

Can “Illiquidity” Explain the Equity Premium Puzzle? The Value of Endogenous Market Trading

Peter L. Swan*

School of Banking and Finance, Faculty of Commerce
University of New South Wales
Version Dated September 16, 2002

ABSTRACT

Yes. I aim to establish empirically that the “equity premium” puzzle, with its 6% excess return per annum over Treasury bills for the last 100 years on the NYSE, can be explained once the value of *endogenous* stock market trading is incorporated into investor preferences. Within my framework, investors enjoy trading. According to my model, the “investor surplus” from trading liquid Treasury bills relative to illiquid equity is exactly compensated for by the expected equity premium. Observed transaction cost and liquidity differentials between equity and bills are consistent with the premium. Extensive tests are carried out on Australian and US NYSE data for 1955-98. The puzzle concerning the volatility of the stochastic discount factor also appears to be explained by trading behavior, which is of comparable volatility. Reasonably accurate estimates of transactions costs are extracted just from daily dividend yields and turnover. Transaction costs would need to be 400% higher to explain the premium by the “amortized spread”, together with *exogenous* trading and habit formation. An ability to create unlimited wealth, implicit in some existing models, no longer applies. Additionally, the model explains the further 15-20% pa discount on illiquid “letter” stock.

Key words: equity premium, asset prices, liquidity, trading, transaction cost, amortized spread.

JEL Classification: G120, G110, G200

*School of Banking and Finance, Faculty of Commerce, University of New South Wales, Sydney NSW 2052 Australia; email: peter.swan@unsw.edu.au.

I wish to thank the Australian Research Council Australian Professorial Fellowship scheme for financial support. I wish to thank Michael Aitken, Henk Berkman, Zhian Chen, Stuart Dennon, Stephen Fisher, Kingsley Fong, Gerald Garvey, Tro Kortian, Narayan Naik and Joakim Westerholm, for encouragement and assistance. SIRCA and the ASX kindly provided access to the underlying intra-day data. I also wish to thank Yakov Amihud, John Campbell, Zhian Chen, Doug Foster and other participants at an AGSM seminar, participants at the Reserve Bank of Australia and UNSW Conference seminars, and at the University of NSW, University of Western Australia, University of Sydney and City University of Hong Kong seminars, Magnus Gammelgard, Mikael Gellback, Ed Prescott, Jeff Sheen and David Simmonds for comments on an earlier draft. Responsibility for both errors and views expressed is entirely my own. George Sofianos kindly provided NYSE turnover data and assisted with the provision of TAQ data. Jeremy J. Siegal kindly provided data on bond yields.

© Copyright 2000, 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.

Can “Illiquidity” Explain the Equity Premium Puzzle? The Value of Endogenous Market Trading

ABSTRACT

Yes. I aim to establish empirically that the “equity premium” puzzle, with its 6% excess return per annum over Treasury bills for the last 100 years on the NYSE, can be explained once the value of *endogenous* stock market trading is incorporated into investor preferences. Within my framework, investors enjoy trading. According to my model, the “investor surplus” from trading liquid Treasury bills relative to illiquid equity is exactly compensated for by the expected equity premium. Observed transaction cost and liquidity differentials between equity and bills are consistent with the premium. Extensive tests are carried out on Australian and US NYSE data for 1955-98. The puzzle concerning the volatility of the stochastic discount factor also appears to be explained by trading behavior, which is of comparable volatility. Reasonably accurate estimates of transactions costs are extracted just from daily dividend yields and turnover. Transaction costs would need to be 400% higher to explain the premium by the “amortized spread”, together with *exogenous* trading and habit formation. An ability to create unlimited wealth, implicit in some existing models, no longer applies. Additionally, the model explains the further 15-20% pa discount on illiquid “letter” stock.

Key words: equity premium, asset prices, liquidity, trading, transaction cost, amortized spread.

JEL Classification: G120, G110, G200

“If the profession fails to make progress in understanding the process driving the equity premium, progress on many of the most important problems in finance—proper asset allocation and hurdle rates—are likely to be pyrrhic victories only”—Peter Tufano, Harvard University.

While it is well known that over the longer-term equity outperforms Treasury bills, the enormous magnitude of this out-performance is less well known. Ibbotson and Associates, the "gold-standard" provider of historical equity premium data, show that an investment of \$1 in 1925 would be worth \$5,116 by 1998, whereas an investment in treasury bills would only be worth \$15 (Tufano, 2000). Between 1900 and 1998 the simple geometric mean equity premium for New York Stock Exchange (NYSE) value-weighted stocks was 6.07% pa, utilizing the government (Treasury) bill rate as a proxy for the ‘riskless’ rate.¹ In a celebrated paper Mehra and Prescott (1985) attempt to account for this premium using simulations of an equilibrium model of intertemporal optimization by a representative investor who consumes aggregate consumption, which abstracts from transaction costs and security market trading, liquidity considerations and other frictions. As Campbell (2000) points out, the risk premium depends on the product of the coefficient of relative risk aversion and the covariance between the return on the asset and growth rate of consumption. But the growth rate of consumption in the US has empirically been very smooth. Mehra and Prescott are able to account for only a negligible proportion of this premium with a maximum of 0.4% explained by risk aversion. Only if the degree of risk aversion were implausibly high would their simulations be able to explain the observed premium. A high degree of risk aversion would also create other problems.

My explanation for the equity premium is based around ‘illiquidity’ and is a very simple one. While financial economists are used to thinking of equity markets as being highly liquid, they are in fact highly illiquid relative to government securities such as bonds and Treasury bills. It is of course the difference in yield between the two types of security that constitutes the equity premium. In the US over the 21 year period, 1980-2000, the average turnover rate (securities traded relative to securities outstanding) for US Government bonds and Treasury bills was 13.9 times pa compared with 0.575 times pa for equity on the New York Stock Exchange (NYSE), a relative rate of 24.58 times. See Table 1 for the derivation of these ratios. Most other countries for which data is available also indicate far greater liquidity, i.e., turnover, for bonds than for equity. My basic idea is that the equity premium is no more than compensation to

¹ I update NYSE data used by Fisher (1995) which in turn is based on Schwert (1990). Bill Schwert also provided additional data from his website.

equity holders for this greater illiquidity with turnover and transaction costs acting as the main proxies for liquidity. Of course, informational and adverse selection costs arising from the trading process underlie these differences.

Insert Table 1 here

Investors seem to have a significant desire for liquidity and appear willing to pay more for an asset which is both easy to acquire and dispose of. This is a natural consequence of people's desire to trade and is related to Black's (1986) description of "noise" traders.² I take up Black's challenge when he concludes: "we may need to introduce direct utility of trading to explain the existence of speculative markets" (footnote, p.531). I am able to formulate the precise valuation placed on the ability to trade at specified terms by expressly incorporating the utility benefits from trading into my expression for asset prices and utilizing our knowledge that for the marginal trade, the investor's benefits and costs must be equal. This condition together with the trading relationship relating trading costs and turnover is used to infer the precise benefits accruing from intra-marginal trades. My model endogenizes trading in that I specify precisely the preferences that investors must have for trading, given empirically estimated trading demand functions, and relate such trading behavior to transaction costs. Investor preferences are not arbitrary. In fact by construction they generate precisely the demand for trading relationship that has been found best to describe the data in almost all empirical studies.

Related to the equity premium puzzle is the stock market volatility puzzle (Campbell, 2000). The high volatility in stock prices should be reflected in high volatility in dividends and consumption growth but the volatility in stock returns can best be explained by changes in the equity premium and stochastic discount factor itself. As Campbell (2000) points out, the conditions for the existence of a stochastic discount factor "are so general, they place almost no restrictions on financial data." Since I utilize a representative investor model, which does impose testable restrictions on financial data, I am able to overcome this objection.³ My model incorporating preferences for trading also provides a solution to this companion volatility puzzle.

² "People who trade on noise are willing to trade even though from an objective point of view they would be better off not trading. Perhaps they think the noise they are trading on is information. Or perhaps they just like to trade" (p. 351).

³ Kurz and Beltratti (1997) and Kurz and Motolese (2000) put forward a "rational beliefs equilibrium" in which both market volatility and the equity premium are "explained" by particular endogenous expectations which contradict the ruling paradigm of rational expectations and depart from the representative investor who lies at the heart of the puzzle.

Before I describe my model in detail I review some earlier approaches. While advocates of a psychological approach to decision-making such as Benartzi and Thaler (1995) suggest that “myopic loss aversion” might explain the puzzle, it requires investors to be both irrational and to monitor wealth at annual intervals. Nor is it clear as to how such theories can be empirically tested. Constantinides (1990) has proposed habit formation as the explanation for the premium but Otrok, Ravikumar and Whiteman (1998) point out that it implies an extreme form of risk aversion in high frequencies that is counter-factual. Campbell and Cochrane (1999) propose an alternative model of habit formation that helps to explain stock market volatility but not the equity premium itself because of its reliance on extreme risk aversion (Campbell, 2000; see also Constantinides, 1990).

Not long after this issue was highlighted by Mehra and Prescott, Amihud and Mendelson (1986*b*), (see also Kane, 1994), developed a theory based on variations in exogenous trading levels by different investor classes or clienteles with the potential to explain the equity premium. Equilibrium in the model requires the existence of a marginal investor who is indifferent between trading the same quantity of equity and bonds, even though the costs of trading such securities may differ substantially. The investor must earn a return net of transaction costs given by the yield on transaction-cost free bonds. Hence the gross yield on equity must equal the product of the marginal investor’s trading level or turnover rate and the equity transaction cost. This product is known as the “amortized spread”. Amihud and Mendelson (1986*b*) show that the “equity premium”, or gross return net of the bond yield, is an increasing and concave function of transaction costs as one moves along the spectrum of equity securities from low to high transaction costs. A representative investor version of the Amihud-Mendelson model is developed in Section III below as a limiting case.⁴ Jacoby, Fowler and Gottesman (2000) extend the approach of Amihud-Mendelson by deriving the Capital Asset Pricing Model (CAPM) in after-transaction cost terms. No attempt is made to take account of the benefits from trading. My model presented in Section II below incorporates terms derived from investor preferences for trading which precisely capture these benefits of trading. It is shown in Section III below that recognition of transaction costs without explicit recognition of trading benefits can give rise to internal contradictions.

A number of studies have introduced transaction costs while treating trading as exogenous and thus not determined within the model. The introduction of transaction costs introduces a wedge between gross and net asset returns. Adopting the Generalized Method of Moments

⁴ I provide a fuller treatment of the Amihud and Mendelson (1986) and Kane (1994) models in the companion paper, Swan (2001).

(GMM) technique, Fisher (1994) uses the actual turnover rate which averages about 60% per annum and historic returns from the NYSE over the period 1900 to 1985 to simulate the required transactions cost rate to explain the observed premium. He finds that the contribution of risk aversion is small but the implied transactions cost is implausibly high at between 9.4 and 13.6 %. Marquering and Verbeek (1999) extend Fisher's contribution using deciles constructed from monthly US stock market return data for 1959-1993 while at the same time adopting a consumer habit persistence approach to estimate the implied bid-ask spread of approximately 10%. This is still considerably higher than Stoll and Whaley's (1983) estimate of the two-sided trading cost for the value-weighted NYSE portfolio of 2.79%.

Turnover is treated as exogenous by Amihud and Mendelson (1986b), Fisher (1994) and Marquering and Verbeek (1999) because existing models which incorporate a motivation for trading are unable to account for more than a tiny proportion of actual trading behavior. The high level of actual turnover acts as a severe tax on investor portfolios within this conventional framework. As Barber and Odean (2000) point out, and as shown by Grossman and Stiglitz (1980) for investors who obtain information, rational investors will only trade when the marginal benefit of doing so equals or exceeds the marginal costs. Barber and Odean (2000) attribute the "excessive" trading that reduces returns of individual investors and active mutual funds below a passive benchmark to a behavioral bias, namely "overconfidence". The present study provides a rational explanation based on the desire for liquidity.

The literature I have reviewed so far treats trading as fixed and immutable or exogenous. It is imposed like manna from heaven from outside the model and is therefore unresponsive to transaction costs. Constantinides (1986) computes the illiquidity premium using numerical simulations based on Merton's (1973) inter-temporal asset pricing model with a generalization provided by Lynch (1996). Trading is endogenous to the extent that transactions are motivated by portfolio rebalancing following shocks, investors accommodate increases in transaction costs by "drastically reducing the frequency and volume of trade" (p. 859). The transaction cost premium becomes of second-order magnitude as the responsiveness of turnover to transaction costs, namely the transactions cost elasticity, approaches unity in absolute magnitude. Sensitivity in this model is not high to begin with as the model predicts relatively low turnover rates even with zero transaction costs. There are also two other studies in this tradition by Aiyagari and Gertler (1991) and Bansal and Coleman (1996). Vayanos (1998) models turnover as endogenously generated by life-cycle considerations. He shows that an increase in transaction costs can actually raise stock price. He acknowledges that his results are "surprising and

contrary to conventional wisdom” (1998, p. 26). Since stock price reflects wealth and hence consumption, raising an important element in the cost of consumption perversely enhances it.

Thus the general thrust of the transaction cost literature with induced or endogenous trading is that transactions costs are not very important for asset pricing because the transaction cost elasticity of demand for trading is high and the demand for trading induced by portfolio rebalancing or lifecycle considerations is very weak. The present study reverses these findings by showing that low transaction costs do appear to explain the equity premium for the very reason that earlier studies reject the link: because the absolute value of the transaction cost elasticity is so large. However, instead of endogenous trading arising from portfolio rebalancing or lifecycle reasons, investors are endowed with the precise preferences that engender the trading patterns that have been empirically observed by financial econometricians. Amihud and Mendelson (1986a, 1986b, 1989, 1992) and Amihud *et. al.* (1997) present convincing empirical evidence that transaction costs seem to be important in terms of asset pricing. Moreover, the present study finds that the average equity premium over the period 1900-1998 is quite consistent with Stoll and Whaley’s estimate and the (arithmetic) average turnover rate of 64.6 percent. This is, even with potentially risk-neutral investors and quite substantial Treasury bill trading costs that are assumed zero by Fisher and Marquering and Verbeek.

Clearly, if trading activity is taken as exogenous it cannot explain the observed premium. As Kocherlakota (1996, p.64) points out, if transaction costs for bills were negligible and stocks were on average turned over at an annual rate of 218% then the observed premium of 6.07% could be explained if stock transactions costs were 2.78% because amortized transactions costs could account for the premium. However, the average turnover rate on the NYSE over the period 1900-1998 is far lower at 64.6%. It is simple arithmetic to show that transaction costs must be 9.396%, corresponding to both Fisher (1994) and Marquering and Verbeek (1999), to account for a sufficiently large amortized spread of 6.07%. These implied transaction costs are implausibly high.

Following Amihud and Mendelson (1986b), a large literature has developed explaining the impact of transaction costs on asset prices. Eleswarapu and Reinganum (1993) find only limited evidence of a relationship. Brennan and Subrahmayan (1996) find evidence of a significant effect due to the variable cost of trading after controlling for factors such as firm size and the market to book ratio. Recognizing that there is considerable variation in turnover rates, Chalmers and Kadlec (1998) use the amortized spread instead of the bid-ask spread for Amex and NYSE stocks over the period 1983-1992. They find stronger evidence that the amortized spread is priced than for the simple bid-ask spread.

