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

Peter L. Swan∗

UNSW

Version Dated December 10, 2005

ABSTRACT

Following Constantinides’ (1986) seminal approach and introducing transaction costs in the Pagano (1989) model, conventional CARA investors with heterogeneous endowments trade to construct optimal portfolios. We calibrate to the 1896-1994 equity and bond markets to show that gains from trade are high and, thus, investors require a high illiquidity premium even for a modest transactional charge. Excluding risk premia, exchange of equity and bonds by $N$ strategic investors, as $N \to \infty$, under a mere 1% round-trip transaction cost induces a 6% illiquidity (equity) premium. Unlike existing literature, our findings are consistent with most stylized empirical facts. We recover the elasticity of trading demand from the excess equity return to confirm a major implication of the model. Among many other, so called, anomalies, we appear to explain the apparent “irrational exuberance” of equity markets, the 600% price premium for otherwise identical “A” stock over “B” stock in China, the low risk-free rate, the 20% letter stock premium and the lower return on “on the run” bonds. Because illiquidity premia do not necessarily imply consumption volatility, variance bounds tests become irrelevant.

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 (ARC) Australian Professorial Fellowship scheme for financial support. I also wish to thank Yakov Amihud, Kerry Back, 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, Lars Hansen, Ravi Jagannathan, Craig Lewis, Debbie Lucas, Narayan Naik, Ananth Madhavan, Michael McAleer, Rajnish Mehra, Robert Merton, Ed Prescott, Ioanid Rosu, Brian Routledge, Jeff Sheen, David Simmonds, George Sofianos, Barbara Spencer, Dimitri Vayanos, Joakim Westerholm, Arnold Zellner, seminar participants at AGSM, Australian National University, Finance seminar, Carnegie Mellon University, City University of Hong Kong, EFMA Conference, London, Nanyang Technological University, Finance seminar, Northwestern University, the Reserve Bank of Australia, Singapore Management University, Finance seminar, GSB, University of Chicago, University of NSW, University of Queensland, UC Santa Barbara, University of Sydney, Finance seminar, Texas A&M University, Finance seminar, Vanderbilt, 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.

A previous version of this paper was named: “Can ‘Illiquidity’ Explain the Equity Premium Puzzle? The Value of Endogenous Market Trading”.

© Copyright 2000-2005 by Peter L. Swan.
“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”—Welch (2000)

Between 1896 and 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), hereafter (M&P), attempt to account for this premium using simulations of an inter-temporal equilibrium model with a 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.¹

M&P 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 modeling of transaction costs requires trading between investors and, clearly, there can be no trading between representative 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. Establishing asset prices by modeling the exchange of assets is urged by O’Hara (2003) in her Presidential Address as obviously an asset is worth only what someone is willing to purchase it for.

While the asset pricing literature, unlike that of international trade specialists, often ignores the gains to investors from security market transactions and even evident rapid turnover of equity and bond portfolios, I show that in the absence of transactions costs otherwise identical investors with differing endowments gain substantially from trading equity. The gains to trade arise from more effectively sharing risks. These gains are eliminated, or at least substantially reduced, by proportional transaction costs that create a spread between the bid and the ask. In an equilibrium with just one risky asset and no alternative for trading purposes, I show that these costs do not discriminate between buyers and sellers and nor do they impinge on the

¹ For an up to date review of the puzzle see Constantinides (2005), Heaton and Lucas (2005) and other papers presented at the UC Santa Barbara Equity Premium Conference, October, 2005.
fundamental value of the asset as represented by its midpoint price, even though both buyer and
seller are worse off relative to trade with no hindrances and the supply price has fallen due to the
transactions cost wedge. The intuitive reason for this is that trading, or the lack of it, does not
impinge on the valuation process when investors are provided with no alternative.

This simple model describes a world in which the only asset for which there is a motive for
trade is subject to transactions costs in a symmetric fashion and does not address the
fundamental question posed by Constantinides (1986) in his seminal contribution. His concern
lay with expected utility comparisons across equilibria. In particular, his focus was the impact
on the equilibrium rate of return of a proportional transaction cost when investors are guaranteed
a minimum utility level fully incorporating all gains from trade provided by the ability to trade
an otherwise identical completely liquid asset with no transaction costs. Speculators cannot
directly arbitrage across equilibria except in expected utility terms, preferences matter, and
assets with the same risk and expected dividend differ in terms of a tradability factor. His
concern, as does mine, lay with the impact on the equilibrium rate of return of a proportional
transaction cost. Consequently, he defined the illiquidity (or liquidity) premium as the increase
in the asset’s mean return, i.e., dividend, that combined with the introduction of transaction costs
leaves unchanged the investor’s expected utility across the two equilibria, without and with
transactions costs. In other words, his focus is on the loss in gains from trade due to the
transaction charge that must be compensated for in expected utility terms by the rise in dividend,
i.e., rate of return, on the asset with a positive transactions cost.

Since, in a closed system no one is in a position to pay a higher dividend, I compute the required
price fall for the asset with trading costs so as to effectively provide the higher required rate of
return for matched trading counter-parties. Thus it is no longer true that the midpoint price is
unaffected by transaction costs as the expected dividend is higher in the equilibrium with
positive transaction costs so as to equate expected utilities across the two equilibria, or
continuum of equilibria as transaction costs vary. The asset price changes with respect to
alterations in transactions costs are defined with respect to the downward-sloping in transactions
costs constant utility (i.e., constant real income) asset pricing function constructed across these
equilibria. This is akin to compensating differentials in job markets with workers indifferent
between unpleasant/dangerous and pleasant/safe occupations. Hence my methodology and that
of Constantinides abstract from complications arising from the ability in multi-asset worlds for
investors to undertake some transactions with cheap-to-trade assets, and to only trade expensive-
to trade-assets when these opportunities have been exhausted once a corner solution in the
cheap-to-trade asset has been reached. For both Constantinides (1986) and present purposes, it
makes no sense to consider complex worlds in which hierarchies of demands for risky assets with differing transaction costs coexist for trading purposes with trading taking place in sequence, low before high cost. Moreover, in such a complex world it is difficult, if not impossible, to capture the compensating price differences required to equate expected utility across assets with varying transactions costs. I build on Constantinides’ approach to show that a simple two-period model of trading, realistically calibrated to reflect actual turnover rates for equity and bonds, induces large gains from trade and compensation for modest transaction costs in the form of a substantial price fall for equity and the observed equity (illiquidity) premium of six percent or more.

In my illiquidity/loss of gains from trade approach there need be no relationship between returns volatility and the volatility of the growth rate in consumption as the driving force is trading volume, no matter how motivated, and not risk. Hence, there is no role for an explanation for the low consumption volatility or for the variance bounds tests of Hansen and Jagannathan (1991). The irrelevance of consumption volatility comes naturally out of my model. Firstly, differences in endowments means that we cannot just look at aggregate consumption and, more importantly, the 6 to 8% illiquidity premium in my approach is not a risk premium and therefore does not necessarily imply volatility in consumption. I just happen to have trading demand motivated by risk sharing, hence I appear to imply some consumption volatility. If for example, trading were motivated by something else (differences in beliefs for example) then the volatility in asset markets is explained by variations in trading demand of individual investors (turnover) as empirically I find is the case for Nokia in Finland with no implications for consumption volatility.

