OPTIMAL PORTFOLIO BALANCING UNDER CONVENTIONAL PREFERENCES AND TRANSACTION COSTS EXPLAINS THE EQUITY PREMIUM PUZZLE

Peter L. Swan[*]

UNSW

Version Dated April 6, 2005

ABSTRACT

For the equity market, 1896-1994, optimal exchange of equity and bonds by *N* strategic investors motivated by heterogeneous endowments, under CARA preferences, low risk aversion (CARA $b = 1$), moderate risk ($\sigma^2 = 3.24\%$), and 1% round-trip transaction cost explains a 7% illiquidity (equity) premium, 98 to 184-fold higher than estimates provided by previous literature. It also explains observed equity and bond yearly turnover rates of 38% and 880%, with 2% bond yield. The cost of forgone equity trading is 18 times higher than the transaction outlay. This paper's findings are consistent with most stylized empirical facts as well as new empirical tests.

Key words: equity-premium puzzle, asset prices, liquidity, trading, transactions cost

JEL Classification: G12, G11, G310, C61, D91, D92

[*] School of Banking and Finance, Faculty of Commerce, UNSW, 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 also wish to thank Yakov Amihud, Henk Berkman, John Campbell, Zhian Chen, George Constantinides, Stuart Dennon, David Colwell, David Feldman, Stephen Fisher, Kingsley Fong, Doug Foster, Gerald Garvey, Magnus Gammelgard, Mikael Gellback, Narayan Naik, Ananth Madhavan, Michael McAleer, Ed Prescott, Jeff Sheen, David Simmonds, George Sofianos, Dimitri Vayanos, Joakim Westerholm, seminar participants at AGSM, City University of Hong Kong, EFMA Conference, the Reserve Bank of Australia, University of NSW, University of Sydney, and University of Western Australia for comments on earlier versions. SIRCA and the ASX kindly provided access to the underlying intra-day data. Responsibility for both errors and views expressed is entirely my own.
The former title of the paper, prior to the introduction of the present model, was: "Can 'Illiquidity' Explain the Equity Premium Puzzle? The Value of Endogenous Market Trading".

"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 … are likely to be pyrrhic victories only"—Peter Tufano

Between 1896 to 1994 the yearly simple geometric mean equity premium for New York Stock Exchange (NYSE) value-weighted stocks was six percent Campbell, Lo and MacKinlay (1997) and has been approximately eight percent for the last fifty years (Cochrane, 2005). In a celebrated paper Mehra and Prescott (1985) attempt to account for this premium using simulations of an inter-temporal equilibrium model with a single representative consumer/investor, abstracting from transaction costs, security market microstructure, liquidity considerations, and other frictions. They are able to account for only a negligible proportion of this premium with a maximum of 0.4% explained by risk aversion.

Mehra and Prescott and the subsequent literature surveyed by Cochrane (2005) and Campbell, Lo and MacKinlay (1997) focus on representative agent equilibria with agents identical in all respects, including endowments. Do these equilibria appropriately represent issues that require heterogeneity such as trading activity? Any meaningful modelling of transaction costs requires trading between investors. Thus, to motivate trade investors must differ in at least one respect, here endowments. I show that, preserving all the standard assumptions of rational utility maximization, and even identical preferences, a simple exchange model with quite small transactions costs explains the major stylized and empirical facts about equity and bond market returns and trading turnover over the last 100 years. Heterogeneous endowments and frequent trading in the form of a short investor-trading horizon are sufficient for the model to predict observed trading activity for equities and bonds together with a very sizeable percent equity (illiquidity) premium. I replicate the established theoretical finding that negligible compensation is required for bearing transactions costs by extending the trading horizon from a short to long interval, but only at the cost of reducing the predicted level of trading to a counter-factual negligible amount.

I explain the importance of the mutually advantageous exchange of equity shares and T-bills in shaping the performance of the world's financial markets and, in particular, the NYSE and the T-bill market, 1896-1994. I fully agree with the existing literature in its finding that "observed" transactional cost outlays could not explain the equity premium. Instead, I focus on an invisible cost, so far neglected by the literature, of stock market trading that my simulations indicate is about 18.5 times higher than all the observed costs of trading, such as spreads and commissions, combined. This is the cost of foregone trades. I find that investors optimally consummate only a very small fraction of the trades they would undertake if transactions costs were zero because of slight but positive trading costs, even though their wellbeing suffers a decline relative to the zero transactions cost ideal. This loss of welfare is manifest in a much higher required return on equity relative to hypothetical identical assets with no transactions costs. A very substantial fall in the price of the asset accomplishes this much higher required return. My simulations indicate that when I add invisible transactions costs to the observed costs, these overall costs do explain the equity premium, the two percent yield on T-bills, and most other stylized facts besides.

These invisible costs do not receive recognition in the national accounts as such, but are manifest in the high cost of equity capital. Think of equity shares as simply claims on an underlying risky asset. Perhaps the simplest way to describe these costs is in terms of the optimal sharing of risks stemming from this underlying asset. One investor has an “excessive” equity endowment in his portfolio while another is “deficient”. Costless transacting, in the form of mutual but oppositely signed optimal portfolio rebalancing trades, would equalize the burdens at the margin, leading to societal optimal risk sharing. Even apparently insignificant transactions costs impose welfare losses on both parties via inefficient risk sharing, even though the (common) degree of risk aversion displayed by both investors is low and volatility is moderate. I find that the inability to transact equity shares as cheaply as T-bills, requires compensation of just over seven percent per annum for one-way total transactions costs of under half of one percent. While these differential trading costs could be due to asymmetric information or microstructural problems impinging more on equity that T-bills, why this is so lies outside the scope of the paper. The components of the equity premium are actual transactions (resource or cash flow) costs of 0.37 percent and costs of forgone trades of 6.78 percent. In my simulation, the optimal equity turnover rate is 38 percent per annum. Whereas for identical investors trading three-month T-bills with the same relative endowments, the required compensation is only about 0.32 percent per annum. The optimal turnover rate for T-bills is 880 percent, which is 24 times as rapid (see Table II below). As perhaps a surprising and little known historical fact, over the period, 1980-2004, Treasury securities have turned over at a rate which on average is 26 times higher than equity (see Table I below). I find an investment horizon of only a fortnight is required to explain observed equity and bond turnover rates. This contrasts with an investment horizon of 20 years that is required for my model to reproduce the illiquidity premium results of Constantinides (1986) with negligible equity and bond turnover. My simulations (Tables II below) indicate that the required equity compensation is almost 20-fold higher than the bond compensation; given investors with identical preferences, endowments, and investor horizons are trading both equity and bonds over the period, 1896-1994.

Building on the seminal contribution of Pagano (1989), I develop a simple and transparent "closed-form" trading model for risky assets, incorporating both proportional transactions costs and market impact "costs". There is less understanding of market impact costs than (say) a stamp duty or tax. They arise in the Nash equilibrium generated by my model because of "thin" markets. That is, investors are strategic, and thus rationally recognize that when they trade they turn the terms of trade against themselves by forcing down the market-clearing price if a seller and up if a buyer. For many real world stocks, the number of potential buyers and sellers of large block trades at any given moment is quite small, making modelling of strategic trading empirically relevant, while emphasising the gains from market integration. I find that the main impact of strategic investor behaviour when market thinness increases is to significantly reduce stock turnover but not to have a major deleterious impact on other market fundamentals.

Because in my model all investors in equity shares and bonds have the same preferences and investor horizons, including a constant absolute risk aversion (CARA) coefficient, endowment heterogeneity motivates trading activity while rebalancing optimal portfolios. All investors also share the same (complete) access to information concerning the mean and variance of the normally distributed shareholder returns. This simple heterogeneous endowment framework enables me to calibrate the model precisely to generate as equilibria the historically observed turnover rates for equity and bonds, as well as the observed T-bill yield and equity premium. These arise endogenously in my model, which is essentially general equilibrium in nature, rather than imposed in an *ad hoc* fashion. I do take as exogenous, however, the levels of transactions costs and the volatilities of the equity and bond markets.

In common with Constantinides (1986), my investors are indifferent between (say) trading bonds or equities with negligible transactions costs and trading an otherwise identical asset with transactions costs in place. This is equivalent to trading the same asset, with and without transactions costs in place. I compute the difference in the expected dividend per share required to equate the expected utility of an investor in the same asset, with or without transactions costs. I then compute the fall in the price of the asset, which is expensive to trade, necessary to generate the increase in expected dividend yield, so that the same investor can include both assets in his portfolio without suffering a utility loss. Over the period, 1896-1994, the simulations indicate that equity prices would have been 400 percent higher (a discount of 75 percent on the price of the equivalent transactions-cost-free asset), if it were possible to eliminate equity-transacting costs of less than one percent (on a round-trip basis). Here, the incidence falls essentially on the firm's founders who face a high cost of capital when raising equity capital. In the limiting case at the other extreme in which investors have no substitutes for an asset subject to transactions costs or an equivalent stamp duty (e.g., Tobin tax), the cost/tax incidence falls entirely on trading investors with no alteration in the price of the asset measured at the mid-point of the bid-ask spread. At either extreme, whether investors are indifferent or bear the entire incidence of transactions costs, the popular belief that the price of an asset equals the present value of the dividend less cost outlay stream, is untrue.

It might seem that the endowment heterogeneity required to calibrate my model so as to explain the returns and trading history of equity and bonds over the last 100 years is relatively high (see Tables II below). However, I could interpret the endowment heterogeneity as no more than a transparent device to generate observed trading demands under identical preferences, identical beliefs, absence of asymmetric information, absence of trades for purely consumption purposes (liquidity trades), absence of intergenerational trading, and so on. While relaxing any of these assumptions could place less reliance on endowment heterogeneity, none of these requirements including CARA utility, which dispenses with income or wealth effects, is actually required to explain observed equity premium and trading patterns.

I show that all that is required is a "well-behaved" stock or bond turnover function, consistent with observed relationships between trading patterns and transactions cost, which is integrable over the transacting range to obtain an expression for the precise compensation required to make investors indifferent between trading any asset bearing trading costs and an equivalent asset costless to trade. From a theoretical perspective, I could also adopt Black's (1986) call for trading models incorporating some generalized trading benefit, or Goettler, Parlour and Rajan (2004) who assign to traders valuable private information that generates endogenous trading decisions. Either way, integrating the implied security demand functions over the range of transaction costs from zero to its observed value yields the implied "consumer surplus" loss arising from transaction costs. This in turn equals the implied illiquidity (equity) premium loss.

I find a great deal of supporting evidence for my model and simulations. In addition, I undertake several new tests. The findings of most empirical studies of "illiquidity" premia are consistent with it, as are my own empirical tests of the model.

I present a brief literature review in section I and the risky security exchange model in section II. Simulations of equity and bond markets, 1896-1994, reinterpretations of existing empirical findings and two studies of my own are in section III. Section IV concludes.

I. Literature Review

Amihud and Mendelson (1986) model discounted cash flow maximization by risk neutral agents such that the gross yield on equity must equal the product of the marginal investor's trading level or turnover rate and the equity transactions costs. Under these assumptions there are no forgone trades due to transaction costs, with compensation for illiquidity higher than actual transactional cost outlays if the marginal investor's desire to trade exceeds the average. A number of studies have introduced transaction costs while treating trading as exogenous and thus not determined within the model. Fisher (1994) uses the actual turnover rate 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.