Datar, Naik, and Radcliffe (1998) use the turnover rate as a proxy for liquidity, instead of the bid-ask spread or amortized spread, on monthly return data for NYSE stocks over the period 1962 to 1991. After controlling for the Fama and French (1992) factors consisting of firm size and the book-to-market ratio and CAPM firm betas, they establish that liquidity plays a significant role in the cross-section of returns. As liquidity rises from an illiquid stock (the 10th percentile) to a liquid stock (90th percentile) the difference in turnover is about 72% per annum and in returns is around 3.25% per annum. The liquidity effect is stronger than the size effect. In section V below it is demonstrated that the magnitude of the liquidity effect is more than sufficient to explain the equity premium given the far greater liquidity of Treasury bills. In developing countries which have raised liquidity by opening up domestic stock markets to foreign ownership there has followed a substantial investment boom (Henry, 2000). This is consistent with a substantial lowering of the equity premium and hence the cost of capital.

While several of the empirical studies I have reviewed set out to test an implication of the clientele model of Amihud and Mendelson (1986b), their findings are not dependent on a classification of investors into short-term and long-term traders with perfectly inelastic trading demands which is the basis of the original model. That is, none of the empirical tests of the Amihud and Mendelson model have established the existence of investor clientele effects or the existence of investor/traders with inelastic trading demands.

There is a considerable literature on transaction costs and security turnover which is contrary to the notion from Amihud and Mendelson (1986b) that trading is immutable and unresponsive to transaction costs. It begins with Demsetz (1968) who found that transaction costs are inversely related to measures of trading volume. Others who obtained similar results include Epps (1976), Jarrell (1984), Jackson and O'Donnell (1985), Umlauf (1993), Aitken and Swan (1993, 1999), and Atkins and Dyl (1997). The latter attribute their result to Amihud and Mendelson's (1986b) clientele effect due to differences in investor horizons. Equally plausibly, investors simply trade more when transaction costs are lower. Barclay, Kandel and Marx (1998) use evidence from the transfer of stock from one exchange to another to show that volume is inversely related to spreads but the price effects are not significant. The transaction cost sensitivity literature is closely paralleled by the one on the impact of transaction tax changes. See, for example, Keifer (1990), Campbell and Froot (1994) and Schwert and Seguin (1993), for surveys.

Despite the huge literature which has been spawned by the Mehra and Prescott findings, the equity premium is still "largely a mystery to economists" (Kocherlakota, 1996, p. 42). The main innovation of Mehra and Prescott was to focus on a representative investor holding the entire

market consisting of equity and treasury bills. The equity premium puzzle arose, according to the model developed below, because the investor is not representative in terms of voluntary trading in both equity and Treasury bills. In my framework the representative investor *trades* the entire market for equities and bonds as well as holding them.

The plan of the paper is as follows: Section I incorporates liquidity into the representative investor's utility function to derive a liquidity-based model of endogenous turnover. The testable and comparative-static implications of the model are developed in II. III provides a representative agent model based on Amihud and Mendelson (1986b) that is contrasted with the liquidity based model. IV reinterprets the existing empirical results of Amihud and Mendelson (1986a) while V provides an estimate of the equity premium based on the relationship between returns on the New York Stock Exchange (NYSE) and liquidity, 1962-1991. VI applies the model to a large cross-section of monthly returns on the Australian Stock Exchange (ASX). In VII both the endogenous trading and exogenous trading models are simulated based on the results from ASX returns and trading. In VIII evidence from discounts on illiquid "letter stock" is used to provide further support for the model of endogenous trading. In IX 43 years of daily dividend yields on the NYSE are used to fairly accurately extract the implied transaction costs over the period 1955-1998. In X it is demonstrated from a cross-section of returns from 90 daily series on the ASX that the same turnover elasticity can be derived from both turnover and returns data. Conclusions are presented in the final section while special cases of the liquidity-based utility functions are derived in the Appendix.

I. A Utility-based Model of the Illiquidity Premium with Endogenous Trading

The aim of this section is to develop a model of asset pricing which takes proper account of transaction costs. Transaction costs, $0 < c < 1$, as a proportion of the asset price for assets could just consist of the competitively-determined relative bid-ask spread in the securities market as in Amihud and Mendelson (1986b). Additionally, it could also include the two-sided brokerage charge, the sum of market impact costs on the buy and sell side, a government-imposed equity market transaction tax and any other impediments to transacting. Effectively, the supply side of the model is provided by the brokerage sector with free-entry and constant returns to scale which generates transactions at unit cost cp^a , where p denotes the asset price and the superscript a indicates the ask-price.⁵ Consider an asset-pricing model such that there is a single representative investor type and just two assets. Both assets generate the same perpetual cash flow of payments/earnings/dividends $\$D$ per unit of time ($D > 0$). Let r denote the constant

⁵ Since the ask-price is endogenous and ultimately a function of transaction costs it is really the dollar transaction cost given by the unit transaction cost, cp^a , which is exogenous rather than the proportional cost, c .

return on the numeraire asset (Treasury bills). Hence $D = r$ and the perpetuity value of Treasury bills, $p_b^a \equiv (D/r) \equiv 1$.

The equity security to be valued incurs a constant round-trip transaction cost, c_e , as a proportion of its asset ask-price, p_e^a . The numeraire asset (liquid Treasury bills) incurs a low transaction cost, c_b , with $c_b \leq c_e$, as a proportion of the numeraire asset price, $p_b^a \equiv 1$. In general, an asset's quoted asking price is $p^a \equiv p(1+a) \equiv p[1+c/(2-c)]$ and bid price $p^b \equiv p^a(1-c) \equiv p(1-a) \equiv p[1-c/(2-c)]$, where p is the mid-point price and the half-spread relative to the mid-point price is $a = c/(2-c)$.

The representative investor expects to purchase a proportion s/n of the fixed number s of shares on issue of the asset at the quoted asking price p_e^a , where n is the (large) number of identical investors. Similarly, the investor purchases a proportion b/n of the total number of Treasury bills, b . For every representative purchasing investor there is an identical representative selling investor so that the net trading demand is zero.⁶ The arrival of investors follows a Poisson process with inter-arrival times and holding periods being stochastically independent. These assets are held over a random, exponentially-distributed horizon time, h , with the amount realized from sale depending on the bid-price, $p^a(1-c)$. The mean horizon of the representative investor is $E_h(h) = 1/\tau(c)$, where E_h is the expectations operator, the turnover rate, τ , is the mean proportion of securities turned over by investors in each period. This proportion is a function of transaction costs, $\tau = \tau(c)$, with $\tau_e \leq \tau_b$ since $c_b \leq c_e$.⁷ The expected investor horizon is increasing in transaction costs. Hence the turnover rate is diminishing in transaction costs, $d\tau/dc = \tau'(c) < 0$, at a diminishing rate, $\tau''(c) > 0$. Since there are n identical investors, each with an expected holding s/n of the available supply of s shares, the expected per period sales/purchases of each investor will be $\tau(c)s/n$ in the expectations equilibrium. The amortized spread is $\tau(c)c$ since the per-period transaction cost per share is c/h over the horizon, h . The expected market clearing ask-price, p^a , is given to each of the large number of investors so that investors act atomistically.

⁶ Although not attempted here, an overlapping generations model could be developed to rationalise this assumption of matching buyers and sellers.

The utility benefit of a trade must at least equal its cost. Logically, the *costs* of transacting cannot be incorporated into an asset valuation model without at the same time also incorporating the *benefits* of transacting. By how much does an investor value the liquidity of an asset, or, alternatively, discount illiquidity relative to a more liquid asset? Just as it is possible to put the liquidity of money into a utility function, it is also possible to express the utility derived by the representative investor for assets with positive transaction costs. Utility can be expressed equivalently in each of two ways: first, as a “direct” function of its ultimate arguments, or, second, as its “indirect” dual involving prices and wealth. In direct form the representative risk adverse investor’s utility function defined over current and future values of consumption is

$$\max_{s, b, \tau_e, \tau_b} u(y_t, y_{t+1}) = u(k_t) + E_t \{u[k_{t+1} + si(\tau_e)p_{e,t+1}^a + bi(\tau_b)p_{b,t+1}^a]\}, \quad (1)$$

where $y_t = k_t$ and $y_{t+1} = k_{t+1} + si(\tau_e)p_{e,t+1}^a + bi(\tau_b)p_{b,t+1}^a$ are the aggregate or augmented flow of consumption inclusive of any benefits derived from trading securities in periods t and $t+1$, E_t is the expectations operator, k_t is the flow of ordinary goods and services from other sources of income and from the investment, τ_e, τ_b are the respective random turnover rates for equity and Treasury bills with $\partial u / \partial k > 0$, s and b , respectively, are the number of equity shares and Treasury bills held in the representative investor’s portfolio and $i(\tau_e), i(\tau_b)$, are the investor liquidity benefit functions for equity and bonds respectively, with $i'(\tau) > 0$. Expected utility of augmented consumption can take the convenient power utility form:

$$u(y_t) = \frac{1}{1 - \rho} y_t^{1 - \rho}, \quad (2)$$

where ρ is the coefficient of relative risk aversion.⁸ To keep the notation simple, the respective share and bond holdings are not deflated by the number of identical investors, n .

⁷ See Table 1 above and section V below for evidence that the turnover rate for government bonds and gilts is many times that of equity in the US, Australia and the UK. Moreover, the transaction costs such as the bid-ask spread incurred in government bond and Treasury bill trading are exceedingly small relative to equity in most markets.

⁸ While risk adverse preferences are assumed, empirically it may be the case that not a great deal would be lost with an assumption of risk neutrality. It would be consistent with the findings of Mehra and Prescott (1985) and Fisher (1994) who found no significant role for risk aversion in explaining the equity premium. It is also consistent with Fama and French (1992) who find no evidence that stocks with a higher covariance with the market portfolio (higher Capital Asset Pricing Model, CAPM, beta) earn a higher return after controlling for firm size and the book to market ratio. Datar, Naik, and Radcliffe (1998) obtain results that are even less favorable to the CAPM model. They control for the CAPM beta when estimating the impact of liquidity on returns. While the effect of beta is

The use of turnover as a general measure of liquidity with its positive partial derivative is justified on the grounds that a higher turnover permits the individual to adopt a shorter investment horizon, h , so that the investment is locked in for a relatively shorter period. Consequently, the liquidity benefit functions in equation (1) could also be expressed as $i'(h) < 0$. Perhaps the closest analogy is with the liquidity services provided by money. The diminution of the “lock in” effect from high transaction charges and low turnover will, by raising liquidity, enable investors to make opportunistic consumption decisions at favorable times and on favorable terms, in much the same way that a liquidity premium attaches to money.⁹ The expected return on real money balances is negative and given by minus the expected inflation rate. Since the liquidity services of money are so valuable, real balances remain positive even when it is very costly to hold money as in periods of high inflation. Government bonds and Treasury bills are less liquid than real money balances and on average earn approximately a zero real return over long time periods. Equity, being less liquid again, earns an expected positive return given by the equity premium. By analogy with money demand in which real money balances appear in the utility function, for example, Friedman (1969) and Samuelson (1947), the specification could also be justified as a reduced form of some more complex model of investor behavior.¹⁰

A related group of models in which liquidity is valuable arises in market microstructure from Kyle (1985) and Holmstrom and Tirole (1993) in which the amount of information released by an informed trader monitoring executive performance depends on the magnitude of the signed order flow stemming from uninformed or ‘noise’ traders. Informed traders can hide

statistically significant, the sign is negative which is the opposite of what theory predicts. Although Datar, Naik and Radcliffe use similar CRSP data to that of Fama-French, their generalized least-squares (GLS) econometric methodology differs.

⁹ The ability to turn over an asset or portfolio may enable higher expected returns to be earned so that consumption k is higher than it otherwise would be. Longstaff (1995) shows that if an investor has perfect stock timing ability, the penalty faced by the holder of a volatile non-marketable asset can be very severe. There is a problem with this argument in the context of a model of the representative investor. Trading ability as reflected in market timing, being a zero-sum game, must result in offsetting losses to the trader taking the other side of the market. Thus a “representative” investor should not be able to gain an advantage in this way. I wish to thank Gerald Garvey for pointing this out.

¹⁰ Feenstra (1986) has shown that entering money into liquidity costs, which appear in the budget constraint, can be functionally equivalent to incorporating real money balances into the utility function.

more effectively in the crowd. Empirically supportive of this liquidity driven view of performance monitoring by informed traders or ‘speculators’ are the findings of Garvey and Swan (2002). The main factor determining the allocation of options and other market-based incentives to executives is stock turnover or liquidity. The liquidity of the company’s stock plays the major role explaining the extent to which boards delegate company monitoring to the market. Thus the more liquid the stock the more precisely stock price reflects fundamentals and the more valuable is its use in incentive contracts.

Utility (or cost of illiquidity) per dollar of the asset is assumed for the sake of simplicity to be additively separable in the twice-differentiable concave benefit of liquidity function, $i(\tau)$, evaluated as $i(\tau_e), i(\tau_b)$ for equity and Treasury bills respectively, with benefits increasing in liquidity, $i'(\tau) > 0$, and diminishing returns to increasing liquidity, $i''(\tau) < 0$. In general, the benefit of liquidity, $i(\tau)$, is additive to regular cash flows. These benefit functions are the embodiment of Black’s (1986, p.531) utility or joy of trading. Hence equity which is less liquid than Treasury bills, $\tau_e < \tau_b$, will exhibit a lower benefit of liquidity, $i(\tau_e) < i(\tau_b)$. Examples of liquidity benefit functions satisfying the requirements includes the power function evaluated for both equity and bonds at τ_e and τ_b respectively:

$$i(\tau_e) = \tau_b c_b + [\alpha^\mu / (1 - \mu)] (\tau_e^{1-\mu} - \tau_b^{1-\mu}) \text{ for } \alpha > 0 \text{ and } \mu > 0 \text{ and } \neq 1$$

$$i(\tau_b) = \tau_b c_b \tag{3a}$$

and the log function:

$$i(\tau_e) = \tau_b c_b + \alpha \ln(\tau_e / \tau_b) \text{ for } \mu = 1 \text{ and } \alpha = \tau_b c_b \equiv \tau_e c_e$$

$$i(\tau_b) = \tau_b c_b \tag{3b}$$

where α denotes the ‘intrinsic liquidity’ parameter and $\mu \equiv 1/\beta$, the inverse of the absolute value of the elasticity of turnover with respect to transaction costs, such that turnover $\tau = \alpha c^{-\beta}$ and $\ln(\cdot)$ denotes the natural logarithm. These two functions are derived in the Appendix from two requirements: (a) that the marginal benefit from trading just declines to unit transaction cost for the marginal trade; and (b) that the turnover demand function take the standard constant elasticity of demand formulation given by (7a) below. Thus an additional, and perhaps the most

telling, case for the approach adopted is that the utility specification generates precisely the observed form of the security turnover demand found in numerous empirical studies.¹¹

For convenience but without loss of generality normalize the price of bonds at unity: $p_b^a \equiv D/r \equiv 1$, so that cash payment/dividend $D \equiv r$. Now consumption of ordinary goods and services in period t is:

$$y_t = k_t = e_t - sp_{e,t}^a - b, \quad (4a)$$

and in period $t+1$:

$$k_{t+1} = e_{t+1} + sp_{e,t+1}^a (1 - c_e \tau_e) + sD + b(1 + D - c_b \tau_b), \quad (4b)$$

where e_t and e_{t+1} are the respective non-investment income available in each period. Period t consumption is reduced by the cost of the purchase of the equity and bond portfolios but in the following period consumption is boosted by the dividend payoffs from both equity and bonds, as well as the values of the equity and bond portfolios in period $t+1$ but is reduced by the cost of the trading activity in both equity and bonds. Thus $\tau_e c_e p_{e,t+1}^a$ and $\tau_b c_b p_b^a \equiv \tau_b c_b$ represents the trading costs or amortized spreads associated with equity and government securities, respectively.¹²

Maximization of (1) with respect to the portfolio and trading choices together with the budget constraint yields four first-order conditions, of which the first with respect to the equity portfolio, s , is:

$$\partial u(\cdot, \cdot) / \partial s = - (\partial u / \partial y_t) p_{e,t}^a + E_t \left((\partial u / \partial y_{t+1}) \{ p_{e,t+1}^a [1 + i(\tau_e) - c_e \tau_e] + D \} \right) = 0, \quad (5a)$$

and the second with respect to bond or T-Bill purchases, b , is:

$$\partial u(\cdot, \cdot) / \partial b = - (\partial u / \partial y_t) + E_t \left((\partial u / \partial y_{t+1}) \{ [1 + i(\tau_b) - c_b \tau_b] + D \} \right) = 0. \quad (5b)$$

Without loss of generality, the numeraire asset, a bond or T-bill with turnover, τ_b , and a lower-bound transactions cost, c_b , has a zero net utility of liquidity, $i(\tau_b) - c_b \tau_b \equiv 0$, forming part of

¹¹ A possible criticism associated with incorporating a preference for trading is that it may be ‘ad hoc’. It is clearly not at all arbitrary and in fact requires a very precise form to generate the findings of numerous empirical studies.

