = (r + t) + \lambda[\gamma_j \operatorname{Var} R_j + \delta_j \operatorname{Cov}(R_j, \hat{R})]. \tag{2}$$

In these equations, R_j denotes the return on asset j; $\hat{R}$ the return for the market; r the safe interest rate; and λ a measure of market risk aversion. Furthermore, t is a measure of marginal transaction costs, and γ_j and δ_j are nonnegative estimable parameters which depend on the relative concentration of holdings of asset j.

The absence of transaction costs tremendously simplifies the theorist's problem of determining endogenously the composition of investors' portfolios since it implies that each investor trades in all assets. When transaction costs (in particular, fixed transaction costs) are introduced, however, it becomes obvious that investors might not trade in all assets.[1] As a result, any particular asset may be held by only a small fraction of investors.

Although the endogeneity of portfolio composition is a paramount issue in the context of transaction costs, this paper follows the approach adopted by Lintner [8] and Levy [6] in taking the composition of investors' partially diversified portfolios as exogenously given.[2] It extends their models, however, by introducing

* This is a revised version of Discussion Paper No. 775 of the Maurice Falk Institute for Economic Research in Israel, Jerusalem. I am indebted to Haim Levy for the initial impetus to this paper and for his comments. For their comments, I am also grateful to the referee Marshall E. Blume and the editors of this journal, to S. Bar-Yosef, D. Galai and Y. Landskroner.

[1] The effects of transaction costs in limiting the size of investors' portfolios were examined, among others, by Leland [4], Brennan [1], and Goldsmith [2].

[2] To recognize the difficulties in deriving equilibrium properties with partial diversification under more general assumptions, see Stapleton and Subrahamanyam [15].

an assumption concerning portfolio composition which makes it possible to derive the pricing equation (2). Equilibrium conditions based on endogenously determined portfolios, though under quite restrictive assumptions, were derived in Mayshar [11]. The different approach adopted here has the advantages of being more general (in particular, by allowing heterogeneity among investors), and more clearly related to the CAPM.

Both Lintner [8] and Levy [6] deduce that in the case of partial diversification, the equilibrium price of a given asset is affected by the expectations, attitude toward risk, and portfolio composition of only those investors who *actually* hold the asset. Levy has further maintained that in this case, in contrast to the CAPM result (1), the own variance of the asset will significantly affect its equilibrium price. Levy's reasoning is essentially the same as that of Tobin and Brainard [17] who have argued:

> Suppose that transaction costs limit the number of assets a typical investor can hold in his portfolio. The "undiversifiable" risk of a particular asset to him then depends on its covariation not with the entire market but with his own portfolio. Obviously an asset will be a higher proportion of the portfolios in which it is held than in the aggregate market "portfolio." Hence its own variance will be more important.

Transaction costs are often narrowly interpreted as including only brokers' fees and losses due to the bid-ask spread. Taxes on transactions and various other obstacles to trade, however, may also be usefully considered as a form of (sometimes prohibitive) transaction costs. These impediments to trade may include the nondivisibility of assets, short-sale restrictions, various institutional restraints, and even the subjective costs of managing one's own portfolio. It is in this broader sense that the term "transaction costs" is used here.[3]

Transaction costs are introduced in this paper in two stages: first, under the assumption that only fixed transaction costs exist, (Sections I–IV), and then under the assumption that there are also transaction costs which depend on the volume of trade (Section V). In Section I, the behavior of individual investors is examined. In Section II, investors' equilibrium conditions are aggregated to produce a market equilibrium condition akin to the one derived by Lintner [8] and Levy [6]. The key result of the paper is then obtained in Section III, under one set of assumptions, and then, in Section IV, it is reestablished in a somewhat more general framework. Some of the possible empirical implications of the model are examined in Section VI.

I. The Individual Investor's Equilibrium

The basic model presented here is a variation of the standard CAPM. In fact, all the key assumption of the CAPM are retained here except those which exclude transaction costs, taxes, and restrictions on borrowing and short sales. Since some of these standard assumptions obtain a somewhat different significance in the context of transaction costs, they will be presented here formally and discussed. For the sake of expediency, however, that presentation is deferred to Section VII.

[3] For a further discussion of the nature of transaction cost in this context, see Mayshar [11].

Let there be n individuals (indexed by i) and $m + 1$ assets (indexed by j). For $j = 1, \cdots, m$, let x_j denote the uncertain gross payout of a unit of asset j whose first period price is p_j. The zero asset is a safe asset offering a unit payout, $x_0 \equiv 1$; its current price is $p_0 = 1/(1 + r)$. M_i denotes the set of indices of risky assets held, long or short, by individual i. His future wealth, y_i, then takes the form

$$y_i = q_{i_0} + \sum_{j \in M_i} q_{ij} x_j, \tag{3}$$

where q_{ij} is the quantity of asset j which he holds (here $q_{ij} > 0$, for an asset actually held; $q_{ij} < 0$ for an asset sold short; and $q_{ij} = 0$ for $j \notin M_i$).

It will be assumed throughout the paper that each investor actually trades in the safe asset. In this section, it is further assumed that only fixed transactions costs for trade in each asset exist. (A more general structure of transaction costs is examined in Section V). For trade (whether long or short) in asset j, investor i is assumed thus to incur a fixed cost c_{ij}. His budget constraint thus takes the form:

$$q_{i0} p_0 + \sum_{j \in M_i} q_{ij} p_j = w_i - c_{i0} - \sum_{j \in M_i} c_{ij}. \tag{4}$$

The initial wealth of each investor, w_i, is taken as exogenously given. This wealth may include the revenue from the sale of previously-held assets. The form of the budget constraint (4) thus precludes the possibility that an investor may not entirely revise his previously-held portfolio as a mean for saving transaction costs.

Investor i will select q_{i0}, M_i and q_{ij} for $j \in M_i$, given (1) and (2), in order to maximize his mean-variance utility $U^i(E_i y_i, \mathrm{Var}_i\, y_i)$. The index i on the expectation operators indicates his use of his own subjective probabilities.