Constantinides (1986) computes the illiquidity premium using numerical simulations based on Merton's (1973) inter-temporal asset pricing model of a single representative agent with constant relative risk aversion (CRRA) preferences and an infinite horizon. It is, in principle, correctly computed as the increment to the required dividend for an asset with transaction costs to make the investor indifferent to an identical asset without transaction costs. The investor accommodates increases in proportional transaction costs by "drastically reducing the frequency and volume of trade" (p. 859), but the required compensation to bear transaction costs is negligible at approximately 0.15 of the one-way transaction cost. In contrast, I find the required compensation in my simulation to be at least seven and possibly 13 times the two-way cost (up to 26 times the one-way cost) when the investment horizon is $1/24$ of a year, *i.e*, a fortnight. An investment horizon that is any longer than a fortnight significantly reduces my ability to calibrate the predicted and actual equity and bond turnover rates. While his model is calibrated according to security yields as well as volatility and the CRRA coefficient, he does not calibrate the model to the stylized facts relating to trading. An investment horizon of 20 years in place of a fortnight would be sufficient to restore his findings with respect to the required compensation but only at the expense of a reduction in equity and bond turnover to unrealistically trivial proportions. For example, the implied bond turnover rate becomes only 0.284 percent of a realistic estimate. In summary, the apparently large differences in the findings of Constantinidies (1986) and my own do not reflect error by either party but simply the difficulty of calibrating a representative investor model to provide realistic estimates of equity and bond trading in the presence of transactions costs.

Pagano (1989) examines issues of concentration and fragmentation with respect to trading volumes and liquidity utilizing a model of trading between counterparties based on conjectures make about the behavior of other investors. While he departs from the representative investor paradigm by allowing differences in endowments to generate portfolio rebalancing trades, he does not consider proportional transaction costs.

Vayanos (1998) models turnover as endogenously generated by investors with CARA preferences based on life-cycle considerations. In common with my model, transactions costs depend on the number of shares traded rather than the dollar value. He shows that within his framework it is possible for asset prices to rise when transaction costs increase. In my setup, I show this to be impossible. He, in common with Constantinides (1986), finds that transaction costs have a negligible effect on asset prices, but attributes his finding to the inability of life-cycle considerations to generate more than a very small turnover. My ability to replicate Constantinidies (1986) results with a long investment horizon and relatively small trading activity is an effective endorsement of Vayanos's conclusion. A brief summary of this literature is provided by Liu (2004) who models multiple assets as well as transactions costs for CARA investors over a continuous time infinite horizon. Jang *et al.* (2004) find that if stochastic regime switching is introduced into the model of Constantinidies (1986) that transactions costs can have a first-order impact. All of these models, with the exception of Pagano, and Vayanos who introduces an age rather than endowment differential between counterparties, differ from mine in that they model a single representative investor making consumption and portfolio choices. The counterparty trades to trades optimal from the perspective of an individual investor are not modeled.

While Kocherlakota (1996) points out that the average resource cost of transacting is too low to explain a six to eight percent premium, as does Jones (2002), nonetheless a large empirical literature has developed explaining the impact of transaction costs on asset prices. A considerable portion of this literature has been motivated by Amihud and Mendelson (1986) who also carry out one of the first empirical investigations. 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) find more evidence that actual (resource) costs are priced than for the simple bid-ask spread. Datar, Naik, and Radcliffe (1998) establish that stock turnover plays a significant role in the cross-section of returns. Extensions in the same vein are provided by Pastor and Stambaugh (2001), Easley, Hvidkjaer and O’Hara (2002), and Easley and O’Hara (2004).

There is also a considerable literature establishing that stock turnover is sensitive to transactions costs. Demsetz (1968) 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), and Atkins and Dyl (1997).

II. The Model

A. Model Specification

My starting point is a simplified two-period model based on Pagano’s (1989) discrete-time model of strategic trading. Investors have identical CARA preferences induced by exponential utility, which together with normally distributed dividends, yields a simple mean-variance approach. In this framework, investors discount expected future dividends less a risk adjustment at the riskless rate of interest whereas in the CRRA case, increases the discount rate incorporate risk. CARA preferences have been standard in the microstructure literature, and they are being increasingly used within asset pricing (for example, Easley and O’Hara, 2004) and in representative agent models of asset prices with transaction costs (for example, Liu, 2004). There are a total of *N* investors, where *N* is an even number, $N \geq 4$, with identical preferences and no asymmetric information in this simple two-period model of strategic investing by risk adverse investors with heterogeneous endowments wishing to maximize mean-variance utility in terminal wealth. I consider a single risky asset. In the initial period, investors differ only in terms of their initial endowments, with half the population, suppliers, $S = 1,...,N/2$, overly endowed with K_0^S units each the perfectly divisible risky asset, *i.e*, equity shares, relative to the other half, demanders, $D = N/2+1,...,N$, their natural counterparties, each with K_0^D units, where $K_0^D < K_0^S$. The total initial endowment of each supplier together with a demander is given by, $K^T \triangleq K_0^S + K_0^D$ and total fixed supply of the risky asset, $NK^T/2$. Due to the random nature of endowments in the original model, Pagano did not model any such simple dichotomy. I define the degree of heterogeneity, *h*, as the relative difference in the endowments of suppliers and demanders, $h \triangleq \frac{K_0^S - K_0^D}{K^T}$, with an upper limit, $h \leq 1$. The initial resource constraint defined by the number of shares initially held by a representative pair, K^T, holds in the second (terminal) period, in which agents how hold assets according to expected utility maximizing choice. That is, $K^T \triangleq K_1^S + K_1^D$, where K_1^S and K_1^D represent the respective asset demands for shares by suppliers and demanders in the second period. I define the "turnover" rate τ at which stock trading occurs, with transacting supply and demand in balance, as the number of shares changing hands relative to the number on issue, $\tau \triangleq \frac{K_1^D - K_0^D}{K^T} \triangleq \frac{K_0^S - K_1^S}{K^T} \triangleq \frac{\Delta K}{K^T}$, where $K_0^S - K_1^S$ is the number of share units placed on the market by each supplier and $K_1^D - K_0^D$ is the identical number of shares purchased by each demander. Even though there is both a "buy" trade and a "sell" trade for every transaction, I adopt the convention of counting a trade only once.

Each of the *N*/2 suppliers is also endowed with w_0^S in risk-free bonds with a unitary price, and *N*/2 demanders are each endowed with w_0^D units of the same bonds, with $w_0^D > w_0^S$. Bonds pay a certain terminating amount, *R*, $R \triangleq 1 + r$, where *R* is termed the cumulation factor or gross return and *r* the per period interest rate, have no trading costs, and in other respects are just like cash. I call the total endowment per representative supplier and counterparty (demander), w_0^T, $w_0^T \triangleq w_0^S + w_0^D$. The length of the calendar period, T, which defines the period over which trading (turnover) occurs and the gross return, *R*, is earned, is specified in Section III below as part of the calibration exercise for the empirical simulations.

In equilibrium, markets clear as follows: at the end of the first period of the two-period model, suppliers sell the risky asset, equity shares, in return for units of the riskless asset, bonds, while demanders take the other side of the transaction. At the end of the final (second) period, investors convert all assets and payoffs costlessly into units of a consumption good with a unitary price. Suppliers consume their random terminal wealth, $\tilde{c},^{S}\ \tilde{c}^{S} = \tilde{w}_1^{S}$, made up of the normally distributed random terminating gross payoff or dividend per share, $\tilde{d}$, on their smaller equity share holdings, K_1^S, $K_1^S \leq K_0^S$, due to the sale of $K_0^S - K_1^S$ shares. In return for this sacrifice suppliers gain the terminating riskless (gross) return on their higher bond holdings, augmented by the sale of shares in return for bonds at the market-clearing supply (bid) price received by sellers, p^S, which excludes transactions costs, to generate their terminal wealth (budget constraint),

$$\tilde{c}^{S} = \tilde{w}_1^{S} = \tilde{d}K_1^S + R\left[w_0^S + p^S\left(K_0^S - K_1^S\right)\right]. \qquad (1)$$

Demanders who consume their terminal wealth, $\tilde{c},^D \tilde{c}^D = \tilde{w}_1^D$, take the other side of the market, buying $K_1^D - K_0^D$ shares from suppliers at the endogenous market-clearing price and selling bonds in return, with the main difference being the higher demand (ask) price, $p^D \triangleq p^S + a$, per share, relative to the supply (bid) price, where a is the fixed dollar transaction cost per share traded. In this model, I treat the dollar spread as exogenous. Demanders consume their terminal wealth, representing their budget constraint, with the amount,

$$\tilde{c}^D = \tilde{w}_1^D = \tilde{d}K_1^D + R\left[w_0^D - \left(p^S + a\right)\left(K_1^D - K_0^D\right)\right]. \tag{2}$$

The mid-point price, $p^{mp} = \dfrac{p^S + p^D}{2} = p^S + \dfrac{a}{2}$, where $\dfrac{a}{2}$ is the dollar half-spread, with $\dfrac{a}{2p^{mp}}$ the relative mid-point half-spread. The dollar amount, a, is the "round-trip" cost, as the investor who buys and then sells a share incurs a total transaction cost of a.

Each investor has an identical CARA utility function defined over terminal consumption with coefficient of risk aversion, $b > 0$. The gross terminating dividend, $\tilde{d}$, is normally distributed with variance, σ^2, and the expected value is $\mathrm{E}\left(\tilde{d}\right) = \mu$, where E is the expectations operator. Each investor chooses his optimal portfolio of risky shares and riskless bonds to maximize his mean-variance utility function in terminal wealth (consumption),

$$\mathrm{E}\left[u\left(c^i\right)\right] = \mathrm{E}\left(c^i\right) - \left(b/2\right)\mathrm{Var}\left(c^i\right), \forall i, i = 1, \ldots, N, \tag{3}$$

with Var the variance operator.

In conventional rational expectations equilibria investors are "schizophrenic" in that they are supposed to take price as given but know that they influence it unless the number of investors is essentially infinite (Kyle, 1989). I provide every investor with some monopsonistic power with respect to his residual demand so that traders are imperfect competitors. Supposing the i^{th} investor is a demander with an initial endowment of equity shares, $K_0 = K_0^D$, he will be competing against the remaining $\left[\left(N/2\right) - 1\right]$ identical demanders whom he correctly conjectures have identical individual linear demand schedules incorporating the fixed per unit transactions cost a,

$$K_1^D = \alpha^D - \beta\left(p^S + a\right), \tag{4}$$

and will be assisted by the *N*/2 suppliers with initial endowments of equity shares, $K_0 = K_0^S$ who face individual demand schedules,

$$K_1^S = \alpha^S - \beta p^S . \tag{5}$$

The differing initial endowments of demanders and suppliers give rise to differences in the intercept parameters, $\alpha^S \geq \alpha^D$. Unlike demanders who pay the ask price, suppliers receive only the bid price, p^S. As in Kyle (1989) and Pagano (1989), if investors maximize a mean-variance (quadratic) objective function subject to linear conjectures in price about the responses of other traders, a unique symmetric in residual demand schedules Nash equilibrium exists. In effect, each investor acts as a Stackelberg leader with respect to his residual demand in a symmetric leader-follower game.

To find the Nash equilibrium to this problem, I construct and simplify the residual demand facing the *i*th demander, after deducting his own demand, by substituting equations (4) and (5), into his residual demand,

$$\left(\frac{N}{2}-1\right)K_1^D + \frac{N}{2}K_1^S = \left(\frac{N}{2}-1\right)\alpha^D + \frac{N}{2}\alpha^S - \left(N-1\right)\beta p^S - \frac{N-2}{2}\beta a \ . \tag{6}$$

Adding the *i*th demander's own demand to both sides of equation (6), I obtain the conjectural variational condition,

$$\frac{N-2}{2}K_1^D + K_{1i}^D + \frac{N}{2}K_1^S = \frac{N}{2}K^T = \frac{N-2}{2}\alpha^D + \frac{N}{2}\alpha^S - \left(N-1\right)\beta p^S - \frac{N-2}{2}\beta a + K_{1i}^D , \tag{7}$$

with the LHS of equation (7) simplifying to $\frac{N}{2}K^T$, where K^T is the total initial endowment of each trading pair, after recognizing that in equilibrium, $K_{1i}^D = K_1^D$. Expressing equation (7) as the implicit function, $f\left(K_i^D, p^S\right) = 0$, I have $\frac{\partial f(\ )}{\partial K_{1i}^D} = 1$, and $\frac{\partial f(\ )}{\partial p^S} = -\beta\left(N-1\right)$. Hence, the impact of the *i*th demander on the residual supply price is adverse from the perspective of the demander,

$$\frac{dp^S}{dK_{1i}^D} = \frac{1}{\beta\left(N-1\right)} > 0 . \tag{8}$$

Equation (8) captures market impact costs, which are recognized and taken into account by the strategic investor, reducing the number of shares he is willing to purchase accordingly.