Hence if I were to put trading directly into the M&P model the loss of trading gains from trade should explain the high variance of asset prices due to potentially random (or business cycle related) trading patterns with stable consumption growth. Thus in my framework, Shiller’s (1981 and 2005) “excessive” volatility in equity markets, or “irrational exuberance” reflects variations in the gains from trade net of transactions costs which rationally factor into asset prices even when dividends or earnings are relatively stable. It should also explain the low to zero return on T-Bills over the last 100 years due to the fact these act essentially as an interest-paying substitute for the convenience (“shoe leather”) yield of money in my model with the entire stock turning over fortnightly and huge compensating gains from trade evident in Figure 3 below. Also the 6 to 8% return on equity due to lower gains from trade relative to T-Bills presumably reflects high asymmetric information relative to bonds giving rise to transaction costs 25 times higher and approx 1/25th of the turnover. Since equity is less “money-like” and
thus less useful in trading for risk sharing and other purposes, the return must be correspondingly higher than either money (expected negative return due to inflation) or bonds (0 to 2% in real terms). My model also explains the high price and lower return on “on the run” bonds as newly issued “on the run bonds” are more liquid and thus trade more frequently. For the same reason, it also provides an explanation for the positive term structure of interest yields as longer dated securities (e.g., 10 and 30 year bonds) that have been issued for some time find more permanent owners and are thus less tradeable than short-dated securities such as T-Bills. Moreover, it explains the cross sectional returns on the NYSE over the 30 years to 1992 as a function of stock turnover (Datar, Naik, et al., 1998), the 20% return on “letter stock” (Sillber, 1991), changes in the 600% premium on “A” relative to “B” stock prices in China when the assets are otherwise identical, the very small returns on “A” relative to “B” stock and the relative daily returns (Chen and Swan, 2005), equity returns in Finland (Swan and Westerholm, 2005) and other anomalies. The Datar, Naik, et al. (1998), Chen and Swan (2005) and Swan and Westerholm (2005) studies all find that illiquidity effects dominate returns even after controlling for the Fama and French (1992) factors.

Constantinides’ (1986) inability to identify a significant illiquidity premium when modeling realistic trading costs with a valid model indicates that the illiquidity premium is not due to transaction costs alone. In fact, trading costs taken in isolation are often irrelevant. Rather, a significant premium only arises when there is some hindrance to what would otherwise be significant gains from trade. Hence, motivation for trading must be strong for a significant illiquidity premium to arise. I show that every security is priced as an autarkic security that never trades plus liquidity or trading benefits that reflects the strength of trading demand which generates a positive turnover. It is this largely random stock turnover that is associated with stock price and return volatility.

Heterogeneous endowments and frequent trading due to portfolio endowment shocks modeled as a short investor-trading horizon are sufficient for the model to predict observed trading activity for equities and bonds together with a very sizeable, six to eight 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 counterfactual negligible amount.

---

2 I wish to thank Ed Prescott for this suggestion.
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 15 times higher than all the observed costs of trading, such as spreads and commissions, combined. This is the cost of foregone trades reflecting a severe decline in gains from trade. 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 severe 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. My simulations indicate that when I add invisible transactions costs to the observed costs, these overall costs do explain the equity premium, the very low two percent or less 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 and maximal gains from trade. 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 about six percent per annum for one-way 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 0.368 percent and costs of forgone trades of 5.7 percent. Thus it would be very easy to treat the actual cost of trading of less than 0.5% as simply a transfer from one investor type to another and still preserve an illiquidity premium of around 6%. In my simulation, the optimal equity turnover rate is 37.6 percent per annum, representing a long-term average. Whereas for identical investors trading three-month T-bills with the same relative endowments, the required compensation is only about 0.31 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 portfolio-rebalancing trades due to portfolio endowment shocks at a fortnightly frequency, i.e., a short investment horizon, is required to explain observed equity and bond turnover rates. This contrasts with portfolio rebalancing trades every 20 years, i.e., a very long investment horizon, which is required for my model to reproduce the illiquidity premium results of Constantinides (1986) with negligible equity and bond turnover. My simulations (Table 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.

It is an implication of my model that the equity premium is a consequence of sizeable gains from trade in combination with a trading friction. However, due to technological and other improvements the cost of trading equity has fallen substantially during the last few decades. Why then has not the equity premium fallen along with the magnitude of the trading friction? This objection fails to recognize that the premium is the premium over and above the bond yield and technological improvements have also lowered the cost of trading bonds. The relative trading data for bonds and equity over the last 25 years presented in Table I below indicate that bond and equity turnover have both grown in response to falling costs. The premium depends only on the relative frictions in the two markets.

Building on the seminal contribution of Pagano (1989), I develop the first simple and transparent “closed-form” trading model for risky assets, incorporating both proportional transactions costs and market impact “costs”, capable of explaining observed trading levels. 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 the market-clearing price down 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 modeling of strategic trading empirically relevant, while emphasizing the gains from market integration. I find that the main impact of strategic investor behavior when market thinness increases is to significantly reduce stock turnover but not to have a major deleterious impact on other market fundamentals. In my equity premium simulations
the number of market participants is set at an exceedingly large figure so that market impact costs are excluded.

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 \textit{ad hoc} fashion. I do take as exogenous, however, the levels of transactions costs and the volatilities of the equity and bond markets from the historical experience.

In common with the fundamental insight from Constantinides (1986), my investors are indifferent in expected utility terms across equilibria between (say) trading bonds or equities with negligible transactions costs and trading an otherwise identical asset with transactions costs in place. While this might appear to be equivalent to sequential trades with a corner solution for the cheap to transact asset such that demander-investor purchases all available shares and then trades the more expensive to trade risky asset, this is not so. Both what Constantinides (1986) and I do is quite different and, I believe, more appropriate. I compute the difference in the expected dividend per share required to equate the expected utility of an investor in the same asset, with and without transactions costs, where each outcome represents a separate equilibrium. Investor counterparty pairs suffer a dramatic loss of investor utility due to even very small transaction costs. It thus has implications for the proposed Tobin tax on security market trading (see Tobin, 1984, Stiglitz, 1989, Summers and Summers, 1990, and Schwert and Seguin, 1993).

It might seem that the endowment heterogeneity and portfolio shock frequency 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 Table II below). However, I could interpret endowment heterogeneity as no more than an illustrative 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, I show that 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 simply require a “well-behaved” downward sloping stock/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 illustrate this far more general perspective by commencing with a downward sloping trading stock turnover relationship that is linear in logarithms as a simple “reduced form” version of the model that is very easy to implement empirically.

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

In a pioneering contribution Amihud and Mendelson (1986) model expected discounted cash flow maximization by risk neutral agents with preferences by each investor type to turnover every asset in that type’s portfolio at an exogenous specified rate per period, which is equal to the inverse of the investment horizon for that type and is irrespective of the bid-ask spread, or rate of transactions costs, incurred on each asset. They show that under these circumstances asset returns are increasing in the relative bid-ask spread and are a concave function of the spread with investor types with longer horizons trading higher-spread securities. There is a small illiquidity premium arising from the existence of multiple securities with differing spreads, together with a chain of indifferent investors linking the returns on these securities. The model abstracts from the question of motivation for trading or why trade counterparties should exist. They find strong empirical support for the predictions of the model at the level of
gross returns and spreads but do not present direct evidence on the existence of ‘clientele effects’. Amihud (2002) provides further cross-sectional and time-series evidence that the “excess” equity return at least partly represents an illiquidity premium.

