A CENTURY OF STOCK MARKET LIQUIDITY AND TRADING COSTS

Charles M. Jones
Graduate School of Business
Columbia University

This version: June 25, 2001

I thank Hans Stoll, Venkat Eleswarapu, and Tarun Chordia for generously providing data. I am grateful to Geert Bekaert, Charles Calomiris, Larry Fisher, Joel Hasbrouck, Raj Singh, Dan Weaver, James Weston, seminar participants at Rutgers University and the University of Florida, and participants at the 2000 Nasdaq-Notre Dame Microstructure Conference, the Fall 2000 NBER Market Microstructure Conference, and the 2001 Western Financial Association meetings for comments and insights. Pamela Moulton provided excellent research assistance. I also thank Callean Henry, Barbara Iyayi, Barbora Jemelkova, Roy Mateus, Jorge Vargas, and especially Anne Caballero for help in entering bid-ask data.

ABSTRACT

I assemble an annual time series of bid-ask spreads on Dow Jones stocks from 1898-1998, along with an annual estimate of the weighted-average commission rate for trading NYSE stocks since 1925. Spreads gradually declined over the course of the century but are punctuated by sharp rises during periods of market turmoil. Proportional one-way commissions rise dramatically to a peak of nearly 1% in the late 1960's and early 1970's, and fall sharply following commission deregulation in 1975. The sum of half-spreads and one-way commissions, multiplied by annual turnover, is an estimate of the annual proportional cost of aggregate equity trading. This cost drives a wedge between aggregate gross equity returns and net equity returns. This wedge can account for only a small part of the observed equity premium, but all else equal the gross equity premium is perhaps 1% lower today than it was early in the 1900's. Finally, I present evidence that these measures of liquidity – spreads and turnover – predict stock returns up to one year ahead. High spreads predict high stock returns; high turnover predicts low stock returns. This suggests that liquidity is an important determinant of conditional expected returns.

1. Introduction

In an effort to understand the behavior of asset prices, financial economists have assembled long time series and panels of asset returns. For example, Schwert (1990) and Siegel (1992a, 1992b) put together various time series for US equity, bond, and/or riskless asset returns going back well into the 1800's. Jorion and Goetzmann (1999) assemble a panel of stock market returns in various countries over the 20th century in order to determine whether selection bias accounts for some or all of the Mehra and Prescott (1986) equity premium puzzle. Froot, Kim, and Rogoff (1995) collect over 700 years of data on the relative prices of various commodities, in an attempt to determine whether purchasing power parity holds in the long run.

However, we currently know very little about the trading environment and the frictions faced by investors in the early years of these time series. Empirical work in market microstructure, for example, has focused almost exclusively on drawing positive and normative conclusions based on the recent trading environment. Empirical asset pricing researchers often use recent trading cost levels in calibration or estimation (see, for example, Heaton and Lucas, 1996). These are sensible decisions given the absence of comprehensive historical transaction cost data.

Despite the lack of comprehensive data to date, economic agents have encountered frictions since the dawn of asset markets. More to the point, these frictions, and variation in these frictions over time, may have far-reaching implications for models of asset pricing. For example, if agents face large transaction costs at certain times, realized equity returns might be considerably lower than the gross equity returns implicit in stock index values. If frictions are substantial, asset price behavior that might initially appear anomalous could be well within transaction cost bounds and thus consistent with efficient markets. And if transaction costs covary with the business cycle, this might account for some of the observed regularities in the cross-section and time-series of equity returns.

To explore some of these issues, I introduce in this paper three annual time series related to US equity market trading frictions and liquidity. The time series include:

(1) quoted bid-ask spreads on large stocks from 1898 to 1998,

(2) the weighted-average explicit costs associated with trading NYSE stocks, including commissions and other fees, since 1925, and

(3) turnover in NYSE stocks since 1900, collected in order to judge the overall incidence of these other frictions.

This work is also closely related to the nascent literature on systematic liquidity, including Hasbrouck and Seppi (2001), Huberman and Halka (1999), and Chordia, Roll, and Subrahmanyam (2001). The last paper examines variation in average NYSE bid-ask spreads since the mid-1980's. Their goal is to predict changes in liquidity at short horizons. In contrast, the goal of this paper is to document systematic, cyclical changes in liquidity over a much longer time period at much longer wavelengths. But Chordia, Roll, and Subrahmanyam (2000) also note that an important issue not investigated in their work "is whether and to what extent liquidity has an important bearing on asset pricing." This paper is concerned with exactly this question. Specifically, by assembling a long time series on liquidity, it becomes possible to explore low frequency time-variation in liquidity. This raises the tantalizing possibility that time-variation in spreads, turnover, and other liquidity measures may be closely associated with time-varying expected returns.

To preview the results, I find that proportional spreads on Dow Jones stocks have declined over time, but the decline has been neither gradual nor smooth, and in fact spreads were as low in the 1920's as they were in the 1980's. There are frequent sharp spikes in spreads. These are usually (but not always) associated with market turmoil.