Because the variance of terminal wealth equals the product of the variance of dividends and the square of period 1 share holdings, $\operatorname{Var}\left(\tilde{c}^{D}\right)=\sigma^{2}\left(K_{1i}^{D}\right)^{2}$, when I substitute equation (2) into equation (3), take the derivative and use equation (8), the demander maximization of expected utility yields the first-order condition,

$$\frac{\partial E\left(u^{D}\right)}{\partial K_{1i}^{D}}=\mu-R\left(p^{S}+a\right)-\frac{R\left(K_{1i}^{D}-K_{0}^{D}\right)}{\beta\left(N-1\right)}-b\sigma^{2}K_{1i}^{D}=0. \tag{9}$$

On solving equation (9) for asset demand, the demander's asset demand in period 1 is,

$$K_{1i}^{D}=\frac{\mu-R\left(p^{S}+a\right)+\dfrac{RK_{0}^{D}}{\beta\left(N-1\right)}}{b\sigma^{2}+\dfrac{R}{\beta\left(N-1\right)}}. \tag{10}$$

The supplier's asset demand is similar, but with the absence of transactions costs,

$$K_{1i}^{S}=\frac{\mu-Rp^{S}+\dfrac{RK_{0}^{S}}{\beta\left(N-1\right)}}{b\sigma^{2}+\dfrac{R}{\beta\left(N-1\right)}}. \tag{11}$$

To establish that the initial linear conjectures were rational and lead to consistent outcomes I substitute equations (10) and (11) into equation (7) by summing up the period 1 asset demands of the $\frac{N}{2}-1$ identical demanders, plus the demands of the $\frac{N}{2}$ identical suppliers, and then add in the demand of the *i*th demander to both sides, as was the case with respect to equation (7),

$$\frac{N-2}{2}K_{1}^{D}+K_{1i}^{D}+\frac{N}{2}K_{1}^{S}=\frac{N}{2}K^{T}$$

$$=\frac{\dfrac{N-2}{2}\left\{\mu-R\left(p^{S}+a\right)+\dfrac{RK_{0}^{D}}{\beta\left(N-1\right)}\right\}+\dfrac{N}{2}\left\{\mu-Rp^{S}+\dfrac{RK_{0}^{S}}{\beta\left(N-1\right)}\right\}}{b\sigma^{2}+\dfrac{R}{\beta\left(N-1\right)}}+K_{1i}^{D}. \tag{12}$$

On differentiating equation (12) expressed as an implicit function with respect to $p^{S}$and K_{1i}^{D} to compute the conjectural response, $\frac{dp^{S}}{dK_{1i}^{D}}$, and equating it to the slope of the market clearing conjectural condition, equation (8), I evaluate the slope term as,

$$\beta = \frac{N-2}{N-1}\frac{R}{b\sigma^2}. \tag{13}$$

Equating the constant terms in (7) and (12) produces the two conjectural intercept coefficients,

$$\alpha^D = \frac{N-2}{N-1}\frac{\mu}{b\sigma^2} + \frac{K_0^D}{N-1} \text{and } \alpha^S = \frac{N-2}{N-1}\frac{\mu}{b\sigma^2} + \frac{K_0^S}{N-1}. \tag{14}$$

Hence, the initial linear conjectures about the intercepts and slope of the demand schedule are correct and therefore self-fulfilling with the market clearing. A variation is consistent if it is equivalent to the optimal response of other investors at the equilibrium defined by that conjecture (Perry, 1982). These conjectures are consistent. Note that neither the intercepts, $\alpha^D$and α^S, nor the slope, β, depends directly on transaction costs. Rather, they depend on the number of market participants, *N*, the coefficient of absolute risk aversion, *b*, risk (volatility), σ^2, expected earnings/dividends, μ, and initial endowments which ultimately reflect endowment heterogeneity, *h*.

Substituting for the parameters in the demander's demand equation (4) and simplifying yields the risky asset holdings of demanders as a function of the demand price, $p^D = p^S + a$,

$$K_1^D = \alpha^D - \beta\left(p^S + a\right) = \frac{K_0^D}{N-1} + \frac{N-2}{N-1}\frac{\mu - R\left(p^S + a\right)}{b\sigma^2}, \tag{15}$$

and, similarly into (5) for suppliers,

$$K_1^S = \alpha^S - \beta p^S = \frac{K_0^S}{N-1} + \frac{N-2}{N-1}\frac{\mu - Rp^S}{b\sigma^2}. \tag{16}$$

Since the sum of the demands equals the initial endowment of the trading pair, $K_0^D + K_0^S \triangleq K_1^D + K_1^S \triangleq K^T$ and the $N/2$ demander demands are identical, as are the $N/2$supplier demands, summing (15) and (16), solving for the market clearing supply price, p^S, and simplifying, yields the demand (ask) and supply (bid) prices, respectively,

$$p^D = p^S + a = \frac{\alpha^S + \alpha^D - K^T}{2\beta} + \frac{a}{2} = \frac{\mu - \frac{b\sigma^2}{2}K^T}{R} + \frac{a}{2}, \tag{17}$$

and

$$p^S = \frac{\alpha^S + \alpha^D - K^T}{2\beta} - \frac{a}{2} = \frac{\mu - \frac{b\sigma^2}{2}K^T}{R} - \frac{a}{2}. \tag{18}$$

These are the Nash equilibrium conditions pertaining to the entire market.

Because of CARA preferences, the economy-wide market clearing price is the certainty equivalent payoff, the expected gross dividend measured net of the risk adjustment and discounted by the gross riskless return with an adjustment for the dollar half-spread, $\frac{a}{2}$, either side of the midpoint price, p^{mp}, where,

$$p^{mp} = p^S + \frac{a}{2} = \frac{\alpha^S + \alpha^D - K^T}{2\beta} = \frac{\mu - \frac{b\sigma^2}{2}K^T}{R}. \tag{19}$$

Remarkably, the midpoint price with transaction costs in place, p^{mp}, given by (19), is completely independent of the dollar spread, *a*, so long as the expected dividend, μ, is independent of transaction costs, and is precisely equal to the bid and ask price in the complete absence of transaction costs, denoted $p^{a=0}$. This is because the transactions cost (or equivalently, tax) does not affect either the intercepts or slopes of the supply and demand functions. Implicit in this result is the inability of investors to be able to continue to trade the asset without the encumbrance of transaction costs. Thus, suppose the dollar spread, *a*, is a tax with all assets subjected to the same tax. The entire tax incidence falls on traders with the mid-point price unaltered. By contrast, the conventional tax incidence story is one of “passing on” in that some or all of the tax takes the form of a higher price.

An important new insight that arises because of the requirement for a market-clearing asset price, absent from most asset pricing models, is the inclusion of the number of shares held in total by the pair of trading investors in the expression for asset price. The greater the risk sharing required between investors, by virtue of higher aggregate supply, K^T, the lower is the asset price. Substituting these market-clearing asset pricing equations, (17) and (18), into the respective demands, equations (15) and (16), yields equilibrium asset holdings for both investor types as a function of transactions costs, a,

$$K_1^D = f^D(a) \triangleq K_1^D(a) = \frac{1}{2}\left(\alpha^S - \alpha^D + K^T - \beta a\right) = \frac{K_0^D}{N-1} + \frac{1}{2}\frac{N-2}{N-1}\left(K^T - \frac{R}{b\sigma^2}a\right), \tag{20}$$

$$K_1^S = f^S(a) \triangleq K_1^S(a) = \frac{1}{2}\left(\alpha^S - \alpha^D + K^T + \beta a\right) = \frac{K_0^S}{N-1} + \frac{1}{2}\frac{N-2}{N-1}\left(K^T + \frac{R}{b\sigma^2}a\right). \tag{21}$$

Unsurprisingly, transactions costs enter into equilibrium asset demands in a symmetric but oppositely signed fashion, both discouraging prospective demanders from buying and encouraging prospective sellers to retain their existing ownership.

Asset stock equilibrium immediately establishes asset flow equilibrium. The equilibrium turnover demand, $\tau = f\left(a\right) \triangleq \tau\left(a\right)$, in the form of identical but differently signed buy and sell orders relative to shares outstanding and obtained from equations (20) and (21), becomes, after simplification,

$$\tau\left(a\right) = \frac{K_1^D - K_0^D}{K^T} = \frac{K_0^S - K_1^S}{K^T} = \frac{1}{2}\left(h - \frac{\alpha^S - \alpha^D - \beta a}{K^T}\right) = \frac{1}{2}\left(\frac{N-2}{N-1}\right)\left(h - \frac{R}{b\sigma^2 K^T}a\right), \quad (22)$$

on computing the difference between the final and initial asset holdings of the demander, ΔK, or supplier since the market clears, while substituting for endowment heterogeneity, *h*. Its maximum value is obtained at *a* =0, with $\tau\left(0\right) = \frac{1}{2}\left(\frac{N-2}{N-1}\right)h \leq \frac{1}{2}$, as the maximum value of *h* is 1. The inverse function, $\tau\left(a\right)^{-1}$, is

$$\tau\left(a\right)^{-1} = a\left(\tau\right) = \frac{b\sigma^2 K^T}{R}\left(h - 2\frac{N-1}{N-2}\tau\right). \quad (23)$$

If stock turnover is defined differently, as it is by some exchanges, with both buy and sell trades counted, then the expression, $\frac{1}{2}$, on the RHS of (22) becomes simply, 1.

B. Comparative Statics

Trading demand is linear in trading costs with a positive intercept which is increasing in the initial degree of asset endowment heterogeneity, *h*, and downward sloping in dollar trading cost *a*. What is remarkable about this finding is that the product of all manifestations of risk, the CARA coefficient, *b*, volatility, σ^2, and the available number of risky shares, K^T, to be traded between the parties, act to overcome the discouraging impact of transactions costs, a, on the propensity to trade, $\tau\left(a\right)$. Thus in richer communities with a greater supply of risky assets per capita, *i.e.*, higher K^T, trading activity should be more intense for a given dollar round-trip spread. Moreover, since all manifestations of risk enter in a multiplicative fashion, they are perfect substitutes in the sense that doubling any one has the same impact as doubling another. While the asset price, as indicated by (19), is unaffected by either transaction costs or the imposition of a specific per unit tax, trading activity is clearly harmed by transaction costs or taxes.

The inverse function, equation (23), also provides new insights. It is pictured in Figure 1, which is drawn to scale and assumes monthly trading, a CARA coefficient, b = 1, annualized $\sigma^2 = 0.1225$, $R = 1.02$, h =1 and K^T =2. A doubling in the number of investors, from four to eight with the same per capita endowment of the risky asset, rotates this function counter-clockwise to the right as the market depth increases, around the autarky point, $b\sigma^2 K^T h / R$. This point provides an upper bound to the observed transaction costs, $\max a \triangleq \overline{a}$. The schedule flattens out as the number of participants increase. Trading activity is increasing in the size of the market due to favourable market externalities. As $N \to \infty$, strategic behaviour evaporates.

In "thin" markets, with few potential participants and little opportunity for risk sharing, there is less trading because "market impact" costs are high. This is due to the recognition by the strategic investor that his own actions turn the terms of trade against himself due to his monopsonistic power. The model, in the way it is specified, does not capture benefits due to the ability to share risk amongst a larger number of participants, as more participants increases the number of risky assets in the same proportion. However, implicitly, for a given supply of risky assets (shares), an increase in the number of investors, N, lowers shareholding per trading pair, K^T, and improves risk sharing thus raising the asset price. However, by making investors more sensitive to trading costs, it reduces trading per investor pair. With more dispersed ownership, transacting plays a less vital role. Increases in risk aversion, volatility, shares on issue, and endowment heterogeneity all shift up the schedule, raising the optimal degree of mutual portfolio rebalancing.