A number of other studies have also 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 representative agent endogenously trading with constant relative risk aversion (CRRA) preferences and an infinite horizon. The premium is 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 portfolio endowment heterogeneity shocks and trading occurs each fortnight, i.e., the investment horizon is 1/24 of a year. Overall, asset prices are about 180 fold more sensitive to transactions costs with my calibrations than with Constantinides (1986) 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. Portfolio rebalancing every 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 Constantinides (1986) and my own simply reflect the difficulty of calibrating a representative investor model to provide realistic estimates of equity and bond trading in the presence of transactions costs. There are also other studies in the tradition of Constantinides (1986), including Davis and Norman (1990), Aiyagari and Gertler (1991), Bansal and Coleman (1996) and Heaton and Lucas (1992, 1996, 2005) who model idiosyncratic labor income shocks.
Heaton and Lucas (1996, p.467) find that a 10% transaction cost at the margin is required to explain a modest 5% equity premium in their framework. They conclude that “moderate trading costs and realistic labor income are not sufficient to resolve the equity premium puzzle” (Heaton and Lucas, 2005). However, none of these contributions follow Constantinides (1986) in equating expected utility across regimes and calibrate to realistic equity and T-Bill turnover rates as I do.

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 and he does consider fixed costs, 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. 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. Huang (2003) also finds a relatively small liquidity effect.

A recent study by Lo, Mamaysky and Wang (2004) like Pagino’s (1989), also models fixed rather than the proportional transaction costs in the context of exponential utility maximizing agents hedging a non-traded risky endowment that is perfectly correlated with the dividends on the traded asset, with zero aggregate risks for the non-traded asset. In the absence of fixed costs of trading, agents trade continuously to eliminate non-traded risk completely with the equilibrium price constant and free of non-traded risk. Trading volume is essentially infinite. With the introduction of fixed transaction costs there is no general closed form solution but a 1% increase in transaction cost decreases trading volume by 0.25. Such a low elasticity is contrary to most empirical studies that indicate elasticity values approximately four times higher at about 0.8 to unity (see, for example, Jones, 2002, and Table IV below). However, if trading demand is sufficiently high in the presence of small but fixed trading costs then the asset price becomes sensitive to trading costs. They do not adopt the Constantinides (1986) approach of finding the magnitude of the dividend change required to offset the utility decline due to loss of the gains from trade.

---

3 See also Vayanos and Vila (1999).
Liu (2004) models multiple assets as well as transactions costs for CARA investors over a continuous time infinite horizon. Jang, Koo, and Lowenstein (2004) find that if stochastic regime switching is introduced into the model of Constantinides (1986) that transactions costs can have a first-order impact. All of these models, with the exception of Pagano, Vayanos who introduces an age rather than endowment differential between counterparties, and Lo, Mamaysky and Wang (2004) who model only fixed costs, differ from mine in that they model a representative investor making consumption and portfolio choices. Of course, the counterparty trades to trades optimal from the perspective of a representative investor are not modeled as there are none.

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 based on NYSE returns over the thirty year period, 1973-1992. Extensions in the same vein are provided by Pastor and Stambaugh (2001), Easley, Hvidkjaer and O’Hara (2002), and Easley and O’Hara (2004).

Chen and Swan (2005) investigate the pricing and returns in China for “A” stock available only to domestic investors and “B” stock with identical dividends traded by international investors when these two markets were completely segmented. “A” stock trading on the Shanghai Exchange turned over 3.8 times more rapidly than the “B” stock trading on the same exchange with a lower relative rate of return of 0.248, were 0.507 cheaper to trade, and the relative price was 5.98. Similarly, “A” stock traded on Shenzhen turned over 5.65 times faster with relative rates of return of 0.4, relative trading costs of 0.52, and relative prices of 5.14. Thus, international investors required a much higher relative return with a correspondingly lower price due to higher transaction costs in the “B” market. Compared with other alternative asset pricing factors, such as firm size, book-to-market ratio, and informativeness of order flow, my liquidity asset pricing model, as exposited in the current paper, is the most successful in terms of
explaining the observed changes in the equity premium on both A-share and B-share markets and changes in the relative price of “A” and “B” stocks using daily panel data.

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 a normally distributed terminating dividend, 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, representative suppliers, $S = 1, \ldots, N/2$, overly endowed with $K_0^S$ units each the perfectly divisible risky asset, i.e., equity shares, relative to the other half, representative demanders, $D = N/2 + 1, \ldots, N$, 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 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 $\tau$ 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, $\tau = \frac{K_1^D - K_0^D}{K^T} = \frac{K_1^S - K_1^S}{K^T} = \frac{\Delta K}{K^T}$, 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 $w_0^D > w_0^S$ with the endowment per pair, $w_0^T = w_0^S + w_0^D$. Bonds pay a certain terminating amount, $R$, $R = 1 + r$, where $R$ is termed the cumulating factor or gross return and $r$ the per period interest rate, have no trading costs, and in other respects are just like cash. 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, $c_1^S c = w_1$, 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),

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

Demanders who consume their terminal wealth, $\tilde{c} = w_1^D$, 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, $p_0^D = p_S + a$, 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,

\[
\tilde{c}^D = w_i = \tilde{d}K_i^D + R\left[w_0^D - \left(p^S + a\right)\left(K_i^D - K_0^D\right)\right].
\]  

(2)

The mid-point price, \( p^\text{mp} = \frac{p^S + p^D}{2} = p^S + \frac{a}{2} \), where \( \frac{a}{2} \) is the dollar half-spread, with \( \frac{a}{2p^\text{mp}} \) 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, \( \sigma^2 \), and the expected value is \( E(\tilde{d}) = \mu \), 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),

\[
E\left[u\left(c'\right)\right] = E\left(c'\right) - \left(b/2\right)\text{Var}(c'), \forall i, i = 1,...,N,
\]  

(3)

with \( \text{Var} \) 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. However, the ability of the model to explain the equity premium does not depend on this refinement. Supposing the \( i^{\text{th}} \) investor is a demander with an initial endowment of equity shares, \( K_0 = K_0^D \), he will be competing against the remaining \( \left(\frac{N}{2} - 1\right) \) identical demanders whom he correctly conjectures have identical individual linear demand schedules incorporating the fixed per unit transactions cost \( a \) between the bid and the ask,

\[
K_i^D = \alpha^D - \beta\left(p^S + a\right),
\]  

(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,

\[
K_i^S = \alpha^S - \beta p^S.
\]  

(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,

$$(N - 1)K_i^d + \frac{N}{2}K_i^s = \left(\frac{N - 1}{2}\right)\alpha^d + \frac{N}{2}\alpha^s - (N - 1)\beta p^s - \frac{N - 2}{2}\beta a.$$ (6)

Adding the $i^{th}$ demander’s own demand to both sides of equation (6), I obtain the conjectural variational condition,

$$N - 2K_i^d + K_i^d + \frac{N}{2}K_i^s = N - 2\frac{\alpha^d}{2} + \frac{N}{2}\alpha^s - (N - 1)\beta p^s - \frac{N - 2}{2}\beta a + K_i^d,$$ (7)

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

$$\frac{dp^s}{dK_i^d} = \frac{1}{\beta(N - 1)} > 0.$$ (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, $\text{Var}(\tilde{c}^D) = \sigma^2(K_i^d)^2$, 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,
\[ \frac{\partial E(u^o)}{\partial K_{i_i}^d} = \mu - R(p^s + a) - \frac{R(K_{i_i}^d - K_{0}^d)}{\beta(N-1)} - b\sigma^2 K_{i_i}^d = 0. \]  

(9)