This optimization problem may be considered to consist of two stages: in the first, he selects the asset composition of his portfolio, i.e. the set M_i; and in the second he selects the optimal holdings q^*_{i0} and q^*_{ij} for each asset j in the portfolio M_i.[4] The first-stage problem involves a discrete optimization for which standard calculus methods are inapplicable. This stage is not explicitly considered here; it is assumed that the set M_i has already been determined. It is the second-stage optimization which concerns us here. For this problem calculus methods *are* applicable. Once the decision to trade in an asset has been made, the fixed transaction costs involved in that trade are sunk costs. These costs should thus have no further effect on the magnitude of trade, which can then be determined by using standard marginal considerations.

Substitution of q_{i0} from (4) into (3), and optimization with respect to q_{ij} yield, then, the first order conditions

$$E_i x_j - A_i \mathrm{Cov}_i(x_j, y^*_i) = (1 + r) P_j \qquad j \in M_i. \tag{5}$$

These conditions (5), together with the budget constraint (4), simultaneously determine the individual's optimal portfolio. Here y^*_i denotes the uncertain

[4] To regard the selection of M_i as a first-stage problem may be somewhat misleading. To find M_i the investor will presumably first solve the second stage problem for every feasible set M_i, and only then, by a discrete comparison of the optimal utility levels obtained, select the best M_i. A framework in which both stages of the optimization problem are explicitly solved, is presented in Mayshar [11].

future wealth of investor i, given his optimal portfolio; A_i is a measure of his risk aversion at this optimal wealth position, given by $A_i = -2U_2^i/U_1^i$, where U_1^i and U_2^i are the respective partial derivatives of U^i, evaluated at the optimum.[5]

II. Aggregating Investors' Equilibrium Conditions

The equilibrium conditions (5) for all individual investors must be aggregated in order to find the corresponding market equilibrium conditions. For that purpose denote, by N_j, the set of (indices of) individual investors who hold asset j, and let n_j be their number. Dividing (5) by A_i and summing over all the investors i who invest in asset j, that is, over $i \in N_j$, we obtain

$$\sum_{i\in N_j} \frac{1}{A_i}[E_i x_j - (1+r)p_j] = \sum_{i\in N_j} \mathrm{Cov}_i(x_j, y_i^*) \qquad j = 1, \cdots, m. \tag{6}$$

To simplify (6), denote first by $\bar{A}_j$ the harmonic mean of the measures of risk aversion A_i of those investors who choose to hold asset j,

$$\bar{A}_j = \left(\frac{1}{n_j}\sum_{i\in N_j} A_i^{-1}\right)^{-1}. \tag{7}$$

$\bar{A}$ (without the index j) will denote the similar harmonic mean for *all* investors.

It is assumed now that all investors share the same expectations concerning assets' variances and covariances, so that the index i may be dropped from the covariance operators. Thus,

$$\sum_{i\in N_j} \mathrm{Cov}_i(x_j, y_i^*) = \mathrm{Cov}(x_j, \sum_{i\in N_j} y_i^*) = \mathrm{Cov}(x_j, Y_j), \tag{8}$$

where Y_j denotes the total future proceeds of the risky assets of the investors who hold asset j; that is,

$$Y_j = \sum_{i\in N_j} [y_i^* - q_{i0}^*(1+r)] = \sum_{i\in N_j} \sum_{k\in M_i} q_{ik}^* x_k.$$

Similarly, Y will denote the total proceeds of all the risky assets in the market. That is, $Y = \sum_{k=1}^m Q_k x_k$, where $Q_k = \sum_{i\in N_k} q_{ik}^*$ is the total number of units of asset k.

Note that it is *not* assumed that investors hold identical expectations of the means of the asset payouts. It is useful, therefore, to denote an average expectation operator $\bar{E}_j$, presenting a weighted average of the expectations of investors in N_j, where

$$\bar{E}_j x_j = \frac{1}{n_j}\sum_{i\in N_j} (\bar{A}_j/A_i) E_i x_j. \tag{9}$$

[5] A much weaker assumption than the one on the existence of a global mean-variance representation of preferences may suffice here. All that is required is that investors "subjective valuation" of the risky assets in which they trade (or, equivalently, investors' marginal rates of substitution between these risky assets and the safe asset) will take the mean-variance form of the left hand side in (5) for some factor A_i. It is quite possible that this would be the case *locally* at y_i^*, at least as a very good approximation, even when there may not exist a *global* mean-variance representation of preference. [For a similar use of the "subjective-valuation" concept see Mayshar [10].

The weighted average expectation of *all* investors, will be denoted as $\bar{E}x_j$.[6]

With the aid of (7), (8) and (9), Equation (6) can finally be expressed as:

$$\bar{E}_j x_j - (1 + r)p_j = (\bar{A}_j/n_j)\text{Cov}(x_j, Y_j). \tag{10}$$

This equation can be regarded as a variant of the basic results of Lintner [8] and Levy [6]. If asset j were held by all investors, Equation (10) would yield the standard CAPM valuation formula (with the index j dropped from the average expectation operator $\bar{E}_j$, and from $\bar{A}_j$, n_j and Y_j). A key to understanding the suggested revision of the CAPM formula is the fact [which was noted already by Levy [6]] that Y_j, the total future risky wealth of investors in N_j, includes *all* the proceeds $Q_j x_j$ of asset j, but may comprise only a small fraction of the *other* risky wealth in the economy, $Y - Q_j x_j$.