Insert Figure 1 about here

The elasticity of turnover demand with respect to transaction costs,

$$\eta_a^\tau = \frac{\tau'(a)}{\tau(a)} a = -\frac{Ra}{b\sigma^2 K^T h - Ra} < 0, \tag{24}$$

found by differentiating (22), becomes more inelastic as trading opportunities increase, *i.e.*, as the degree of risk aversion, volatility, endowment heterogeneity, or supply of the risky asset increases, because the incentive to rebalance the portfolio is now higher. The absolute magnitude of the trading demand elasticity is increasing in transactions costs, so that trading in high transaction cost stocks become even more responsive to changes in transaction costs. The foregone gross yield on the riskless asset, R, reflects the opportunity cost of transactions costs since the dividend occurs only subsequent to trading. Thus, a rise in this yield has exactly the same impact as a rise in transactions costs itself.

C. Compensated Dividend Required to Offset Transactional Cost Impacts

To be willing to hold both the risky asset with transactions costs in place and the identical asset without trading costs, the expected dividend per share on the asset with trading costs must rise by a compensating amount, denoted $c(a)$, to maintain indifference, as pointed out in Constantinides' (1986) seminal contribution. Clearly, the incidence of transactions cost and/or a tax will be entirely different to the situation described by equation (19) above if there remains a perfect substitute for the expensive to transact asset that is free of charges or taxes. For example, the UK Government applies a stamp duty (tax) to trades of equity shares in UK domiciled stocks exclusively while Gilts and T-bills are almost costless to transact and are free of stamp duty. These government securities and foreign-domiciled equity securities are likely to be close substitutes for domestic domiciled equity. Transactions costs reduce the aggregate supply of the riskless asset per investor and counterparty, w^T, in the second period by the amount of the total two-sided costs of trading, $a\tau(a)K^T$. This term represents the actual cost of trades which are "consummated", given the actual spread, *a*. Furthermore, the number of shares held by demanders will be less than the number held by identical suppliers in the post-trading equilibrium, due to the barrier to optimal trading imposed by transactions costs. This reduces efficient risk sharing and thus represents the opportunity cost of "unconsummated" or "forgone" trades that would have occurred with zero transactions costs, requiring additional compensation. Transfers of the riskless asset from the demander to the supplier in exchange for the risky asset simply cancel out, as far as the summed wellbeing of the supplier and demander counterparty is concerned. Thus, with transactions costs in place, aggregate equilibrium utility per supplier and demander counterpart become,

$$U^a \triangleq \left[u\left(c^D\right)+u\left(c^S\right)\ \right]=\left[R\left[w_0^T - aK^T\tau\left(a\right)\right]+\left[\mu+c\left(a\right)\right]K^T-\frac{b}{2}\sigma^2\left[\left(K_1^D\right)^2+\left(K_1^S\right)^2\right]\right],(25)$$

where the function, $\tau(a)$, is specified by equation (22), $a\tau(a)K^T$ is a rectangular area representing the aggregate loss of resources (cash flow) due to transactions costs, and the squared asset demands, $K_1^S(a)$ and $K_1^D(a)$, are found by squaring the transactions-cost sensitive functions, given by equations (20) and (21), respectively. The equivalent of (25), with zero transaction costs at the lower-bound dollar spread, $a = 0$, U^0, is then subtracted from U^a to obtain, after simplification and set to zero,

$$\Delta U = U^{a} - U^{0} = c(a)K^{T} - \frac{Ra}{2}\frac{N-2}{N-1}\frac{N}{N-1}\left(K^{T}h - \frac{R}{2b\sigma^{2}}a\right) = 0. \tag{26}$$

The component of required compensation resulting from forgone trades, due the inability of paired investors to trade as much as they would have liked to do in the absence of trading costs, is the triangular "dead-weight" "equilibrating" or "compensating" utility loss area reflecting the diminution of "consumer surplus" as a result of transactions costs,

$$dwl(a) = \frac{1}{2}\frac{(N-2)R}{(N-1)^{2}}a\left[h + \frac{(N-2)R}{2b\sigma^{2}K^{T}}a\right], \tag{27}$$

which is not a payment to any outside party, and is thus lost to the economy as a whole. Since there are no income effects due to CARA preferences, these three measures are identical.

The required compensating increase in the expected dividend to offset exactly the utility loss is the simple sum of the two sources of investor loss, $dwl(a)$ from (27) plus the actual resource costs, $a\tau(a)$, expressed as,

$$c(a) = \frac{1}{2}\frac{N}{N-1}\frac{(N-2)R}{N-1}a\left(h - \frac{R}{2b\sigma^{2}K^{T}}a\right). \tag{28}$$

The compensating amount is the expected per-period equity premium, expressed in dollar terms, due to illiquidity (*i.e.*, imposition of transactions costs). It is termed the illiquidity premium, or sometimes the liquidity premium, *e.g.*, Constantinides (1986).

With more participants, additional *N*, bringing with them the same endowment of the risky asset per pair of investors, and hence greater market depth, the propensity to trade is greater, as investors trade more aggressively, knowing their own actions are less likely to "spoil the market". Hence, the amount of compensation required for more liquid stocks with higher *N*, $\frac{\partial c(a)}{\partial N} = \frac{2}{(N-1)^{3}} > 0$, is higher. Moreover, the greater the propensity to trade, as indicated by a higher risk aversion coefficient, *b*, higher risk, σ^{2}, more risky assets, K^{T}, requiring sharing between the parties, and greater relative endowment heterogeneity, *h*, the greater the compensation required. To express the dollar illiquidity premium as a yield relative to the mid-point asset price, the expected illiquidity premium rate is $e(a) \triangleq \frac{c(a)}{p^{mp}}$. By setting the expected dividend such that p^{mp} =1, the dollar cost, *a*, and relative transactions costs, $\frac{a}{p^{mp}}$, are equated.

D. The Cost-Change Induced Alteration to the Asset Price

If the transactions charge, *a*, is set to zero for the mid-point price, equation (19), I obtain the market-clearing asset price, $p^{a=0}$, in the absence of the transactional charge,

$$p^{a=0} = \frac{\mu - \frac{b\sigma^2}{2} K^T}{R}. \tag{29}$$

As already noted above, equation (29) is in fact identical to equation (19), the mid-point price itself with the transactions cost charge in place, so long as there is no substitute for the asset bearing transactions costs. Suppose that investors are free to choose the asset described in equation (29) and the identical asset with the transaction charge in place. Clearly, no one would be willing to hold the asset with the charge in place at the same price as before the imposition of the trading charge. By how much, therefore, must the share price fall? To answer this, I need to know by how much the expected dividend must rise to make the investor indifferent between the asset with and without the charge in place, with the precise amount, $c(a)$. However, the only mechanism by which the dividend can rise with the imposition of the charge is if the asset price falls. This fall will raise the expected dividend yield by exactly the right amount to maintain indifference.

To compute the actual midpoint price with the charge in place when a perfect substitute is available, I subtract the present value of the required compensation from the market-clearing price in the absence of the charge,

$$p^{mp} = p^{a=0} - \frac{c(a)}{R} = \frac{\mu - c(a) - \frac{b\sigma^2}{2} K^T}{R}. \tag{30}$$

The perpetuity counterpart of the two-period price expression, equation (30), in which the endowment shock and resulting transaction precisely repeat themselves indefinitely, is given by,

$$p^{mp} = p^{a=0} - \frac{c(a)}{R-1} = \frac{\mu - 1 - c(a) - \frac{b\sigma^2}{2} K^T}{r}, \tag{31}$$

where $\mu - 1$ is the net dividend or per-period expected dividend and $R - 1 = r$ is the net or per period bond yield.

The result, given alternatively by equations (30) or (31), is remarkable because it says, essentially, that the midpoint price of a stock with transactions costs is always lower than the price of its liquid counterpart, by more than the present value of actual transactions cost outlays as $c(a) > a\tau(a)$, utilizing the risk-free gross or net return as the basis for the discount factor. By contrast, in much of the asset pricing literature, it is conventional to focus only on the transactions cost cash outlays, to the neglect of the dead-weight utility losses stemming from trades which "should have" been undertaken but weren't due to transactions costs. A consequence of this neglect is that conventional analysis understates the true illiquidity premium, especially for stocks that are highly illiquid due to prohibitive transactions costs.

Note firstly that the fall in the price of the asset subject to transactions costs or a tax has no effect on the stock turnover equation (22). Transactions costs still discourage trading. Also, the turnover rate equation (22) and compensation for illiquidity, equation (28), is applicable to any trading interval since the horizon of investors in the two-period model is not specified *a priori*. If the interval is of calendar length T years then the annualized gross bond and equity yield are $R^{\frac{1}{\mathrm{T}}}$ and $\mu^{\frac{1}{\mathrm{T}}}$ respectively, the annualized net bond and equity yield are $R^{\frac{1}{\mathrm{T}}}-1$ and $\mu^{\frac{1}{\mathrm{T}}}-1$ respectively, the annualized variance is $\frac{1}{\mathrm{T}}\sigma^2$, the annualized stock turnover rate is $\frac{\tau(a)}{\mathrm{T}}$, and the annualized compensation rate implicit in Constantinidies[1] (1986) is,

$$e(a)^{annual} = \frac{c(a)^{annual}}{p^{mp}} = \left[\left(\frac{\mu}{p^{mp}}\right)^{\frac{1}{\mathrm{T}}} - 1\right] - \left[\left(\frac{\mu - c(a)}{p^{mp}}\right)^{\frac{1}{\mathrm{T}}} - 1\right] \approx \left(\frac{\mu}{p^{mp}}\right)^{\left(\frac{1}{\mathrm{T}}\right)-1} \frac{c(a)/p^{mp}}{\mathrm{T}}. \qquad (32)$$

Since $\frac{\tau(a)}{\mathrm{T}}$ and (32) are diminishing in T, both the annualized stock turnover and compensation rate are falling in the investment horizon of market participants that determines the frequency with which investors trade. Thus, I achieve consistency: the longer the investment horizon, T, the lower the valuation of trading activity with less trading activity and lower compensation required for bearing transactions costs. Consequently, the choice of the investment horizon is not arbitrary. It must be set to calibrate the model's predicted equity and bond turnover rates with the stylized facts relating to observed turnover rates. If, for whatever reason, this calibration is omitted then the adoption of an excessively long investment horizon with result in not one but two counter-factual conclusions; trading activity in the presence of transactions costs is insignificant and the required compensation for bearing transactions costs is vanishingly small.

[1] This was kindly pointed out to me by Constantinides in correspondence.

E. Valuing the Ability to Trade

The maximum value of the proportional two-way dollar trading cost, $a \to \bar{a} = \frac{b\sigma^2 hK^T}{R}$, at which autarky occurs with the inverse function, $a(\tau) = 0$, in equation (23), requires a compensating rise in the expected yield, μ, on the risky asset of,

$$c(\bar{a}) = b\sigma^2 \left(\frac{N-2}{N-1}\right)\left(\frac{N}{N-1}\right)K^T \left(\frac{h}{2}\right)^2, \tag{33}$$

found by evaluating equation (28). Alternatively, $c(\bar{a})$ is a measure of the maximum benefits from being able to freely trade, relative to the prohibitive level of transactions cost.