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

\[ K_{i_i}^d = \frac{\mu - R(p^s + a) + \frac{R K_{0}^d}{\beta(N-1)}}{b\sigma^2 + \frac{R}{\beta(N-1)}}. \]  

(10)

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

\[ K_{i_i}^s = \frac{\mu - R p^s + \frac{R K_{0}^s}{\beta(N-1)}}{b\sigma^2 + \frac{R}{\beta(N-1)}}. \]  

(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),

\[ \frac{N - 2}{2} K_{i_i}^d + K_{i_i}^d + \frac{N}{2} K_{i_i}^s = \frac{N}{2} K^r \]

\[ = \frac{N - 2}{2} \left\{ \mu - R(p^s + a) + \frac{R K_{0}^d}{\beta(N-1)} \right\} + \frac{N}{2} \left\{ \mu - R p^s + \frac{R K_{0}^s}{\beta(N-1)} \right\} + K_{i_i}^d. \]  

(12)

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

\[ \beta = \frac{N - 2}{N-1} \frac{R}{b\sigma^2}. \]  

(13)

Clearly, as \( N \to \infty \), the slope \( \beta \to \frac{R}{b\sigma^2} \) but had I begun by adopting price taking behavior at the outset, I would have not been able to correctly compute the slope of the demand curve as terms
in $\beta$ would have dropped out. This illustrates the importance of realistically only considering price taking behavior as the limiting case of the general solution.

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

$$
\alpha^D = \frac{N - 2}{N - 1} \frac{\mu}{b\sigma^2} + \frac{K^D_0}{N - 1} \quad \text{and} \quad \alpha^S = \frac{N - 2}{N - 1} \frac{\mu}{b\sigma^2} + \frac{K^S_0}{N - 1}.
$$

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 $\alpha^S$, nor the slope, $\beta$, depend directly on transaction costs. Rather, they depend on the number of market participants, $N$, the coefficient of absolute risk aversion, $b$, risk (volatility), $\sigma^2$, expected earnings/dividends, $\mu$, 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, $p^D = p^S + a$,

$$
K^D_i = \alpha^D - \beta (p^S + a) = \frac{K^D_0}{N - 1} + \frac{N - 2}{N - 1} \frac{\mu - R(p^S + a)}{b\sigma^2},
$$

and, similarly into (5) for suppliers,

$$
K^S_i = \alpha^S - \beta p^S = \frac{K^S_0}{N - 1} + \frac{N - 2}{N - 1} \frac{\mu - Rp^S}{b\sigma^2}.
$$

Since the sum of the demands equals the initial endowment of the trading pair, $K^D_0 + K^S_0 \equiv K^D_0 + K^S_0 \equiv K^T$ 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,

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

and
\[ p^* = \frac{\alpha^s + \alpha^d - K^T}{2\beta} - \frac{a}{2} = \frac{\mu - \frac{b\sigma^2}{2} - K^T}{R} - \frac{a}{2}. \] (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,

\[ p^{mp} = p^* + \frac{a}{2} = \frac{\alpha^s + \alpha^d - K^T}{2\beta} = \frac{\mu - \frac{b\sigma^2}{2} - K^T}{R}. \] (19)

Remarkably, in this symmetric equilibrium (or infinite number as there exists a solution for every value of \( a \)) the midpoint price with transaction costs in place, \( p^{mp} \), given by (19), appears independent of the dollar spread, \( a \), so long as the expected dividend, \( \mu \), is independent of transaction costs, and appears precisely equal to the bid and ask price in the complete absence of transaction costs, denoted \( p^{a=0} \). This is because, up to this point at least, investors receive no choice of security variants, such as an identical asset with no transaction cost applicable that would guarantee investors receive a minimum utility level reflecting the maximum gains from trade. Below the methodology of Constantinides (1986) is adapted to reveal the price falls precisely required to induce higher expected returns that make investors indifferent between every otherwise identical security variant with transaction costs ranging from the autarkic or prohibitive level to none at all.

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 the risk component of the 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, \( a \),

\[ K_D^0 = f^D (a) = K_D^0 (a) = \frac{1}{2} \left( \alpha^s - \alpha^d + K^T - \beta a \right) = \frac{K^D_0}{N-1} + \frac{1}{2} \left( \frac{N-2}{N-1} \left( K^T - \frac{R}{b\sigma^2} a \right) - \frac{1}{2} \right), \] (20)
\[ K_i^s = f^s (\alpha) \triangleq K_i^s (\alpha) = \frac{1}{2} \left( \alpha^s - \alpha^d + K_T^s + \beta a \right) = \frac{K_0^s}{N-1} + \frac{1}{2} \frac{N-2}{N-1} \left( K_T^s + \frac{R}{b \sigma^2} a \right). \]  

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, \( \tau = f (\alpha) \triangleq \tau (\alpha) \), 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,

\[ \tau (\alpha) = \frac{K_0^D - K_0^D}{K_T^s} = \frac{K_0^s - K_1^s}{K_T^s} = \frac{1}{2} \left( h - \frac{\alpha^s - \alpha^d - \beta a}{K_T^s} \right) = \frac{1}{2} \left( \frac{N-2}{N-1} \right) \left( h - \frac{R}{b \sigma^2 K_T^s} a \right), \]  

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

\[ \tau (\alpha)^{-1} = a (\tau) = \frac{b \sigma^2 K_T^s}{R} \left( h - \frac{N-1}{N-2} \tau \right). \]  

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, \( \sigma^2 \), and the available number of risky shares, \( K_T^s \), to be traded between the parties, act to overcome the discouraging impact of transactions costs, \( a \), on the propensity to trade, \( \tau (\alpha) \). Thus in richer communities with a greater supply of risky assets per capita, \( i.e. \), higher \( K_T^s \), 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 equation (19), appears 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 counterclockwise to the right as the market depth increases, around the autarky (no trade) point, \( b\sigma^2 K^T h/R \). This point provides an upper bound to the observed transaction costs, max \( a \leq \tilde{a} \).

The schedule flattens out as the number of participants increase. Trading activity is increasing in the size of the market due to favorable market externalities. As \( N \to \infty \), strategic behavior 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.

The elasticity of turnover demand with respect to transaction costs,

\[
\eta_a = \frac{\tau'(a)}{\tau(a)} a = -\frac{Ra}{b\sigma^2 K^T h R - Ra} < 0 ,
\]

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. Compensating Dividend Required to Offset Transactional Cost Impacts**

For investors to be willing to hold both the risky asset with transactions costs in place and an identical asset with perfectly correlated returns and identical variance without trading costs, the expected rate of return, and hence dividend per share on the asset with trading costs, must rise by a compensating amount to maintain indifference. This was a key feature of Constantinides’ (1986) seminal contribution. Clearly, the situation described by the pricing equation (19) above with the mid-point price of the asset preserved as transaction costs rise does not tell the full story if there remains a perfect substitute for the expensive to transact asset that is free of charges or taxes or multiple assets with differing transaction costs. For example, a severe transaction charge, high $a$, will reduce the utility of prospective demanders and suppliers close to reservation levels implied by autarky. Thus, the utility decline for owners of an impacted asset may be severe relative to an asset with zero transaction costs when the gains from risk sharing induced by trading are high. Consider the U.K. situation in which the Government applies a stamp duty (tax) to trades of equity shares in U.K. domiciled stocks exclusively while foreign domiciled stocks and Gilts are free of stamp duty. These government securities and foreign-domiciled equity securities are likely to be close substitutes for domestic domiciled equity so that U.K. stock must sell at a discount relative to otherwise identical foreign stock such that the price reduction yields an implicit dividend increase which compensates for the disability reducing the attractiveness of U.K. stock.