III. Asset Valuation with Fixed Transaction Costs: The Case of the Diagonal Model

For Equation (10) to have empirical content, some assumptions are required concerning the composition of Y_j, the future risky wealth of investors in N_j. In order to examine what assumptions may be reasonable, it is useful to consider first the special case of a market in which asset payouts are distributed according to the "diagonal" (or "single index") model. Thus, assume

$$x_j \equiv b_j \bar{x} + e_j \qquad j = 1, \cdots, m, \tag{11}$$

where the market-wide factor $\bar{x}$ (which is common to all assets) and all the specific factors e_j are independently distributed. The future wealth of investor i can then be expressed in the form

$$y_i = \bar{q}_i \bar{x} + \textstyle\sum_{j \in M_i} q_{ij} e_j + q_{io}, \tag{12}$$

where $\bar{q}_i = \sum_{j \in M_i} q_{ij} b_j$ is the total quantity held of the common factor $\bar{x}$ by investor i, or alternatively, $\bar{q}_i = \text{Cov}(\bar{x}, y_i)/\text{Var}\ \bar{x}$. From (11) it follows directly that

$$\text{Cov}(x_j, Y_j) = b_j \text{Cov}(\bar{x}, Y_j) + Q_j \text{Var}\ e_j. \tag{13}$$

A rather simple assumption in this context is that an investor's optimal total holding $\bar{q}_i^*$ of the common factor $\bar{x}$ would not reveal anything about whether he holds asset j or not. This "independence" assumption for the diagonal model can be restated as:

ASSUMPTION I.1A: *The magnitude of each investor's covariance between his future wealth and the common market factor*, $\text{Cov}(\bar{x}, y_i^*)$,

[6] The existence of heterogeneous expectations among investors which was examined already by Lintner [8], is considered here because it seems particularly relevant in the present context. Differences in investors' expectations are probably the main source of differences in the composition of their portfolios. We may expect thus that in general $\bar{E}_j x_j > \bar{E}x_j$. Note that under the quite plausible further assumption that $E_i x_j$ and $1/A_i$ are independently distributed among the investors in N_j, the weighted average in (9) might be approximated by the simple average $(1/n_j) \sum_{i \in N_j} E_i x_j$.

regarded as a random drawing over the population of investors, is independent of whether the investor holds any given asset j or not.[7]

An application of the law of large numbers for the covariances $\mathrm{Cov}(\bar{x}, y_i^*)$ will suggest, then, that the average covariance for investors in N_j will (almost) equal the average covariance of the investors outside N_j. That is,

$$\frac{1}{n_j}\sum_{i\in N_j}\mathrm{Cov}(\bar{x}, y_i^*) = \frac{1}{n-n_j}\sum_{i\notin N_j}\mathrm{Cov}(\bar{x}, y_i^*), \tag{14}$$

or equivalently, upon noting that $\sum_{i\notin N_j}\mathrm{Cov}(\bar{x}, y_i^*) = \mathrm{Cov}(\bar{x}, Y - Y_j)$ and rearranging,

$$\mathrm{Cov}(\bar{x}, Y_j) = \frac{n_j}{n}\mathrm{Cov}(\bar{x}, Y). \tag{15}$$

Obviously, in the perfect market case of the CAPM, when $n_j = n$ and $Y_j = Y$, (15) is an identity which requires no assumptions.

Substituting from (15) into (13) and then into (10), we obtain

$$(1+r)p_j = \bar{E}_j x_j - \frac{\bar{A}_j}{n}\left[b_j\mathrm{Cov}(\bar{x}, Y) + \frac{n}{n_j}Q_j\mathrm{Var}\ e_j\right]. \tag{16}$$

Since the ratio $\mathrm{Cov}(\bar{x}, Y)/Q_j\mathrm{Var}\ e_j$ is likely to be very large (on the order of magnitude of m), in the perfect market case of $n_j = n$ the effect of the nonsystematic specific risk e_j is likely to be negligible in comparison with the effect of the systematic market-risk. However, in the more realistic case, where n/n_j may also be very large, the effect of the specific risk may no longer be regarded as negligible.

A somewhat tedious calculation makes it possible to substitute the terms $\mathrm{Cov}(x_j, Y)$ and $\mathrm{Var}\ x_j$ for the terms $\mathrm{Cov}(\bar{x}, Y)$ and $\mathrm{Var}\ e_j$ in Equation (16). We then obtain[8]

$$(1+r)p_j = \bar{E}_j x_j - \frac{\bar{A}_j}{n_j}\left[d_j\mathrm{Cov}(x_j, Y) + (1-d_j)Q_j\mathrm{Var}\ x_j\right], \tag{17}$$

where the parameter d_j is defined by

$$d_j = \frac{\bar{Q}\frac{n_j}{n} - Q_j b_j}{\bar{Q} - Q_j b_j} = 1 - \frac{1 - n_j/n}{1 - Q_j b_j/\bar{Q}}, \tag{18}$$

[7] Slightly more formally we can introduce characterizing functions η_{ij} such that $\eta_{ij} = 1$ if $i \in N_j$ and $\eta_{ij} = 0$ if $i \notin N_j$. The assumption here is that for any given asset j, η_{ij} and $\mathrm{Cov}(\bar{x}, y_i)$ can be regarded as two independent variables across i.

[8] From the equations $\mathrm{Var}\ x_j = b_j^2\ \mathrm{Var}\ \bar{x} + \mathrm{Var}\ e_j$ and $\mathrm{Cov}(x_j, Y) = b_j\bar{Q}\ \mathrm{Var}\ \bar{x} + Q_j\mathrm{Var}\ e_j$, we can solve to obtain

$$Q_j\mathrm{Var}\ e_j = (\bar{Q} - Q_j b_j)^{-1}[Q_j\bar{Q}\ \mathrm{Var}\ x_j - Q_j b_j\mathrm{Cov}(x_j, Y)]$$

and

$$b_j\mathrm{Cov}(\bar{x}, Y) = b_j\bar{Q}\ \mathrm{Var}\ \bar{x} = (\bar{Q} - Q_j b_j)^{-1}[\bar{Q}\ \mathrm{Cov}(x_j, Y) - Q_j\bar{Q}\ \mathrm{Var}\ x_j].$$

Substituting in (16) we obtain then (17) and (18).

where $\bar{Q} = \text{Cov}(\bar{x}, Y)/\text{Var}\,\bar{x} = \sum_{j=1}^{m} Q_j b_j$. Again, in the perfect market case of the CAPM, when $n_j = n$ and thus $d_j = 1$, the result (17) obtained here reduces to the standard CAPM formula.