¹² Critics of monetary theorists who place real money balances in the utility function might be happier to see it appear in the budget constraint instead. In the same vein we can remove the terms involving $i(\tau)$ in the inter-temporal utility function (1) and, instead, incorporate them as additive terms in the budget constraint (4b). Either way the first-order conditions (5a) and (5b) below will be unaffected.

the term in square brackets in (5b), by definition.¹³ Moreover, since the bond dividend, D , equals its yield, r , (5b) implies that $p_b^a \equiv 1 = E_t[\frac{u'(y_{t+1})}{u'(y_t)}(1+r)]$ so that the expected marginal rate of substitution in augmented consumption:

$$E_t[\frac{u'(y_{t+1})}{u'(y_t)}] = \frac{1}{(1+r)}. \quad (5c)$$

On substituting for (5c) in (5a) and setting the expected ask price of equity in period $t+1$, $E_t(p_{e,t+1}^a)$ to its initial price, $p_{e,t}^a$, in (5a) and solving for this initial price yields:

$$p_{e,t}^a = D/\{r - [i(\tau_e) - \tau_e c_e]\} \equiv D/(r + ep), \quad (6a)$$

where $ep \equiv -[i(\tau_e) - \tau_e c_e]$ is the expected equity premium. Dropping the time subscript and rearranging the budget constraint (4b) yields an expression for the consumption of conventional goods and services, k :

$$k = e + s(D - c_e \tau_e p_e^a) + b(D - c_b \tau_b) \quad (4c)$$

Hence the expected per-period consumption equals non-investment income plus the continuous dividend stream from the portfolio of equity and bond assets less total transaction cost incurred on the entire portfolio.¹⁴

There is an alternative derivation of the expected equity premium. To incorporate utility derived from liquidity, along with the costs associated with transacting, the liquidity benefit function, $i(\tau)$, from the utility function (1) must be added to cash receipts in the Amihud and

¹³ The normalization is required since the yield on Treasury bills, r , will also incorporate an illiquidity premium relative to an asset with a truly zero transaction cost, if such an asset exists. We wish to define the illiquidity premium for Treasury bills in such a way that it is zero and so that we can treat the return on Treasury bills as the numeraire liquid asset. Of course, in reality government bonds do accrue liquidity premia relative to each other and even with respect to bonds with the same maturity. Thus the most recently issued bond of a given maturity is designated an 'on the run' security while securities issued earlier are 'off the run'. On the run securities are more heavily traded and are also priced at a higher level such that the yield is lower.

¹⁴ This neglects the general equilibrium redistribution of the income of the financial sector back into the economy. From a general equilibrium perspective, receipts from new share and Treasury bill issues plus stock broking income, representing the cost of both stock and bill turnover, is rebated to the community as a lump-sum.

Mendelson (1986b) expression for asset ask-price.¹⁵ The expected equilibrium ask-price, p_e^a , equals the expected discounted value of dividends over the random exponentially-distributed horizon, h , after taking account of the flow valuation of liquidity, $i(\tau_e)p_e^a$, plus the expected net receipts from disposal of the asset at the bid-price, $(1 - c_e)p_e^a$, once the horizon is reached:

$$\begin{aligned} p_e^a &= E_h \left[\int_{s=0}^h e^{-rs} [D + i(\tau_e)p_e^a] ds \right] + E_h [e^{-rh} p_e^a (1 - c_e)] \\ &= [\tau_e + r]^{-1} [D + i(\tau_e)p_e^a + \tau_e p_e^a (1 - c_e)]. \end{aligned} \quad (6b)$$

On solving for p_e^a , the final expression for asset ask-price, (6a) above, is obtained.

The two remaining first-order conditions found by maximizing (1) subject to the budget constraints, (4a) and (4b), with respect to equity and bond turnover or trading yields the two simple marginal benefit equals marginal cost conditions:

$$i'(\tau_e) = c_e, \quad (5d)$$

and

$$i'(\tau_b) = c_b, \quad (5e)$$

for equity and bonds respectively.¹⁶ In other words, the benefit of trading an asset net of transaction costs for the marginal trade must be zero. Transaction costs for both equity and Treasury bills are treated parametrically by the investor, as are the respective ask-prices for equity and Treasury bills, which consequently drop out.¹⁷ We seek an equilibrium in which both equity and Treasury bills are included in the investor's portfolio with trading in both securities entirely voluntary. Voluntary trading requires that incremental trading benefits are equal to the incremental (and average) transaction cost c_e for the last unit of equity and c_b for the last unit of Treasury bills traded, with $c_e \geq c_b$. The second-order conditions for maximum utility requires $\partial^2 u / \partial \tau_e^2 = i''(\tau_e) < 0$ and $\partial^2 u / \partial \tau_b^2 = i''(\tau_b) < 0$. The concavity of $i(\tau)$ ensures that these conditions are satisfied.

¹⁵ There is no expression for the 'benefit' of trading in the Amihud and Mendelson (1986b) expression because trading propensity is assumed immutable and determined like manna from heaven with trading demand the same irrespective of trading costs.

¹⁶ Note that the expected utility function describing risk aversion drops out in the derivation of (5d) and (5e) so that the equity premium due to trading is separable from the conventional premium due to risk aversion.

In the case of the exemplar liquidity benefit functions (3a) and (3b), the first-order conditions imply equality of marginal liquidity benefits and costs: $i'(\tau_e) = (\alpha/\tau_e)^\mu = c_e$ and $i'(\tau_e) = (\alpha/\tau_e) = c_e$, which give solutions to the turnover/liquidity expression:

$$\tau_e = \alpha c_e^{-1/\mu} \equiv \alpha c_e^{-\beta} \quad (7a)$$

and

$$\tau_e = (\alpha/c_e), \quad (7b)$$

respectively, where $\beta \equiv (1/\mu)$. Hence for the power illiquidity function the general turnover relationship takes the simple constant transaction cost elasticity form, $\tau(c) = \alpha c^{-\beta}$.¹⁸ Not coincidentally, the constant elasticity turnover demand function has been the staple of virtually all empirical analysis of security trading commencing with Jackson and O'Donnell (1985).¹⁹ Note that expressions in (7a) and (7b) satisfy the convexity requirements given above as $\tau'(c) < 0$ and $\tau''(c) > 0$. Moreover, a higher elasticity unambiguously enhances liquidity (turnover), $\partial\tau/\partial\beta = -\tau \ln(c) > 0$, since $c < 1$ and $\ln(c) < 0$. Finally, an increase in α , the “intrinsic” liquidity of a security, boosts turnover for a given transaction cost: $\partial\tau/\partial\alpha = c^{-\beta} > 0$.

Since the second-order condition for a maximum is satisfied, in the exemplar liquidity benefit function, $i''(\tau_e) = -\mu\alpha^\mu\tau_e^{-(1+\mu)} < 0$, the respective optimum turnover/liquidity solutions, τ_e and τ_b , are unique. On substituting the optimum values back into the net illiquidity expression for the expected equity premium from (6a), we obtain the equilibrium solutions in terms of liquidity and transaction costs, respectively:

$$\begin{aligned} ep &= -[i(\tau_e) - \tau_e c_e] = -\alpha^\mu \mu \left\{ \tau_e^{1-\mu} / (1-\mu) \right\} - \left\{ \tau_b^{1-\mu} / (1-\mu) \right\} \\ &\equiv \alpha \left\{ c_e^{1-\beta} / (1-\beta) \right\} - \left\{ c_b^{1-\beta} / (1-\beta) \right\} \\ &\equiv [1/(1-\beta)] \{ \tau_e c_e - \tau_b c_b \}. \end{aligned} \quad (8a)$$

¹⁷ Since the equity asset ask-price actually does depend on transaction cost c_e an alternative procedure is to replace the dollar cost $c_e p_e^a$ by the specified dollar cost. However the result following optimization is the same.

¹⁸ Recall that the utility function was constructed via integration from this constant elasticity demand function which in turn was chosen for its empirical relevance and applicability. Hence the recovery of this functional form from the utility function should come as no great surprise.

¹⁹ However, it has never been derived from any utility analysis prior to the present study.

Hence for the power illiquidity function the equilibrium expected equity premium is simply the difference between the amortized spreads for the two assets deflated by a term $(1 - \beta)$, which depends on the elasticity of turnover response to transaction costs. Intuitively, the equation shows that in order to voluntarily hold equity as well as bonds the investor needs to be compensated by the net difference in amortized spread for equity relative to T-bills given by the term in $\{\}$ braces. Moreover, this dollar amount is modified by the cost elasticity term which reflects the investor's response to the transaction costs differential between the two securities and thus incorporates the investor's trading preferences.

For the special case in which $\mu = 1$, we have:

$$ep = -[i(\tau_e) - \tau_e c_e] = \alpha \ln(\tau_b / \tau_e) \equiv \alpha \ln(c_e / c_b). \quad (8b)$$

Both (8a) and (8b) are exact expressions for the net valuation of liquidity with the sign reversed since the premium is no more than compensation for illiquidity.

Note that the premium for equity is zero, as it is for Treasury bills, when $c_e = c_b$ and $\tau_e = \tau_b$, as we require. Moreover, conventional asset pricing theory, which neglects transaction costs and illiquidity altogether, implicitly assumes that the premium due to illiquidity for all assets, and not just T-bills, is identically zero. Models that incorporate transaction costs due to trading into asset prices, while at the same time ignoring the utility gain from trading, risk the creation of logical contradictions since trading is not rational.

II. Testable Implications and Comparative-Static Results

After dropping time subscripts, the direct utility expression (1) in combination with the time-independent budget constraint (4c) can be simplified to

$$\max_{s, b, \tau_e, \tau_b} u(y) \equiv u\{e + (s + b)D + s[i(\tau_e) - \tau_e c_e] p_e^a + b[i(\tau_b) - \tau_b c_b]\}, \quad (9a)$$

where the two terms in square brackets are, respectively, the net liquidity benefit of equity after transaction costs and the net liquidity benefit of T-bills. The first term in square brackets in (9a) is the overall loss due to illiquidity for the equity investor who achieves turnover, τ_e , after taking account of the transaction cost given by c_e per dollar of equity assets. Equivalently, it can be regarded as the utility provided by liquidity net of the costs incurred in achieving that liquidity represented by the amortized equity spread. A similar optimization process takes place with respect to the turnover rate for Treasury bills. These conditions logically imply that the net benefit to the representative investor due to the inclusion of the relatively illiquid equity asset in

the portfolio is negative and increases to zero in the limit, as the asset approaches its numeraire value of unity with the price of equity bounded between zero and unity: $0 < p_e^a \leq 1$. Investors require compensation for illiquidity as reflected in adverse terms of trade.

Substituting the turnover demand function (7a) back into the direct utility function (9a) gives rise to the complementary indirect price dual:

$$u(y) \equiv v(D, c) = e + (s + b)D + s p_e^a \{i[\tau(c_e)] - \tau(c_e)c_e\} + b \{i[\tau(c_b)] - \tau(c_b)c_b\}, \quad (9b)$$

with

$$\partial v / \partial c_e = [i'(\tau)\tau'(c_e) - \tau'(c_e)c_e - \tau(c_e)]p_e^a = -\tau(c_e)p_e^a < 0. \quad (9c)$$

We have established:

Proposition 1: Utility is always diminishing in transaction costs, c_e , with its marginal impact equal to turnover, τ_e . This result is entirely intuitive as the impact of a worsening of the terms on which one trades should depend precisely on the amount of trading that one does. Moreover, it is universal in that it does not depend on the values of the parameters, α and β . That is, the compensation required to hold equity is equal to its trading cost (amortized spread) less the benefits associated with that asset arising from its liquidity (as measured by its turnover). Because equity is illiquid relative to bonds the expected equity premium is always positive and typically greater than the amortized spread.

The equity premium is precisely equal to the “investor surplus” trapezoid from a representative investor who holds and trades T-bills rather than equity. It is the flow of compensation required to exactly compensate the investor/trader for simultaneously holding less liquid equity in their portfolio as well as more liquid T-bills. This area is illustrated in Figure 1.

Place Figure 1 here

In general, an asset exhibiting a higher equity premium as given by $ep = -[i(\tau) - \tau]$, will, everything else equal, have a lower asset price (an illiquidity price discount) and a higher expected rate of return. The existence of higher expected returns due to relative illiquidity does not represent an anomaly. Rather it is the expected compensation for illiquidity, dependent on investor preferences.²⁰

²⁰ Since the utility derived from trading/liquidity is a matter of investor tastes or preferences, that is, the joy of trading, it is not necessarily the case that we can fully explain these tastes any more than we can explain the taste or preferences of consumers for any item or service such as red or green apples. Similarly, the Amihud and Mendelson

Proposition 2: The price of any tradable financial asset such as an equity security is equal to the dividend stream discounted by the cost of capital. The cost of capital consists of the yield on the numeraire asset, Treasury bills, r , plus the expected equity premium, ep , given by the negative of the net liquidity benefit from trading this security relative to the numeraire asset,

$$ep = -[i(\tau_e) - \tau_e c_e] \equiv [1/(1 - \beta)]\{\tau_e c_e - \tau_b c_b\} > 0.$$

In turn, this is equal to the difference in amortized spread between equity and bonds deflated by a term which depends on the responsiveness of trading turnover to transaction costs.

(For Proof: see (8a) above.)

Proposition 3: Since the equity premium depends only on trading conditions as reflected in the amortized spread from trading equity and bonds, as well as on the responsiveness elasticity β , these conditions can be derived directly from stock returns.

Proposition 3 provides scope for numerous refutable and hence testable propositions. Conventional financial models based on CAPM predict that no information pertaining to transaction costs and trading demand can be extracted from asset returns and the Amihud and Mendelson (1986b) model predicts that only the amortized spread representing the cost of trading can be extracted from asset returns.