Transactions costs reduce the aggregate supply of the riskless asset per investor and counterparty, $wT$, in the second period by the amount of the total two-sided costs of trading, $aT(a)K^T$. 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 suppliers with identical preferences 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,

\[
U^a \triangleq \left[ u(e^d) + u(e^s) \right] = \left[ R [w_0^T - aK^T \tau(a)] + [\mu + c(a)]K^T - \frac{b}{2} \sigma^2 \left( (K_i^o)^2 + (K_i^s)^2 \right) \right],
\]

where \(c(a)\) denotes the compensating increase in the required dividend to offset the utility loss from the trade restriction induced by transaction costs, 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, \(K_i^S(a)\) and \(K_i^D(a)\), are found by squaring the transactions-cost sensitive functions, given by equations (20) and (21), respectively. The equivalent of equation (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,

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

The component of required compensation resulting from forgone trades, due to 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,

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

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, on adopting the methodology pioneered by Constantinides (1986), is the simple sum of the two sources of investor loss, \(dwl(a)\) from (27) plus the actual resource costs, \(a \tau(a)\), and expressed as,

\[
c(a) = \frac{1}{2} \frac{N}{N-1} \frac{(N-2)R}{N-1} a \left( h - \frac{R}{2b\sigma^2 K^T a} \right).
\]
With many price-taking participants as $N \to \infty$, then the simpler expression, 
$$c(a) \approx \frac{1}{2} Ra \left( h - \frac{R}{2b\sigma^2 K^T} a \right),$$

is obtained.

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 the liquidity premium by its originator, Constantinides (1986), and by construction it represents the amount by which the per-period return on an asset with transaction costs $a$ must rise to make each investor pair indifferent between trading an asset with a zero transaction cost and one with a positive transaction cost.

In order to implement the compensating rise in the rate of return methodology of Constantinides (1986), the required higher expected dividend is $c(a)$ per share but in a closed economy, there is no additional income source to pay the higher dividend. Consequently, the mid-point asset price burdened by transaction costs must fall to create an equivalent utility-equalizing dividend: the required price fall on the expensive to trade asset with price $P_{a=0}^{\text{mp}}(a)$ when invested at the riskless rate must equal the required additional dividend, $R(p_{a=0}^{S} - P_{a=0}^{\text{mp}}(a)) = c(a)$, where $P_{a=0}^{S}$ is the price of the asset with no transactions cost. Hence, the price with transaction costs is

$$P_{a=0}^{\text{mp}}(a) \triangleq p_{a=0}^{S} - \frac{c(a)}{R} = \frac{\mu - c(a) - \frac{b}{2}\sigma^2 K^T}{R}.$$  \hspace{1cm} (29)

It differs from equation (19) by the subtraction of the discounted compensating rate of return term, $\frac{c(a)}{R}$. Hence, the expected dividend of $E(\tilde{d}) = \mu$ in the initial model described by the pricing equation (19) above, where utility is only bounded below by the autarky level provided by prospective returns on risky and safe endowments, is no longer applicable when matched trading-pair investors are guaranteed the utility level provided by maximal trading gains from an asset with zero transaction costs. The new expected dividend on an asset with transaction costs $a$ is effectively lower at $\mu - c(a)$ in equation (29) compared with equation (19) and the corresponding random asset price, $P_{a=0}^{\text{mp}}(a)$ is also correspondingly lower by the present value, $\frac{c(a)}{R}$. Equation (29) describes the constant utility (i.e., real income) asset pricing function across all the equilibria described by each possible value of transaction costs $a$. The observed
rates of return across these equilibria differ by the compensating amount, \( c(a) \), deflated by the midpoint price, \( P^{mp}(a) \), and thus increase in \( a \).

The perpetuity counterpart of the two-period price expression, equation (29), in which the endowment shock and resulting transaction precisely repeat themselves indefinitely, is given by,

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

(30)

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

With more participants, \( i.e. \), a higher \( N \), 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, \( \sigma^2 \), 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 midpoint 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.

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 were not 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 when the underlying transactional demand is high. It is important to recognize that transactions costs \( per se \) do not necessarily result in illiquidity premia. Rather, it is the loss of the gains from trade, when fundamentally strong reasons to trade exist, which is at the heart of the occurrence of illiquidity premia.
D. The Investment Horizon and Frequency of Portfolio Rebalancing

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 \textit{a priori}. If the interval is of calendar length $T$ years then the annualized gross bond and equity yield are $\frac{1}{T} R$ and $\frac{1}{T} \mu$ respectively, the annualized net bond and equity yield are $\frac{1}{T} R - 1$ and $\frac{1}{T} \mu - 1$ respectively, the annualized variance is $\frac{1}{T} \sigma^2$, the annualized stock turnover rate is $\frac{\tau(a)}{T}$, and the annualized compensation rate implicit in Constantinides$^4$ (1986) is,

\[
e(a)_{\text{annual}} = \frac{c(a)_{\text{annual}}}{p_{mp}} = \left( \frac{\mu}{p_{mp}} \right)^{\frac{1}{T}} - \left[ \left( \frac{\mu - c(a)}{p_{mp}} \right)^{\frac{1}{T}} - 1 \right] \approx \left( \frac{\mu}{p_{mp}} \right)^{\left[1 - \frac{1}{T}\right]} \frac{c(a)/p_{mp}}{T} \tag{31}
\]

Since $\frac{\tau(a)}{T}$ and equation (31) are diminishing in $T$, both the annualized stock turnover and compensation rate are falling in the investment horizon, and hence frequency of the portfolio endowment shock of market participants, that determines how often investors trade. Thus, the model is consistent: 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 or is unsuccessful, 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. It is an unpleasant fact that if one wishes to explain observed security trading volume by endowment shocks for investors living for two periods, then frequent shocks and implicitly short-lived investors are required.

E. Valuing the Ability to Trade

$^4$ This was kindly pointed out to me by George Constantinides in correspondence.
The maximum value of the proportional two-way dollar trading cost, \( a \rightarrow \bar{a} = \frac{b\sigma^2 hK^T}{R} \), at which autarky occurs with the inverse function, \( a(\tau) = 0 \), in equation (23), requires a compensating rise in the expected yield, \( \mu \), i.e., rate of return, on the risky asset of,

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

found by evaluating equation (28), or equivalently, a maximal price fall to \( P^{mp}(a) = p^s - \frac{c(\bar{a})}{R} \), where \( p^s \) is the price in the absence of trading costs. 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.

The asset pricing equation (29) and compensation, equation (32), indicate the traditional view as expressed, for example, in Amihud and Mendelson (1986) and in Vayanos and Vila (1999), that the price of a stock is equal to the present value of dividends less the present value of transaction costs, is only part of the story. In an autarky regime the equity premium given by equation (32) is at its highest since the maximal welfare loss, \( c(a) \), for all \( a < \bar{a} \), is sustained, yet by definition no transaction costs are incurred. The compensation required for the imposition of trading costs, \( c(a) \), 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 at zero cost, the benefit from being able to transact at rate \( 0 < a < \bar{a} \), rather than at the prohibitive cost, \( \bar{a} \), is

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

(33)

Since \( P_{mp}(\bar{a}) = p^s - \frac{c(\bar{a})}{R} \) and \( P_{mp}(a) = p^s - \frac{c(a)}{R} \), a relatively liquid asset with transaction cost \( a \) sells for a premium of \( P_{mp}(a) - P_{mp}(\bar{a}) = \frac{c(\bar{a}) - c(a)}{R} = \frac{B(a)}{R} \) over an asset that does not trade at all. Clearly, the pricing benefits from liquidity, \( \frac{B(a)}{R} \), 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, \( \frac{\partial B}{\partial \tau} > 0 \), stock volatility, \( \frac{\partial B}{\partial \sigma^2} > 0 \), and shares held by counterparties, \( \frac{\partial B}{\partial K^T} > 0 \) for all \( a < \bar{a} \). This establishes that the price of any asset depends on a lot more than just expected dividends net of actual trading costs, \( a \tau(a) \) and a risk adjustment.