Equation (17), obtained here in the context of the diagonal model, is the basic result of this paper. It formalizes the assertion that in the absence of full diversification by investors, the own-variance of each asset will directly affect its price. Indeed, if we multiply both sides of (17) by Q_j, the market value $p_j Q_j$ of the total proceeds of asset j can be obtained as dependent on a weighted average of the own-variance of the proceeds $Q_j x_j$ and their covariance with aggregate future wealth. The weight d_j is some measure of how widespread the ownership of the asset is.

IV. Asset Valuation with Fixed Transaction Costs: the General Case

Before attempting to go beyond the diagonal case, it is worthwhile to examine one potential bias in Assumption I.1a. If investors who invest in asset j differ systematically from other investors in the level of their initial wealth, it is likely that they will differ also in the scale of their final wealth y_i^*, and thus, contrary to I.1a, in the covariance $\text{Cov}(\bar{x}, y_i^*)$. A normalization for the scale of initial wealth invested in risky assets can correct this bias. For that purpose define by $W_j = \sum_{i \in N_j} \sum_{k \in M_i} q_{ik}^* p_k$ and $W = \sum_{k=1}^{m} Q_k p_k$, the market value of the total risky assets held by investors in N_j and by all investors respectively. Equation (15) may then be assumed to hold with the ratio W_j/W replacing the ratio n_j/n. The basic equation (17) will still hold, with W_j/W replacing n_j/n in the definition of d_j in (18).

In trying to extend Assumption I.1a to a general nondiagonal model, it is further useful to note that $\text{Cov}(\bar{x}, y_i^*) = (1/b_j)\,\text{Cov}(x_j, y_i^* - q_{ij}^* e_j)$. Assumption I.1a can thus be alternatively interpreted as requiring that knowledge of the covariance of a given asset j with $y_i^* - q_{ij}^* e_j$ would yield no distinguishing information on whether investor i holds asset j or not.

A simple extension would seem then to assume that no feature of the distribution $y_i^* - q_{ij}^* x_j$ (in particular, its covariance with x_j) would yield such distinguishing information. One obvious bias in such an assumption, however, is that the scale of $y_i^* - q_{ij}^* x_j$ may be expected to be generally lower for investors in N_j, for the simple reason that it comprises only some of their future risky wealth, whereas for investors not in N_j all their future risky wealth is included. One way of correcting that bias is by normalizing $y_i^* - q_{ij}^* x_j$ for the (average) initial wealth invested in assets other than j. Thus, define a two-valued variable ω_j^i taking the value of $(W_j - Q_j p_j)/n_j$ for investors in N_j and taking the value of $(W - W_j)/(n - n_j)$ for investors not in N_j. The alternative to Assumption I.1a is now:

ASSUMPTION I.1B: *Given any asset j, the magnitude of the covariance* $\text{Cov}[x_j, (y_i^* - q_{ij}^* x_j)/\omega_j^i]$ *for each investor i, considered as a random drawing over the population of investors, is independent of whether the investor holds asset j or not.*

The law of large numbers can be used here once more as implying that the averages of these covariances over investors inside N_j and outside N_j will be equal,

$$\frac{1}{n_j}\sum_{i\in N_j}\mathrm{Cov}\left[x_j, \frac{1}{\omega_j^i}(y_i^* - q_{ij}^* x_j)\right] = \frac{1}{n-n_j}\sum_{i\notin N_j}\mathrm{Cov}\left[x_j, \frac{1}{\omega_j^i}(y_i^* - q_{ij}^* x_j)\right]. \tag{19}$$

Since the scaling variable ω_j^i is constant on both sides of (19), it can be taken outside the summation operator. The summation can then be applied directly to the terms $y_i^* - q_{ij}^* x_j$, which add up to the aggregate future wealth from risky assets other than j, of the investors inside and outside N_j. Equation (19) can thus be expressed as

$$\frac{1}{W_j - V_j}\mathrm{Cov}(x_j, Y_j - Q_j x_j) = \frac{1}{W - W_j}\mathrm{Cov}(x_j, Y - Y_j), \tag{20}$$

where $V_j = Q_j p_j$ denotes the total market value of asset j. Equation (20) can be further modified into

$$\mathrm{Cov}(x_j, Y_j) = d_j \mathrm{Cov}(x_j, Y) + (1 - d_j)Q_j \mathrm{Var}\, x_j, \tag{21}$$

where the parameter d_j is here redefined by

$$d_j = \frac{W_j - V_j}{W - V_j} = 1 - \frac{1 - W_j/W}{1 - V_j/W}. \tag{22}$$

Substituting (21) into (10) we then obtain the basic equation (17) again.

The formulation of (22) clearly sets the weight d_j between zero and one. In the CAPM case when $W_j = W$, d_j equals one. On the other hand, d_j equals zero [and in (17) only the own-variance matters] when $W_j = V_j$. In this case asset j is the only risky asset held by each of the investors in N_j. Under this extreme form of nondiversification, as noted already by Tobin and Brainard [17], there is indeed no reason for the market price of asset j to depend on its covariance with other assets.