For the empirically realistic case of a constant turnover response elasticity, equation (8a) above, the equity premium consists of the difference between the equilibrium amortized spread for the equity security and for Treasury bills, deflated by unity minus the turnover elasticity. For a less than perfectly liquid security such as equity, the amortized spread, $\tau_e c_e$, differs from Treasury bills, $\tau_b c_b$, but as transaction costs on the security declines to that of Treasury bills, the amortized spreads become the same with the pricing of the security and Treasury bill identical. In the limiting case of a unitary elasticity the equity premium is proportional to the log of relative equity and Treasury bill transaction costs.

(1986b) model with rather peculiar investor trading preferences turns out to be a special case of my model. We cannot simply reject the evident preference for investors to trade simply because (say) many finance academics might believe that investors should only buy and hold. Like most preferences it is difficult to test directly the proposition that investors gain utility from trading. Likewise, it is virtually impossible to directly test the fundamental assumption made in virtually all asset-pricing models that investors are risk averse. However, we can test our model based on its ability to predict relative to competing models such as the CAPM and with respect to the Fama–French factors. Often apparently simple and “unrealistic” behavioral assumptions lead to better predictions in the light of Milton Friedman’s methodology than do complex and more “realistic” ones.

A simple numerical example is provided in Table 2 for the limiting case of unitary elasticity, $\beta = 1$, to help with understanding of the model. Three cases are investigated. A base case with plausible transaction costs and turnover rates, the limiting case in which equity is as liquid as government bonds and an example of highly illiquid equity. The bond transaction cost and turnover remain the same in each example, as does the intrinsic liquidity parameter, α , and the bond yield of $r = 0.05$, i.e., 5% pa. A value of $c_e = 0.02$, i.e., 2%, and $c_b = 0.001$ in the base case yields an equity premium of 0.05 and an equity price, $p_e^a = 0.5$, which is half the price of the more liquid bond. Assuming a single bond worth \$1 and two equity shares also worth a total of \$1, the total investor wealth, $w = \$2$, with a cash yield of 0.15 pa. This is sufficient for the investor to consume goods worth 0.117 with an equity share contributing a negative utility of -0.033 which is partially offset by the bond providing a liquidity benefit of 0.017. The investor's overall augmented consumption inclusive of trading securities is 5% of the wealth of \$2, namely $y = 0.1$ pa. The equity share yields 0.1 so that the equity premium is 0.05 or 5% pa. This is 2.5 times the transaction cost of 0.02.

Insert Table 2 here

In the liquid equity case lowering the equity transaction cost to that of debt raises the share price to \$1 and the value of the overall portfolio to \$3. The 5% return on this portfolio ensures a utility of 0.15 and a zero illiquidity premium on equity. Equity and bonds are now identical. The consumption of goods, k , is reduced to only 0.1 since trading securities is now a more attractive pastime.

Finally, a transaction cost of 0.1 for illiquid equity not only lowers turnover drastically but also reduces the price of equity to 0.394 and the value of the overall portfolio to $w = 1.788$. High trading costs results in a substitution towards consuming goods with $k = 0.12$, with the utility of trading equity diminished even below that of the base case. Consequently, overall augmented consumption falls to 0.089 and the illiquidity premium is raised to 0.077 that is higher than in the base case. However, the premium proportional to transaction costs is lower at 0.769 indicating a concave relationship between the equity premium and transaction costs.

An increase in the turnover elasticity raises turnover (see equation (7a) above). It also raises the amortized spread for all positive values of β since $\tau c = \alpha e^{(1-\beta)\ln(c)}$ and $\partial \tau c / \partial \beta = -\tau c \ln(c) > 0$ as $c < 1$ and $\ln(c) < 0$. The rate of change of the amortized spread is also increasing in β , $\partial^2 \tau c / \partial \beta^2 = \tau c \ln^2(c) > 0$. By contrast, the impact of a higher transaction cost, c , on the amortized spread switches at $\beta = 1$, $\partial \tau c / \partial c = (1 - \beta)\tau > 0$ if $\beta < 1$ and < 0 if $\beta >$

1. The difference between the amortized spreads for equity and Treasury bills in (8a) can be expressed as $\alpha c_b^{(1-\beta)} [(c_e/c_b)^{(1-\beta)} - 1]$ on substituting for turnover using (7a). This is positive for $\beta < 1$ since $c_b < c_e$ but the sign is reversed for $\beta > 1$ and the differential amortized spread becomes negative but continues to rise in absolute value as the turnover elasticity increases. Since the multiplicative factor, $1/(1-\beta)$, also reverses sign when the numerator does, the equity premium term, $ep = [1/(1-\beta)](\tau_e c_e - \tau_b c_b)$, in (8a) is always positive so long as transaction costs for equity exceeds that of Treasury bills and is also increasing in the transaction cost elasticity.

This establishes:

Proposition 4: The expected equity premium is positive for all positive values of the turnover/transaction cost elasticity. Moreover, it is an increasing function of the turnover elasticity for all values of the elasticity.

Despite the complexities in the equity premium expression, the intuition here is relatively straightforward: The greater is the turnover elasticity and hence the more sensitive is liquidity (turnover) to changes in transaction cost, the greater is both the liquidity of equity and the liquidity of bonds since both markets are impacted. Moreover, the amortized spread for both security types is raised. The budgetary impact on equity is most severe because of the higher transaction costs and this effect dominates. Hence the compensation that investors require to hold a relatively illiquid asset such as equity relative to liquid Treasury bills is also higher. *Ceteris paribus*, a higher turnover elasticity will unambiguously raise turnover. But this does *not* mean because the asset turns over more frequently, that it will be more highly valued.

The result given by (8a) can be generalized to the Gordon growth model case on the assumption that the asset price is expected to grow at the constant geometric rate, g . The expected present value of the bid-price in (6b) now becomes $E_h[e^{-(r-g)h} p_e^a (1 - c_e)]$, with (8a) expressed in earnings/dividend yield form as:

$$D/p^a = \{r - g + [1/(1-\beta)](\tau_e c_e - \tau_b c_b)\} > 0. \quad (10)$$

The impact of a rise in unit transaction cost on the price of the asset in (6a), after taking into account the dependence of turnover on transaction cost, is unequivocally negative for all elasticity values:

$$\frac{\partial p_e^a}{p_e^a} \bigg/ \frac{\partial c_e}{c_e} = -\tau_e c_e \bigg/ \left(\frac{D}{p_e^a} \right) \equiv -\tau_e c_e / \{r + [(\tau_e c_e - \tau_b c_b)/(1-\beta)]\} < 0, \quad (11)$$

where we designate $\gamma \equiv \frac{\partial p_e^a}{p_e^a} / \frac{\partial c_e}{c_e}$ as the asset price elasticity with respect to transaction costs.

The asset price elasticity γ is consistently negative for all β values. The elasticity has a nice intuitive interpretation as the negative of the amortized spread, which has been converted to its present value by discounting at the security's cost of capital. The cost of capital is, in equilibrium, equal to the dividend yield, which is given by the Treasury bill yield plus the investor's valuation of the illiquidity of the equity security relative to Treasury bills.

This establishes:

Proposition 5: The equity premium is always increasing in the transaction cost rate, c_e , for all values of the turnover elasticity, β .

It is clear from the expression for the equity premium in (8a) that increasing the “intrinsic” liquidity parameter, α , raises the equity premium for a given cost of transacting. While this may seem counter-intuitive at first blush, after all, more transacting and hence liquidity is valued, a higher intrinsic liquidity parameter raises the liquidity value of both low-transaction-cost bonds and high-transaction-cost equity. By enhancing the role of trading, the cost differential between equity and bonds is exacerbated. Moreover, the asset price elasticity, γ , given by (11) increases in absolute value in the intrinsic liquidity parameter, α , incorporated in the amortized spread, $\tau_e c_e = \alpha c_e^{1-\beta}$, in the numerator of γ . While this is partly offset by the higher cost of capital, *i.e.*, earnings/dividend yield, for securities with higher “intrinsic” liquidity, the numerator term dominates. Therefore, the negative price impact of a higher transaction cost will be greater for more heavily traded stocks whose amortized spread is higher for a given transaction cost. This establishes:

Proposition 6: More heavily traded stocks with higher “intrinsic” liquidity but the same unit transaction costs will have a higher equity premium and a higher absolute value of the asset price elasticity, γ , given by (11) above.

As we have seen, for a given unit transaction cost a higher intrinsic liquidity raises the cost of trading given by the amortized spread and thus accentuates the trading cost disability associated with equity relative to Treasury bills. While higher liquidity/turnover is beneficial if this is induced by lower trading costs, it does not follow that higher intrinsic liquidity that raises the amortized spread and lowers the consumption of other goods and services will have a positive impact on the price of equity securities.

A rise in transaction cost, c_e , has an impact on the equity premium, ep , given by turnover, $\partial ep / \partial c_e = \tau_e$, with the proportionate change, $\left(\frac{\partial ep}{ep} / \frac{\partial c_e}{c_e} \right) = \tau_e c_e / ep > 0$.

Differentiating this elasticity with respect to transaction cost, c_e , reveals that the sign depends on the sign of $[(1 - \beta)ep - \alpha c_e^{1-\beta}] \equiv$ the sign of $-\alpha c_e^{1-\beta} < 0$ on substituting for ep using (8a). Consequently, this elasticity falls with increases in transaction cost. Thus we have established:

Proposition 7: Less liquid stocks with higher transaction costs will have a higher equity premium and the proportionate impact on the premium falls with higher transaction costs.

This establishes rigorously the intuition in Amihud and Mendelson (1986b) in which a higher transaction cost has a smaller impact due to the reduced significance of trading volume. This result is the opposite of the CAPM model incorporating transaction costs of Jacoby *et al.* (2000) in which gross returns increase in transaction costs but at an *increasing* rate. Since theirs is entirely a one-period model, as they point out, it lacks these dynamic forward-looking effects.

Utilizing (11) the asset price elasticity, γ , is equal to the amortized spread discounted at the cost of equity capital. For a turnover elasticity which is less than unity, $\beta < 1$, the direct positive impact of a higher elasticity on the amortized spread exceeds the depressing effect on the cost of capital represented in Proposition 5. However, for values of $\beta > 1$, the reverse is true. This establishes:

Proposition 8: The absolute value of the asset price elasticity increases in the turnover elasticity until the turnover elasticity is unity. It then diminishes in the turnover elasticity.

This proposition implies that asset prices are most sensitive to transaction costs when the turnover elasticity is unity in absolute value. Given that many empirical estimates suggest that the absolute value of the turnover elasticity is approximately unity, this is an important result. It is also the exact opposite of Constantinides (1986) who finds in his model that the sensitivity is negligible in this case.

Based on (8a) the general asset pricing expression (6a) can be expressed in terms of the turnover rate τ :

$$p_e^a = D / \{ r + \alpha^\mu [\mu / (\mu - 1)] [\tau_e^{1-\mu} - \tau_b^{1-\mu}] \}. \quad (12)$$

The impact of an increase in turnover (liquidity) on asset price is found by differentiating the asset price expression (12):

$$\frac{\partial p_e^a}{p_e^a} / \frac{\partial \tau_e}{\tau_e} = \left(\tau_e c_e / \beta \left(\frac{D}{p_e^a} \right) \right) > 0. \quad (13)$$

This establishes

Proposition 9: Greater liquidity, as represented by a higher equity turnover rate, τ_e , raises the asset price and lowers the equity premium for all values of the turnover elasticity, β .

Intuitively, this result seems to contradict Propositions 4 and 5 regarding the impact of a turnover elasticity and higher intrinsic liquidity. However, here we are holding the parameters of the utility function constant and consequently implicitly lowering transaction costs to achieve higher liquidity. Thus it is consistent with Proposition 7 indicating the impact of a change in transaction costs.

We have established via (13) that the asset price rises by the amortized spread, $\tau_e c_e$, deflated by the absolute value of the turnover elasticity, β , and capitalized using the dividend yield (cost of equity capital) for stock of that liquidity. This makes a great deal of intuitive sense. Since more liquid stocks with a higher turnover have a higher price, both the yield and cost of capital are lower. This is consistent with the finding of Barber and Odean (2000): individual investors who trade more earn lower rates of return. “Irrational” over-confidence is not necessary to explain this finding. Rather investors who enjoy greater liquidity can expect to pay the price in terms of lower expected financial returns.

III. Alternative Asset Pricing Models Incorporating Transaction Costs

The model developed in the previous section extends the representative agent model used so effectively by Mehra and Prescott (1985) by making security transactions and turnover representative as well. The agent not only holds the entire stock of both equity and Treasury bills but also turns over both types of stocks at the actual rates that are observed. By contrast the classic asset pricing model incorporating transaction costs of Amihud and Mendelson (1986b) assigns a preordained exogenous investment horizon to each investor along with its inverse, the investor’s turnover rate. In order for this model to be appropriately specified, there must exist potential marginal investors over the entire spectrum of turnover rates. Accordingly, it is appropriate to assume a continuum of investors with perfectly inelastic turnover rates τ ranked from low, $\underline{\tau} = 0$, to high, $\bar{\tau}$, in the range $\tau \in (\underline{\tau}, \bar{\tau})$. For simplicity, the distribution of investors is

uniformly distributed over this range. Both assets provide the same constant flow of cash payments, $\$D$, per period forever.

Equilibrium in this model requires the gross rate of return on equity, r_e , to exceed the Treasury bill yield, r_b , by the differential amortized spread, $\tau^\circ(c_e - c_b)$, of the marginal investor with a turnover rate of τ° (see, for example, Kane, 1994, or Bodie, Kane and Markus, 2002, pp. 279-284). If the differential amortized spread is less than the equity premium, $r_e - r_b$, the investor will specialize and just hold equity whereas if the spread is greater than the premium, only Treasury bills will be held in the investor's portfolio. For the marginal investor who is just indifferent the net of differential transaction cost equity return will precisely equal the yield on bills. Apart from the fact that trading demand by any individual investor is entirely inelastic and thus occurs irrespective of trading cost, the model fails to identify the actual marginal investor out of the infinity provided by the continuum (Swan, 2002). Since this is consistent with all possible values of the turnover rate at which the marginal investor is indifferent, no unique equilibrium solution exists. This implies that without modification the model is incapable of explaining the impact of a change in transaction costs.

It is therefore necessary to modify the assumption of a distribution of investor-types with perfectly inelastic investor horizons and hence pre-specified turnover demands so that a unique equilibrium is now achieved within the Mehra and Prescott (1985) representative investor framework. While the implications of allowing investors to be responsive to transaction costs are now explored, the aim is to show that no rational investor could have such an objective. In effect, I argue that so long as security turnover is responsive to transaction costs, this style of model has its limitations for comparative-static purposes.

Consider our asset-pricing model (6a) and (8a) above such that there is a single representative investor type who incurs transaction costs c_e but there is no illiquidity premium, $i(\tau) \equiv 0$, as is conventionally (and implicitly) assumed to be the case. Equation (6a) now becomes:

$$p_e^a = [\tau(c_e)c_e + r]^{-1} D, \quad (14)$$

on recognizing the dependence of turnover on transaction cost, $\tau(c_e)$; $\tau'(c_e) < 0$; $\tau''(c_e) > 0$. The case analyzed by Amihud and Mendelson is a special case of (14) with $\tau'(c_e) \equiv 0$ so that demand is perfectly inelastic. The equity premium is now simply the amortized spread on equity, $ep = \tau(c_e)c_e$ rather than the differential amortized spread deflated by the turnover

elasticity expression, $1 - \beta$, as in (8a) for the power function illiquidity premium expression (3a).²¹

Following the lead of Amihud and Mendelson (1986b, p. 228), the market-clearing ask-price of the asset in (14) can also be expressed as:

$$p_e^a = D/r - \tau(c_e)c_e p_e^a / r, \quad (15)$$

so that the price of the equity asset is equal to the expected capitalized dividend stream attributable to the perfectly liquid asset less the expected capitalized stream of the observed amortized spread. Other studies besides Amihud and Mendelson (1986b) obtain results similar to the asset-pricing equation (14) used to compute the effect of a change in transaction costs on asset price. See, for example, Jackson and O'Donnell (1985), Amihud and Mendelson (1990 and 1992), Keifer (1990), and Hubbard (1993).