In contrast to spreads, average proportional commissions on NYSE stocks climbed steadily from 1925 to the late 1960's and early 1970's to a high of almost 1%. Of course, commissions plummeted shortly thereafter as the SEC broke the NYSE commission cartel.

Finally, turnover in NYSE stocks varies widely over time. Turnover exceeds 200% in the early years of the 20th century, plunges to single digits following the Great Depression, and has been steadily increasing since.

When I calibrate a very simple model of gross vs. net equity returns, there is nothing that can account for much of the observed equity premium in US stocks. However, the calculations suggest that the gross equity premium might have fallen about 1% after the first third of the century.

Last, and perhaps more important, I take these time series of liquidity variables and investigate whether liquidity, broadly defined, might account for some of the apparent time-variation that has been observed in expected stock returns. I find that spreads and turnover both predict excess stock returns up to one year ahead. Over the course of the 20^{th} century, these liquidity variables dominate traditional predictor variables, such as the dividend yield.

The paper is organized as follows. Section 2 discusses the time series of bid-ask spreads, commission rates and other fees, and turnover, and proposes a combined measure of aggregate annualized trading costs. Section 3 uses this measure of annualized trading costs to contrast the gross and net equity premium over the last 100 years. Section 4 estimates predictive regressions and a VAR using bid-ask spreads and turnover as broad measures of liquidity. Section 5 recaps and discusses potential future work.

2. Data

2.1. Bid-ask spreads

There is no shortage of data on quoted spreads for recent years. A complete record of intraday trades and inside quotes on NYSE and AMEX stocks has been available since the mid-1980's. From 1987 to 1992, intraday data are available from the Institute for the Study of Securities Markets ("ISSM").[1] Beginning in 1993, the Trades and Quotes ("TAQ") database is available directly from the New York Stock Exchange.

All bid-ask data prior to this date must be and have been collected by hand. For the period from 1960 to 1979, Hans Stoll generously provided annual proportional bid-ask spreads on all NYSE stocks.[2] These are hand collected from the periodical *Stock Quotations on the NYSE* published by Francis Emory Fitch. Stoll and Whaley (1983) use these data and find that transaction costs erode some of the return differential between small and large stocks. These data are also used by Amihud and Mendelson (1986) and Brennan, Chordia, and Subrahmanyam (1998). Both sets of authors find that bid-ask spreads explain some of the cross-section of expected returns.

Eleswarapu and Reinganum (1993) collect similar data for 1980-1989. I use their 1980-1986 data to bridge the gap between the earlier Fitch data and the intraday ISSM data.

Prior to the end of 1961, the Commercial and Financial Chronicle (hereafter the "C&FC") provides month-end bid and ask prices on *all* NYSE and Curb stocks, as well as a

[1] ISSM data are available beginning in 1984, but I do not have access to the first three years of data.
[2] In this dataset, the observation for year t is the average of quoted spreads at the end of year t and the end of year t-1, reflecting the notion of the spread as an average spread throughout the calendar year.

large number of over-the-counter stocks. Between 1928 and 1961, these quotes are published in the Bank and Quotation Record, a separate publication of the C&FC. Prior to 1928, the Bank and Quotation Section is published once a month in the C&FC itself. Closing bid-ask data are also available in many daily newspapers, including the Wall Street Journal and New York Times, up until around 1950. Closing bid-ask data have been used in other contexts by Arnold et al. (1999), Calomiris (1999) and Fisher and Weaver (1999), but their existence is not well known.

I use all three of these sources to collect monthly bid-ask spread data on a subset of stocks in the Dow Jones averages. From 1928 to 1961, I collect data on all 30 Dow Jones Industrial Average (DJIA) stocks. Focusing on the DJIA has several advantages. First, the index has a relatively small number of stocks, which makes data collection easier. Second, historically at least, it has closely tracked the performance of a broader value-weighted index of stocks. In fact, during this time period these thirty stocks alone account for between one-third and one-half of the total market capitalization of all NYSE stocks.

Prior to October 1, 1928, the DJIA had fewer than 30 stocks. At its inception on May 26, 1896, the average consisted of 12 stocks with a heavy emphasis on commodities (examples include American Cotton Oil, American Sugar, and Tennessee Coal and Iron, as well as General Electric, the only original DJIA component still in the average). The average also included up to two preferred stocks in this early period. On October 4, 1916, preferred stocks were removed, and the average expanded to 20 common stocks. Because there are so few stocks in the DJIA during this period, I also include the common stocks that were components of the Dow Jones Railroad Average (the predecessor to today's Dow Jones Transportation Average). The Railroad Average consists of 20 stocks throughout. After eliminating preferred stocks, the overall sample during the 1896-1928 period contains a minimum of 25 and a maximum of 40 stocks.