This expression indicates the traditional view, that the price of a stock is equal to the present value of dividends less the present value of transaction costs, is at best only part of the story. In an autarky regime an asset's price is at its lowest since the maximal welfare loss, $c(\bar{a}) > c(a)$, for all $a < \bar{a}$, is sustained, yet by definition no transaction costs are incurred. The compensation required for the imposition of trading costs or stamp duty (trading tax), $c(a)$, is the sum of two components, the actual transaction resource costs and the compensation required for the inability to choose the desired portfolio or make the preferred trade. As actual transactions costs rise above the point that maximizes the transactional cost outlay, the second cost term begins to dominate the first. Thus, even though resources consumed actually transacting may be zero because of prohibitive transaction costs, the compensation required under autarky, as the elasticity of trading demand approaches infinity, will exceed the maximum rate of transaction costs at the point of unitary elasticity of trading demand. Many asset-pricing models incorporate transactions costs via "frictions" which typically only marginally reduce asset returns. They reflect the traditional perspective that only transactions costs actually incurred affect stock returns and asset prices, with the asset price equalling the present value of dividends plus the present value of transactions costs. More commonly, the main costs are not actual costs but rather the neglected opportunity cost of foregone trades. Hence, the almost universal (and misleading) conclusion that transactions cost cannot account for more than a small fraction of the equity premium.

Since the illiquidity cost, $c(a)$, represents the loss to the investor from trading at the transaction cost rate a rather than zero, the benefit from being able to transact at rate 0 < a < $\bar{a}$, rather than zero, $B(a) \triangleq c(\bar{a}) - c(a)$, is

$$B(a)=\frac{N}{2}\frac{N-2}{(N-1)^2}\left[h\left(\frac{b\sigma^2K^Th}{2}-Ra\right)+\frac{(Ra)^2}{2b\sigma^2K^T}\right]. \tag{34}$$

Clearly, the gains from trade, $B(a)$, are diminishing in a for all $a < \bar{a}$, *i.e.*, $B'(a)=-c'(a)<0$. They are also increasing in the size of the market, $\frac{\partial B}{\partial N}=\frac{2}{(N-1)^3}>0$, the intrinsic potential demand for trading, h, representing relative endowment heterogeneity, the degree of risk aversion, $\frac{\partial B}{\partial b}>0$,stock volatility, $\frac{\partial B}{\partial \sigma^2}>0$, and shares held by counterparties, $\frac{\partial B}{\partial K^T}>0$ for all $a < \bar{a}$.

F. *Transaction Cost Rate to Maximize Transactional Outlays*

The problem of choosing a proportional dollar transactional cost amount, $a^{\max}$, which maximizes the transaction cost outlay is the solution to the problem: $\max_a a\tau(a)K^T$, where $\tau(a)$ is given by equation (22) above, with solution,

$$a^{\max}=\frac{b\sigma^2K^Th}{2R}\equiv\frac{1}{2}\bar{a}\,. \tag{35}$$

Thus, the entity wishing to maximize the present value of the transaction cost outlay will choose a level that is exactly half the autarky level, at the point with unitary elasticity of trading demand. A monopoly-specialist who is truly a value-maximizing monopolist will set the commission accordingly.

G. The Illiquidity Compensation Function, Stock Price and Trading Demand

The slope of the dollar compensation function found by differentiating (28) is,

$$c'(a)=\frac{N}{N-1}R\tau(a)>0\ , \tag{36}$$

with an elasticity value,

$$\eta_a^c=\frac{N}{N-1}\frac{Ra\tau(a)}{c(a)}, \tag{37}$$

which depends on the ratio of the resource cost of trading to the illiquidity premium itself.

This means that the incremental illiquidity premium is approximately equal to the stock turnover rate, with the relationship exact for a price-taking investor, after taking account of the delayed benefit following the incurring of transactions costs. Moreover, the midpoint price elasticity of response to a higher transactions cost is approximately equal to the present value of the transactions cost outlays deflated by the midpoint stock price, with an exact relationship as $N \to \infty$,

$$\eta_a^{p^{mp}} = -\frac{1}{R}\frac{c'(a)a}{p^{mp}} = -\frac{N}{N-1}\frac{a\tau(a)}{p^{mp}} < 0, \tag{38}$$

utilizing (36). An increase in transactions cost unambiguously reduces the stock price irrespective of the elasticity of demand for trading so long as investors are free to trade the identical transactions-cost-free asset. This makes perfect sense. Investors cannot benefit from having to pay more to participate in any market via higher transactional costs and stamp duties.

If the dead-weight utility loss triangle, $dwl(a)$, given by (27), is neglected in the specification of $c(a)$ in (28), then transaction cost cash flow, $a\tau(a)$, replaces the compensating amount, $c(a)$, in the pricing equation (30), as it does in much of the conventional asset pricing and tax literature. The price elasticity with respect to transactions costs now becomes,

$$\eta_a^{p^{mp}} = -\frac{1}{R}\frac{a\tau(a)}{p^{mp}}\left(1 - \left|\eta_a^{\tau}\right|\right), \tag{39}$$

which is positive if the absolute value of the turnover demand elasticity with respect to transactions costs, $\left|\eta_a^{\tau}\right|$, is greater than one (elastic). For an illiquid asset with a sufficiently high transactions cost, *a*, to eliminate trading, the stock price, p^{mp}, is maximized at the point where the present value of the transactional cost outlays becomes zero. Hence, in the conventional literature, the stock price falls with higher transactions costs or stamp duty if, and only if, the elasticity of share turnover with respect to transactions costs is smaller than one in absolute value. Note how the conventional analysis implies something quite counterintuitive: adding a transaction cost or imposing a tax on an asset raises its price (value to an investor), the more trading demand declines in response to the imposition of the cost or tax. Thus, if this theory were correct assets for which trades are non-existent because transaction costs are too high, should be the most highly priced and thus the most valuable!

The conventional elasticity, equation (39), is only approximately the same as the true asset price elasticity in equation (38) if, and only if, turnover demand is perfectly inelastic. Think of a clientele model with two investors, one is patient with an investment horizon of two periods and the impatient investor has a one period horizon. The impatient investor turns over his portfolio of the risky asset and bonds once each period and the patient one, by half, irrespective of the absolute and relative costs of trading the two assets. Hence, only in a limiting case in which investor's trading horizon is completely unresponsive to transactional charges, is the investors' objective of maximizing the present value of net cash flows from the portfolio of stocks over this horizon, appropriate. Hence, I validate the consistency of the clientele model of Amihud and Mendelson (1986) based on these assumptions.

Another important finding which stems from (36) is that the compensation function is simply the area under the trading (turnover) demand function over the range of opportunity cost of transacting from 0 to the actual value, a ,

$$c(a) = \frac{RN}{N-1}\int_{x=0}^{a} \tau(x)\,dx\,. \tag{40}$$

In the case of thin trading with relatively small *N*, the area slightly understates the illiquidity premium. I can thus interpret the illiquidity premium as the sum of two components, the transaction cost outlay, $a\tau(a)$, and the triangular dead weight cost area, $dwc(a)$, reflecting the diminution in trading activity due to the imposition of transaction costs. See Figure 2 below. The intuitive reason for this simple relationship between trading demand and the illiquidity premium is that points on the trading demand schedule represent the incremental trading benefit. Due to the absence of income or wealth effects, investor utility remains constant along the schedule. By summing these points over the range denied investors due to trading costs, I capture the compensating return (*i.e.*, consumer surplus variation) necessary to offset the utility loss.

Since $c'(a) > 0$ and $c''(a) < 0$, the compensation function is concave. It is also increasing in the "intrinsic liquidity" of the stock, i.e., stocks with higher endowment heterogeneity, *h*, or a higher intercept, for a given transaction cost, will have a higher "illiquidity" premium, and is hence a "value stock" with a higher expected yield and lower asset price, as a result of being more heavily traded. This result depends crucially on higher turnover for given transactions costs. The finding that the illiquidity premium is increasing in investor endowment heterogeneity is the key to understanding the traditional result that only negligible compensation is required for bearing transactions costs. In traditional representative investor models, and variants that depart only marginally from this paradigm, there is no or insufficient investor endowment heterogeneity to stimulate neither a desire for trading nor any concomitant compensation requirement for bearing transactions costs.

Since stock trading turnover from (22), $\tau(a)$, is itself a function of transactions costs, the illiquidity compensation premium can be expressed directly as a function of stock turnover, $c\left[a(\tau)\right]$, by substituting the inverse function given by (23) into (28). The illiquidity premium is diminishing in stock turnover,

$$c'(\tau) = c'(a)a'(\tau) = -2\frac{N}{N-2}b\sigma^2 K^T \tau(a) < 0\,, \tag{41}$$

as lower transaction costs result in higher turnover and hence reduced required compensation for bearing transactions costs. Remarkably, increased turnover raises the illiquidity premium when it is due to higher endowment heterogeneity but lowers it when due to falling transactions costs for given endowment heterogeneity. Hence, it is important to distinguish between to two alternative means by which trading activity may increase.

The stock turnover function, equation (22), derived from utility maximization is linear in the dollar transactions cost, a. From the perspective of empirical estimation, a more plausible specification, and its inverse is linear in logarithms rather than the level,

$$\tau(\varphi) = \alpha\varphi^{-\gamma} \text{and } \varphi = \tau(\varphi)^{-1} = \left(\frac{\tau}{\alpha}\right)^{-\left(\frac{1}{\gamma}\right)}, \gamma \neq 1, \tag{42}$$

where $\varphi \triangleq \dfrac{Ra}{p^{mp}}$ is the relative transactions cost in opportunity cost terms and γ is the absolute value of the (constant) elasticity of demand for turnover with respect to transactions costs. Even without knowing the exact form of the utility function and the budget constraint of portfolio rebalancing investors from which I have implicitly derived equation (42), I can obtain all the information required directly from the empirically or theoretically specified stock turnover function by simple integration. The illiquidity premium, expressed in terms of both opportunity costs and stock turnover, becomes,

$$c(\varphi) \cong \int_{\varepsilon}^{\varphi} \alpha x^{-\gamma} dx, \text{ as } \varepsilon \to 0, \cong \frac{\alpha\varphi^{1-\gamma}}{1-\gamma} = \frac{\tau\varphi}{1-\gamma}, \tag{43}$$

$$c(\tau) \triangleq c\left[\varphi(\tau)\right] = \frac{\alpha}{1-\gamma}\left(\frac{\tau}{\alpha}\right)^{-\frac{1-\gamma}{\gamma}}, \gamma \neq 1, \tag{44}$$

where ε is a vanishingly small positive amount, $\alpha = \left[Ra\right]^{\gamma}\tau$, and $\gamma = 1 - \frac{\tau Ra}{c(a)}$ are parameters obtained from the transactions costs and illiquidity premium from the linear equation (22) and (28). Finally, the marginal impact of stock turnover becomes,

$$c'(\tau) = -\frac{1}{\gamma}\left(\frac{\tau}{\alpha}\right)^{-\left(\frac{1}{\gamma}\right)} = -\frac{\varphi}{\gamma}, \qquad (45)$$

on substituting for transactions cost, as the higher turnover is converted into a falling illiquidity premium, along with the lower opportunity cost of trading. Note that the slope is independent of τ and α. It is thus independent of the time interval.

III. Supporting Evidence

A. *Why Is the Illiquidity Premium Not as High on Bonds as it is on Equity?*

Over the 25 year period, 1980-2004, for which data is available the average turnover rate on marketable US Treasury securities was 16.56 times per annum while on the NYSE the comparable rate for equity was 0.64 (see Table I). The ratio of the security turnover rate to the equity turnover rate was, on average, 25.87 times over this period.

Insert Table I about here

The much greater liquidity of Treasury securities is not an exclusive property of the US market. Data is also available for two other relatively comparable markets, namely 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).[2] 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 have higher liquidity than Gilts. Over the same five-year period equity turnover on the ASX rose from about 0.25 to 0.5 times per annum,

[2] 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.

so that the ratio of bond to equity turnover fell from 33.3 to 17.2 times over this period.[3] The average experience over this period is quite similar to the US. Consequently, government bonds, including Gilts, are exceedingly more liquid, *i.e*., higher turnover, than equity in all three countries. Thus, if I were to explain the demand for trading Treasury securities by investors with identical preferences to those trading equity, I would expect to find significant differences in transactions costs between the two markets, with a much lower illiquidity premium for liquid bonds.

