Will the true marginal investor please stand up?: Asset prices with immutable security trading by investors[+]

Peter L. Swan[*]

University of New South Wales

Version: August 20, 2002

ABSTRACT

A 'Comment' is offered on the important and highly influential Amihud and Mendelson (1986) model of asset pricing incorporating immutable security trading by a continuum of investors/traders. It is made less opaque by overcoming a problem with the numerical simulation. The model has a surprising feature because the marginal investor, on whom the analysis rests, is reluctant to reveal herself. More fundamentally, the model gives rise to a fascinating paradox once it is generalized to allow investor/traders to respond to transaction costs so as to overcome the marginal investor problem: governments benefit everybody by taxing trading out of existence. Amihud and Mendelson (1992) provides a critical signpost required to resolve this paradox.

Key words: asset pricing, liquidity, security trading, transactions cost, equity premium

JEL Classification: G120, G110, G200

[+] © Copyright 2001, 2002 by Peter L. Swan. All rights reserved. Short sections of text, not exceeding two paragraphs, may be quoted without explicit permission provided that full credit is given to the source.

[*] I wish to thank the Australian Research Council Australian Professorial Fellowship and Discovery Grant Program for financial support and Jay Muthuswamy for drawing to my attention to the relevance and importance of the Kane (1994) and Bodie, Kane and Marcus (2002) contributions. I also wish to thank Yakov Amihud for comments. Contact details: Peter L. Swan, School of Banking and Finance, University of New South Wales, Sydney NSW 2052 Australia. Tel. 61 (0)2 9385 5871; Fax. 61 (0)2 9385 6347; Email: peter.swan@unsw.edu.au.

1. Introduction

The path breaking study by Amihud and Mendelson (1986), hereafter A&M, of the impact of transaction costs on asset prices is famous in its own right. The Social Sciences Citation Index indicates that it has been cited 204 times up until August 2002 and it is recognized on the JFE All Star Papers, 1974-1995, Website List with 11.13 cites per year. In addition it has motivated a large body of empirical work. See for example, Eleswarapu and Reinganum (1993), Brennan and Subrahmayam (1996), Barclay, Kandel, and Marx (1998), Chalmers and Kadlec (1998), and Datar, Naik and Radcliffe (1998). It is justly renowned as the first successful attempt to derive asset prices from a specified set of preferences by investors for trading. Alternative approaches have tried to find a motivation for trading based on portfolio rebalancing or life-cycle considerations, for example, Constantinides (1986) and Vayanos (1998), but without attracting the same degree of empirical interest as the A&M model.

This success has come despite the characterization of the model by Kane (1994) as 'quite opaque'. In fact, Kane and numerous editions of Bodie, Kane, and Marcus (2002), hereafter, BKM, aim to provide a simple closed form solution to A&M, so as to exposit the path-breaking contribution to a wider audience. Why is the A&M model opaque? Primarily I believe because A&M provide a misleading characterization of their own numerical solution that is critical to an understanding of their contribution. At first blush their solution appears to include all possible values of the required return on the perfectly liquid asset. That is, one that has no transaction costs. In fact, doubtless to keep things simple, it only pertains to a liquid asset with a dividend stream of $1 per period and an asset price of $1. It is certainly a peculiar asset when the required return is 100% per period. I hasten to add that this simplification, which unbeknownst to the reader forms the basis of their numerical simulations, does not affect their basic conclusions.

Kane's contribution along with BKM is to place the A&M solution explicitly within the context of a continuum of investor types, each with perfectly inelastic trading demands, so as to provide a closed form solution. Unfortunately, the curvilinear relationship they derive I show to be inconsistent with the solution of A&M that is strictly linear. This is because Kane (1994) in his desire for expositional clarity apparently thought it unnecessary to incorporate the actual optimizing solution of A&M into his reformulation of the problem. Consequently, their model has only a tangential relationship to A&M.

This linear solution is strictly defined by the trading characteristics of the marginal investor. If, as is the case, the identity of this marginal investor is unknown, the asset price of each partially illiquid asset is also unknown unless the unique marginal investor can be induced to reveal herself. Thus, not only is there no closed form solution, there is no solution to the problem analyzed by A&M at all unless the marginal investor is arbitrarily chosen. Alternatively and equivalently, there is a solution for each of the infinite number of potential marginal investors. Not surprisingly, if the identity of the marginal investor remains unknown and unless she can be induced to reveal herself, we cannot choose between these myriad solutions. As an illustration, I make one of the four investors in the A&M simulation the marginal investor across all ten asset types in the original simulation. Now only the zero-transaction-cost asset and highest-transaction-cost asset are held in equilibrium by other than the marginal investor and the solution is quite different from the original A&M solution. The relationship between transaction costs and returns is no longer strictly concave as in the original simulation but, being linear, it remains concave/convex.

What can be done to overcome these issues with the A&M approach? One solution commonly used is to dispense with the assumption of a continuum of investor-types and adopt a representative investor approach where by definition the marginal investor is known since she corresponds to the representative average investor. This still leaves the problem that such an investor has a perfectly inelastic trading demand. The inability to obtain a unique solution could also be due to the inflexibility assumption by A&M that trading demand by an investor-type is immutable: each investor trades a specified amount per period irrespective of transaction costs. Perfectly inelastic preferences for trading overall ensures that trading costs are irrelevant for the *amount* of trading activity but the marginal investor is still savvy enough to trade the most cost-effective asset with the lowest transaction costs available. Since most empirical studies show that security turnover is responsive to transaction costs[1], it would appear that the model should be extended to ensure that trading does react to trading costs. This should (a), improve realism and (b), possibly make the solution amenable to a unique solution without requiring an investor to identify herself as the critical marginal investor. To this end, when I introduce well-behaved downward-sloping investor trading demand functions into the A&M model I generate a paradox: a government which the more it taxes investors by raising transaction costs imposed on security traders, the higher are asset prices and the better off are investors. This because there is an interesting feature embedded within the A&M model. It is asymmetric in terms of its treatment of the costs and benefits of trading. Costs are incorporated but there is no need to provide offsetting benefits for trading since the demand to trade is infinite up until the given pre-specified trading quantity at which point it falls to zero. This model characteristic is disguised by the assumption that the desire for trading is so powerful that it is immutable, even in the face of trading costs that may approach infinity.

In fact, a rationality implication is needed within the A&M approach: a requirement that each investor trades until the expected marginal benefit of an additional trade declines to equal the marginal cost of a trade. Subsequently Amihud and Mendelson (1992) hint at how such a

---

[1] See, for example, Demsetz, 1968; Epps, 1976; Jarrell, 1984; Jackson and O'Donnell, 1985; Umlauf, 1993; Aitken and Swan, 1993; and Atkins and Dyl, 1997.

utility maximizing relationship might be obtained. An implication is that the original A&M model cannot be meaningfully generalized by simply incorporating investors with elastic trading propensities without explicitly incorporating the benefits as well as the costs of trading.

Section 2 derives the A&M model by incorporating the insights of Kane (1994). Section 3 considers the solution provided by Kane (1994) and by numerous editions of BKM. Section 4 addresses the identity of the marginal investor. Section 5 relaxes the assumption of immutable trading propensities while some concluding remarks are contained in Section 6.

2. The model of A&M

Following Kane (1994) and BKM there is a continuum of investor types distributed uniformly over a range with an investment horizon h ranging from an upper bound of $\overline{h}$ to a lower bound of $\underline{h}$. The corresponding immutable security turnover μ rates for this distribution of investors ranges from $\underline{\mu}$ to $\overline{\mu}$, where the turnover rate $\mu \equiv 1/h$ with $\mu \in \left(\underline{\mu}, \overline{\mu}\right)$. To keep it as elementary as possible, there are two assets, one a bond or T-bill with zero transaction costs and rate of return ρ and another, equity, with positive transaction costs $\$a \equiv p_e c_e$, given by the product of the asset ask price p_e and the proportional transaction costs or bid-ask spread, c_e. A&M (1986) derive an expression for asset ask-price. The arrival of investors follows a Poisson process with interarrival times and holding periods being stochastically independent. The expected ask-price, p_e, for equity securities equals the expected discounted value of dividends at the constant dividend rate d = *$1* over the random exponentially-distributed horizon, $h \equiv 1/\mu$, plus the expected net receipts from disposal of the asset at the bid-price, $(1 - c_e)p_e$, once the horizon is reached:

$$p_e = E_h\left[\int_{s=0}^{h} e^{-\rho s} d\, ds\right] + E_h\left[e^{-\rho h} p_e (1 - c_e)\right] = \left[\mu + \rho\right]^{-1}\left[d + \mu p_e (1 - c_e)\right] \tag{1}$$

Solving (1) for the ask price results in the A&M ask price for equity of

$$p_e = d/(\rho + \mu c_e), \tag{2}$$

where the turnover rate μ is yet to be specified by an appropriate investor class, and for zero transaction cost bonds with the same dividend rate *d* the perpetuity is valued at

$$p_b = d/\rho , \tag{3}$$

which is both a bid and ask price since the two coincide.