The two alternative definitions of d_j in (18) and (22) are quite similar in structure. Apart from the implicit assumption behind (18) that $W_j = (n_j/n)W$, which was discussed above, the major difference between these formulations is the substitution in (22) of the ratio V_j/W for the ratio $b_j Q_j/\bar{Q}$ in (18). The formulation in (22) may seem to have an advantage in that d_j is then more readily estimable; it has the disadvantage, however, of including on the right hand side of the price equation (17) a variable, V_j, which belongs in fact on the left hand side. Furthermore, while in the context of the diagonal model Assumption I.1a, which resulted in (18), seems to be quite sensible (at least as a first approximation), it should be readily admitted that it is much less transparent whether I.1b, which resulted in (22), is as good an assumption in the more general framework of a nondiagonal model.[9]

[9] One possible bias in I.1b is suggested by a generalized diagonal model in which an additional industry-specific factor is incorporated in (11). If investors who hold asset j tend to diversify their portfolios by avoiding other assets in the same industry, then the right-hand side of (19) may exceed the left hand side. If, however, they tend to concentrate their holdings in a given industry (about

V. Asset Valuation with Volume-Related Transaction Costs

In the preceding sections, only fixed transaction costs were assumed to exist. The results can readily be extended, however, to the case where there exists transaction costs which depend on the amount of trade.

A quite general form of the budget equation for investor i is the separable form

$$C_{i0}(q_{i0}) + \sum_{j=1}^{m} C_{ij}(q_{ij}) = w_i, \tag{23}$$

where the cost functions $C_{ij}(q_{ij})$ measure the total cost to this investor of obtaining q_{ij} units of asset j. These cost functions are assumed to be convex and differentiable, *except* at the origin, and to satisfy $C_{ij}(0) = 0$.

A typical cost function $C_{ij}(q_{ij})$ is depicted in the solid curve in Figure 1. To purchase asset j (i.e., for $q_{ij} > 0$), a fixed charge c_{ij}^{+} which is independent of the size of trade is incurred. For a short sale (i.e., for $q_{ij} < 0$), the fixed charge is c_{ij}^{-}. As drawn in Figure 1, both fixed and variable costs are likely to be significantly greater for a short sale, representing the various restrictions and collateral requirements on such transactions. The standard cost function which does not take into account transaction costs is represented in Figure 1 by the dashed line through the origin.

An important special case of (23), represented by the semidotted line in Figure 1, is that of proportional transaction costs where

$$C_{ij}(q_{ij}) = \begin{cases} c_{ij}^{+} + q_{ij}p_j(1 + t_j^{+}) & q_{ij} > 0 \\ 0 & q_{ij} = 0 \\ c_{ij}^{-} + q_{ij}p_j(1 - t_j^{-}) & q_{ij} < 0 \end{cases} \tag{24}$$

Here, in addition to fixed trading costs, there are proportional costs which apply at a rate t_j^{+} to a purchased asset and at a rate t_j^{-} to an asset sold short.

A cost structure $C_{i0}(q_{i0})$ similar to (24) can also be applied to the safe asset, i.e. for $j = 0$. Thus, differences in transaction costs are assumed here to capture the entire differences which may exist between lending and borrowing, or between purchasing and short-selling.

The investor's optimization problem, when the budget constraint (23) replaces (4), is quite similar to the one considered earlier.[10] Given the investor's trade set M_i, and given the assumption of actual trade in the riskless asset, the local first-order conditions corresponding to (5) are

$$E_i x_j - A_i \mathrm{Cov}_i(x_j, y_i^{*}) = C'_{ij}(q_{ij}^{*})/C'_{i0}(q_{i0}^{*}) \qquad j \in M_i. \tag{25}$$

Here the price ratio $p_j/p_0 = (1 + r)p_j$, which appears on the right-hand side of (5), is replaced by the more general ratio between the marginal total cost of obtaining asset j and the marginal total cost of obtaining the safe asset.

For the proportional cost function (24) the term at the right-hand side of (25)

---

which they may have some informational comparative advantage), the inequality might be in the opposite direction.

[10] One added complication, however, is that in the discrete problem of selecting the assets to trade in, investor will have to distinguish between assets according to whether they are to be purchased or to be sold short. Since the discrete part of investors' optimization is not examined here, the existence of this subdivision of assets within M_i is not presented here formally.