B. Numerical Simulation of the Equity and Bond Markets, 1889-1994.

I take up the challenge to explain all important features of both equity and bond markets over nearly a hundred year period, such that all investors have identical mean-variance (i.e., exponential) utility functions with identical CARA coefficients. A simple numerical example based on simulating the model is provided in Table II, to help with understanding of the model and to replicate all the important features known about the performance of the US stock and Treasury Bill market over the period, 1889 to 1994, which was the subject of Mehra and Prescott's (1985) equity premium study. The key facts are a mean six percent per annum equity premium over the Standard and Poors stock index, a 18% annual standard deviation of returns (3.24 percent variance) and a two percent per annum riskless real return on T-bills with a standard deviation slightly under six percent (Campbell, Lo and MacKindlay, 1997). Cochrane (2005, p.21) reports an equity premium over the last 50 years in the US of eight percent so I compromise at seven percent. Jones (2002) computes the average round-trip relative transactions cost (spread plus commission relative to price) of approximately 1.68 percent over the period 1925-2000 for the largest and presumably most liquid Dow Jones stocks on the NYSE. Costs over the period, 1889-2024, are likely to have been similar or higher. The annual average stock turnover rate is conservatively 38 percent per annum. Jones's estimated annual average round trip resource cost is approximately 0.76 percent per annum. To be on the conservative side, and to represent more recent experience following deregulation of commission rates in the 1970s, a round trip relative rate of 0.0098, or just under 1 percent, is assumed. Luttmer (1996) computes the two-way spread between the bid and ask for 3-month T-bills to be far lower than for equity at only 0.03 percent, although higher on an annualized basis. The coefficient of absolute risk aversion, *b*, is assumed to equal 1 for all equity and bond investors. The endowment of each seller, buyer pair is $K_0^S = 7.5, K_0^D = 0$, with $K^T = 7.5$ and relative endowment heterogeneity, *h*, is at its maximum of 1. This is true for all equity and bond

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

investors. The number of investors is set as a large number (N = 10,000) to ensure price-taking behavior with negligible market impact costs. The investment horizon, T, is set at $1/24$ of a year. Hence, fortnightly portfolio rebalancing is assumed for all equity and bond investors, so that all the reported values generated by the model, such as the expected returns and illiquidity premium from (32), are annualized values generated from the fortnightly model.

Insert Table II about here

The first column shows the solution for a liquid asset with no transaction costs (a = 0), but with the net expected return or dividend, $\mu - 1$= 14.15 percent, set to generate an asset price, $p^{a=0}$ =1, when there is a 2 percent per annum yield on riskless bonds. Thus, even in the absence of an illiquidity premium, the model predicts quite a high risk premium just to compensate for the average risk level observed over the 100 year period. The annualized turnover rate at 12 fold is also very high in the absence of transactions costs. Moreover, with a negligible trading cost, the compensation required per unit of transactions costs, $c(a)/a$, is exceedingly high at 13.75 fold, indicating the significant deleterious impact of even apparently insignificant transactional costs, or taxes for that matter. In the second column the same asset and still without transactions costs has been given the same net expected return, $\mu - 1$= 22.28 percent, as the base-case, costly to trade, asset in column 3. The expected return on the base-case asset is just sufficient to generate a price, p^{mp} = 1, with the very moderate transactions costs in place. Similarly, the zero transaction cost bond in column 5 is paired with the base-case bond in column 6, reflecting the observed transaction costs on T-bills with a common net return of $\mu - 1$ = 2.08 percent. In every column the utility of investors in the same asset with and without transactions costs are shown to be the same. In particular, the asset in column 2 without transactions costs but with the same dividend stream as the base-case asset, column 3, yields the same utility as the base-case asset as its one-period asset price, $p^{a=0}$, has appreciated by exactly the amount of the fortnightly illiquidity premium, 25 basis points. On a perpetuity basis utilizing equation (31), this amounts to an asset price in the absence of transactional costs that is over 400 percent higher, $p^{a=0}$ = 4.0635. The base case yields an annualized equity premium due to transaction costs of 7.15 percent and the average turnover rate of 0.38 or 38 percent. The optimal turnover rate for bonds shown in column 6 is 880 percent per annum, or 24 times higher than the equity turnover rate. This approximately matches the historical record over the period, 1980-2004 (Table I above). Given the stylized facts to be explained, calibration of the model sets a maximum value for the investment horizon, $\mathrm{T} = 1/24$ of a year. Even a slightly higher value would result in an annualized turnover rate for bonds of less than 8.8 times per annum, which would then conflict with the historical record. Moreover, the equity turnover rate would be too low as well. In correspondence, Constantinides has pointed out to me that an investment horizon of (say), $\mathrm{T} = 20$ years would generate an annualized illiquidity premium consistent with Constantinidies (1986). He is perfectly correct in this respect as the annualized premium is reduced to the negligible value of 0.0001. However, calibration is impossible as the annualized equity turnover rate is reduced to 2.5 percent and, more significantly, the annualized bond turnover rate is also reduced to 2.5 percent, or 0.284 percent of my conservative historical estimate.

The 75 percent reduction in the price of the base-case asset, relative to the liquid asset, is not accounted for by the present value of transaction costs which are quite small at about 50 cents. Rather, it is because the asset demander would like to purchase 3.75 units of the risky asset per fortnight (column 2) but can only purchase 0.117 and the seller is optimally required to bear excessive risks while holding an undesirably large balance. This inability to equalize the marginal impact of risk via exchange has a considerable disutility cost reflected in the low market clearing price for the illiquid asset, even though it may seem strange that an apparently insignificant transactions cost of less than 1 percent round trip cost actually reduces the asset price by 75 percent. At the point of unitary elasticity of trading demand in the forth column marked “max outlay”, the premium is slightly lower at 5.33 percent and the cost rate is half the autarky amount at 0.513 percent.

A possible objection to the realism of the moderate volatility base-case solution is the high reported elasticity of trading demand with respect to transaction costs, η_a^τ , of -30.96. However, the corresponding parameter of the constant elasticity demand specification, equations (42) and (43), is the constant absolute value of the demand elasticity, $\gamma = 0.9393$, such that the transactions cost, turnover rate, and illiquidity premium are identical, indicating a plausible average transaction cost elasticity of slightly less than 1 in absolute value. Jones (2002) provides an estimate of this elasticity for Dow Jones stocks over the period, 1926-2000, which is slightly higher at 1.13. Hence, the estimate obtained from my simulation is quite conservative.

Columns 5-7 are similar to the previous columns, 2-4, except that they model the illiquidity premium for three-month T-bills instead of equity using the basic facts for T-bills provided above and the identical investor preferences and endowments as for the equity case. The notional riskless rate for T-bills, and also free of transactions costs, has been set at an annualized rate of 0.4 percent. In the base case, column 6, and also for T-bills with no transactions costs, column 5, the annualized net expected return is 2.08 percent, very close to the rate of two percent utilized for T-bills in the first four columns. This rate is the sum of the notional riskless rate, the required compensation for risk given the known volatility of T-bills and the illiquidity premium for T-bills. The annualized illiquidity premium for T-bills is 0.32 percent in the base case. Thus, in summary, I find that the entire difference in the performance of the T-bill market relative to equity is accounted for by differences in volatility and transactions costs alone since investors have identical preferences, endowments, and investor horizons. The base case equity premium solution is drawn to scale in Figure 2. The equity premium corresponds to the sum of the transactional outlay rectangle plus the dead-weight loss triangular area. In this simulation the latter is 18.4 times bigger than the former. The base case T-bill simulation is drawn to scale in Figure 3.

Insert Figures 2 about here

Insert Figure 3 about here

To illustrate the impact of market "thinness" on the base-case outcome, the number of participants, *N*, is reduced from 10,000 to only four. Depending on the stock and time of day, at any one time the number of institutional investors prepared to undertake a large block trade could be quite small. Preserving the same price of $1, the expected dividend falls marginally by only 3.34 percent, the illiquidity premium by 11.1 percent, and the stock turnover rate by a very significant 33.3 percent. In the absence of economies of scale and scope, the degree of market "thinness" only has a small impact on yields but a much larger impact on stock turnover, as was indicated in Figure 1 above.

Perhaps the most remarkable finding from the model simulation is that the ratio of the illiquidity premium to the transactional charge causing it, $\frac{e(a)}{a}=\frac{c(a)}{p^{mp}}\frac{1}{a}$, is approximately 14-fold for relatively small values of *a*, and over seven-fold in the base case. This is 98 to 184-fold higher that the value obtained by Constantinides (1986) of approximately $\delta(k)/k$ = 0.15, where δ is the annualized compensation according to equation (32) and $k=0.5a$ is the half-spread, adopting the notation in his classic paper. In my model investor endowment heterogeneity, *h*, and the number of equity shares to optimally held by each natural trading pair is specified, enabling calibration of my model to precisely replicate the observed annualized equity premium and trading intensity, $\tau(a)/T$, for equity and bonds.

Another significant finding shown in the Table II simulation is the negative impact of transactions costs on stock price with an elasticity of -0.18 in the base case equity simulation.

This is just slightly less than the finding by Jackson and O'Donnell (1985) measuring the response of the LSE stock price to a reduction in the rate of stamp duty. The Jackson and O'Donnell finding indicates that there are close substitutes to UK equity securities not subject to stamp duty, e.g., Gilts and foreign domiciled equity securities.

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

Datar, Naik, and Radcliffe (1998) conclude that a one percent drop in the monthly percentage stock turnover rate for non-financial firms on the NYSE increases the cross-sectional monthly return by 4.5 basis points over the period, 1962-1991, conditional on the Fama-French (1992) factors, size, book to market, and CAPM beta. Evaluating the slope of the compensation function with respect to turnover using the base case information in the equity simulation (Table II, column 3), the monthly turnover rate from the empirical study and equation (25) above, $c'(\tau) = 0.000377$, or approximately four basis points per month. Hence, my base case equity simulation almost perfectly duplicates the Datar, Naik, and Radcliffe (1998) empirical finding.

D . 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 a form of private placement that 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 years[4]. 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 base case simulation with moderate volatility using the equivalent constant elasticity specification. The imposition of a prohibitive relative transactions cost yields an illiquidity premium of 19.5% per annum. Hence, the model requires no special assumptions in order to generate a premium consistent with the evidence of Silber and Pratt.

E. 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, 1994-1998, inclusive, from the Security Industry Research Centre of Asia-Pacific (SIRCA's) trade by trade database.[5] 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

[4] 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.

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

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, φ_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 percent 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,

$$c\left(\tau_{et}\right) = \alpha^{\rho}\left[\rho/\left(\rho - 1\right)\right]\left[\tau_{et}^{1-\rho} - \tau_{b}^{1-\rho}\right], \tag{46}$$

and

$$\tau_t = \alpha\varphi_t^{-\left(1/\rho\right)} \tag{47}$$

where the T-bill turnover rate, τ_b, is set at an annualized rate of eight based on the Australian evidence.[6] The first of the two equations to be simultaneously estimated, (46), is the general equity premium result, (44) above with the implied Treasury bill transaction cost, φ_b, solved for in terms of the known turnover rate, τ_b, the intrinsic liquidity parameter, α, and the inverse of the turnover elasticity, $\rho \triangleq \left(1/\gamma\right)$. The second equation, (47), is simply the equity turnover relationship, (42) above. Simultaneous estimation ensures that consistency is maintained between the estimates of the equity premium and turnover regressions. In particular, it forces the turnover elasticity values in the specification of the equity premium and the turnover equation to be the same, as the theory predicts.

The model was first estimated for the full data set consisting of 24,350 monthly observations. Summary statistics are shown in Table III. 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.

Insert Table III about here

[6] 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 III.A above).

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. The high volatility of stock turnover helps provide an explanation for the "volatility puzzle", just as trading and transaction costs provide an explanation for the equity premium.