It is now demonstrated that if the representative investor observes this standard investment rule, a tax agent, such as the government, could extract some of the wealth from investors by taxing transactions with *no* harmful, and potentially beneficial, consequences for asset prices and investor wealth and utility. While the demonstration of peculiar results holds only for a special case, a formulation that recognizes transaction costs without at the same time recognizing trading benefits has the potential to mislead since trading is either irrational or utility reducing. Transacting is not motivated by differences in expected returns within these models.

In the particular case of a constant turnover elasticity (14) becomes

$$p_e^a = D/[r + \alpha c_e^{(1-\beta)}], \quad (16)$$

after substituting for the amortized spread, $\tau(c_e)c_e$, using (7a). Since $c_e^{(1-\beta)} \rightarrow 1$ as $\beta \rightarrow 1$, $p_e^a \rightarrow D/[r + \alpha]$, and price is independent of the cost proportionality factor, c_e , for $\beta = 1$. On evaluating the effect of transactions cost increase, $\Delta c \cong dc_e$, the impact is given by:

$$dp_e^a / dc_e = -(1 - \beta)\tau_e p_e^a / (r + \tau_e c_e) < 0 \text{ for } \beta < 1, \text{ or } = 0 \text{ for } \beta = 1, \text{ or } > 0 \text{ for } \beta > 1,$$

with asset price elasticity:

²¹ Since the Amihud and Mendelson (1986b) model sets aside the amortized spread due to Treasury bills it contains only the equity amortized spread.

$$\frac{dp_e^a}{p_e^a} \bigg/ \frac{dc_e}{c_e} = -(1 - \beta) [\tau_e c_e / (r + \tau_e c_e)]. \quad (17)$$

To see the intuition, let the intrinsic liquidity, α , alter when the turnover elasticity, β , is varied so that the term in square brackets in (17) is constant. This term in square brackets is simply the perpetuity value of the amortized spread and is not greatly affected by alterations to the elasticity β , especially since the amortized spread is a constant in the vicinity of unitary elasticity. With this adjustment it is easy to recognize that the maximum negative impact of a unit transaction cost increase on equilibrium price occurs when $\beta \rightarrow 0$ from above; the impact is progressively reduced until it becomes zero for an elasticity β value of unity. Further increases in the transaction costs elasticity have the perverse and counter-intuitive effect of raising asset price. Although the representative investor appears to have been penalized in terms of a per-unit transaction cost rise, total transaction costs fall for an increase in c when $\beta > 1$ and the asset price rises rather than falls the greater is this penalty. Within this framework, the amortized spread, $\tau(c)c$, is viewed as being unequivocally harmful so that a transactions cost *increase* Δc which *reduces* the amortized spread, $\tau(c)c$, when the response is elastic ($\beta > 0$), is perversely regarded as beneficial. In a “well behaved” economic framework it must always be the case that an increase in transaction cost, c , alone must lower asset price, $dp^a/dc < 0$.

My peculiar result, that asset prices rise in response to higher transaction costs, obviously cannot be the case. As already noted, while Vayanos (1998) obtained such a result, he recognized that it could not be correct. Suppose that $\beta > 1$, which empirically turns out to be the case for several securities markets that have been investigated. The asset with a positive transaction cost c will sell at a discount relative to the numeraire asset with low or zero transaction costs. The observed gross rate of return, D/p_e^a , will exceed the return on the numeraire asset, Treasury bills, r_b . Now suppose that the taxing authority imposes a tax on transactions that raises c_e to $c_e + \Delta c_e$. This not only raises revenue for the taxing authority, but also raises the asset ask-price, p_e^a , and consequently investor wealth and the present value of investor consumption. This “magic pudding” process can continue until both turnover and total transaction cost have been driven down to zero with the premium for illiquidity removed.²² Such

²² The Australian artist, Norman Lindsay, wrote and illustrated a famous book about a magic pudding which could not only change flavour according to customer preferences but would rejuvenate itself every time a slice was cut so that supply was inexhaustible without a requirement for ingredients.

a model is rejected in favor of the liquidity model presented in section I above incorporating endogenous trading in the empirical tests presented below.

We have established:

Proposition 9: Cost of trading models without explicit recognition of trading benefits (in representative investor form), in which investors discount the cost of transactions (the amortized spread) at the cost of capital r for Treasury bills, do not hold if individual investors respond to alterations in transaction costs by varying the holding period.

The fatal error in my “representative investor” version of the model which recognizes trading costs but not benefits is the failure to recognize that trading is endogenous rather than exogenous. That is, it is responsive to transaction costs and embodies utility maximizing behavior. The “rents” earned by investors on “intra-marginal” units by the opportunity to trade up to the very last share, for which marginal benefits and costs are equal, are ignored. While transaction costs are recognized, the fact that there are benefits which more than offset the costs are ignored. Costs are incurred on the marginal trade for which there are exactly offsetting benefits. This problem did not arise in the “investor-types” model actually used by Amihud and Mendelson (1986b) because the marginal investor is assumed to trade (fixed) equal quantities of both Treasury bills and the equity security. Hence, implicitly, the same liquidity benefits are assumed for both types of assets and only the trading costs differ.

The thrust of one stream of the transaction cost literature is that transaction costs are *not* very important for asset pricing because of the high cost elasticity of demand for trading. In fact, no asset price changes occur at all when this elasticity is unity in absolute value. While Amihud and Mendelson (1986a, 1986b, 1989, 1992) show empirically that transaction costs are important, when the realistic case in which transaction demand is approximately unitary elastic is relevant, they lack a model such as the one developed in section I.²³

Moreover, as $\beta \rightarrow 0$, the limiting value of the equity premium is $\tau_e c_e - i(\tau) \equiv (\tau_e c_e - \tau_b c_b)$. This is the only case in which the (modified) Amihud and Mendelson

²³ Amihud and Mendelson (1992) present numerical examples in which small changes in transaction cost have a substantial effect on asset prices but the holding period response to the change in transaction tax is ignored. As they point out (1992, p. 481), “it is also necessary to know how the holding periods of each stock will be adjusted in response to the tax”. This problem is particularly severe when a simple log-linear specification of their equity premium model is applied because the premium is concave in the bid-ask spread. Since this log specification implies a unitary elasticity (see (3b), (7b) and (8b)), the “representative investor” version of the model actually implies that asset prices should be unaffected by the spread.

(1986*b*) model provides a close to correct solution, namely the difference in the simple amortized spreads.²⁴ Once again, the reasoning is simple: the assumption underlying the Amihud and Mendelson (1986*b*) approach is that the demand for transactions by each and every investor is in-built and is thus perfectly inelastic, i.e., that $\beta \rightarrow 0$.

A very significant additional contribution made by Amihud and Mendelson (1992, section 5, p. 489) is to 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. In effect, the present study takes up this challenge and derives the rather radical consequences.

IV. Empirical Results: Reinterpreting Amihud and Mendelson

In their first study of the impact of transaction costs on asset returns Amihud and Mendelson (1986*a*) utilize 20 years of NYSE returns and the bid-ask spread as a percentage of the stock price measured on the last trading day of the year.²⁵ They regress the gross monthly return R_j for seven portfolios on the natural log of the average spread S_j over the 20 years to obtain:

$$R_j = 0.031 + 0.00412 \ln(S_j) + u_j, \quad (18)$$

(6.16)

where the t value is in brackets and the u is the residual error term. The gross return is an approximation to the equity premium and the bid-ask spread, S_j can be regarded an estimate of the transaction cost of equity relative to Treasury bills, c_e/c_b , given the reasonable assumption that the bid-ask spread on Treasury bills remains fairly constant. Hence the expression estimated by Amihud and Mendelson is practically identical to the expression derived in equation (11*b*) above in which the equity premium is shown to be a function of the log of the relative transaction cost.

This formulation with the coefficient of $\ln(S_j)$ equal to $\alpha = \pi$, or amortized spread, is precisely true if the absolute value of the turnover elasticity $\beta = 1$. Many of the empirical

²⁴ Neither Amihud and Mendelson (1992) nor Kane (1994) explicitly consider that Treasury bills also have an amortized spread.

²⁵ These estimates are calculated as the simple average of the bid-ask spread measured relative to stock price computed at the beginning and end of each year. These values are admittedly crude in terms of the much better bid-ask spread data which is available today, but were state-of-the-art at the time.

turnover elasticity estimates are about unity. The precise values for the amortized spread over this period are not known, but the average turnover rate for NYSE stocks over this period is approximately 0.25 or 25% pa. Hence the implied average spread is 0.01648 or about 1.5% given $\alpha = \tau_e c_e = 0.00412$. This seems quite plausible. Moreover, it is also slightly smaller than Stoll and Whaley's (1983) estimate of the overall two-sided transaction cost of 2.79%. Hence these results provide considerable support for the model of endogenous trading.

In applying results similar to these to simulate the impact of a transaction tax of 0.5% Amihud and Mendelson (1992, 478-480) assume that the impact of the tax on trading volume is zero. Had they used the more realistic assumption of unitary elasticity with respect to turnover they would have obtained a zero price response to the tax increase instead of the substantial impact that they simulate with their simple NPV analysis of the increase in total turnover cost. Of course, in the model of section I with *endogenous* trading, the price change can be substantial even when the turnover elasticity is unitary. In fact, the asset price elasticity impact of an increase in transaction costs is maximized at unitary elasticity.

V. An Estimate of the Equity Premium Based on NYSE Returns, 1962-1991

Datar, Naik, and Radcliffe (1998) conclude that a 1% increase in the monthly percentage turnover for non-financial firms on the NYSE reduces the cross-sectional monthly return by 4.5 basis points over the period, 1962-1991, conditional on Fama-French (1992) factors and CAPM beta (see above). The question can now be posed: is this magnitude sufficient to account for the equity premium? Over the period 1980-1991 for which data is available the average turnover rate on US Government securities was 12.56 times per annum while on the NYSE the comparable rate for equity was 0.5 (see Table 1 above). Hence the liquidity of US Government securities was 25 times higher than for equity as a whole over this period. The differential return is 0.01126 per month or 0.1351 per annum. This estimate at over 13% is high as an estimate of the equity premium and suggests that the Datar, Naik, and Radcliffe (1998) estimate is perhaps on the high side.

Data is also available for two other relatively comparable markets, the UK and Australia. In 1992 the annual turnover of Gilts (all UK Government Bonds) by final customers was 3.6636 times and for the equity of UK and Irish companies, 0.4308, on the London Stock Exchange (LSE). If the intra-market turnover of both Gilts and equity is included the rate for Gilts rises to

7.125 times annually and for equity, 0.6948 (London Stock Exchange, 1992).²⁶ Gilts are reasonably liquid with the entire stock turning over every 1.68 months, but less so than the US or UK. Since professional market makers may not be as sensitive to transaction costs it is better to focus on final customer trades.

The turnover rates for Australian Commonwealth Government bonds were 8.33 in 1993-94, 11.58 in 1994-95, 9.22 in 1995-96, 10.77 in 1996-97 and 8.61 in 1997-98 (Briers, Cuganesan, Martin and Segara, 1998, p.42). Hence these bonds are even more liquid than Gilts. Over the same five-year period equity turnover on the ASX rose from about 0.25 to 0.5 times per annum so that the ratio of bond to equity turnover fell from 33.3 to 17.2 times over this period.²⁷ The average experience over this period is quite similar to the US. Consequently, government bonds or Gilts are far more liquid than equity in all three countries.

VI. Further Empirical Results: Monthly Returns on the Australian Stock Exchange, 1994-1998

A monthly database with a total of 24,350 observations was constructed for approximately 576 stocks over a five-year period from the Security Industry Research Centre of Asia-Pacific (SIRCA's) trade by trade database.²⁸ Since many of the smaller stocks are relatively illiquid, bid-ask spreads were computed only when stocks traded so as to avoid the problem of stale quotes. Monthly returns were computed with the inclusion of dividends and also the volatility measure, the average daily high-price minus low-price deflated by the average daily price, was computed along with the monthly volume of shares traded and shares on issue. The monthly equity premium for each stock was computed by deducting the monthly return on three-month Australian Treasury bills from the overall return. Transaction cost, c_e , as a proportion of the ask-price was computed using the sum of the actual bid-ask spread, stamp duty and brokerage. Stamp duty fell from 0.6% on a two-sided transaction to 0.3% on July 1, 1995. Brokerage was assumed to be 0.4% on all two-sided transactions.

Both the equity premium and equity turnover rate are estimated simultaneously using non-linear Ordinary Least Squares (OLS). The two simultaneous equations are:

²⁶ In 1992 the total value of trades in Gilts was 1,238,791 billion British Pounds with an estimated valuation of 173,865.3 billion Pounds. This estimate was computed from data supplied by the London Business School. The total value of trades in equities was 433,858.9 billion Pounds for British and Irish companies traded on the LSE with an estimated value outstanding of 624,393.3 billion Pounds for British and Irish companies.

²⁷ This relative change is largely due to the halving of stamp duty on stock exchange transaction from July 1995.

²⁸ I wish to thank SIRCA and Kingsley Fong for the construction of this data base.

$$ep_t = \{ \alpha^\mu [\mu / (\mu - 1)] [\tau_{et}^{1-\mu} - \tau_b^{1-\mu}] \}, \quad (19a)$$

and

$$\tau_t = \alpha c_t^{-1/\mu}, \quad (19b)$$

where the Treasury bill turnover rate, τ_b , is set at an annualized rate of eight based on the Australian evidence.²⁹

The first of the two equations to be simultaneously estimated, (19a), is the general equity premium result, (12) above with the implied Treasury bill transaction cost, c_b , solved for in terms of the known turnover rate, τ_b , the intrinsic liquidity parameter, α , and the inverse of the turnover elasticity, $\mu \equiv (1/\beta)$. The second equation, (19b) is simply the equity turnover relationship, (7a) above. Simultaneous estimation ensures that consistency is maintained between the estimates of the equity premium and turnover regressions.

The model was first estimated for the full data set consisting of 24,350 monthly observations. Summary statistics are shown in Table 3. The mean and median annualized equity premium was negative over this period and transaction costs were quite high on average because of the inclusion of illiquid stocks. The mean turnover rate is approximately the same as the market as a whole over this period. The transaction costs, turnover and market capitalization variables all show an indication of skewness.

Of note is the comparable volatility of the equity premium and turnover. The standard deviation of the premium is 0.83756 and turnover, 0.39685. As Campbell (2000) and other commentators have pointed out, neither dividends nor consumption growth are sufficiently volatile to be consistent with the volatility in asset prices and the equity premium itself. The high volatility of liquidity (turnover) helps provide an explanation for the “volatility puzzle” as well as the equity premium puzzle.

The results are summarized in column 1 of Table 4. The following results were obtained: the intrinsic liquidity coefficient is both positive and highly significant ($\alpha = 0.01455$, t -stat. = 41.807), the turnover elasticity is close to unity in absolute value and also highly significant ($\beta = 0.78137$; $\mu = 1.2798$, t -stat. = 139.63), and implied Treasury bill transaction cost, $c_b = 0.000311$. Both the estimated coefficients are significant at an extraordinarily low probabilistic value. The estimated turnover elasticity is consistent with most empirical studies.