The marginal investor with a horizon of h^m and a turnover rate of μ^m who is indifferent between one equity share or one bond will value the equity at precisely

$$p_e = d/(\rho + \mu^m c_e), \tag{4}$$

on substituting μ^m for μ into Eq. (2). Suppose she devotes her entire wealth \$w to one equity security at a cost of p_e and turns it over, by trading with a person with similar preferences but also the desire to take the other side of the market, at the rate, μ^m. Her present value cost of trading becomes the perpetuity, $\mu^m c_e p_e/\rho$, and her outlay of wealth in its entirety is constituted by

$$w = p_e(1 + \mu^m c_e)/\rho = d/\rho = p_b , \tag{5}$$

on substituting for p_e using Eq. (4) and for p_b using Eq. (3). Hence for the marginal investor the perpetuity cost of the equity and bond purchase are the same. It is necessary to establish this as a requirement for the marginal investor to be capable of purchasing and trading either the equity or bond security.

The gross return or yield on the equilibrium value of equity is

$$\text{Gross return} \equiv d/p_e = \rho + \mu^m c_e , \tag{6}$$

found by rearranging Eq. (4). It depends only on the preferences of the marginal investor with respect to trading equity as well as on the general discount rate and equity transaction cost. The net yield on the equity security takes into account the trading costs per unit of the outlay. The net yield of intra-marginal investors whose preferences dictate less trade than the marginal investor, $\mu \leq \mu^m$, depends on individual trading preferences and is given by an amount in excess of the bond yield:

$$\text{Net yield} \equiv d/p_e - \mu c_e = \rho + c_e\left(\mu^m - \mu\right);\ \mu < \mu^m, \tag{7}$$

and simply the bond yield, $d/p_e - \mu c_e = \rho\,;\ \mu = \mu^m$, for the marginal investor. Note for future reference that the net yield Eq. (7) is linear in trading preferences given by the investor class, μ .

The marginal investor's portfolio is either all equity or all bonds with equity selling at a discount relative to bonds. However, the wealth devoted to equity and bonds is the same and the net return after transaction costs of ρ is also the same. Intra-marginal investors in equity lie to the left of the marginal investor and earn a premium net of transaction costs shown by the downward sloping linear net yield line shown in Fig. 1 with a slope of c_e. Since the premium becomes negative for investors with a greater propensity to trade beyond that of the marginal investor, only bonds free of transaction costs are held to the right of this point. If the marginal investor lies further to the right, indicating a greater propensity to trade, the net yield line will shift vertically to the right but will maintain the same slope with intra-marginal investors receiving a greater net return. However, the observed equity premium or gross return over bonds is the product of equity transaction costs and turnover rate of the marginal investor given by what A&M call the amortized spread, $c_e\mu^m$.

Table 1 shows the solution values for the numerical example provided by A&M (1986, p.229). It includes the set of prices for each security relative to the bond with zero transaction costs assuming a bond yield $\rho = 0.05$. This is the only series I have computed that is not in the original Table. Transaction costs c range from 0 to 0.045 for the ten different asset types. The shaded rectangles represent the highest net yield or return for each of the four investor types with investment horizons ranging from five years for Investor #1 to one month for Investor #4 and with corresponding mandated and immutable turnover propensities ranging from 0.2 to 12. Within the shaded regions each investor-type is indifferent between the respective asset-types. Hence for these assets the investor is the marginal investor indifferent between adjacent assets with different transaction costs and prices. Every column except column 3 is taken directly from A&M (1986, Table 1, p.229). Thus in this solution each of the four investor-types is a marginal investor for at least two assets. The A&M solution is depicted graphically in Fig. 2 for the investors with the highest trading propensity and in Fig. 3 for all investors and the ten asset types. Note that the rays for each of the ten asset types, depicting the net rate of return according to the investor-type located along the horizontal axis, are linear and radiate out from the point of indifference for each marginal investor. The negative return segments will clearly never be chosen by any investor-type.

A&M appear to obtain one set of market clearing prices for the ten assets irrespective of the bond discount rate ρ so long as all prices for assets with positive transaction costs are expressed relative to the price of the asset with zero transaction costs. It is apparent from the ratio of the equity price to the bond price (Eqs. (2) and (3) above) that, apparently contrary to the statements of A&M, the ratio is dependent on the bond discount rate, ρ. Moreover, for a 5% bond return, that is $\rho = 0.05$, the relative prices are as shown in column 3, not as given by A&M in column 4. I find that the price of the bond with zero transaction costs is \$20, the next asset \$9.09 and the last asset \$4.83 with relativities for the 2nd and last of \$0.4545 and \$0.157 respectively. The A&M relativities specified at 0.943 and 0.864, respectively, are much higher and in fact implicitly assume a value for $\rho = 1$ or a required return of 100% per period for transaction cost free bonds.

To see this we need to examine how A&M derive their Table. The relationship between the gross yield and net yield, as derived by A&M, is given by Eq. (6). They use the formula

$$d/p_e \equiv 1/p_e = 1 + \mu^m c_e, \qquad (8)$$

where the dividend d = \$1 and the asset price is defined relative to the price of the asset with zero transaction costs. Hence an asset price of \$1 implies a net return of zero corresponding to the zero transaction cost, an asset price of 0.943 implies a net return of 0.06, and so on. Since Eq. (8) differs from Eq. (6) effectively only by the substitution of 1 for the unknown bond return, ρ, their relationship is only true if the required return on transaction cost free bonds is $\rho = 1$ or 100% per period with a bond paying \$1 in perpetuity with no transaction costs worth only \$1. This explains why an asset with a substantial return of 0.157 in excess of the return on transaction-cost-free bonds sells for a relatively high discounted price of 0.846 rather that the more realistic value of 0.24 with $\rho = 0.05$. This defect in the main simulation table of A&M has contributed to its opacity. Opacity is the main subject of Kane's (1994) comment. Fortunately, it does not affect the validity or otherwise of their findings. More substantive issues are addressed below.

3. The model of Kane and Bodie, Kane and Marcus

A less complex and more intuitive version of A&M (1986) is the aim of Kane (1994) and BKM (2002, 5th ed., pp.279-284, and a number of earlier editions) so as to achieve the ultimate goal of a closed form solution to the A&M problem. The framework and notation is identical to that described above with a continuum of investor types with increasing propensities to trade. They suppose that the gross return on equity and bonds/T-bills is higher by unknown constant amounts, x_e, and x_b, that are proportional to the transaction costs on each class of asset, c_e and c_b. Thus on our equity asset with transaction cost c_e the gross return for investor class μ is

$$\text{Gross return} \equiv d/p_e = \rho + x_e c_e \,. \tag{9}$$

The proportionate amount must be less than unity, $x_e < 1$, according to the authors (Kane, 1994, and BKM, Fourth ed., 1999, p. 269).[2] These authors argue that this is because diversified stock portfolios would dominate the asset with zero transaction costs in terms of net returns. We shall see below that this cannot be correct and in fact in the Fifth edition they change their minds.