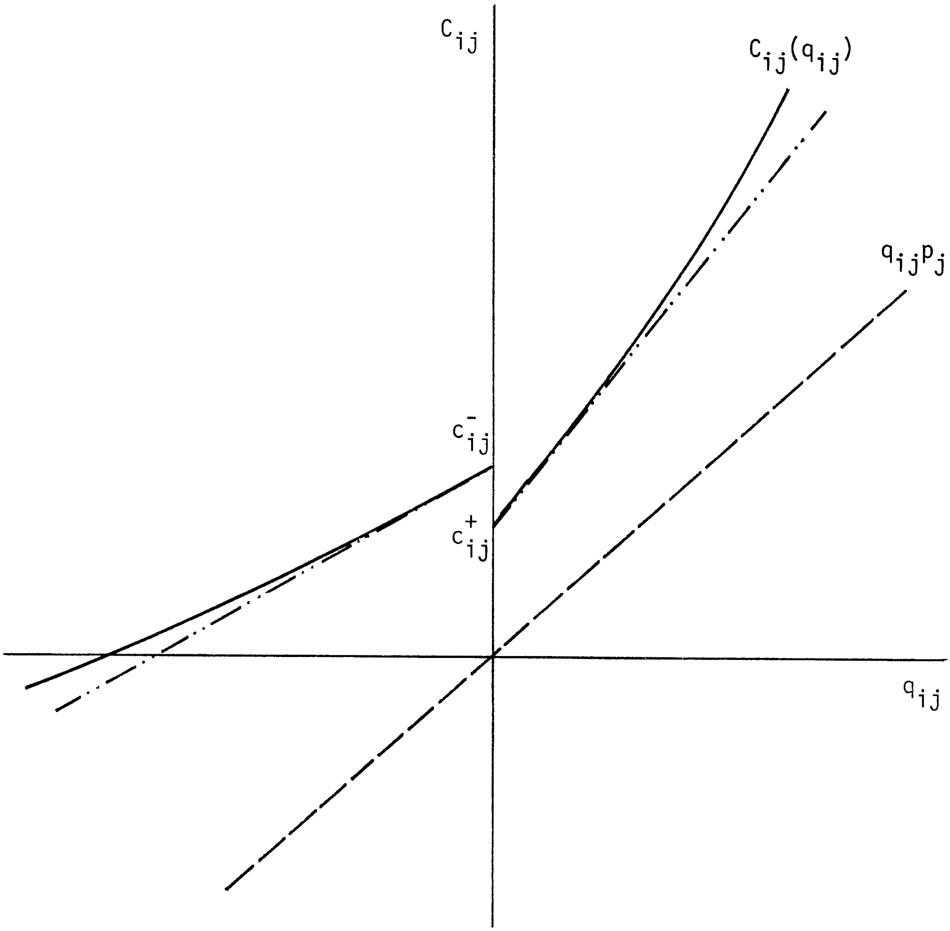

Figure 1. Cost of Trade in an Asset

will depend on whether investor i is a lender or a borrower, and on whether he purchases asset j or sells it short. For a purchased asset, for example, this term will be $(1 + r)p_j(1 + t_j^+)/(1 - t_0^+)$ for a lender, and $(1 + r)p_j(1 + t_j^+)/(1 - t_0^-)$ for a borrower. A useful way of expressing these results is to define new parameters t_{ij}, representing differential transaction costs, such that (25) can be rewritten in the form:

$$E_i x_j - A_i \mathrm{Cov}_i(x_j, y_i^*) = p_j(1 + r + t_{ij}) \qquad j \in M_i. \tag{26}$$

For the proportional cost structure (24) and for an individual i who purchases asset j, we can approximate by $t_{ij} \simeq t_j^+ - t_0^+$ if he is a lender, and by $t_{ij} \simeq t_j^+ + t_0^-$ if he is a borrower. In either case, it is reasonable to assume that $t_{ij} > 0$. (Similar considerations for an asset sold short will indicate that we should expect $t_{ij} < 0$).

At this point it may be significant to note that the assumption that transaction costs occur only at the time of purchase is not restrictive. The costs of holding and selling an asset may be discounted to the initial trading period and incorporated in the purchase costs. Since taxes can be considered as a form of such costs, it can be instructive to examine how taxes can be incorporated in the analysis, and how they might affect the differential transaction costs t_{ij} in (26).

The aftertax payout of a risky asset j may be expressed as $x_j - \tau(x_j - p_j)$, and of the safe asset as $1 - \tau_0(1 - p_0)$, where τ and τ_0 are the respective proportional tax rates (which may clearly depend also on i). The term on the left hand side of (5) and (25) will have to be replaced by $\{(1 - \tau)[E_i x_j - \mathrm{Cov}_i(x_j, y_i^*)] + \tau p_j\}/[1 - \tau_0(1 - p_0)]$. If we were to ignore other value-related transaction costs, the equilibrium condition (5), may then be revised by replacing the right-hand side with $p_j[1 - r(1 - \tau_0)(1 - \tau)^{-1}]$. For the most common case of a lender purchaser, the combined effect of transaction costs and taxes may be represented in the form of (26), with the differential costs t_{ij} approximated by $[t_j^+ - t_0^+ + r(\tau - \tau_0)]/(1 - \tau)$. Even if τ_0 were larger than τ, the total effect of taxes might still be to increase t_{ij}.

The equilibrium conditions (26) can be interpreted as though investors equate the certainty-equivalent yield of each asset to the effective yield of the safe asset. In addition to the riskless interest rate, this effective yield includes a premium t_{ij} which is a measure of the marginal saving in transaction costs due to investing in the safe asset rather than in the alternative risky asset.

The equilibrium conditions (26) for each investor may be aggregated over investors in a manner similar to the one adopted in Section II. One can obtain then a slight variation of Equation (10), in which the term $(1 + r)p_j$ is replaced by $(1 + r + \bar{t}_j)p_j$, where $\bar{t}_j$ is a weighted average of the differential transaction costs t_{ij} incurred by the investors who trade in asset j:

$$\bar{t}_j = \frac{1}{n_j} \sum\nolimits_{i \in N_j} t_{ij}(\bar{A}_j/A_i). \tag{27}$$

The assumptions of Sections III and IV may then be reapplied. This would result in the basic pricing relation,

$$(1 + r + \bar{t}_j)p_j = \bar{E}_j x_j - \frac{\bar{A}_j}{n_j}[d_j \mathrm{Cov}(x_j, Y) + (1 - d_j)Q_j \mathrm{Var}\ x_j]. \tag{28}$$

in which the value-related transaction costs are reflected by the parameter $\bar{t}_j$.

VI. Reformulations

In order to simplify (28) and provide it with a more verifiable content, two further assumptions will now be introduced:

ASSUMPTION I.2: *There are no systematic differences between assets with respect to the degree of risk aversion exhibited by their holders.*

ASSUMPTION I.3: *There are no systematic differences between assets with respect to the marginal differential transaction costs incurred by their holders.*

Assumptions I.2 and I.3 are used here to justify the introduction of a measure of average risk aversion A,[11] and a measure of marginal differential transaction

[11] If instead of assuming the existence of mean-variance preferences we were to *define* parameters A_{ij} so that (5) holds, and then define $\bar{A}_j$ in (7) as the harmonic average of A_{ij}, for $i \in N_j$, all the results up to here would still apply, but then Assumption I.2 would have to be entirely reinterpreted.

costs t, such that

$$\bar{A}_j = A; \bar{t}_j = t \qquad \text{for} \quad j = 1, \cdots, m. \tag{29}$$

Assumptions I.2 and I.3 are both somewhat less than compelling. The possibility that investors with high aversion to risk may predominate among the holders of "blue chips," or that marginal transaction costs may be higher for assets whose volume of trade is small, are only two examples of objections which might be raised against the acceptance of (29).