The results are summarized in column 1 of Table IV. 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 less than unity in absolute value and also highly significant (β = 0.78137; $\rho = 1.2798$, *t*-stat. = 139.63), and implied T-bill transaction cost, φ_b = 0.000311, is three basis points and therefore in agreement with the US evidence from Luttmer (1996) and the T-bill simulations in Table II above. Both the estimated coefficients are significant at better than the one percent level. The implied equity premium is approximately three percent. The estimated turnover elasticity is consistent with most empirical studies and clearly rejects the implicit assumption made in some theoretical models of the illiquidity premium that the turnover rate is unresponsive to transactions costs.

Insert Table IV about here

The model was re-estimated in column 2 of Table IV including only stocks of above median market capitalization given by $57.45m. This increases the value of the intrinsic liquidity parameter, α, by approximately one-third while the estimated transaction cost elasticity falls to γ= 0.72145. For smaller than median stocks in column 3 the intrinsic liquidity parameter is lower and this is accompanied by a higher elasticity, γ = 0.87138. Hence smaller stocks tend to have higher turnover elasticities, compensated for by a lower intrinsic liquidity parameter, α.

The model using the full data set is re-estimated in column 4 allowing for the respective equity premium and turnover elasticity inverses, ρ_e and ρ_τ, to differ between equations (46) and (47). The implied turnover elasticity estimated from the first equation is γ_e= 0.798021 and from the second, γ_τ = 0.760572. Hence the differences are very slight. This is yet further confirmation of the strength of the model.

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

Equations (46) and (47) above imply two paths by which the turnover elasticity, γ, can be estimated. The two paths should generate the same γ outcome. Firstly, γ is identified with respect to transaction costs, φ_e, directly from turnover information, τ_e, and, secondly, γ indirectly defined from the equity premium, $c(\tau)$, and the resource cost, $\tau_e\varphi_e$. 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,

$$\tau_e = \alpha\varphi_e^{-\gamma}, \tag{48}$$

and

$$c(\tau) = \alpha_0 + \tau_e\varphi_e/(1-\gamma-\alpha_1), \tag{49}$$

with the γ estimate the same in both equations if α_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 ten times a day. Thus, 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 γ elasticity estimates from each equation. Hence, the hypothesis of equal γ estimates is accepted for 76 of the 90 equations. There were also 17 instances in which the estimate of γ itself failed to be both positive and have a significant *t* value. However, for ten 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 γ in these cases. Consequently, the hypothesis of a zero γ elasticity is rejected for 73 of the 90 stocks. Moreover, the joint hypothesis of the same γ estimate from both equations and a positive and significant γ 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 supportive of the model.

IV. Conclusions

In this paper I develop the first model of strategic exchange of risky assets incorporating proportional transactions costs and the requirement that investors be indifferent between an asset with transactions costs and an identical one without. Exceedingly simple, attractive and tractable closed-form solutions are obtained in the model. These explain asset turnover as a linear function of transactions costs and the strategic behavior of an arbitrary number, *N*, of investors who recognize that their own trading behavior results in market impact costs that impinge adversely on themselves. Mutual exchange in the form of portfolio rebalancing between risk adverse investors with identical CARA preferences increases with endowment heterogeneity. It is more resilient to transactions costs the greater is aversion to risk, volatility and the amount of risk (number of shares) that have to be held by any pair of traders. These three aspects of risk act as perfect substitutes in terms of encouraging resiliency. Establishing asset prices by modeling the exchange of assets is urged by O'Hara (2003) in her Presidential Address.

The model explains how investors with standard CARA preference in common, low risk aversion and facing the actual volatilities of equity and T-bills and very conservative measures of actual transactions costs over the period 1896-1994 will generate as an equilibrium a seven percent equity (illiquidity) premium and two percent bond yield with precisely the same security turnover rates for equity and bonds as actually observed, 38 percent and about 880 percent per annum, respectively. The model also explains the cross-sectional returns on the NYSE with respect to stock turnover found by Datar, Naik and Radcliffe (1998), the letter stock puzzle (Silber, 1991), the impact of changes to the rate of stamp duty tax on the London Stock Exchange on stock prices, the equity premium in monthly returns on the Australian Stock Exchange and the identity of estimates of the stock turnover relationship found directly from transactions data and implied by the equity premium itself. My findings suggest that the welfare cost of the imposition of Tobin taxes (stamp duties) on financial markets is likely to be very high. They also show that the imposition of a higher transaction cost or trading tax on an asset possessing an untaxed substitute must always reduce rather than enhance its value. This remains true even when trading demand is elastic. Finally, the model provides a tractable vehicle for investigating a whole range of related issues by incorporating informational effects and other features of actual markets into a model of asset prices determined by mutual exchange.

References

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

Atkins, Allen B. and Edward A. Dyl, 1997, "Transaction Costs and Holding Periods for Common Stocks," *Journal of Finance* 52 (1), 309-325.

Barclay, Michael J., Eugene Kandel and Leslie M. Marx, 1998, "The Effect of Transaction Costs on Stock Prices and Trading Volume," *Journal of Financial Intermediation*, 7 (2) (April), 130-150.

Black, Fischer, 1986, "Noise," *Journal of Finance* 41 (3) (July), 529-543.

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., Andrew W. Lo and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Campbell, John Y., 2000, "Asset Pricing at the Millennium," *Journal of Finance* 55 (4) (August), 1515-1567.

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

Cochrane, John H., 2005, *Asset Pricing,* Revised Edition, Princeton University Press, Princeton.

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

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

Demsetz, Harold, 1968, "The Cost of Transacting," *Quarterly Journal of Economics* 82, 33-53.

Eleswarapu 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.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, "Is Information Risk a Determinant of Asset Returns?", *Journal of Finance* 57, 2185-2222.

Easley, David and Maureen O'Hara, 2004, "Information and the Cost of Capital", *Journal of Finance* 59 (4), 1553-1583.

Fama, Eugene F. and Kenneth R. French, 1992, "The cross-section of expected stock returns," *Journal of Finance* 47, 427-465.

Fisher, Stephen J., 1994, "Asset Trading, Transaction Costs and the Equity Premium," *Journal of Applied Econometrics* 9, S71-S94.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan, 2004, "Equilibrium in a Dynamic Limit Order Book Market". Forthcoming, *Journal of Finance*.

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.

Jang, Bong-Gyu, Hyeng Keun Koo, Hong Lui and Mark Lowenstein, 2004, Transaction Cost can have a First-Order Effect on Liquidity Premium, Ohlin School of Business, Washington University in St Louis, November 20.

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

Jones, Charles M., 2002, A Century of Stock Market Liquidity and Trading Costs, Working Paper, Columbia University.

Kocherlakota, Narayana, 1996, “The Equity Premium: It’s Still a Puzzle,” *Journal of Economic Literature* 34 (March), 42-71.

Kyle, Albert S., 1989, Informed Speculation with Imperfect Competition, *The Review of Economic Studies* 56 (3), 317-355.

Liu, Hong, 2004, “Optimal Consumption and Investment with Transaction Costs and Multiple Risky Assets”, *Journal of Finance* 59 (1), 289-338.

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.

O’Hara, Maureen, 2003, “Presidential Address: Liquidity and Price Discovery”, *Journal of Finance* 58 (4), 1335-1354.

Pagano, Marco, 1989, Trading Volume and Liquidity, *The Quarterly Journal of Economics* 104 (2), 255-274.

Pastor, Lubos and Robert Stambaugh, 2001, “Liquidity Risk and Stock Returns”, working paper, University of Pennsylvania.

Perry, Martin K., 1982, “Oligopoly and Consistent Conjectural Variations, *The Bell Journal of Economics* 13 (1), 197-205.

Pratt, S. P., 1989, *Valuing a Business,* (Irwin, Homestead, Il.).

Schwert, G. W., 1990, “Indexes of U.S. Stock Prices from 1802 to 1987,” *Journal of Business,* 63 (3), 399-426.

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.

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.

Table I: Derivation of Turnover Rates for US Treasury Securities and NYSE Equities, 1980-2004.

Year	1	2	3	4	5	6	7	8
1980	616	11.4	6.9	7.72	11,352	31,871	0.36	21.67
1981	683	13.3	11.2	9.32	11,854	36,004	0.33	28.32
1982	824	17.4	14.8	10.16	16,458	38,907	0.42	24.01
1983	1,024	23.3	18.8	10.69	21,590	42,317	0.51	20.94
1984	1,247	28.5	24.3	11.01	23,071	47,105	0.49	22.47
1985	1,438	39.6	35.8	13.64	27,511	50,759	0.54	25.16
1986	1,619	53.3	42.3	15.35	35,680	56,024	0.64	24.11
1987	1,725	64.6	45.6	16.61	47,801	65,711	0.73	22.84
1988	1,821	63.0	39.2	14.59	40,850	73,989	0.55	26.43
1989	1,945	69.8	43.1	15.09	41,699	79,574	0.52	28.79
1990	2,196	68.7	42.5	13.17	39,665	86,852	0.46	28.83
1991	2,472	78.5	49.0	13.41	45,266	95,177	0.48	28.20
1992	2,754	95.7	56.4	14.36	51,376	107,731	0.48	30.11
1993	2,990	107.7	65.9	15.10	66,923	123,446	0.54	27.85
1994	3,126	116.1	75.2	15.91	73,420	136,667	0.54	29.62
1995	3,307	112.7	80.5	15.19	87,218	148,500	0.59	25.86
1996	3,460	117.3	86.4	15.31	104,636	165,832	0.63	24.26
1997	3,457	120.9	91.2	15.95	133,312	192,017	0.69	22.98
1998	3,356	126.5	100.1	17.56	169,745	223,196	0.76	23.09
1999	3,281	101.3	85.3	14.79	203,914	260,116	0.78	18.86
2000	2,967	98.6	108.0	18.11	262,478	297,433	0.88	20.52
2001	2,968	138.8	159.1	26.10	307,509	327,723	0.94	27.82
2002	3,205	170.8	195.6	29.72	363,136	345,709	1.05	28.30
2003	3,575	200.8	232.7	31.53	352,398	354,784	0.99	31.74
2004	3,846	226.3	270.0	33.55	365,352	369,042	0.99	33.89
Average				16.56			0.64	25.87

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, Daily trading in Treasury Securities $US Billions
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 Turnover Rate to Equity Turnover Rate (Col. 4/Col.7)

Table II: Simulation of the Illiquidity Premium Model Replicating both Returns and Turnover for Equity and T-Bill Trading in the US, 1894-1994.