By contrast, A&M (1986) specify the gross return as

$$\text{Gross return} \equiv d/p_e = \rho + \mu^m c_e \,, \tag{6}$$

from Eq. (6) above. Since the assumptions of the two sets of authors is for all practical purposes identical, Eq. (9) can only be logically correct if and only if the unknown proportionate amount in Eq. (9) is identically equal to the trading propensity of the marginal investor, $x_e \equiv \mu^m$. Since this marginal trading propensity can take on any value including numbers greater than unity, it is true in this framework that for certain investors with low trading propensities the net return on positive transaction cost assets can dominate the zero transaction cost asset. After all, the positive transaction cost asset is priced at a discount precisely because of its disadvantageous trading costs. Such an outcome is not surprising in, what is after all, a model of specialization, and hence should not be ruled out *a priori*. The net

[2] This inappropriate inequality restriction is maintained in Kane (1994) and Bodie, Kane and Marcus up to the Forth edition, but is relaxed without comment in the latest (Fifth) edition, 2002.

return on taking into account transaction costs for an investor with trading propensity μ now becomes

$$\text{Net return} = d/p_e - \mu c_e = \rho + c_e(x_e - \mu), \tag{10}$$

which should be equal to the A&M expression, Eq. (7), given the identity $x_e \equiv \mu^m$.

For the investor indifferent between the equity asset and bonds or T-bills with zero transaction costs the authors correctly identify that the net return in (10) must equal the bond yield ρ and thus $x_e \equiv \mu^m$, even though this may violate the restriction imposed by the authors that $x_e < 1$ (that is, until the most recent 2002 edition). However, things go awry when the authors equate using Eq. (10), the net return for the equity asset and a bond asset with a low but positive transaction cost c_b:

$$\rho + c_e\left(x_e - \mu_{eb}^m\right) = \rho + c_b\left(x_b - \mu_{eb}^m\right), \tag{11}$$

where μ_{eb}^m is the trading propensity of an investor indifferent between the higher and lower transaction cost asset. They cannot (or should not) do this since the expression on the left in Eq. (11) is actually $\rho + c_e\left(x_e - \mu_{eb}^m\right) \equiv \rho + c_e\left(\mu_{eb}^m - \mu_{eb}^m\right) \equiv \rho$ and the expression on the right is also identically equal to ρ since $x_b \equiv \mu_{eb}^m$. The authors then solve for the apparent unknown x_e in terms of the second unknown x_b as:

$$x_e = \mu_{eb}^m + \frac{c_b}{c_e}\left(x_b - \mu_{eb}^m\right), \tag{12}$$

which is then substituted into the expression for the gross return on the equity asset given by their equation (BKM, 2002, 5th ed., Eq. (9.10), p. 283).

Of course, since we know that $x_e \equiv x_b \equiv \mu_{eb}^m$, Eq. (11) is satisfied as an identity, it is unnecessary to solve for x_e in terms of x_b and other variables. The A&M difference on gross returns comes directly from Eq. (6) above for equity and its counterpart for bonds with transaction costs yielding:

$$d/p_e - d/p_b = \mu_{eb}^m (c_e - c_b), \tag{13}$$

which is the difference in transaction cost times the propensity to trade of the marginal investor. Given that this reflects the period over which the transactions are amortized, the return difference can be thought of as the difference in amortized transaction costs for the marginal investor.

Fig. 4a depicts a graphical representation of the 'closed form' solution of Kane (1994, p. 1181) and BKM (2002, p. 282) for the net yield transformed such that turnover rather trader horizon is displayed on the horizontal axis. Note that the simulated relationship is curvilinear rather than the strictly linear relationship implied by the A&M (1986) solution. The traders endowed with the highest demand to trade are indifferent between bonds/T-bills and the Liquid asset at $\mu_{\rho L}^m$ while investors endowed with a lesser propensity to trade are indifferent at μ_{LI}^m. Investors who have a greater propensity to trade than $\mu_{\rho L}^m$ will trade only T-bills. Intermediate ones trade only the Liquid asset between μ_{LI}^m and $\mu_{\rho L}^m$ whilst those with the least propensity to trade, only invest in and trade the Illiquid asset.

In Fig. 4b, I show the correct linear relationship of A&M (1986) consistent with the simple linear relationships shown in Figs. 1, 2 and 3, rather than the curvilinear one of Kane and BKM. Everything else remains the same. The inability of Kane and BKM to solve for the two unknowns, x_e and x_b , as well as the strange characterization of the solution as curvilinear rather than linear must be of concern. However, they also fail to recognize that their proposed solution involves a third unknown, μ_{eb}^m.

4. Who is the marginal investor?

So far it has been assumed that the marginal investor exists and can be identified out of the continuum of investor types with different trading propensities along the interval, $\underline{\mu}$ to $\overline{\mu}$. If the marginal investor cannot be identified then the analysis is rather like the Cheshire cat in Lewis Carroll's, *Alice in Wonderland*, which disappears on closer examination. The simplest case would be one in which there exists a fixed supply of (say) the asset with zero and positive trading costs respectively. If these assets were to be discreet, such as different types of widgets or products such as motor vehicles, with a discreet number of investors, then supply restrictions could be used to identify the marginal investor. Claims such as financial assets and securities are typically not limited in supply in this way and are generally divisible. Hence it is the case that supply restrictions cannot be used in general to identify the marginal investor.

In A&M's numerical example the marginal trader is identified by a restriction on the analysis to just four investors, each of whom is assumed to be a marginal investor for two or more asset classes. The zero transaction cost asset is in unlimited supply while all other assets are restricted to just one unit. However, there is no need to restrict the analysis in this way.

Table 2 provides an equally valid set of results to those shown in Table 1. This indicates that the results using this model are not unique, and in fact there are an infinite number of potential solutions with a continuum of investor types. Once again the bond yield for zero transaction cost asset is $\rho = 0.05$. Transaction costs c range from 0 to 0.045 for the ten different asset types. The investment horizons range from five years for Investor #1 to one month for Investor #4 and with corresponding mandated turnover propensities ranging from 0.2 to 12. There is by assumption a single shaded region for investor of Type #3 with a turnover rate of 2 (see column 6). This investor-type is indifferent between all the respective asset-types. Equally, investor types #1, #2 or #4 could have become the marginal investor. It can be seen that the results are quite different from those of A&M in Table 1 above.

All investors along the continuum from μ = 12 to μ < 2 hold and trade only the zero transaction cost asset. Investors with turnover rates μ < 2 hold and trade only the highest transaction cost asset since it yields the highest return premium in excess of the bond return. The market return (column 2) now increases linearly with transaction costs so that the strictly concave diminishing marginal returns no longer pertains. The bond security with zero transaction costs yields a return of $\rho = 0.05$. The prices of all assets normalized relative to this asset are shown in column 3. Relative asset prices decline uniformly with higher transaction costs with the rate of decline far more gradual than in Table 1.

The solution provided in Table 2 is depicted graphically in Fig. 5. There are nine equity securities with positive transaction costs and the most costly to trade with a transaction cost c_e=0.045. There is also a bond with zero transaction costs. The marginal investor is assumed to have a horizon of 0.5 and a turnover rate of 2. She is assumed to be indifferent between all ten asset classes. Her portfolio is either all equity or all bonds. The net return after transaction costs of $\rho = 0.05$ is also the same. Intra-marginal equity investors in equity lie to the left of the marginal investor, $\mu^m = 2$, and earn a premium net of transaction costs shown by the downward sloping net yield line with a slope of c_e = 0.045. These investors hold and trade only the highest transaction cost asset that yields the highest return net of transaction costs. No investor with the possible exception of the marginal investor holds or trades equity assets with transaction costs in the range, > 0 or < 0.045. Since the premium becomes negative for investors with a greater propensity to trade beyond that of the marginal investor, only bonds free of transaction costs are held to the right of this point. If the marginal investor lies further to the right indicating a greater propensity to trade the net yield line will shift vertically to the right but will maintain the same slope with intra-marginal investors receiving a greater net return. However, the observed equity premium or gross return over bonds is the product of equity transaction costs and turnover rate of the marginal investor given by the amortized spread, $c_e\mu^m = 0.045 * 2$.

The relations between the returns on the ten assets with transaction costs ranging from zero to 0.045 are exactly linear and are therefore concave but not strictly so. The strictly concave result obtained by A&M is a product of their assumption of a succession of marginal investors. However, as succession of marginal investors allocated in the opposite way to A&M such that investors with higher propensities to trade prefer higher transaction cost securities is not possible. Hence a strictly convex relationship can be ruled out. This finding is supportive of A&M. There are an infinite number of such concave or quasi-concave solutions depending on the assumptions made about the marginal investor.