**F. Transaction Cost Rate to Maximize Transactional Outlays**

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

\[
a_{\text{max}} = \frac{b \sigma^2 K^T h}{2R} = \frac{1}{2} \bar{a}.
\]

(34)

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,

\[ c'(a) = \frac{N}{N-1} R \tau(a) > 0 \tag{35} \]

with an elasticity value,

\[ \eta'_a = \frac{N}{N-1} \frac{R \tau(a)}{c(a)} \tag{36} \]

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 \( N \to \infty \),

\[ \eta_a^{\text{mp}} = -\frac{1}{R} \frac{c'(a) a}{p^{\text{mp}}} = -\frac{N}{N-1} \frac{a \tau(a)}{p^{\text{mp}}} < 0 \tag{37} \]

utilizing equation (35). 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 (29), as it does in much of the conventional asset pricing and tax literature. The price elasticity with respect to transactions costs now becomes,

\[ \eta_a^{\text{mp}} = -\frac{1}{R} \frac{a \tau(a)}{p^{\text{mp}}}, \tag{38} \]

which is positive if the absolute value of the turnover demand elasticity with respect to transactions costs, \( |\eta'_a| \), is greater than one (elastic). For an illiquid asset with a sufficiently high transactions cost, \( a \), to eliminate trading, the stock price, \( p^{\text{mp}} \), in equation (38) is now 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: increasing the 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 so that actual trading costs fall due to the reduction in trading. 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! This is clearly completely nonsensical.

The conventional elasticity, equation (38), is only approximately the same as the true asset price elasticity in equation (37) 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 expected present value of net cash flows from the portfolio of stocks over this horizon, appropriate. Thus, I validate the internal consistency of the clientele model of Amihud and Mendelson (1986) based on these assumptions.

Another important finding which stems from equation (35) 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, $a$, 

$$c(a) = -\int_{x=0}^{\tau(x)} dx = \sum_{x=0}^{\tau(x)} dx$$

In the case of thin trading with relatively small $N$, the area slightly understates the illiquidity premium. The illiquidity premium is 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, it is possible to 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; only negligible compensation is required for bearing transactions costs when trading demand is negligible. In traditional representative investor models, and variants that depart only marginally from this paradigm, there is no or insufficient investor endowment heterogeneity to stimulate either a desire for trading or 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, \( c[a(\tau)] \), by substituting the inverse function given by (23) into (28). The illiquidity premium is diminishing in stock turnover,

\[
c'(\tau) = c'(a)a'(\tau) = -2\frac{N}{N-2}b\sigma^2K^\tau\tau(a) < 0, \tag{40}
\]

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, or any factor that intrinsically raises trading demand and hence the gains from trade, but lowers it when due to falling transactions costs for given endowment heterogeneity or trade motivation. Hence, it is important to distinguish between the two alternative means by which trading activity may increase.

**H. A Simple Reduced-Form Empirical Specification**

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 is linear in logarithms rather than the level,

\[
\tau(\varphi) = \alpha\varphi^{-\gamma} \text{ and } \varphi = \tau(\varphi)^{-1} = \left(\frac{\tau}{\alpha}\right)^{\frac{1}{\gamma}}, \gamma \neq 1, \tag{41}
\]
where \( \varphi = \frac{R_a}{p^\varphi} \) is the relative transactions cost in opportunity cost terms and \( \gamma \) 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 (41), 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,

\[
c(\varphi) \equiv \int_{\varepsilon}^{\varphi} ax^{-\gamma} \, dx, \quad \text{as } \varepsilon \to 0, \quad \gamma \to 0, \quad \frac{\alpha \varphi^{1-\gamma}}{1-\gamma} = \frac{\tau \varphi}{1-\gamma},
\]

(42)

\[
c(\tau) \equiv c[\varphi(\tau)] = \frac{\alpha}{1-\gamma} \left( \frac{\tau}{\alpha} \right)^{-\gamma}, \quad \gamma \neq 1,
\]

(43)

where \( \varepsilon \) is a vanishingly small positive amount, \( \alpha = [R_a]^{\gamma} \), and \( \gamma = 1 - \frac{\tau R_a}{c(\alpha)} \) 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,

\[
c'(\tau) = -\frac{1}{\gamma} \left( \frac{\tau}{\alpha} \right)^{\gamma-1} = -\frac{\varphi}{\gamma},
\]

(44)

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 \( \tau \) and \( \alpha \). 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 U.S. 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 U.S. market. Data is also available for two other relatively comparable markets, namely the U.K. and Australia. In 1992, the annual turnover of Gilts (all U.K. Government Bonds) by final
customers was 3.6636 times and for the equity of U.K. 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). Gilts are reasonably liquid with the entire stock turning over every 1.68 months, but less so than the U.S. 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, so that the ratio of bond to equity turnover fell from 33.3 to 17.2 times over this period. The average experience over this period is quite similar to the U.S. Consequently, government bonds, including Gilts, are exceedingly more liquid, i.e., with 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 U.S. stock and Treasury Bill market over the period, 1889 to 1994, which was the subject of M&P’s equity premium study. The key facts are a mean six percent per annum equity premium over the Standard and Poors stock index, a 18 percent 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 U.S. of eight percent. Jones (2002) computes the average round-trip relative transactions cost (spread plus commission relative to

---

5 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.

6 This relative change is largely due to the halving of stamp duty on stock exchange transaction from July 1995.
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, 1889-1924, 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 \( K_o^S = 7.5, K_o^D = 0 \), with \( K^T = 7.5 \) and relative endowment heterogeneity, \( h \), is at its maximum of 1. This is true for all equity and bond 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 equation (28), 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 12 fold, indicating the significant deleterious impact of even apparently insignificant transactional costs, or taxes for that matter. The same expected return for the base-case in column 2 is just sufficient to generate a price, \( p^{\text{vw}} = 1 \), with the very moderate transactions costs in place of less than 1% on a round trip basis. Similarly, the zero transaction cost bond/T-bill in column 4 is paired with the base-case bond in column 5, reflecting the observed transaction costs on T-bills with a common net return of \( \mu - 1 = 2 \) percent. In every column the summed utility of investor counterparty pairs in the same asset with and without transactions costs are shown to be the
same once the compensating dividend is taken into account. 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 (30), this amounts to an asset price in the absence of transactional costs that is over 400 percent higher, $p^{a-0} = 4.0635$. The annualized equivalent of the compensating amount is shown as the illiquidity premium due to transactions costs. The base case yields an annualized equity premium due to transaction costs of 6.2 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, $T=1/24$ 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 that an investment horizon of (say), $T=20$ years would generate an annualized illiquidity premium consistent with Constantinides (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 illiquidity (equity) premium comes about 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, indicating the huge gains from the ability to trade freely when there are no transactions costs. At the point of unitary elasticity of trading demand in the third column marked “max outlay”, the premium is slightly lower at 4.556 percent and the cost rate is half the autarky amount at 0.506 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, $\eta^T_0$, of $-30.96$. However, the corresponding parameter of the constant elasticity demand specification, equations (41) and (42), 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 5, and also for T-bills with no transactions costs, column 4, the annualized net expected return is 1.75 percent, or 2.06 percent once the illiquidity premium is added, and is very close to the rate of two percent utilized for T-bills in the first three 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 0.312 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 15.5 times bigger than the former. The base case T-bill simulation is drawn to scale in Figure 3.