In order to express the pricing equation in terms of rates of return denote: $1 + R_j = x_j/p_j$ and $1 + \hat{R} = \sum_{k=1}^{m} Q_k x_k/W$. With ER_j standing for the same weighted average of $E_i R_j$ as in (9), and using (29), the pricing condition (28) takes the form (2), where $\lambda = AW/n$ and where the parameters γ_j and δ_j are given by

$$\gamma_j = (1 - d_j)\,\frac{V_j/n_j}{W/n},$$

$$\delta_j = d_j(n/n_j). \tag{30}$$

The standard CAPM result (1) is obtained if $\gamma_j = 0$ and $\delta_j = 1$. To examine the relative magnitude of γ_j and δ_j, we can use (30) and the definition of d_j in (22) to obtain that $\gamma_j + \delta_j = nW_j/n_jW$. On the average, we should thus expect that γ_j and δ_j would sum to unity. The expected return is thus determined by a weighted average of the covariance $\mathrm{Cov}(R_j, \hat{R})$ and the own-variance $\mathrm{Var}\, R_j$. Under the assumption that $W_j = Wn_j/n$ and when V_j/W is sufficiently small, we may approximate d_j of (22) by $d_j \simeq n_j/n - V_j/W$. We can then obtain the further approximation

$$\gamma_j \simeq (1 - \delta_j) \simeq \frac{V_j/n_j}{W/n}. \tag{31}$$

The parameter γ_j may thus be interpreted as some measure of the concentration of ownership of asset j, since its numerator V_j/n_j defines the value of holdings of asset j for its average investor. From (31), $\sum \gamma_j n_j \simeq n$. A crude estimate of the average γ_j can be provided thus by $n/\sum_{j=1}^{m} n_j$, or alternatively, by the inverse of the number of risky assets in the portfolio of the average investor.[12]

In conformity to the standard representation of the CAPM result (1) in terms of "beta," Equation (2) can also be rewritten as

$$ER_j = r + t + \lambda'(\delta_j\beta_j + \gamma_j\alpha_j) \tag{32}$$

where $\lambda' = \lambda\, \mathrm{Var}\, \hat{R}$, $\beta_j = \mathrm{Cov}(R_j, \hat{R})/\mathrm{Var}\, \hat{R}$ and where the new component $\alpha_j = \mathrm{Var}\, R_j/\mathrm{Var}\, \hat{R}$ measures the relative own-variability of asset j.[13]

For empirical purposes, it may be significant to note the possibility of a positive correlation across assets between the parameters γ_j and α_j. As approximated in

[12] Note that if $\bar{m}_i$ is the average number of different risky assets in investors' portfolios, both $\bar{m}_i n$ and $\sum_{j=1}^{m} n_j$ represent the total number of holdings of different risky assets by investors.

[13] Whereas for the CAPM, $\lambda' = E\hat{R} - r$, here, by multiplying (35) by V_j/W and summing over j, one obtains $\lambda' = [E\hat{R} - (r + t)]/(\bar{\delta} + \overline{\gamma\alpha})$, with $\bar{\delta}$ and $\overline{\gamma\alpha}$ as weighted averages of δ_j and $\gamma_j\alpha_j$ respectively. Since $t > 0$ and since on average we should expect $\alpha_j > 1$ and $\gamma_j + \delta_j \simeq 1$, λ' of (32) can be expected to be smaller than the CAPM's $E\hat{R} - r$.

(31), γ_j would be relatively low for companies for whom the size of the average holding is low. These are likely to be companies like utilities and blue chips whose ownership is diffuse over many small investors and whose return is relatively less volatile. On the other hand, there may be no correlation, or (for similar reasons) even a negative correlation, across assets between δ_j and β_j. If this were the case and if (32) were to hold, then in a cross-section regression of ER_j on β_j and α_j (however estimated), when γ_j and δ_j are presumed to be constants, we may find the effect of the α_j factor amplified by the missing parameter γ_j, while the effect of β_j may be dampened by the missing parameter δ_j.

VII. The Assumptions

Except for the introduction of transaction costs, the model here has not departed from the standard CAPM. Existent transaction costs, however, indicate basic weaknesses in other of the CAPM assumptions. It may be instructive to present these assumptions, as they were applied here, more formally.

The key assumptions about the behavior of individuals, which underlined the individual's equilibrium conditions (5) are:

ASSUMPTION A.1: *Each investor is risk averse and seeks to maximize his mean-variance utility of wealth in a single future period.*[14]

ASSUMPTION A.2: *The entire future wealth of each investor consists of the proceeds of assets purchased in the market.*

ASSUMPTION A.3: *Each investor purchases his* entire *portfolio in a single planning and trading period.*

ASSUMPTION A.4: *Each investor purchases every asset in his portfolio separately; (that is, he does not acquire combinations of assets through financial intermediaries).*

ASSUMPTION A.5: *There* exists *a bond market in which each investor either lends or borrows at a given risk-free rate of interest.*

The assumptions about common properties of investors, used for the aggregation formula (10), are:

ASSUMPTION A.6: *All investors trade in a common period and share a common planning horizon.*

ASSUMPTION A.7: *All investors share the same expectations concerning the covariance between assets.*

In addition, it was assumed here that assets are perfectly divisible and that the equilibrium is one of competitive exchange.

All of these assumptions hold also in the CAPM [see, e.g. Jensen [3]]. When there are no transaction costs, Assumptions A.3 and A.4 on the direct purchase of the entire portfolio of assets do not need any further justification. On the other hand, with no transaction costs, investors' investment horizon will have to be taken as very short (infinitesimal), which is clearly not the case.[15]

[14] As noted in note 5, the mean-variance utility assumption may be weakened significantly. What is needed is that at his chosen wealth position, each investor will have mean-variance subjective valuations for all risky assets in which he invests.

[15] The data on trading by individual investors presented by Schlarbaum, Lewellen, and Lease [13], and the examination by Levhari and Levy [5] of the time-horizon which provides the best empirical

As recognized by Markowitz [10] and Tobin [16] and as analyzed recently by Magill and Constantinides [9]. It is transaction costs which cause investors to revise their portfolios only periodically. As these writers further emphasize, transaction costs also cause investors to revise only *part* of their holdings each time.