The key equation describing the A&M equilibrium is the gross rate of return relationship, Eq. (6). The asset price of the equity security is known once the return on bonds is known, together with the dividend rate, the cost of transacting on the equity asset and, most importantly, the turnover rate for the marginal investor type. Equivalently, knowing the asset price and gross yield, together with bond yield, dividend and transaction cost, the trading propensity of the marginal investor and investor-type can be identified. Thus if we know the unique equilibrium we can identify the unique marginal investor. However, unfortunately in its present form the model is not capable of yielding either the asset price or the identity of the marginal investor. Paradoxically, there are too many potential marginal investors and there is nothing in the model to require the true marginal investor to stand up and reveal herself. Hence there is difficulty carrying out comparative-static or similar analyses. Nor can it be used to compute the equity premium in terms of the amortized spread of the marginal investor or to describe how the premium varies as transaction costs increase. In essence, not only is there no ‘closed form’ solution, as Kane (1994) indicated he has found, but I have difficulty finding any solution at all!

5. Dispensing with immutable trading propensities

A possible reason for this inability to solve the model for a unique marginal investor and asset price is the inflexible, stark, and counter-intuitive assumption of a fixed propensity to trade irrespective of the cost of trading. An investor’s propensity to trade must surely depend on the cost of trading such that as trading costs fall as a proportion of the security price, the propensity to trade intensifies. A number of authors have followed this path to generalize A&M including Hubbard (1993). In fact, a closed form solution can be generated with this refinement but, as we shall see, a new issue emerges.

A simple linear (in logarithms) specification that has gained considerable empirical support is the simple constant elasticity of turnover demand formulation:

$$\mu = \alpha c^{-\beta}, \ \mu \in \left(\underline{\mu}, \overline{\mu}\right), \ \alpha \in \left(\underline{\alpha}, \overline{\alpha}\right), \text{ and } \beta > 0, \tag{14}$$

where turnover demand depends on a parameter, α, and a constant demand elasticity, β, for each investor type along the continuum $\mu \in \left(\underline{\mu}, \overline{\mu}\right)$. The demand elasticity remains the same, but the intrinsic liquidity parameter, α, varies along the continuum so that for any given transaction cost parameter, *c*, only the intrinsic liquidity parameter will vary to alter turnover. The A&M and BKM analyses are now special limiting cases as $\beta \to 0$ and demand becomes perfectly inelastic.

Substituting for the marginal equity investor in (14), we have

$$\mu^m = \alpha^m c_e^{-\beta}, \tag{15}$$

and incorporating Eq. (15) into Eq. (6) I have at last derived a unique closed form solution to the A&M (1986) and Kane (1994) problem in terms of transaction costs for the marginal equity investor:

$$\text{Gross yield } d/p_e = \rho + \alpha^m c_e^{1-\beta}. \tag{16}$$

This result preserves the basic result of A&M (1986) in that the equity premium given by $\mu^m c_e \equiv \alpha^m c_e^{1-\beta}$ from Eq. (16) is still the amortized spread of the marginal investor.

It is also possible to carry out comparative-static analyses so long as we treat the intrinsic liquidity parameter, α^m, of the marginal investor as approximately constant or replace it by some average propensity in empirical work. Thus the gross equity asset yield, d/p_e, is increasing in the bond yield parameter, ρ, the intrinsic liquidity parameter, α^m, and at the rate, $1-\beta$, in transaction costs, c_e, on differentiating Eq. (16) with respect to the parameters.

These last results with respect to transaction costs seem sensible at first blush so long as the demand elasticity is bounded from below by zero and from above by unity: $0 \le \beta < 1$. Thus higher transaction costs result in higher yields. However, for the case in which demand becomes elastic, $\beta \to 1$ or $\beta > 1$, either there is no effect or the yield falls and the asset price increases with further increases in transaction costs. This is because the principle underlying the model is that amortized transaction costs act as a discount to the asset price. The higher the amortized spread, $\mu^m c_e = c_e^{1-\beta}$, the lower the asset price. The asset price reflects the present value of an infinite sequence of asset trades, with expected transaction costs being incurred on a continuous basis.

This modified A&M model solution is illustrated in Table 3 and Fig. 6. Simulations using Eq. (16) are made of the turnover rate μ, gross yield and asset price for two different values of the intrinsic liquidity parameter α and trading responsiveness parameter β, with the dividend set at \$1 and the return on assets with zero transaction costs, $\rho = 0.05$. The simulated outcomes with $\alpha = 0.2$; $\beta = 0.5$ at least appear sensible with the yield increasing in transaction costs while the inverse, the asset price, diminishes in transaction costs. The required yield increases in transaction costs and the asset price declines so long as the elasticity $\beta < 1$. The simulated outcomes with $\alpha = 0.002$; $\beta = 1.5$ are peculiar and violate the law that consumption opportunities are finite with the yield reducing in transaction costs while the inverse, the asset price, increasing in transaction costs. The required yield reduces in transaction costs and the asset price increases so long as the elasticity $\beta > 1$. These simulation results indicate the peculiar possibility that a government, which eliminates trading by imposing close to infinite transaction taxes or stamp duty, can boost the consumption stream of investors by raising asset prices.

With a demand elasticity greater than unity, this amortized spread falls as the transaction costs are raised, resulting in a lower present value of transaction costs and hence a higher asset price. Surprisingly, investors become wealthier as transaction costs increase and in the limit as transaction costs approach infinity and trading ceases the amortized spread approaches zero. Thus the price of an asset with infinite trading costs converges to the price of a bond with zero transaction costs as trading costs approach infinity, so long as $\beta > 1$. Such a paradoxical outcome clearly cannot exist. In generalizing the model to enable an internal closed form solution to be obtained and in the process making it considerably more realistic, a more interesting fundamental issue has been exposed.

In fact the A&M model does not generalize for any value of the transaction cost elasticity β, not just the values of $\beta > 1$ which are empirically very likely but have implausible consequences. This is because the model does not set out to express the benefits from trading along with the costs of trading. By assuming that trading propensities are immutable, and hence the desire to trade up to a specified amount is essentially infinite, the opportunity to obtain more powerful results has been missed.

6. Concluding remarks

While there are a number of difficulties in the exposition of the A&M model, both in the original article and in the attempt by Kane (1994) and BKM (2002) to improve on the original exposition, two insurmountable problems remain even when expositional problems are overcome:

(a), the marginal investor on whom the analysis is based cannot be identified unless trading is made responsive to transaction costs, and

(b), explicit attention is paid to the cost of trading but no consideration is paid to the benefits of trading once it ceases to be immutable.

Thus an attempt to overcome the identification of the marginal investor issue via recognition of the fact that trading volume is responsive to transaction costs solves one problem but exposes another. The inability of the model to capture the benefits of trading gives rise to peculiar outcomes.

This means that as matters stand we cannot rely on the main conclusions from the A&M model. In general the equity premium for assets is not given by the difference between the amortized spread of the equity security relative to the bond or T-bill security for the marginal investor, even if the investor can be identified. Moreover, a valid proof is lacking that the relation between transactions costs such as the bid-ask spread and the equity premium is concave. However, it cannot be strictly convex under the A&M assumptions. In a companion piece, Swan (2002), it is shown how the impact of transaction costs on asset prices can be computed, without the serious problems encountered by A&M model, by explicitly modeling the benefits of trading as well as the costs. This paper builds on a very significant additional contribution made by Amihud and Mendelson (1992, section 5, p. 489) when they recognize that their model "underestimates the real cost of the tax to investors" because it neglects the "welfare loss", i.e., the decline in consumer surplus, from increased transactional taxes. This is akin to reduced welfare for users of a toll road who end up taking a costly detour on an untaxed road. Swan (2002), together with Swan and Westerholm (2002), contains a number of empirical tests of the endogenous trading model incorporating the benefits of trading.

References

Aitken, Michael J., Peter L. Swan, 1993. Principal trading and agency trading by brokers is complementary: The welfare effects of deregulating agency trading and banning principal trading. Accounting and Finance 33 (2), 19-42.

Amihud, Yakov, Haim Mendelson, 1986. Asset pricing and the bid-ask spread. Journal of Financial Economics 17, 223-249.

Atkins, Allen B., Edward A. Dyl, 1997. Transaction costs and holding periods for common stocks. Journal of Finance 52 (1), 309-325.

Barclay, Michael J., Eugene Kandel, and Leslie M. Marx, 1998. The effect of transaction costs on stock prices and trading volume. Journal of Financial Intermediation 7 (2) (April), 130-150.