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 12-fold for relatively small values of \( a \), and over six-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 $\delta$ is the annualized compensation according to equation (28) 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.

**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 2), 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 comes close to duplicating the Datar, Naik, and Radcliffe (1998) empirical finding.

**D. A Test of the Model 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. 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. It also reinforces a key point of the model that the illiquidity premium is at its maximum when the explicit or implicit costs of trading are infinite so that the observed or resource costs of trading, $ar(a)$, are zero.


---

7 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.
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. Since many of the smaller stocks are relatively illiquid, bid-ask spreads were computed only when stocks traded so as to avoid the problem of stale quotes. Monthly returns were computed with the inclusion of dividends and also the volatility measure, the average daily high-price minus low-price deflated by the average daily price, was computed along with the monthly volume of shares traded and shares on issue. The monthly equity premium for each stock was computed by deducting the monthly return on three-month Australian Treasury bills from the overall return. Transaction cost, $\varphi$, 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,

$$c(\tau_\alpha) = \alpha^\rho \left[ (\rho_1 - \rho) \right] \left[ \tau_\alpha^{1-\rho} - \tau_b^{1-\rho} \right] + \epsilon_i, \quad (45)$$

and

$$\tau_\alpha = \alpha \varphi^{-\left(1/\rho\right)} + \epsilon_i, \quad (46)$$

where $\epsilon_i$ is the iid distributed error term and the T-bill turnover rate, $\tau_b$, is set at an annualized rate of eight based on the Australian evidence. The first of the two equations to be simultaneously estimated as a system of simultaneous equations, (45), is the general equity premium result, (43) above with the implied Treasury bill transaction cost, $\varphi_b$, solved for in terms of the known turnover rate, $\tau_b$, the intrinsic liquidity parameter, $\alpha$, and the inverse of the turnover elasticity, $\rho \triangleq (1/\gamma)$. The second equation, (46), is simply the equity turnover relationship, (41) 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

---

8 I wish to thank SIRCA and Kingsley Fong for the construction of this database.

9 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 3.1 above).
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

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 \( \alpha = 0.01455, t\text{-stat.} = 41.807 \), the turnover elasticity is less than unity in absolute value and also highly significant
\( \gamma = 0.78137, \rho = 1.2798, t\text{-stat.} = 139.63 \), and implied T-bill transaction cost,
\( \phi_b = 0.000311 \), is three basis points and therefore in agreement with the U.S. 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, \( \alpha \), by approximately one-third while the estimated transaction cost elasticity falls to
\( \gamma = 0.72145 \). For smaller than median stocks in column 3 the intrinsic liquidity parameter is
lower and this is accompanied by a higher elasticity, \( \gamma = 0.87138 \). Hence smaller stocks tend to
have higher turnover elasticities, compensated for by a lower intrinsic liquidity parameter, \( \alpha \).

The model using the full data set is re-estimated in column 4 allowing for the respective equity
premium and turnover elasticity inverses, \( \rho_e \) and \( \rho_r \), to differ between equations (45) and (46).
The implied turnover elasticity estimated from the first equation is \( \gamma_e = 0.798021 \) and from the
second, \( \gamma_c = 0.760572 \). Hence the differences are very slight. This is yet further confirmation of the strength of the model.

**F. Identity of Elasticity Estimates from Turnover and Equity Premium Data**

Equations (45) and (46) above imply two paths by which the turnover elasticity, \( \gamma \), can be estimated. The two paths should generate the same \( \gamma \) outcome. Firstly, \( \gamma \) is identified with respect to transaction costs, \( e_\tau \), directly from turnover information, \( \tau_e \), and, secondly, \( \gamma \) indirectly defined from the equity premium, \( c(\tau) \), and the resource cost, \( \tau_0 \). 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,

\[
\tau_{et} = \alpha \phi_{et}^{-\gamma} + \varepsilon_e, \quad (47)
\]

and

\[
c(\tau) = \alpha_0 + \tau_0 \phi_{et}/(1 - \gamma - \alpha_i) + \varepsilon_e, \quad (48)
\]

with iid error terms \( \varepsilon_e \) and the \( \gamma \) estimate the same in both equations if \( \alpha_i = 0 \).

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, \( \alpha_i \), 14 had absolute \( t \) values of 1.96 or better meaning that there is a statistically significant difference between the \( \gamma \) elasticity estimates from each equation. Hence, the hypothesis of equal \( \gamma \) estimates is accepted for 76 of the 90 equations. There were also 17 instances in which the estimate of \( \gamma \) 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 \( \gamma \) in these cases. Consequently, the hypothesis of a zero \( \gamma \) elasticity is rejected for 73 of the 90 stocks. Moreover, the joint hypothesis of the same \( \gamma \) estimate from both equations and a positive and significant \( \gamma \) 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.

The model explains how portfolio rebalancing by investors will generate as an equilibrium a six percent equity (illiquidity) premium and two percent bond yield with precisely the same security turnover rates for equity and bonds as actually observed over the period 1896-1994, 38 percent and about 880 percent per annum, respectively. These investors have identical CARA preference, exhibit low risk aversion, and face the actual volatilities of equity and T-bills observed over this period. Very conservative measures of transactions costs have been adopted. The equity premium is explained by the high gains to trade evident from observed equity and bond turnover rates and the significant effect that even modest transactions costs have in reducing these gains from trade. 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. It also explains the differential returns and prices of otherwise identical “A” and “B” stocks trading in China (Chen and Swan, 2005) My findings suggest that the welfare cost of the imposition of Tobin taxes (stamp duties) on financial markets is likely to be very high. 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


Jang, Bong-Gyu, Hyeng Keun Koo, Hong Liu 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.


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

Key:
2. Transactions with Interdealer Brokers; Daily trading in Treasury Securities, U.S. Billions
3. Transactions with Others, Daily trading in Treasury Securities U.S. 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 U.S., 1889-1994.