The existence of a noninfinitesimal investment horizon suggests the necessity of explaining the length of that critical time period, or, more generally, of explaining investors' choice of timing for trade. Here this issue is avoided by the assumption of a predetermined (and identical[16]) investment period. Moreover, also the partial-revision issue is avoided here by Assumption A.3 that investors fully revise their previously-held portfolios. (Otherwise, trade would have been complicatedly dependent on the structure of prior asset ownership).

Another problem created by transaction costs is the need to determine endogenously the composition of investors' partially diversified portfolios. That issue is circumvented here (rather than solved) by the introduction of the independence assumption I.1.

Finally, it should be recognized that it is transaction cost which explains the prevalent practice of indirect purchase of asset-combinations through financial intermediaries (such as mutual funds). The activities of such intermediaries, however, which would have to be made endogenous to the model, has been also ignored here.

Many assumptions of the CAPM which seem innocuous enough when transaction costs do not exist, obtain thus a rather different dimension of significance, and complexity, when such costs are introduced.

VIII. Concluding Comments

A considerable research effort has been directed toward testing the empirical validity of the CAPM pricing formula. Jensen [3], in his survey of that effort, concluded (p. 391) that "the currently available empirical evidence seems to indicate that the simple version of the asset pricing model does not provide an adequate description of the structure of expected security returns." The most damaging empirical evidence, recently cited and reestablished by Levy [6] is that the β factor does not seem to provide a significant explanation of asset prices while the own-variability factor does.[17] The theory presented in this paper provides a potential explanation for this result.[18]

---

fit for the asset pricing equation, suggest that an annual horizon may be considered as generally representative.

[16] The assumption of an identical horizon is used here in justifying the homogeneity of expectations on covariances. Note that for empirical purposes it might be possible to specify different investment horizons for different assets, related for example, to their turnover periods.

[17] In many of the tests, the residual variance was used instead of the variance. If α_j^e denotes the relative variance of the residual return $R_j^e = R_j - \beta_j\hat{R}$, Equation (32) could be rewritten with the term $\delta_j\beta_j + \gamma_j\alpha_j$ replaced by $(\delta_j + \gamma_j\beta_j)\beta_j + \gamma_j\,\alpha_j^e$. In this case one may find, however, that the coefficient of the β factor is statistically significant even if $\delta_j = 0$, that is, even when the β factor has *no* effect on the valuation of assets.

[18] The theory here also provides a potential explanation (which may be considered a variant of existing explanations) to another damaging empirical finding, that the intercept in the regression estimates of the CAPM formula is found to be higher then the risk free interest rate.

The purpose of this paper was to examine the implications of introducing transaction costs to the CAPM. As the discussion above points out, however, many of the complex issues opened up by the existence of transaction costs have not been adequately treated here. Much further research is needed to overcome these difficult problems. More disaggregated data on investors' characteristics and the structure of asset ownership will have to be used. What this paper indicates, I hope, is that this research direction, in spite of the difficulties, has important potential for improving our understanding of financial markets.

REFERENCES

1. M. J. Brennan. "The Optimal Number of Securities in a Risky Asset Portfolio When There Are Fixed Costs of Transacting: Theory and Some Empirical Results." *Journal of Financial and Quantitative Analysis* 10 (September 1975), 483–96.
2. D. Goldsmith. "Transaction Costs and the Theory of Portfolio Selection." *Journal of Finance* 31 (September 1976), 1127–39.
3. M. C. Jensen. "Capital Markets: Theory and Evidence." *Bell Journal of Economics* 3 (Autumn 1972), 357–98.
4. H. E. Leland. "On Consumption and Portfolio Choices with Transaction Costs." In *Essays on Economic Behaviour Under Uncertainty*, eds. M. Balch, D. McFadden, S. Wu. Amsterdam: North-Holland, 1974.
5. D. Levhari and H. Levy. "The Capital Asset Pricing Model and the Investment Horizon." *Review of Economics and Statistics* 59 (February 1977), 92–104.
6. H. Levy. "Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio." *American Economic Review* 68 (September 1978), 643–58.
7. J. Lintner. "Security Prices, Risk, and Maximal Gains from Diversification." *Journal of Finance* 20 (December 1965), 587–615.
8. ———. "The Aggregation of Investors' Diverse Judgments and Preferences in Purely Competitive Security Markets." *Journal of Financial and Quantitative Analysis* 4 (December 1969), 347–400.
9. M. J. P. Magill and G. M. Constantinides. "Portfolio Selection with Transaction Costs." *Journal of Economic Theory* 13 (October 1976), 245–63.
10. H. M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments.* New York: John Wiley, 1959.
11. J. Mayshar. "Investors' Time Horizon and the Inefficiency of Capital Markets." *Quarterly Journal of Economics* 92 (May 1978), 187–208.
12. ———. "Transaction Costs in a Model of Capital Market Equilibrium." *Journal of Political Economy* 87 (August 1979), 673–700.
13. G. G. Schlarbaum, W. G. Lewellen, and R. C. Lease. "Realized Returns on Common Stock Investments: The Experience of Individual Investors." *Journal of Business* 51 (April 1978), 299–325.
14. W. F. Sharpe. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance* 19 (September 1964), 425–42.
15. R. C. Stapleton and M. G. Subrahmanyam. "Market Imperfections, Capital Market Equilibrium, and Corporation Finance." *Journal of Finance* 32 (May 1977), 307–19.
16. J. Tobin. "The Theory of Portfolio Selection." In *The Theory of Interest Rates*, eds. F. H. Hahn, F. P. R. Brechling. London: Macmillan, 1965.
17. ——— and W. R. Brainard. "Asset Markets and the Cost of Capital." In *Economic Progress, Private Values, and Public Policy: Essays in Honor of William Fallner*, eds. B. Balassa, R. Nelson. Amsterdam: North-Holland, 1977.