²⁹ Recall that the turnover rate for Australian Government bonds is approximately eight fold annually making it approximately 32 times higher than for equity over this period (see section V above).

The implied transaction costs for Treasury bills seem to be a little on the low side but are reasonably consistent with industry-based information.

The original Amihud and Mendelson (1986b) model with perfectly inelastic trading response to transaction costs for individual investors is nested in (19a) and (19b) with $\beta \equiv 0$.³⁰ That model is rejected at the 99% confidence level with $\beta > 0$. The model was re-estimated in column 2 including only stocks of above median market capitalization given by \$57.45m. This increases the value of the intrinsic liquidity parameter by approximately one-third while the estimated transaction cost elasticity falls to 0.72145. For smaller than median stocks in column 3 the intrinsic liquidity parameter is lower and this is accompanied by a higher elasticity, 0.87138. Hence smaller stocks tend to have higher turnover elasticities compensated for by a lower intrinsic liquidity parameter.

The model using the full data set is re-estimated in column 4 allowing for the respective equity premium and turnover mue's, μ_e and μ_τ , to differ between equations (19a) and (19b). The implied turnover elasticity estimated from the first equation is $\beta_e = 0.798021$ and from the second, $\beta_\tau = 0.760572$. Hence the differences are very slight.

VII. Simulation of the Endogenous and Exogenous Trading Models

A comparison can be made between the expression for the equity premium and asset prices taking into account the utility benefits from owning and trading equity in expressions (6a) and (8a) above with the Amihud and Mendelson (1986b)-type expression (14) for asset ask-price. Both incorporate dividends and the amortized spread, but only the utility expression incorporates the liquidity benefits from equity trading captured by $i(\tau)$. That is, the utility from holding and trading liquid assets relates to the net benefits, $i(\tau) - \tau c$, after deducting the amortized spread. It is not at all related to the amortized spread taken in isolation as in the conventional model. This contrasts with the conventional model in which turnover is exogenous and values of the turnover elasticity, $\beta > 1$, result in higher asset values as transaction costs are raised (see (17) above).

The impact of changes in the turnover elasticity β are simulated in Table 5 with both the endogenous trading model, equation (10) inclusive of a Gordon dividend growth factor term

³⁰ It is conceivable that the aggregate or market demand could be sensitive to transaction costs even if the individual participants are not.

together with the asset pricing elasticity (11), and the conventional model, equations (16) and (17). The base-case assumptions are set out in the top half of Table 5 and are based on the estimates from column 1 of Table 4 with the estimated base-case turnover elasticity $\beta = 0.721449$. In the lower half of Table 5 (and illustrated in Figure 2) the impact of varying the turnover elasticity β is simulated in the range from zero to 1.2. Since a higher elasticity results in a higher turnover of both equity and Treasury bills, and this accentuates the illiquidity of equity relative to Treasury bills, the value of the equity premium gradually rises to 22.84% for an elasticity value of 1.2 with endogenous trading. In the conventional model with exogenous trading the premium depends only on the amortized spread. Consequently, the premium is far lower and reaches a maximum of 2.75% for an elasticity of 1.2.

The more interesting results relate to the elasticity of the asset price with respect to transaction costs, γ . With endogenous trading, this elasticity is always negative, reaching an absolute maximum at $\beta = 1$ of -14.68% . In the conventional model the absolute value of the elasticity remains small as β increases to reach -4.82% at $\beta = 0.5$. It increases slightly more then declines in absolute value to zero at $\beta = 1$ before becoming substantially positive at 9.93% for $\beta = 1.2$. The sign alteration for values of $\beta > 1$ illustrates the “magic pudding” effect with higher trading taxes underpinning both higher transaction costs and higher asset values.

VIII. A Test of the Model with Endogenous Trading Based on “Letter Stock” Returns.

Silber (1991) estimates the magnitude of the illiquidity premium by estimating the discount on “letter” stock. Letter stock is issued by firms under SEC Rule 144 and is identical to regular stock except that it cannot be traded for a period that is typically two years³¹ The annual discount of 14.5 to 18% that he identifies is in addition to the conventional equity premium. Pratt (1989) summarizes the results of eight separate studies of the discount that ranges from 17.5 to 20% per annum. This additional illiquidity premium can be generated in my simulation by a reduction in the turnover rate from 0.6 per annum to virtually zero at 0.004, so that the return rises to 21% (15% plus 6%) per annum. The “shadow” transaction cost also rises from 0.02 to 10.49891 to curtail trade by the requisite amount. Hence the model requires no special assumptions in order to generate a premium of 21% per annum, consistent with the evidence of Silber and Pratt.

³¹ In Australia there is a similar concept relating to shares owned by the founders at the time of an IPO. There is typically an *escrow* period of two years.

This contrasts with the conventional model derived in the previous section in which the equity premium is simply given by the amortized spread with turnover established by some *ad hoc* mechanism external to the model rather than derived from utility maximization subject to a budget constraint. Transacting is not endogenized within the conventional model. Hence the benefits of transacting are ignored and the illiquidity function consists of only the amortized spread per dollar of the asset, $-\tau(c)c$. As pointed out above, the value of the derivative of this expression with respect to transaction costs initially rises and then falls as the elasticity β increases from less than to greater than unity.

IX. Further Empirical Results: Daily Returns on the NYSE, 1955-1998

The data base consists of daily total and capital gains value weighted returns on the NYSE from Schwert (1990) and CRSP over the period 1954-1998, inclusive. The contemporaneous dividend yield is computed as $D_t/p_{t-1} = R_t - (p_t - p_{t-1})/p_{t-1}$ where R_t is the total return and p_t represents the closing price. To remove the seasonal factor in dividend yield the average daily dividend was computed in place of the contemporaneous dividend in the numerator of the dividend price ratio, D_t/p_{t-1} . Daily turnover on the NYSE was computed from the daily volume supplied by the NYSE and daily shares on issue interpolated from data on shares on issue at annual intervals supplied by the NYSE. The daily Federal Funds Treasury bill rate was added from July 1, 1954 and the Three-Month Treasury bill rate from February 1962. Since transaction costs are not available over this period, the endogenous turnover model was estimated using turnover and the non-linear dividend yield specification:

$$D_t/p_t = \{\alpha_1 r_t + \alpha^\mu [\mu/(\mu-1)] [\tau_e^{1-\mu} - \tau_b^{1-\mu}] + u_t\}, \quad (20)$$

based on (12) above, where τ_t is the daily equity turnover rate, τ_b the (fixed) bill turnover rate which was assumed to be eight based on Australian evidence, r_t the daily Treasury bill yield (Federal Funds rate for estimation starting in 1955) and u the residual error term.

Estimation was carried out over the period 1955-1998 and sub-intervals using non-linear least squares with the results shown in Table 6. To ensure that daily turnover is exogenous its value was predicted from the turnover rates on the preceding 20 days. The adjusted R^2 was typically of the order of 85 to 90% for this predictive equation. The level of explanation for the dividend yield is very high for all five sub-periods and all coefficients are significant at the 1% level with very high t values. The first column is for the period 1955-74 inclusive with a remarkable 64% of the variation explained by just the relative equity/Treasury bill turnover rate

and the Federal Funds rate. The estimated turnover elasticity at 2.768 was higher than expected and the intrinsic liquidity parameter, α , lower. The predicted equity transaction costs rate of 2.775%, shown below the mean values in Table 6, corresponds almost exactly to Stoll and Whaley's (1983) estimate of 2.79%.

Estimates are provided for the whole period, 1955-98 inclusive, in column 2. Both the turnover elasticity estimate of 0.74717 and liquidity parameter estimate look reasonable but the predicted equity transaction costs at 3.476% are higher than the Stoll and Whaley (1983) estimates. The highest implicit transaction cost estimates at 6.327% are for the period 1975-1988 and would appear to be upward biased. The estimates for the period 1975-98 at 2.935% are in agreement with the earlier Stoll and Whaley (1983) estimates. The final column for the period 1989-98 indicates that transaction costs, as perceived by traders, are now much lower at less than 1% (*i.e.*, 0.958%).

X. Identity of Elasticity Estimates from Turnover and Equity Premium Data

My liquidity model of the equity premium and asset pricing implies two paths by which the turnover elasticity, β , can be estimated. The two paths should generate the same β outcome. Firstly, β is identified with respect to transaction costs, c_e , directly from turnover information, τ_e , and, secondly, β indirectly defined from the equity premium, ep , and the amortized spread, $\tau_e c_e$. Consequently, the following two-equation simultaneous equation model was estimated separately using non-linear least squares for ninety individual stocks using 505 days of daily data for each stock:

$$\tau_e = \alpha c_e^{-\beta} \quad (21a)$$

and

$$ep = \alpha_0 + \tau_e c_e / (1 - \beta - \alpha_1) \quad (21b)$$

with the β estimate the same in both equations if $\alpha_1 = 0$.

The data consist of estimates of the daily equity premium, turnover rate and transaction cost, made up of the bid-ask spread, market impact costs, brokerage charge and stamp duty using 90 Australian (ASX) stock returns, 1994/95 to 1996/97. To be included a stock must trade a minimum of 10 times a day. Hence only the most liquid stocks are included. Of the 90 separate estimates of the additive constant term, α_1 , 14 had absolute t values of 1.96 or better meaning that there is a statistically significant difference between the β elasticity estimates from each equation. Hence the hypothesis of equal β estimates is accepted for 76 of the 90 equations.

There were also 17 instances in which the estimate of β itself failed to be both positive and have a significant t value. However, for 10 of the 17 insignificant estimates the average equity premium was negative. This may, perhaps, have contributed to the failure of the hypothesis of a positive and significant β in these cases. Consequently, the Amihud and Mendelson (1986b) assumption of a zero β elasticity is rejected for 73 of the 90 stocks. Moreover, the joint hypothesis of the same β estimate from both equations and a positive and significant β elasticity is satisfied for 56 of the 90 equations. According to risk-based theories such as the CAPM and its variants, there should be no relationship between the equity premium and the turnover elasticity. In these circumstances, the success rate of 76/90 or 56/90 is highly supportive of the model.

XI. Conclusions

In this paper I construct, and estimate using a wide variety of data sources, a model in which trading is fully *endogenous* in the sense that it is responsive to transaction costs and stems from rational utility maximization subject to a budget constraint. Accepting a challenge by Black (1986) to model investors who enjoy trading, “liquidity”, as proxied by the security turnover rate, is incorporated into the investor’s utility function with the result that the equity premium depends on the amortized spread for equity less the amortized spread for Treasury bills. This difference is deflated by a term that depends on the responsiveness of turnover to transaction costs. This frees the expected premium predicted by the illiquidity model to take on realistic values instead of tiny values stemming from the amortized spread alone. As a natural consequence of the trading process and demand for liquidity, the volatility of returns is intimately associated with both the volatility of turnover demand and its magnitude. My empirical evidence demonstrates that illiquidity is priced. “Investor surplus” from trading equity relative to more liquid Treasury bills is shown to be capable of accounting for the equity premium in security returns.

In the original Amihud and Mendelson (1986b) model based on “investor types” every investor has a perfectly inelastic demand for trading. It is shown to imply an infinite number of solutions. A more conventional representative-investor asset-pricing model that overcomes these problems was developed in III. It concludes that the “amortized spread”, the product of turnover and unit transaction cost, represents the premium of equity securities over the Treasury bill rate. However, these models imply that asset prices rise in response to an increase in unit transaction cost when share turnover is highly responsive to transaction costs. This is because asset prices

are computed net of the present value of the amortized spread and this diminishes if security turnover demand is elastic. Many estimates indicate that turnover is elastic. The “magic pudding” effect is a consequence of not incorporating benefits from trading and implies that governments can create additional wealth by imposing security taxes that reduce security turnover by more than the tax-imposed transaction tax increase.

The reason for this strange result is that many extant models treat share turnover as *exogenous*, or very nearly so. Even the models with endogenous trading predict turnover rates of only a few percent when the current rate on the NYSE is about 88% per annum. Consequently, asset-pricing models that incorporate the present-value cost of trading typically fail to meet the requirement that the incremental benefits from transacting must exceed the costs. Yet this contradicts our knowledge that trading is voluntary rather than mandated. The original Amihud and Mendelson (1986a) regression results based on 20 years of NYSE returns and bid-ask spreads are reinterpreted to support the present model with a turnover elasticity of unity and to provide realistic estimates of both transaction costs and the amortized spread.

Perhaps the most striking comprehensive evidence for the significance of liquidity comes from Datar, Naik, and Radcliffe (1998). It is based on 30 years of monthly NYSE data on security returns and turnover for 880 stocks. After controlling for stock size, book to market ratio and CAPM beta, annualized returns diminish at the rate of 0.54% per 1% increase in stock turnover. Information on the differential liquidity between equity and US Government securities traded over the last 21 years indicates that the turnover of the Government securities is approximately 25 times higher. This liquidity differential is more than adequate to account for the observed equity premium. The equity premium model is tested using a large sample of approximately 576 stocks and 24,350 monthly returns quoted on the ASX over the period 1994-99 with all the parameters of the model estimated using comprehensive data on turnover and transaction costs. The new model is successful in explaining the equity premium, while the nested conventional model is statistically rejected with an exceedingly high Student *t* value. The volatility (standard deviation) of the equity premium and turnover are comparable. This suggests that the model accounts for Campbell’s (2000) “volatility puzzle” and well as the equity premium puzzle. When the estimation procedure permits differing turnover elasticities to be obtained from the equity premium and turnover relationship, these coefficients are very close together. The model also explains why “illiquid” letter stock can obtain a return of 15-20% per annum over identical stock that can be traded. The simulation is accomplished either by lowering the turnover rate to almost zero or simulating a very large but temporary increase in transaction costs.

The model is also used to explain the variation in the dividend yield on the NYSE using daily returns and turnover data for periods from the 1955 to the 1998. In excess of 60% of the daily variation in the dividend yield is explained by only two variables, equity turnover and interest rates. The model is able to successfully predict equity transaction costs over a range of different intervals just from the dividend yield and the two independent variables. The model shows for the first time that perceived total transaction costs on the NYSE have fallen to less than 1% during the last decade. By contrast, existing models that use transaction costs to explain the equity premium require transaction costs on a two-sided transaction of 10% or more to explain the premium. Typically, relatively high estimates of actual costs are of the order of 2.8% prior to the last decade, that is, about one quarter of the required magnitude. Finally, the daily returns for two years of trading 90 liquid Australian stocks is used to successfully test the model's prediction that the turnover elasticity implied by the equity premium is the same as that directly estimated from turnover data. Further tests are carried out on Finnish data (Swan and Westerholm, 2002) and on Chinese data (Chen and Swan, 2002). Since the new theory of asset pricing is the first to embody endogenous equity and Treasury bill trading in the model, I expect that a range of additional tests of the model will be forthcoming.

In a variety of ways the paper raises more questions than it answers. If my main proxy for liquidity, stock turnover, explains so much of what lies behind 'the joy of trading' captured by incorporating liquidity into investor preferences? Why have previous explicit attempts to find a motivation for the exceedingly high trading levels we observe fallen so short of the mark? Liquidity in the form of signed order flow is intimately associated with information content in security prices and hence the driving of security prices closer to fundamentals. This link between liquidity and information content needs to be explored further so as to provide a fuller explanation for why liquidity is so valuable. At the very least this paper indicates that the links between market microstructure and asset pricing are far stronger than have previously been recognized. An even bigger challenge arises in explaining the relatively small differentials in returns between highly liquid governmental bonds and T-bills on the one hand and relatively illiquid corporate bonds on the other. If returns on 'junk' bonds are often comparable with equity why is the return on highly rated but apparently fairly illiquid "corporates" not also comparable?