<table>
<thead>
<tr>
<th></th>
<th>Equity</th>
<th>T-Bill</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zero Tcost</td>
<td>Base Case</td>
<td>Max Outlay</td>
<td>Zero Tcost</td>
<td>Base Case</td>
<td>Max Outlay</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Annualised Return Variance, $\sigma^2$</td>
<td>3.24%</td>
<td>3.24%</td>
<td>3.24%</td>
<td>0.36%</td>
<td>0.36%</td>
<td>0.36%</td>
</tr>
<tr>
<td>Annualized Riskless Rate, $r$</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Coefficient of absolute risk aversion, $b$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Initial endowment of each seller, $K_{0,S}$</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>Initial endowment of each buyer, $K_{0,D}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Relative Endowment Heterogeneity, $h$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Initial Bond Endow, Supplier, $w_{0,S}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Initial Bond Endow, Demander $w_{0,D}$</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>Total Population of Traders, $N$</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>Annualised Expected Net Return, $\mu - \lambda$</td>
<td>14.15%</td>
<td>14.15%</td>
<td>14.15%</td>
<td>1.75%</td>
<td>1.75%</td>
<td>1.75%</td>
</tr>
<tr>
<td>Two sided prop dollar spread, $a$</td>
<td>0%</td>
<td>0.98%</td>
<td>0.506%</td>
<td>0%</td>
<td>0.03%</td>
<td>0.056%</td>
</tr>
<tr>
<td>Prohibitive level of Trading Cost, $a_{bar}$</td>
<td>1.03%</td>
<td>1.03%</td>
<td>1.03%</td>
<td>0.11%</td>
<td>0.11%</td>
<td>0.11%</td>
</tr>
<tr>
<td>TCost Prop of Mid-Point Price, $a/p_{mp}$</td>
<td>0%</td>
<td>0.980%</td>
<td>0.506%</td>
<td>0%</td>
<td>0.030%</td>
<td>0.056%</td>
</tr>
<tr>
<td>Equilibrium Demand by Seller, $K_{1,S}$</td>
<td>3.750</td>
<td>7.383</td>
<td>5.625</td>
<td>3.75</td>
<td>4.75</td>
<td>5.625</td>
</tr>
<tr>
<td>Equilibrium Demand by Buyer, $K_{1,D}$</td>
<td>3.750</td>
<td>0.117</td>
<td>1.875</td>
<td>3.75</td>
<td>2.75</td>
<td>1.875</td>
</tr>
<tr>
<td>Equilibrium Amnt Sold per Forntight, $\Delta K$</td>
<td>3.750</td>
<td>0.117</td>
<td>1.875</td>
<td>3.75</td>
<td>2.75</td>
<td>1.875</td>
</tr>
<tr>
<td>Annualised Turnover Rate, $\tau_{a}(\cdot)$</td>
<td>12</td>
<td>0.38</td>
<td>6</td>
<td>12</td>
<td>8.80</td>
<td>6</td>
</tr>
<tr>
<td>Annual Illiq Prem per unit Tcost, $c(a)/a$</td>
<td>12.010</td>
<td>6.193</td>
<td>9.007</td>
<td>12.002</td>
<td>10.401</td>
<td>9.001</td>
</tr>
<tr>
<td>Annual Illiq Prem due Tcost, $c(a)$</td>
<td>0%</td>
<td>6.069%</td>
<td>4.556%</td>
<td>0%</td>
<td>0.312%</td>
<td>0.506%</td>
</tr>
<tr>
<td>Ratio Unobserved to Observed Tcost</td>
<td>NA</td>
<td>15.494</td>
<td>0.501</td>
<td>NA</td>
<td>0.182</td>
<td>0.500</td>
</tr>
<tr>
<td>Ratio Equity to Tbill Premium</td>
<td>NA</td>
<td>19.449</td>
<td>9.000</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Utility of Trading Pair with TCosts</td>
<td>NA</td>
<td>15.0315</td>
<td>15.0315</td>
<td>15.0046</td>
<td>15.0046</td>
<td>15.0046</td>
</tr>
<tr>
<td>Utility of Trading Pair with No TCost</td>
<td>15.0315</td>
<td>NA</td>
<td>NA</td>
<td>15.0046</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Annual Gains fm Trade rel Autarky, $B(a)$</td>
<td>6.075%</td>
<td>0.006%</td>
<td>1.519%</td>
<td>0.675%</td>
<td>0.363%</td>
<td>0.169%</td>
</tr>
<tr>
<td>Equil Price without Tcost, $p_{mp}^{net}$</td>
<td>1</td>
<td>0.9951</td>
<td>0.9975</td>
<td>1</td>
<td>0.9999</td>
<td>0.9997</td>
</tr>
<tr>
<td>Equil Mid-Point Price with Tcost, $p_{mp}^{new}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Trading Elasticity wrt Tcosts, $\eta_{a}$</td>
<td>0</td>
<td>-30.957</td>
<td>-1</td>
<td>0</td>
<td>-0.364</td>
<td>-1</td>
</tr>
<tr>
<td>Elastic of Req Compen wrt Tcosts, $\phi_{a}$</td>
<td>1</td>
<td>0.061</td>
<td>0.667</td>
<td>1</td>
<td>0.846</td>
<td>0.667</td>
</tr>
<tr>
<td>Annual Gains fm Tcosts, $B(a)$</td>
<td>-0.2430</td>
<td>-0.0076</td>
<td>-0.1215</td>
<td>-0.0270</td>
<td>-0.0198</td>
<td>-0.0135</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Statistic</th>
<th>Equity Premium, ( c(\varphi) )</th>
<th>Trans. Cost, ( \varphi )</th>
<th>Turnover, ( \tau )</th>
<th>Market Cap.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.061887</td>
<td>0.041707</td>
<td>0.27575 pa</td>
<td>$726.4 m.</td>
</tr>
<tr>
<td>Stand. Dev.</td>
<td>0.83756</td>
<td>0.039103</td>
<td>0.39685</td>
<td>0.26118\times 10^{10}</td>
</tr>
<tr>
<td>Median</td>
<td>-0.063089</td>
<td>0.029113</td>
<td>0.159 pa</td>
<td>$68.1 m.</td>
</tr>
</tbody>
</table>

Table IV: Regression results for sample of approximately 576 Australian securities listed on the Australian Stock Exchange, 1994-98.

Estimating two simultaneous equations: 

\[ c(\tau)_i = \alpha \rho \left( \frac{1}{\rho - 1} \right) \left[\tau_i \rho - \tau_i \rho \right] \]

and 

\[ \tau_i = \alpha \varphi_i \left( \frac{1}{\rho} \right) \]

for the equity premium and the stock turnover rate, respectively.

<table>
<thead>
<tr>
<th>Column 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>Full Sample</td>
<td>Above Median Market Cap.</td>
<td>Below Median Market Cap.</td>
</tr>
<tr>
<td>Intrinsic Liquidity Coeff. (( \alpha ))</td>
<td>0.01455 (41.807)</td>
<td>0.021293 (31.217)</td>
<td>0.009559 (14.379)</td>
</tr>
<tr>
<td>Inverse of Turnover Elastic, ( \rho )</td>
<td>1.2798 (139.63)</td>
<td>1.3861 (86.199)</td>
<td>1.1476 (53.674)</td>
</tr>
<tr>
<td>Equity Inv of Tover Elastic ( \rho_e )</td>
<td>1.2531 (85.396)</td>
<td>1.3148 (76.412)</td>
<td>1.0909 (53.673)</td>
</tr>
<tr>
<td>Inverse of Turnover Elastic ( \rho_t )</td>
<td>0.78137</td>
<td>0.72145</td>
<td>0.87138</td>
</tr>
<tr>
<td>Turnover Elasticity (( \rho ))</td>
<td>0.78137</td>
<td>0.72145</td>
<td>0.87138</td>
</tr>
<tr>
<td>Elasticity estimated from the Equity Premium (( \rho_e ))</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Elastic est from Turnover (( \rho_t ))</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Implied T-bill TCost (( \varphi_b ))</td>
<td>0.000311</td>
<td>0.00027</td>
<td>0.000443</td>
</tr>
</tbody>
</table>

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 1 is drawn to scale and assumes monthly portfolio endowment shocks and trading, i.e., a investment horizon of one month, a CARA coefficient, $b = 1$, annualized variance, $\sigma^2 = 0.1225$, 2% annual interest rate, $R = 1.02$, endowment heterogeneity, $h = 1$, and 2 shares held by each trading pair, $K^T = 2$. 
Figure 2: Six 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