Bodie, Zvi, Alex Kane, and Alan J. Marcus, 2002. Investments, (fifth ed.). McGraw Hill, Boston. Also fourth ed., 1999.

Brennan, Michael J., Avanidhar Subrahmayam, 1996. Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. Journal of Financial Economics 41, 441-464.

Chalmers, John M. R., Gregory B. Kadlec, 1998. An empirical examination of the amortized spread. Journal of Financial Economics 48, 159-188.

Constantinides, George C., 1986. Capital Market Equilibrium with Transaction Costs. Journal of Political Economy 94, 842-862.

Datar, Vinay T., Narayan Y. Naik, and Robert Radcliffe, 1998. Liquidity and stock returns: An alternative test. Journal of Financial Markets 1, 203-219.

Demsetz, Harold, 1968. The cost of transacting. Quarterly Journal of Economics 82, 33-53.

Eleswarapu, Venkat R., Mark R. Reinganum, 1993. The seasonal behavior of the liquidity premium in asset pricing. Journal of Financial Economics 34, 373-386.

Epps, T.W., 1976. The demand for brokers' services: The relation between security trading volume and transaction cost. The Bell Journal of Economics 7 (1) 163-194.

Hubbard, R. G., 1993. Security transaction taxes: Can they raise revenue? Working paper, Catalyst Institute.

Jackson, P. D. and A. T. O'Donnell, 1985. The effects of stamp duty on equity transactions prices in the U.K. Stock Exchange, Bank of England working paper.

Jarrell, G. A., 1984. Changes at the Exchange: The Causes and Effects of Deregulation. Journal of Law and Economics 27, 273-312.

Kane, Alex, 1994. Trading cost premiums in capital asset returns-a closed form solution. Journal of Banking and Finance 18, 1177-1183.

Swan, Peter L., 2002. Can 'illiquidity' explain the equity premium puzzle?: The value of endogenous market trading. Working Paper, School of Banking and Finance, Faculty of Commerce, UNSW, Sydney.

Swan, Peter L., P. Joakim Westerholm, 2002. Asset prices and liquidity: The impact of endogenous trading. Working Paper, School of Business, University of Sydney.

Umlauf, S. R., 1993. Transaction taxes and the behaviour of the Swedish stock market. Journal of Financial Economics 33, 227-240.

Vayanos, Dimitri, 1998. Transaction costs and asset prices: A dynamic equilibrium model. Review of Financial Studies 11 (1) (Spring), 1-58.

Figure 1: The equilibrium solution to the A&M and Brodie, Kane and Marcus problem with a single equity security and a bond with zero transaction cost. The marginal investor's portfolio is either all equity or all bonds with the market value of the equity portfolio less than that of the bond portfolio. However, the wealth required to buy and maintain each portfolio is the same. The net return after transaction costs of ρ is also the same. Intra-marginal investors in equity lie to the left of the marginal investor and earn a premium net of transaction costs shown by the downward sloping net yield line with a slope of c_e. Since the premium becomes negative for investors with a greater propensity to trade beyond that of the marginal investor, only bonds free of transaction costs are held to the right of this point. If the marginal investor lies further to the right indicating a greater propensity to trade the net yield line will shift vertically to the right but will maintain the same slope with intra-marginal investors receiving a greater net return. The observed equity premium, i.e., gross return on equity over bonds, is the product of equity transaction costs and turnover rate of the marginal investor given by the per-period trading costs (i.e., amortized spread), $c_e\mu^m$.

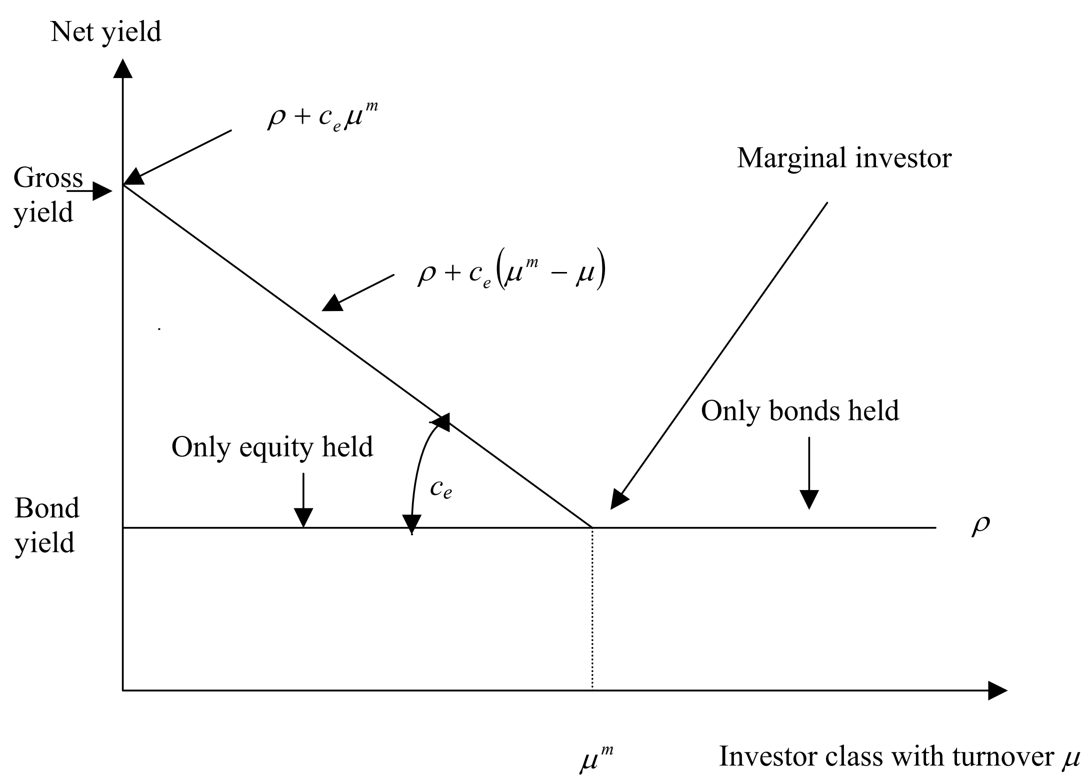

Table 1: The derivation of the numerical solution provided by A&M (1986, p.229) including the corrected set of prices for each security relative to the bond with zero transaction costs assuming a bond yield ρ = 0.05. Transaction costs c range from 0 to 0.045 for the ten different asset types. The shaded rectangles represent the highest net yield or return for each of the four investor types with investment horizons ranging from five years for Investor #1 to one month for Investor #4 and with corresponding mandated turnover propensities ranging from 0.2 to 12. Within the shaded regions each investor-type is indifferent between the respective asset-types so that for these assets the investor is the marginal investor indifferent between adjacent assets with different transaction costs and prices. Every column except column 3 is taken directly from Table 1 of A&M (1986, p.229). Hence in this solution each of the four investor-types is a marginal investor for at least two assets. A&M appear to obtain one set of market clearing prices for the ten assets irrespective of the bond discount rate ρ so long as all prices for assets with positive transaction costs are expressed relative to the price of the asset with zero transaction costs. It is apparent from the ratio of the equity price to the bond price (Eqs. 2 and 3 in the text) that the ratio is dependent on ρ. Moreover, for a ρ = 0.05 the relative prices are as shown in column 3, not as given by A&M in column 4. I find that the price of the bond with zero transaction costs is \$20, the next asset \$9.09 and the last asset \$4.83 with relativities for the 2[nd] and last of 0.4545 and 0.157 respectively. The A&M relativities specified at 0.943 and 0.864, respectively, are much higher and in fact implicitly assume a value for ρ = 1 or a return of 100% per period.