References

- Aitken, Michael J. and 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.
- Aitken, Michael J. and Peter L. Swan, 1999, "Impact of a Transaction Tax on Security Market Traders: The Case of Australia's Tax Reduction," Department of Finance, University of Sydney Working Paper.
- Aiyagari, S. Rao and Mark Gertler, 1991, "Asset returns with transaction costs and uninsured individual risk," *Journal of Monetary Economics* 27, 311-331.
- Amihud, Yakov and Haim Mendelson, 1986a, "Liquidity and Stock Returns," *Financial Analysts Journal* (May-June), 43-48.
- Amihud, Yakov and Haim Mendelson, 1986b, "Asset pricing and the bid-ask spread," *Journal of Financial Economics* 17, 223-249.
- Amihud, Yakov and Haim Mendelson, 1989, "The Effects of Beta, Bid-Ask Spread, Residual Risk, and Size on Stock Returns," *Journal of Finance*, 44 (2) (June), 479-486.
- Amihud, Yakov and Haim Mendelson, 1992, "Transaction Taxes and Stock Values," in Kenneth Lehn and Robert Kamphuis (eds), *Modernizing U.S. Securities and Regulations: Economic and Legal Perspectives*, Irwin, Illinois.
- Amihud, Yakov, Haim Mendelson and B. Lauterback, 1997, "Market microstructure and securities values: evidence from the Tel Aviv Stock Exchange," *Journal of Financial Economics* 45, 365-390.
- Atkins, Allen B. and Edward A. Dyl, 1997, "Transaction Costs and Holding Periods for Common Stocks," *Journal of Finance* 52 (1), 309-325.
- Bansal, Ravi and Wilbur John Coleman, 1996, "Premium, Term Premium, and Risk-Free Rate Puzzles," *Journal of Political Economy* 104 (6), 1135-1171.
- Barber, Brad M. and Terrance Odean, 2000, "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *Journal of Finance* 55 (2) (April), 773-806.
- 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.
- Benartzi, Schlomo and Richard H. Thaler, 1995, "Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics* 111 (February), 73-92.
- Black, Fischer, 1986, "Noise," *Journal of Finance* 41 (3) (July), 529-543.
- Bodie, Zvi, Alex Kane and Alan J. Markus, 2002, *Investments*, (fifth ed.), Boston, McGraw Hill.
- Briers, Michael, Suresh Cuganesan, Paul Martin and Reuben Segara, *Australian Financial Market Review: Towards a Regional Financial Centre*, 1998, An AFMA-SIRCA joint study, UNSW Press, 1998.
- Brennan, Michael J. and Avanidhar Subrahmayam, 1996, "Market microstructure and asset pricing: On the compensation for illiquidity in stock returns," *Journal of Financial Economics* 41, 441-464.

- Campbell, John Y., 2000, "Asset Pricing at the Millennium," *Journal of Finance* 55 (4) (August), 1515-1567.
- Campbell, John Y. and John H. Cochrane, 1999, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy* 107 (2), 205-251.
- Campbell, J. and K. Froot, 1994, "The international experience with securities transaction taxes," in J. Frankel (ed.), *The Internationalization of Equity Markets*, University of Chicago Press, Chicago.
- Chalmers, John M. R. and Gregory B. Kadlec, 1998, "An empirical examination of the amortized spread," *Journal of Financial Economics* 48, 159-188.
- Chen, Zhian and Peter L. Swan, 2002, "Empirical Test of the Liquidity Asset Pricing Model on the Segmented Chinese Equity Market," School of Banking and Finance, University of New South Wales, Working Paper, June.
- Constantinides, George C., 1986, "Capital Market Equilibrium with Transaction Costs," *Journal of Political Economy* 94, 842-862.
- Constantinides, George C., 1990, "Habit Formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy* 98, 519-43.
- 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 and 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.
- Fama, Eugene F. and Kenneth R. French, 1992, "The cross-section of expected stock returns," *Journal of Finance* 47, 427-465.
- Feenstra, Robert C., 1986, "Functional Equivalence Between Liquidity Costs and the Utility of Money," *Journal of Monetary Economics* 17, 271-291.
- Fisher, Stephen J., 1994, "Asset Trading, Transaction Costs and the Equity Premium," *Journal of Applied Econometrics* 9, S71-S94.
- Friedman, Milton, 1969, "The Optimum Quantity of Money", in: *The Optimum Quantity of Money and other Essays* (Aldine Publishing Co., Chicago IL)
- Garvey, Gerald T. and Peter L. Swan, 2002, "What Can Microstructure Contribute to Explaining Executive Incentive Pay?: Liquidity and the Use of Stock-Based Compensation," working paper, School of Banking and Finance, University of New South Wales, June.
- Grossman, Sanford J. and Joseph E. Stiglitz, 1980, "On the impossibility of informationally efficient markets," *American Economic Review* 70, 393-408.
- Henry, Peter Blair, 2000, "Do stock market liberalizations cause investment booms?," *Journal of Financial Economics* 58, 301-334.
- Holmstrom, Bengt R. and Jean Tirole, 1993, "Market liquidity and performance monitoring," *Journal of Political Economy* 101, 678-709.
- Hubbard, R. G., 1993, "Security transaction taxes: Can they raise revenue?," 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.
- Jacoby, Gady, David J. Fowler and Aron A. Gottesman, 2000, "The capital asset pricing model and the liquidity effect: A theoretical approach," *Journal of Financial Markets* 3, 69-81.
- 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.
- Keifer, Donald W., 1990, "The securities transactions tax: An overview of the issues," Congressional Research Service, The Library of Congress, Report No. 87-278 S. July.
- Kocherlakota, Narayana, 1996, "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature* 34 (March), 42-71.
- Kurz, Mortecai and Andrea Beltratti, 1997, "The Equity Premium is No Puzzle," in M. Kurz (ed.) *Endogenous Economic Fluctuations: Studies in the Theory of Rational Belief*, Chapter 11, pp. 283-316. Studies in Economic Theory No. 6, Berlin and New York: Springer-Verlag.
- Kurz, Mortecai and Maurizio Motolese, 2000, "Endogenous Uncertainty and Market Volatility", Working Paper, Department of Economics, Stanford University, January.
- Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 54, 1315-1336.
- London Stock Exchange, 1992, *Quality of Markets Review*, Winter Edition, London.
- Longstaff, Francis A., 1995, "How Much Can Marketability Affect Security Values?," *Journal of Finance* 50 (5) (December), 1767-1774.
- Lynch, Anthony W., 1996, "Decision Frequency and Synchronization Across Agents: Implications for Aggregate Consumption and Equity Return," *Journal of Finance* 51 (4) (September), 1479-1497.
- Marquering, Wessel and Marno Verbeek, 1999, "An empirical analysis of intertemporal asset pricing with transaction cost and habit persistence," *Journal of Empirical Finance* 6, 243-265.
- Mehra, R. and Edward C. Prescott, 1985, "The Equity Premium: A Puzzle," *Journal of Monetary Economics* 15 (March), 145-162.
- Merton, R. C., 1973, "An Intertemporal Capital Asset pricing Model," *Econometrica* 41 (September), 867-887.
- Otrok, Christopher, B. Ravikumar and Charles H. Whiteman, 1998, "Habit Formation: A Resolution of the Equity Premium Puzzle?," Working Paper, Department of Economics, University of Iowa.
- Pratt, S. P., 1989, *Valuing a Business*, (Irwin, Homestead, Il.).
- Samuelson, Paul A., 1947, *Foundations of Economic Analysis* (Harvard University Press, Cambridge MA)
- Schwert, G. W., 1990, "Indexes of U.S. Stock Prices from 1802 to 1987," *Journal of Business*, 63 (3), 399-426.
- Schwert, G. W. and P. J. Seguin, 1993, "Security transaction taxes: An overview of the costs, benefits and unresolved questions," *Financial Analysts Journal* 49 (September-October), 27-35.

- Siegel, Jeremy J., 1992, "The real rate of interest from 1800-1900: A study of the U.S. and the U.K.," *Journal of Monetary Economics* 29, 227-252.
- Silber, William L., 1991, "Discounts on Restricted Stock: The Impact of Illiquidity on Stock Prices," *Financial Analysts Journal* 47 (July-August), 60-64.
- Stoll, Hans R. and Robert Whaley, 1983, "Transaction costs and the small firm effect," *Journal of Financial Economics* 12, 57-80.
- Swan, Peter L., 2002, "Will the true marginal investor please stand up?: Asset prices with immutable security trading by investors," Working Paper, School of Banking and Finance, Faculty of Commerce, University of New South Wales.
- Swan, Peter L. and P. Joakim Westerholm, 2002, "Asset prices and liquidity: The impact of endogenous trading," Working Paper, School of Business, University of Sydney, January.
- Tufano, Peter, 2000, Social Science Research Network (SSRN), FEN Educator: Discussion Forum on the Equity Premium Puzzle.
- 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.

Appendix: How are our “benefit of liquidity” functions derived?

We seek to construct the entire set of liquidity functions for which the marginal benefit of a trade is simply the transaction cost incurred. The respective liquidity premiums can be expressed as $i(\tau_e) \equiv \alpha_0 + \lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^{\tau_e} i'(x) dx \equiv \alpha_0 + \lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^{\tau_e} \tau^{-1}(x) dx$ and $i(\tau_b) \equiv \alpha_0 + \lim_{\varepsilon \rightarrow 0} \int_{x=0}^{\tau_b} i'(x) dx \equiv \alpha_0 + \lim_{\varepsilon \rightarrow 0} \int_{x=0}^{\tau_b} \tau^{-1}(x) dx$, where α_0 is (approximately) a constant, and $c_e = \tau^{-1}(\tau_e)$ and $c_b = \tau^{-1}(\tau_b)$ are the unique inverses of the downward-sloping and convex turnover function, $\tau = \tau(c)$, with $\tau'(c) < 0$ and $\tau''(c) \geq 0$. To elucidate this definition of the illiquidity function in more concrete terms, the constant elasticity form of the turnover function, which has proved to be highly successful in empirical research, is adopted for expository simplicity. The inverse function transaction cost function, $c = \tau^{-1}(\tau) = (\tau/\alpha)^{-(1/\beta)} \equiv (\alpha/\tau)^\mu$ with $\mu \equiv (1/\beta) > 0$ and $\mu \neq 1$ is integrated to yield a special case of the illiquidity premium function, $i(\tau_e) \equiv \alpha_0 + \alpha^\mu \lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^{\tau_e} (x)^{-\mu} dx \equiv \alpha_0 + \alpha^\mu \lim_{\varepsilon \rightarrow 0} [x^{1-\mu}/(1-\mu)]_{\varepsilon}^{\tau_e} \equiv \alpha_0 + \alpha^\mu \tau^{1-\mu}/(1-\mu)$, where α_0 is the constant as before, with $\partial i/\partial \tau = \alpha^\mu \tau^{-\mu} \equiv (\tau/\alpha)^{-(1/\beta)} = c > 0$ and second-order condition for a maximum, $\partial^2 i/\partial \tau^2 < 0$, satisfied since $\mu = 1/\beta > 0$. A similar condition holds for Treasury bills: $i(\tau_b) \equiv \alpha_0 + \alpha^\mu \lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^{\tau_b} x^{-\mu} dx \equiv \alpha_0 + \alpha^\mu \tau_b^{1-\mu}/(1-\mu)$, so that the net benefit of liquidity, $i(\tau_b) - \tau_b c_b \equiv \alpha_0 + \alpha^\mu \tau_b^{1-\mu}/(1-\mu) - \alpha^\mu \tau_b^{1-\mu} \equiv \alpha_0 + [\mu/(1-\mu)] \alpha^\mu \tau_b^{1-\mu}$. Note that the two liquidity expressions, $i(\tau_e)$ and $i(\tau_b)$, only differ because transaction costs differ with $c_e > c_b$. If equity has the same transaction costs as Treasury bills, the liquidity benefit of the two securities must be identical. Consequently, the constants, α_0 , in both expressions are the same. Since by definition the interest yield on Treasury bills, r , incorporates any “Treasury bill premium”, the net Treasury bill liquidity benefit, $i(\tau_b) - \tau_b c_b \equiv 0$. Hence, $i(\tau_b) \equiv \tau_b c_b$, and consequently, $\alpha_0 \equiv -[\mu/(1-\mu)] \alpha^\mu \tau_b^{1-\mu} \equiv \alpha c_b^{1-\beta}/(1-\beta)$ on simplifying. Thus we have derived (3a) in Section I above. Of course, when (3a) is differentiated we recover our starting point, the constant elasticity turnover function in (7a).

The liquidity benefit given by (3a) is not defined for the constant elasticity turnover function in the limiting case where the elasticity $\beta \equiv 1/\mu = 1$, yet many empirical studies have estimates which are clustered around unity. The liquidity expression in this case utilizes $\tau = \alpha/c \Leftrightarrow c = \alpha/\tau$ to become $i(\tau_\varepsilon) \equiv \alpha_0 + \lim_{\varepsilon \rightarrow 0} \int_{x=\varepsilon}^{\tau_\varepsilon} i'(x)dx \equiv \alpha_0 + \alpha \lim_{\varepsilon \rightarrow 0} \int_{x=\varepsilon}^{\tau_\varepsilon} (1/x)dx \equiv \alpha_0 + \alpha \lim_{\varepsilon \rightarrow 0} [\ln(\tau_\varepsilon) - \ln(\varepsilon)]$, where $\ln(\)$ denotes the natural log, for a small number, ε , since $\ln(0)$ is not defined. Similarly, $i(\tau_b) \equiv \alpha_0 + \alpha [\ln(\tau_b) - \ln(\varepsilon)]$ with the condition, $i(\tau_b) - \alpha \equiv 0$, yielding $\alpha_0 \equiv -\alpha [\ln(\tau_b) - \ln(\varepsilon) - 1]$. Hence we have derived (3b) in Section I above from our starting point (7b).

Table 1: Derivation of Turnover Rates for US Treasury Securities and NYSE Equities, 1980-2000.