	Equity				Tbill		
	Zero Tcost	Zero Tcost	Base Case	Max Outlay	Zero Tcost	Base Case	Max Outlay
	1	2	3	4	5	6	7
Annualised Return Variance, σ^2	3.24%	3.24%	3.24%	3.24%	0.36%	0.36%	0.36%
Annualized Riskless Rate, *r*	2%	2%	2%	2%	0.4%	0.4%	0.4%
Coefficient of absolute risk aversion, *b*	1	1	1	1	1	1	1
Initial endowment of each seller, K_0^S	7.5	7.5	7.5	7.5	7.5	7.5	7.5
Initial endowment of each buyer, K_0^D	0	0	0	0	0	0	0
Relative Endowment Heterogeneity, *h*	1	1	1	1	1	1	1
Initial Bond Endow, Supplier, w_0^S	0	0	0	0	0	0	0
Initial Bond Endow, Demander w_0^D	7.537	7.537	7.537	7.537	7.537	7.537	7.537
Total Population of Traders, *N*	10,000	10,000	10,000	10,000	10,000	10,000	10,000
Annualised Expected Net Return, $\mu-1$	15.13%	22.28%	22.28%	20.46%	2.08%	2.08%	2.28%
Supplier Intercept, α^S	745.0343	746.9072	746.9072	746.4404	6,671.7280	6,671.7280	6,672.2674
Demander Intercept, α^D	745.0335	746.9065	746.9065	746.4396	6,671.7273	6,671.7273	6,672.2667
Slope of Demand Schedules, β	741.2839	741.2839	741.2839	741.2839	6,667.1109	6,667.1109	6,667.1109
Two sided prop dollar spread, *a*	0%	0%	0.98%	0.506%	0%	0.03%	0.056%
Prohibitive Level of Trading Cost, a^{bar}	1.03%	1.03%	1.03%	1.03%	0.11%	0.11%	0.11%
TCost Prop of Mid-Point Price, a/p^{mp}	0%	0%	0.980%	0.513%	0%	0.03%	0.057%
Equilibrium Demand by Seller, K_1^S	3.750	3.750	7.383	5.625	3.750	4.750	5.625
Equilibrium Demand by Buyer, K_1^D	3.750	3.750	0.117	1.875	3.750	2.750	1.875
Equil Amnt Sold per Fortnight, ΔK	3.750	3.750	0.117	1.875	3.750	2.750	1.875
Annualised Turnover Rate, $\tau(\phi)=\Delta K/K^T$	12.00	12.00	0.38	6.00	12.00	8.80	6.00
Annual Illiq Prem per unit Tcost, *c(a)/a*	13.75	13.71	7.30	10.54	12.20	10.59	9.18
Annual Illiquidity Prem due Tcost, *c(a)*	0%	0%	7.15%	5.33%	0%	0.32%	0.52%
Ratio Unobserved to Observed Tcost	0	0	15.494	0.501	0	0.182	0.5
Utility of Trading Pair with TCosts	NA	15.0872	15.0872	15.0825	15.0423	15.0423	15.0429
Utility of Trading Pair with No TCost	15.0683	15.0872	15.0872	15.0825	15.0423	15.0423	15.0429
Annual Gains fm Trade rel Autarky, *B(a)*	6.756%	7.158%	0.007%	1.802%	0.686%	0.370%	0.172%
Trading Elasticity wrt Tcosts, η_a^τ	0	0	-30.96	-1	0	-0.36	-1
Stock Price Elasticity wrt Tcosts, $\eta_a^{p^{mp}}$	0	0	-0.18	-1.52	0	-0.66	-0.84
Elastic of Req Compen wrt Tcosts, η_a^c	1	1	0.061	0.667	1	0.846	0.667
Annual Slope Compn Fn wrt Tover, $c'(\tau)$	-0.243	-0.243	-0.008	-0.122	-0.027	-0.020	-0.014
Equil Price without Tcost, $p^{a=0}$	1	1.0025	1.0025	1.0019	1.0001	1.0001	1.0002
Equil Mid-Point Price with Tcost, p^{mp}	1	NA	1	1	1	1	1
Equil Price without Tcost for Perpetuity, $p^{a=0}$	1	4.0635	4.0635	3.2999	1.7816	1.7816	2.2680
Equil Md-Pt Price with Tcost-Perpetuity, p^{mp}	1	4.0635	1	1	1.7816	1	1

This simulation is generated from the equations in the text explaining turnover demand, $\tau(a)$, the required compensation for transactions costs, $c(a)$, and the pricing equation explaining the mid-point price of an asset as a function of the required compensation and a variety of other variables, as well as other equations. I take the rate of transaction costs for equity and T-bills (bonds) conservatively from historical information, as also is the volatility of the two types of securities.

Table III: 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, $c(\varphi)$	Trans. Cost, φ	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 IV: Regression results for sample of approximately 576 Australian securities listed on the Australian Stock Exchange, 1994-98.

Estimating two simultaneous equations: $c(\tau)_t = \{ \alpha^{\rho}[\rho/(\rho-1)][\tau_{et}^{1-\rho} - \tau_b^{1-\rho}]\}$ and $\tau_t = \alpha\varphi_t^{-(1/\rho)}$ 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)
Inverse of Turnover Elastic, ρ	1.2798 (139.63)	1.3861 (86.199)	1.1476 (53.674)	NA
Equity Inv of Tover Elastic ρ_e				1.2531 (85.396)
Inverse of Turnover Elastic ρ_τ				1.3148 (76.412)
Turnover Elasticity (ρ)	0.78137	0.72145	0.87138	NA
Elasticity estimated from the Equity Premium (ρ_e)	NA	NA	NA	0.798021
Elastic est from Turnover (ρ_τ)	NA	NA	NA	0.760572
Implied T-bill TCost (φ_b)	0.000311	0.00027	0.000443	0.000279

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

Two equations, one representing the linear in logs stock demand function for stock turnover as a function of the relative transaction costs, and the other, the area under the stock turnover demand function over the range as the cost of transacting goes from zero to its observed value representing the equity premium (required compensation for bearing transaction costs) are estimated simultaneously. The model is based on equations (46) and (47) in the text.

Figure 1: Increasing Market Depth by Adding Investors Rotates the Turnover Demand Anti-Clockwise Around the Autarky Point

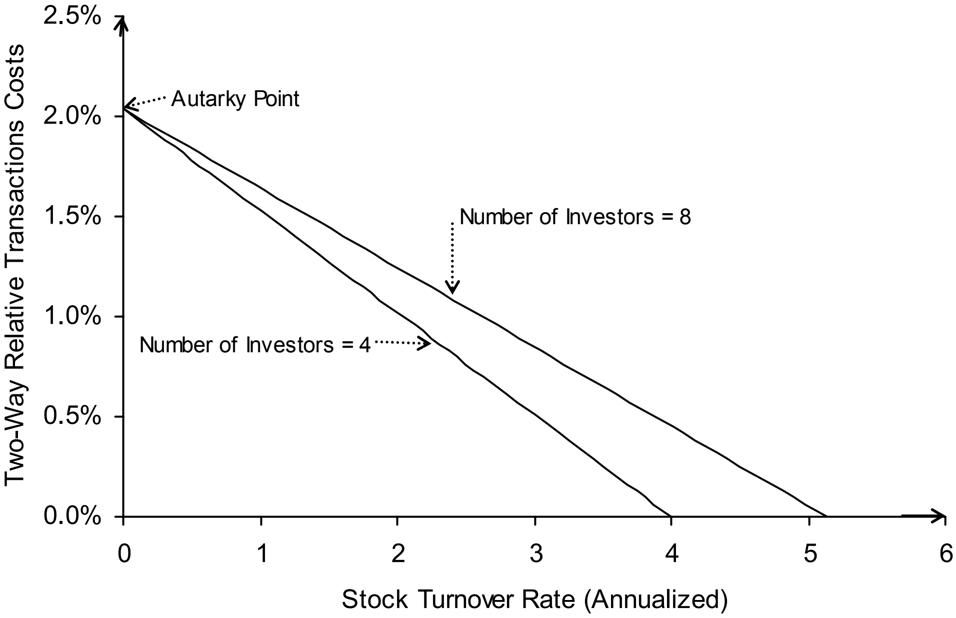


Figure 2: Seven Percent Illiquidity Premium with Moderate Volatility, Base-Case Equity Simulation from Table II

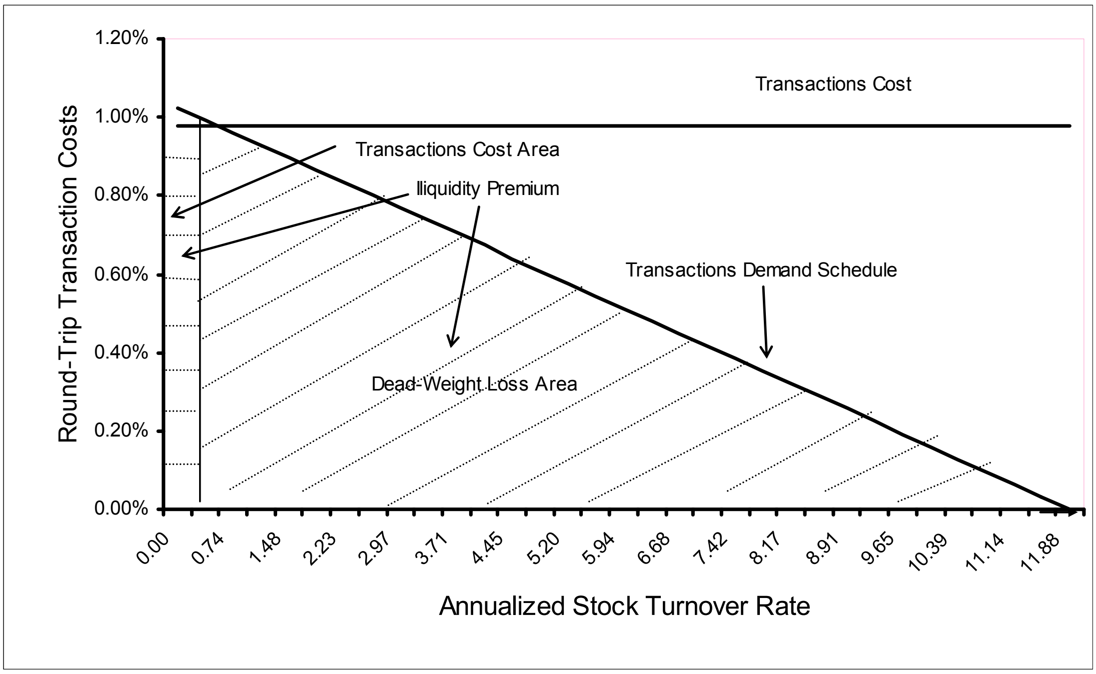
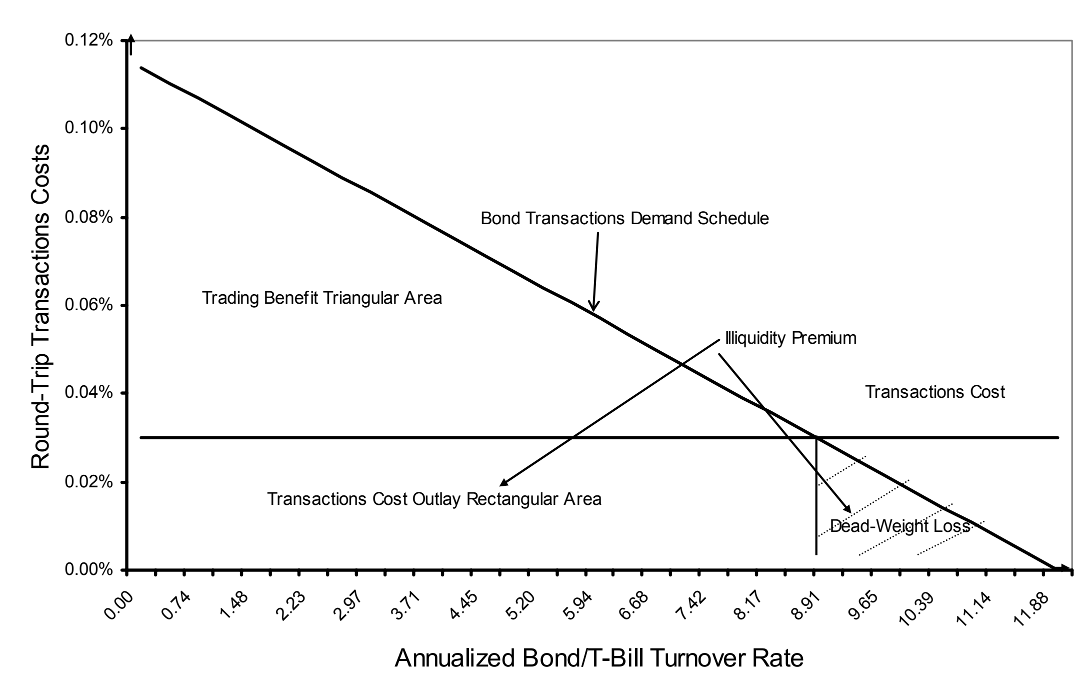


Figure 3: Simulation of High Bond/TBill Turnover Rate; Base-Case Bond Simulation from Table II


Round-Trip Transactions Costs
Bond Transactions Demand Schedule
Trading Benefit Triangular Area
Illiquidity Premium
Transactions Cost
Transactions Cost Outlay Rectangular Area
Dead-Weight Loss
Annualized Bond/T-Bill Turnover Rate