1	2	3	4	5	6	7	8
	Investor number			#1	#2	#3	#4
	Turnover rate			0.2	1	2	12
Trans cost	**Market return in excess of bond yield**	**Price of equity rel. to bond value**	**Price of equity rel to bond value as shown by A&M**	**Net yield for each investor type**			
0	0	1	1	0	0	0	0
0.005	0.06	0.454545	0.943	0.059	0.055	0.05	0
0.01	0.12	0.294118	0.893	0.118	0.11	0.1	0
0.015	0.13	0.277778	0.885	0.127	0.115	0.1	-0.05
0.02	0.14	0.263158	0.877	0.136	0.12	0.1	-0.1
0.025	0.145	0.25641	0.873	0.14	0.12	0.095	-0.155
0.03	0.15	0.25	0.87	0.144	0.12	0.09	-0.21
0.035	0.155	0.243902	0.866	0.148	0.12	0.085	-0.265
0.04	0.156	0.242718	0.865	0.148	0.116	0.076	-0.324
0.045	0.157	0.241546	0.864	0.148	0.112	0.067	-0.383

Figure 2: The Amihud and Mendelson numerical solution to the asset pricing problem for investors with trading/turnover demands ranging from 0.1 to 12 for three asset classes with transaction costs ranging from 0 to 0.01

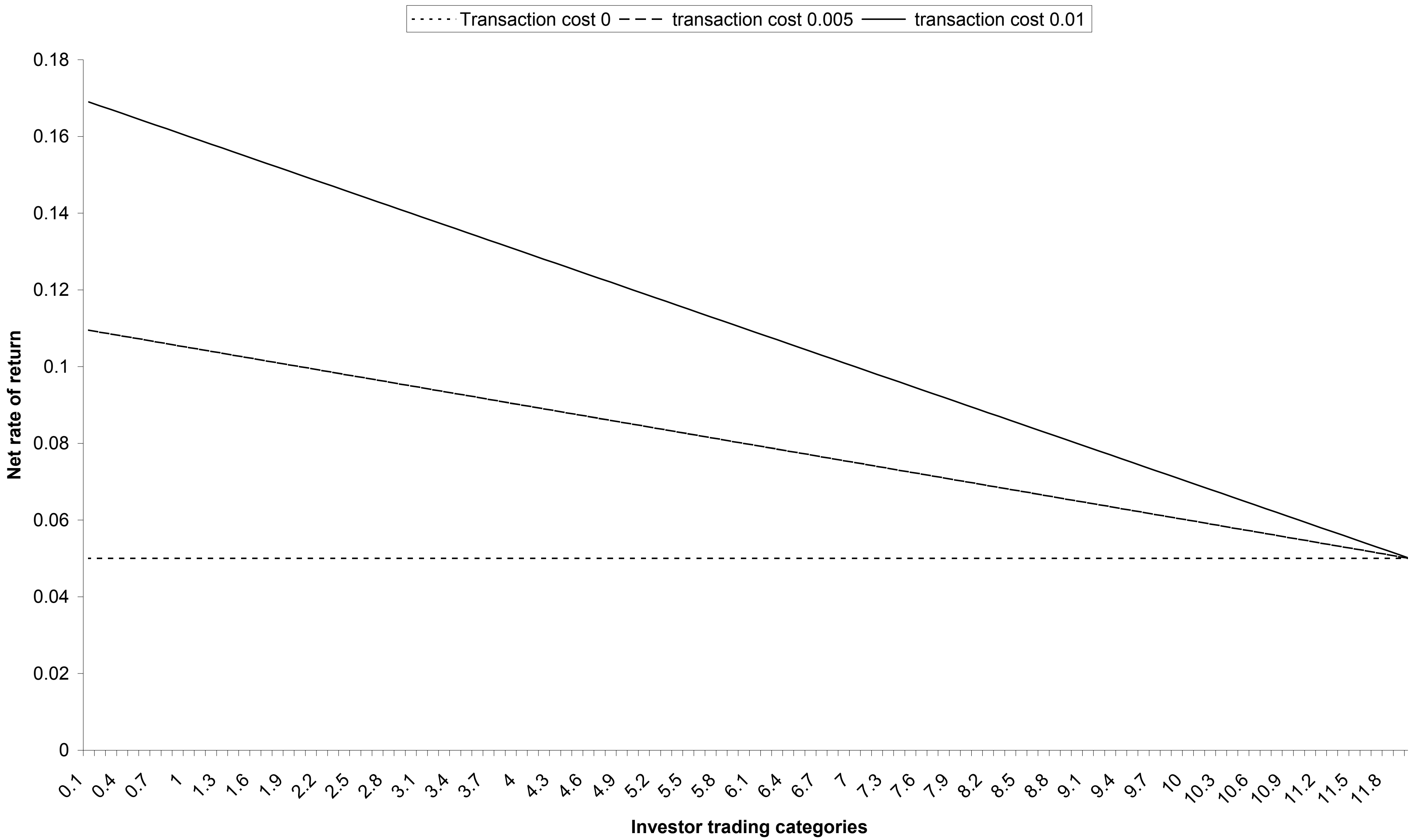

Figure 3: Net rates of return for the 10 asset classes and four investor-types in the Amihud and Mendelson numerical solution. Where the lines cross indicates an investor who is indifferent between the intersecting net returns.

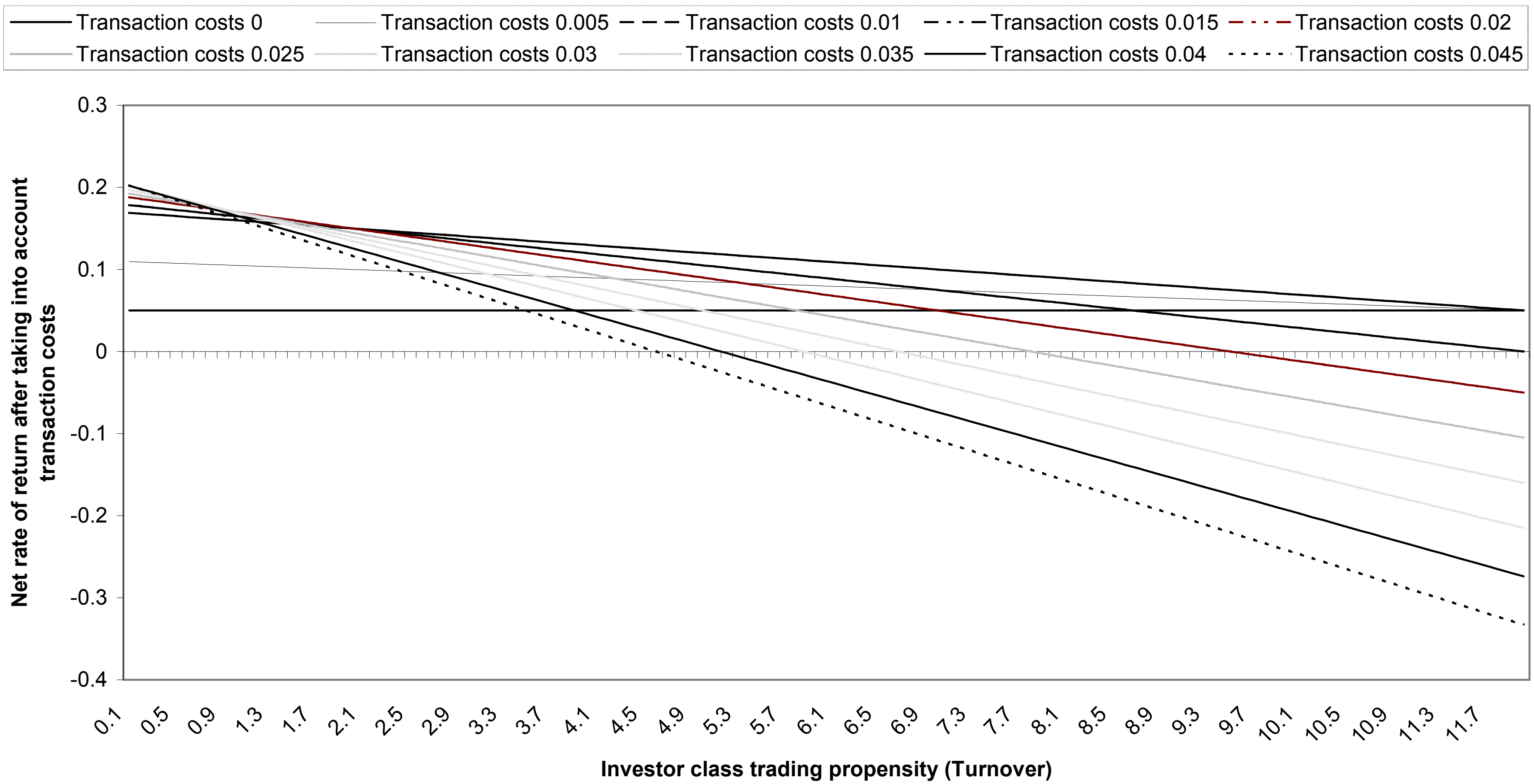

Figure 4a: A graphical representation of the 'closed form' solution of Kane (1994, p. 1181) and BKM(2002, p. 282) for the net yield is transformed such that turnover rather trader horizon is displayed on the horizontal axis. Note that the simulated relationship is curvilinear rather than the strictly linear relationship implied by the A&M (1986) solution. The traders endowed with the highest demand to trade are indifferent between bonds/T-bills and the Liquid asset at $\mu_{\rho L}^{m}$ while investors endowed with a lesser propensity to trade are indifferent at μ_{LI}^{m}. Investors with a greater propensity to trade than $\mu_{\rho L}^{m}$ trade only T-bills. Intermediate ones trade only the Liquid asset between μ_{LI}^{m} and $\mu_{\rho L}^{m}$ whilst those with the least propensity to trade, trade only the Illiquid asset.