Year	1	2	3	4	5	6	7	8
1980	616.4	11.4	6.9	7.719014	11,352.3	31,871	0.356195	21.67
1981	683.2	13.3	11.2	9.32377	11,853.7	36,003.5	0.329237	28.32
1982	824.4	17.4	14.8	10.15526	16,458	38,907	0.423009	24.01
1983	1,024.4	23.3	18.8	10.68528	21,589.6	42,316.9	0.510189	20.94
1984	1,247.4	28.5	24.3	11.00529	23,071	47,104.8	0.489780	22.47
1985	1,437.7	39.6	35.8	13.63567	27,510.7	50,759.4	0.541982	25.16
1986	1,619	53.3	42.3	15.35269	35,680	56,023.8	0.636872	24.11
1987	1,724.7	64.6	45.6	16.61274	47,801.3	65,711.4	0.727443	22.84
1988	1,821.3	63	39.2	14.58958	40,849.5	73,988.5	0.552106	26.43
1989	1,945.4	69.8	43.1	15.08893	41,698.5	79,573.5	0.524025	28.79
1990	2,195.8	68.7	42.5	13.16696	39,664.5	86,851.8	0.456692	28.83
1991	2,471.6	78.5	49.0	13.41236	45,266	95,177.1	0.475598	28.20
1992	2,754.1	95.7	56.4	14.35896	51,375.7	107,730.5	0.476891	30.11
1993	2,989.5	107.7	65.9	15.09818	66,923.3	123,446	0.542126	27.85
1994	3,126	116.1	75.2	15.91107	73,420.4	136,667	0.537221	29.62
1995	3,307.2	112.7	80.5	15.18868	87,217.5	148,499.8	0.587324	25.86
1996	3,459.7	117.3	86.4	15.30826	104,636.2	165,831.8	0.630978	24.26
1997	3,456.8	120.9	91.2	15.9529	133,312.1	192,016.6	0.694274	22.98
1998	3,355.5	126.5	100.1	17.55804	169,744.6	223,195.5	0.760520	23.09
1999	3,281	101.3	85.3	14.78696	203,914.3	260,115.9	0.783936	18.86
2000	2,966.9	98.6	108.0	18.10509	262,483.4	298,276.6	0.880000	20.57

Key:

1. US Treasury Securities Outstanding supplied by the Bond Market Association website.
<http://www.bondmarkets.com/Research/tsyos.shtml>. In \$US billions
2. Transactions with Interdealer Brokers; Daily trading in Treasury Securities \$US Billions
3. Transactions with Others
4. Annual bond turnover rate assuming 260 trading days pa
5. Number of Shares Traded on the NYSE Annually in Millions
6. Average Number of Shares On Issue in Millions from NYSE
7. Annual Turnover Rate on the NYSE from NYSE Annual Reports
8. Ratio of Treasury Securities to Equity Turnover (Col. 4/Col.7)

**Figure 1: The Equity Premium Corresponds to the Net Welfare Trapezoid Area
Welfare Loss Arising from the Difference in Transaction Cost between Equity
and Treasury Bills**

Figure 1 shows the compensation required to induce a representative investor/trader to hold both equity with a transaction cost of c_e and turnover τ_e and Treasury bills with a transaction cost of c_b and turnover of τ_b in their portfolio. The endogenous trading model shows that the investor must be compensated by exactly the large shaded trapezoid area made up of a rectangle of height $c_e - c_b$ and width τ_e plus the shaded triangular area between τ_e and τ_b . It represents the investor/trader gain from trading at transaction cost c_b rather than the higher cost c_e . This area is commonly known as the consumer surplus change or the equivalent/compensating variation. By contrast the conventional model incorrectly attributes the equity premium to the amortized spread for equity given by the small rectangle with width τ_e and height c_e .

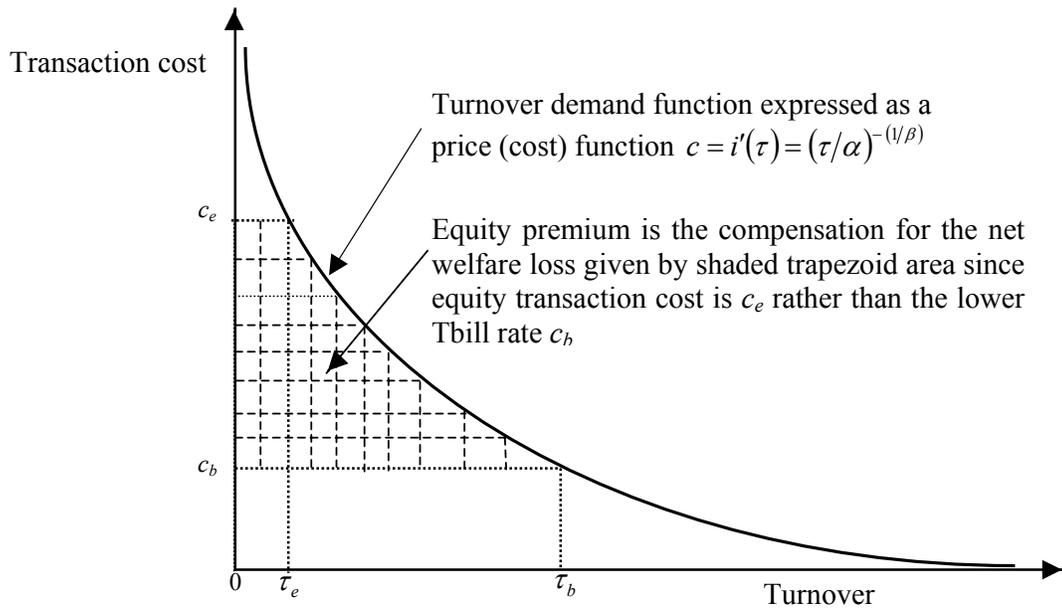


Table 2: Simulation of the equity premium for the unitary transaction cost elasticity case, $\beta = 1$, for three cases with different transaction costs

Item	Base Case	Liquid Equity	Illiquid Equity
Bond yield, $D = r$	0.05	0.05	0.05
Equity transaction cost, c_e	0.02	0.001	0.1
Bond transaction cost, c_b	0.001	0.001	0.001
Turnover coefficient, α	0.0167	0.0167	0.0167
Turnover rate pa Equity, τ_e	0.8335	16.670	0.1667
Turnover rate pa Bonds, τ_b	16.670	16.670	16.670
\$ Amortized spread for equity share	0.0083	0.0167	0.0066
Amortized spread per bond	0.0167	0.0167	0.0167
Equity premium, $ep = \alpha \ln(c_e/c_b)$	0.050	0.000	0.077
Price of equity, p_e^a	0.5003	1	0.3944
Number of Bonds, b	1	1	1
Number of shares, s	2	2	2
Wealth constraint, w	2	3	1.788
Cash return on wealth (Total dividend)	0.15	0.15	0.15
Consumption of goods, k	0.117	0.100	0.120
Equity utility per share, $i(\tau_e)$	-0.033	0.017	-0.060
Bond utility per bond, $i(\tau_b)$	0.0167	0.0167	0.0167
Augmented consumption, y	0.1	0.15	0.089

Table 3: Summary Statistics for the Sample of Approximately 576 Australian Securities Listed on the Australian Stock Exchange, 1994-98, with 24,350 Monthly Observations.

Statistic	Equity Premium (ep)	Trans. Cost (c)	Turnover (τ)	Market Cap.
Mean	-0.061887	0.041707	0.27575 pa	\$726.4 m.
Stand. Dev.	0.83756	0.039103	0.39685	0.26118×10^{10}
Median	-0.063089	0.029113	0.159 pa	\$68.1 m

Table 4: Regression results for sample of approximately 576 Australian securities listed on the Australian Stock Exchange, 1994-98, estimating two simultaneous equations:

$ep_t = \{ \alpha^\mu [\mu / (\mu - 1)] [\tau_{et}^{1-\mu} - \tau_b^{1-\mu}] \}$ and $\tau_t = \alpha c_t^{-1/\mu_\tau}$ for the equity premium and the stock turnover rate, respectively.

Column	1	2	3	4
Item	Full Sample	Above Median Market Cap.	Below Median Market Cap.	Separately Estimated Elasticities
Intrinsic Liquidity Coeff. (α)	0.01455 (41.807)	0.021293 (31.217)	0.009559 (14.379)	0.015847 (24.266)
Mue (Inverse of Turnover Elasticity) μ	1.2798 (139.63)	1.3861 (86.199)	1.1476 (53.674)	NA
Equity Mue (Inverse of Turnover Elasticity) μ_e				1.2531 (85.396)
Turnover Mue (Inverse of Turnover Elasticity) μ_τ				1.3148 (76.412)
Turnover Elasticity (β)	0.78137	0.72145	0.87138	NA
Elasticity estimated from the Equity Premium (β_e)	NA	NA	NA	0.798021
Elasticity estimated from Turnover (β_τ)	NA	NA	NA	0.760572
Implied Treasury bill Transaction Cost (c_b)	0.000311	0.00027	0.000443	0.000279

Student's t statistics are in parentheses. All coefficients are significant at the 1% level.

Table 5

Simulation of Equity Turnover, Equity Premium, Dividend Yield and Asset Price Elasticity for Different Turnover Elasticities based on the Rational Investor Model with Endogenous Turnover and the Conventional Asset Pricing Model with Exogenous Turnover.

General Assumptions/Parameters	
Estimated turnover elasticity, β	= 0.7814
Estimated turnover intercept, α	= 0.01455
Equity turnover rate, τ_e	= 17.42%
Treasury Bill turnover rate, τ_b	= 800%
Unit cost for equity as % of price, c_e	= 4.1707%
Unit cost of bill trading, c_b	= 0.031121%
Dividend growth rate, g	= 3.5%
Treasury bill yield, r	= 6.2835%

Variation in Turnover Elasticity					
β Turnover Elasticity	Equity Turnover	Treasury Bill Turnover	Equity Premium	Discount Factor/ Dividend Yield	Asset Price Elasticity
Utility Based Model with Endogenous Trading					
0.0	1.46%	1.46%	0.06%	2.84%	- 2.13%
0.1	2.0%	3.26%	0.09%	2.88%	- 2.90%
0.5	7.12%	82.48	0.54%	3.33%	- 8.93%
0.7814	17.42%	800.00%	2.18%	4.97%	- 14.62%
1.0	34.89%	4,675.3%	7.13%	9.91%	- 14.68%
1.2	65.86%	23,500%	22.84%	25.63%	- 10.72%
Conventional Model with Exogenous Trading					
0.0	1.46%	N. A.	0.06%	2.84%	- 2.13%
0.1	2.0%	N. A.	0.08%	2.87%	- 2.62%
0.5	7.12%	N. A.	0.3%	3.08%	- 4.82%
0.7814	17.42%	N. A.	0.73%	3.51%	- 4.52%
1.0	34.89%	N. A.	1.45%	4.24%	0.00%
1.2	65.86%	N. A.	2.75%	5.53%	9.93%

Figure 2: Equity Premium and Price-Cost Elasticity with Endogenous Trading and Conventional Model Simulated with Transaction Cost Elasticities Ranging from Zero to 1.3

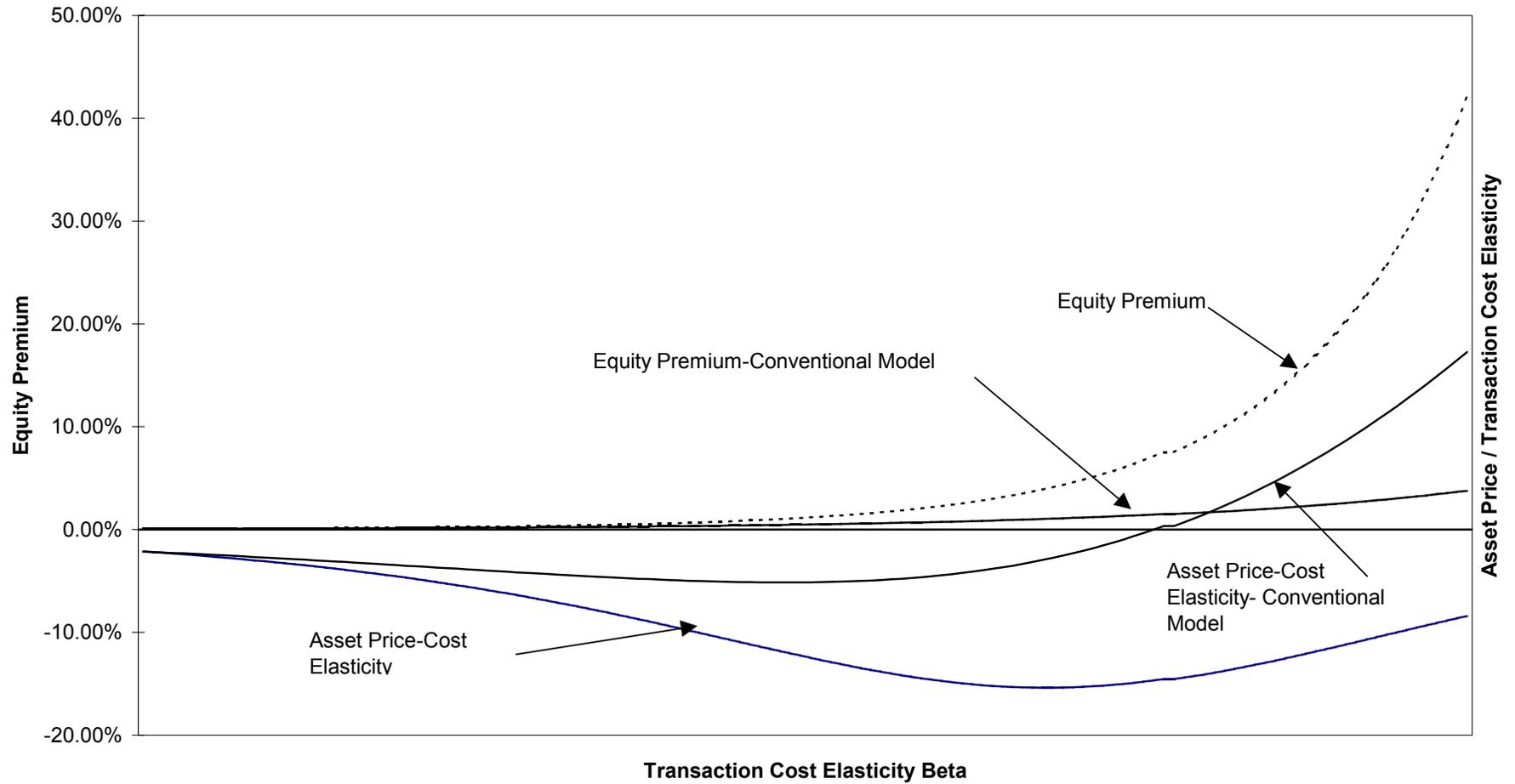


Table 6: Estimates of the Parameters of the Equity Premium Liquidity Model Using Non-Linear Equation Estimation and Annualized Daily Value Weighted NYSE Dividend Yield and Turnover, January 3, 1955 to August 30, 1998

Estimation Period	1955-1974	1955-98	1975-88	1975-98	1989-98
Variable	Mean	Mean	Mean	Mean	Mean
Dividend Yield	0.05379	0.04908	0.05843	0.04514	0.02589
NYSE Turnover	0.16950	0.31982	0.39409	0.44577	0.52062
Tbill/Federal Funds	0.04463	0.06179	0.08097	0.06925	0.05227

Predicted Values Based on the Estimated Coefficients and Mean Values

Pr. Equity Tr. Cost	0.02775	0.03476	0.06327	0.02935	0.00958
Pr. Tbill Trans. Ct.	0.00689	0.00047	0.00048	0.00042	0.00048
Pr. Dividend Yield	0.05371	0.04611	0.05270	0.04326	0.02554

Depend. Variable	Dividend Yield	Dividend Yield	Dividend Yield	Dividend Yield	Dividend Yield
Intrins. Liq., Alpha	0.83167*10 ^{-5*} (5.9348)	0.025991* (36.742)	0.071747* (70.694)	0.040544* (45.090)	0.00739* (12.548)
Turnover El., Beta	2.7680* (60.087)	0.74717* (90.953)	0.61712 (107.50)	0.67944* (102.98)	0.91541* (49.253)
Tbill/Federal Coef.	0.56413* (135.42)	0.27413* (55.598)	-0.029198 (3.5938)	0.18627* (25.118)	0.23693* (39.224)
No. of Observ.	5,012	10,994	3,539	5,982	2,443
R Squared (Adj.)	0.6424	0.3638	0.6783	0.4308	0.4932

Absolute *t* values are in brackets. * denotes significant at the 1% level.