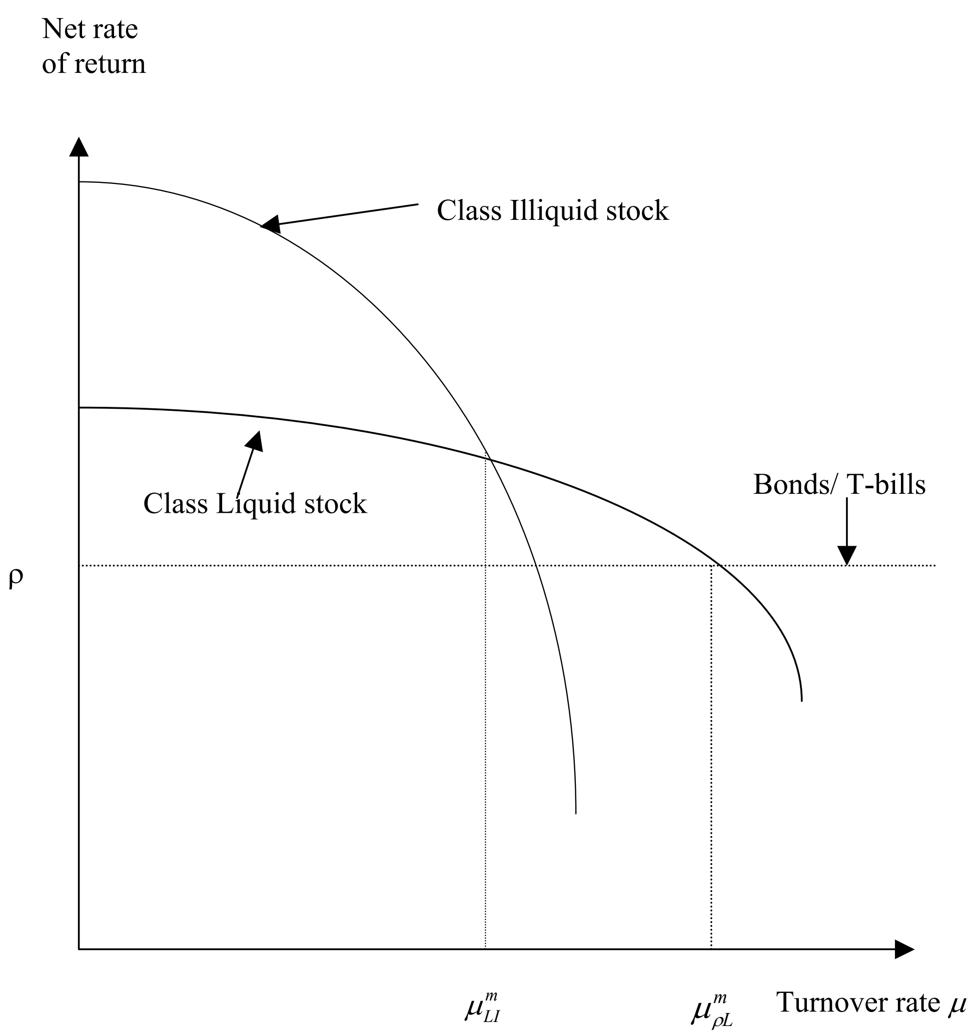

Figure 4b: A graphical representation of the 'closed form' solution of Kane (1994, p. 1181) and BKM (2002, p. 282) for the net yield transformed such that turnover rather trader horizon is displayed on the horizontal axis. I show the correct linear relationship of A&M (1986) rather than the curvilinear one of Kane and BKM. The traders endowed with the highest demand to trade are indifferent between bonds/T-bills and the Liquid asset at $\mu_{\rho L}^{m}$ while investors endowed with a lesser propensity to trade are indifferent at μ_{LI}^{m}. Investors with a greater propensity to trade than $\mu_{\rho L}^{m}$ trade only T-bills. Intermediate ones trade only the Liquid asset between μ_{LI}^{m} and $\mu_{\rho L}^{m}$ whilst those with the least propensity to trade, only invest in the Illiquid asset.

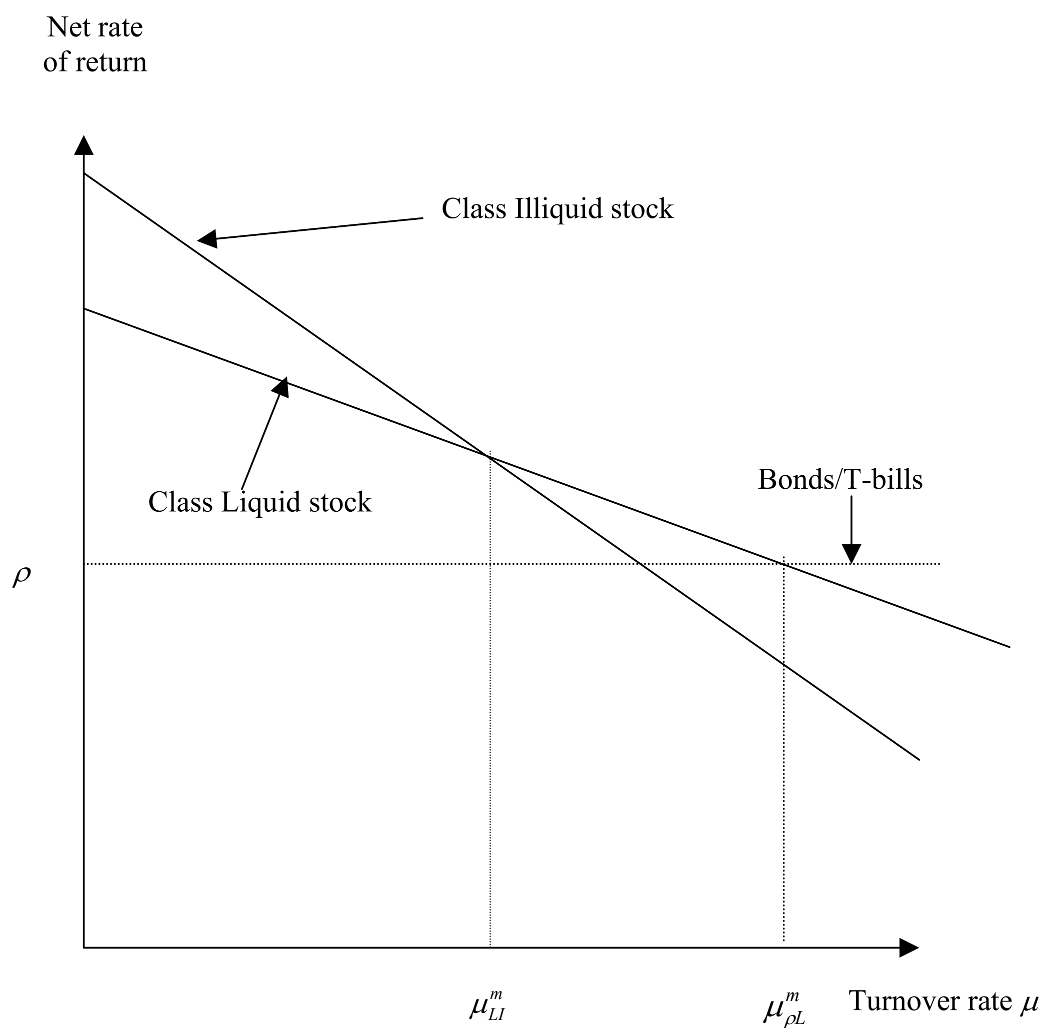

Table 2: This Table provides an equally valid set of results to those shown in Table 1. This indicates that the results using this model are not unique. The derivation of the numerical solution provided by A&M (1986, p.229) including the corrected set of prices for each security relative to the bond with zero transaction costs assuming a bond yield $\rho = 0.05$. Transaction costs c range from 0 to 0.045 for the ten different asset types. The investment horizons range from five years for Investor #1 to one month for Investor #4 and with corresponding mandated turnover propensities ranging from 0.2 to 12. There is by assumption a single shaded region for investor of Type #3 with a turnover rate of 2 (see column 6). This investor-type is indifferent between all the respective asset-types. Equally, investor types #1, #2 or #4 could have become the marginal investor. It can be seen that the results are quite different from those of A&M in Table 1 above. All investors along the continuum from $\mu = 12$ to $\mu < 2$ hold and trade only the zero transaction cost asset. Investors with turnover rates $\mu < 2$ hold and trade only the highest transaction cost asset since it yields the highest return premium in excess of the bond return. The market return (column 2) now increases linearly with transaction costs so that the strictly concave diminishing marginal returns no longer pertains. The bond security with zero transaction costs yields a return of $\rho = 0.05$. The prices of all assets normalized relative to this asset are shown in column 3. Relative asset prices decline uniformly with higher transaction costs with the rate of decline far more gradual than in Table 1.

1	**2**	**3**	**4**	**5**	**6**	**7**
	Investor number		**#1**	**#2**	**#3**	**#4**
	Turnover rate μ		**0.2**	**1**	**2**	**12**
Transaction cost	**Market return in excess of bond yield**	**Price of equity rel. to bond value**				
0	0	1	0	0	0	0
0.005	0.01	0.833333	0.009	0.005	0	-0.05
0.01	0.02	0.714286	0.018	0.01	0	-0.1
0.015	0.03	0.625	0.027	0.015	0	-0.15
0.02	0.04	0.555556	0.036	0.02	0	-0.2
0.025	0.05	0.5	0.045	0.02	0	-0.25
0.03	0.06	0.454545	0.054	0.02	0	-0.3
0.035	0.07	0.416667	0.063	0.02	0	-0.35
0.04	0.08	0.384615	0.063	0.04	0	-0.4
0.045	0.09	0.357143	0.063	0.045	0	-0.45

Figure 5: This Figure illustrates another apparent equilibrium solution to the A&M and Brodie, Kane and Marcus problem describing Table 2. There are nine equity securities with positive transaction costs and the most costly to trade with a transaction cost c_e=0.045. There is also a bond with zero transaction costs. The marginal investor is assumed to have a horizon of 0.5 and a turnover rate of two. He is assumed to be indifferent between all ten asset classes. His portfolio is either all equity or all bonds. The net return after transaction costs of ρ = 0.05 is also the same. Intra-marginal equity investors in equity lie to the left of the marginal investor, μ^m=2, and earn a premium net of transaction costs shown by the downward sloping net yield line with a slope of c_e = 0.045. These investors hold and trade only the highest transaction cost asset that yields the highest return net of transaction costs. No investor with the possible exception of the marginal investor holds or trades equity assets with transaction costs in the range, > 0 or < 0.045. Since the premium becomes negative for investors with a greater propensity to trade beyond that of the marginal investor, only bonds free of transaction costs are held to the right of this point. If the marginal investor lies further to the right indicating a greater propensity to trade the net yield line will shift vertically to the right but will maintain the same slope with intra-marginal investors receiving a greater net return. The observed equity premium or gross return over bonds is the product of equity transaction costs and turnover rate of the marginal investor given by the amortized spread, $c_e\mu^m = 0.045 * 2$. There are an infinite number of such solutions depending on the assumptions made about the marginal investor.

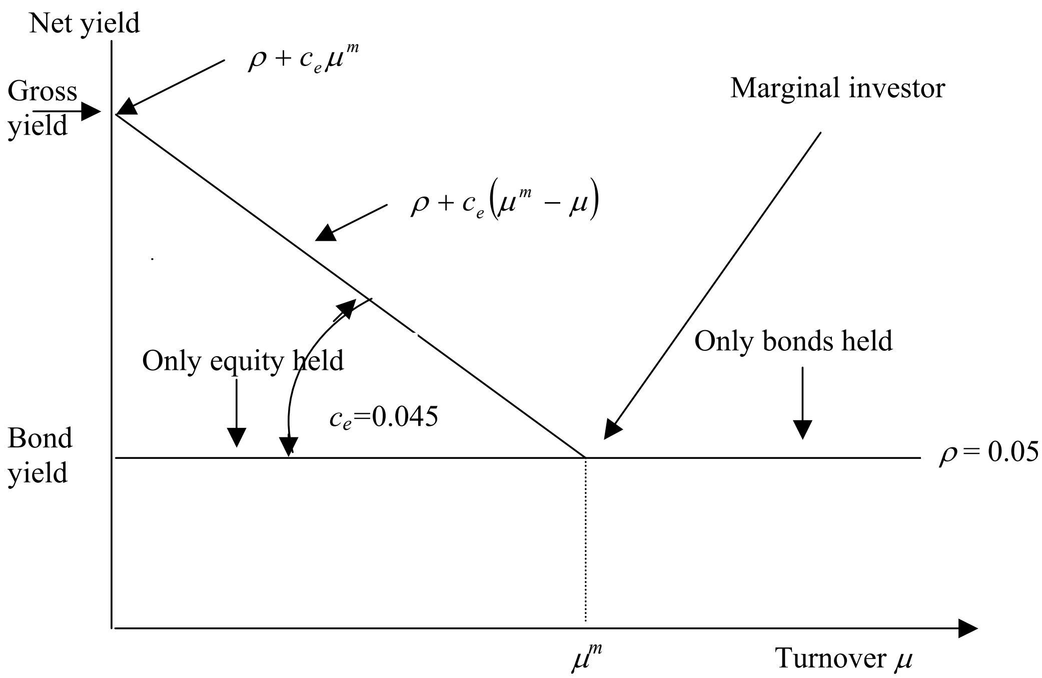

Table 3: **Modeling investors whose trading is responsive to transaction costs.** Simulation of the turnover rate μ, yield and asset price for two different values of the intrinsic liquidity parameter α and trading responsiveness parameter β with the dividend set at $1 and the return on assets with zero transaction costs, ρ = 0.05. The simulated outcomes with α = 0.2; β = 0.5 at least appear sensible with the yield increasing in transaction costs while the inverse, the asset price, diminishes in transaction costs. The required yield increases in transaction costs and the asset price declines so long as the elasticity $\beta < 1$. The simulated outcomes with $\alpha = 0.002$; $\beta = 1.5$ are peculiar and violate the law that consumption opportunities are finite with the yield reducing in transaction costs while the inverse, the asset price, increasing in transaction costs. In fact, the asset price approaches zero and the yield infinity as the transaction costs approach zero for the peculiar case in which $\beta > 1$. The required yield reduces in transaction costs and the asset price increases so long as the elasticity $\beta >$ 1. These results indicate serious flaws in the model.

	Simulated outcomes with α = 0.2; β = 0.5			Simulated outcomes with α = 0.002; β = 1.5		
Transaction Cost, c	**Turnover Rate μ**	**Yield**	**Price**	**Turnover Rate μ**	**Yield**	**Price p_e**
0	Very large	0.0500	20.0000	Very large	Very large	Very low
0.005	2.8284	0.0641	15.5904	5.6569	0.0783	12.7740
0.01	2.0000	0.0700	14.2857	2.0000	0.0700	14.2857
0.015	1.6330	0.0745	13.4237	1.0887	0.0663	15.0762
0.02	1.4142	0.0783	12.7740	0.7071	0.0641	15.5904
0.025	1.2649	0.0816	12.2515	0.5060	0.0626	15.9619
0.03	1.1547	0.0846	11.8146	0.3849	0.0615	16.2477
0.035	1.0690	0.0874	11.4395	0.3054	0.0607	16.4771
0.04	1.0000	0.0900	11.1111	0.2500	0.0600	16.6667
0.045	0.9428	0.0924	10.8194	0.2095	0.0594	16.8271

Figure 6: Simulated asset yields with transaction cost responsiveness of 0.5 and 1.5 based on Table 3.