Another advantage of using Dow stocks is that the components change infrequently over most of the sample period. For example, there are no changes to the industrials between March 14, 1939 and July 3, 1956 (over 17 years) and also between June 1, 1959 and November 1, 1972 (more than 13 years). During the Great Depression, there is greater turnover in the Dow components, which largely reflects the tumult in American industry at the time. During the 1930's, for example, there are 23 additions/deletions to the DJIA. More details on the composition of the Dow Jones averages can be found at http://averages.dowjones.com.

For each month, I calculate the bid-ask spread for each Dow stock as a proportion of its bid-ask midpoint and aggregate up to an equal-weighted cross-sectional mean. The original data source was compiled by hand, and as a result there are a small number of obvious typographical errors. Filters remove all observations with spreads that are negative or zero. Filters also flag any proportional spread that is greater than 10%. In each such case, other data sources and/or nearby months were examined in an effort to determine whether the large spread was representative of trading conditions in that stock. In each case, it appeared that the quote was erroneous, and all such observations were deleted. As a result, for some months the average bid-ask spread is calculated using one or two fewer stocks.

Between 1960 and 1987, the spread data are annual. An annual observation for each year prior to 1960 is calculated using the median of the 12 monthly average spreads. The median is calculated because it is more robust in the presence of errors in the recording of bid and ask prices that might have escaped the filters.

From 1987 to 1998, I use intraday data and standard filters to calculate the time-weighted average proportional quoted spread for each DJIA stock on each trading day. The pooled mean proportional spread for the calendar year is used in the annual time series.

The resulting annual time series of Dow Jones bid-ask spreads 1898-1998 is displayed in Figure 1. There are several things to note. First, bid-ask spreads are more volatile in the first third of the 20th century. In these early years, there are several instances when average spreads on Dow Jones stocks increase or decrease by 40 basis points in a single year. It is perhaps surprising that spreads on Dow Jones stocks were around 0.60% for sustained periods around 1910 and in the 1920's and were at similar levels in the 1950's and in the 1980's. Spreads have fallen dramatically over the last twenty years.

Spread levels also appear to depend on stock market movements. This is apparent from Figure 2, which provides a comparable monthly time series of stock index levels based on the data in Shiller (2000). Spreads skyrocketed in the Great Depression, and they rose somewhat during the bear market in the first half of the 1970's. Spikes in spreads often coincide with or closely follow market downturns; examples include 1903, 1907, and 1913-1914. However, the relationship is far from perfect. There are a number of market downturns that do not appear to be associated with higher spreads (e.g., 1920, 1937, 1957, and 1962).

2.2 Commissions

Of course, spreads are not the only cost associated with trading stocks. Equity investors must also pay brokerage commissions as well as certain fees and taxes. Commissions are now quite small, especially for the institutions that dominate the US market today. For example, Jones and Lipson (2001) find that one-way institutional commissions on NYSE-listed stocks during 1997 are about 0.12% of the amount transacted. However, this was not true for most of the 20th century. Prior to May 1, 1975, the NYSE and other exchanges set minimum commissions that were almost always binding. The commission schedules changed several times over this period, but as an example, NYSE commissions between March 3, 1959 and December 5, 1968 were set according to the following schedule:

Money Involved	Minimum commission per 100 shares
\$100 to \$400	\$3 + 2% of amount traded (with a \$6 minimum)
\$400 to \$2,400	\$7 + 1% of amount traded
\$2,400 to \$5,000	\$19 + 0.5% of amount traded
Over \$5,000	\$39 + 0.1% of amount traded

Odd lots of less than 100 shares were subject to a slightly different schedule. Appendix A summarizes the commission schedules in effect since 1925.

At the end of 1962, the average NYSE share price was \$40. Trading 100 shares of such a stock would result in a one-way commission of \$39, or 0.975% of the money involved. This is a substantial fraction. It is also important to note that, prior to 1968, the NYSE commission schedule was always linear: a trade of 3,000 shares incurred a commission 30 times as large as a trade of 100 shares. Thus, one can think of commissions as a proportional tax on transactions, where the tax rate depends on the share price.[3]

Since the proportional commission depends only on the share price, it is possible to estimate the weighted average commission rate during the fixed commission regime by looking only at the cross-sectional distribution of share prices and the total volume of trade. Other than ignoring odd lot transactions, which are a negligible part of total volume over most of the

[3] Commission rebates were strictly prohibited. However, other arrangements were made. In addition to the bundling of research and execution services, as well as other soft dollar arrangements that continue today (see, for example, Conrad, Johnson, and Wahal (1997)), the give-up was particularly popular until it was outlawed by the SEC in 1968. This involved a NYSE member giving up part of his commission to another member or even to a non-member, if the order was executed on a regional exchange or in the upstairs market. In a typical give-up, an

sample, one does not need information on the distribution of order sizes. This is fortunate, since such data are not readily available prior to the advent of intraday data. CRSP volume data begins in July 1962, and from that date forward, it is possible to calculate the weighted average commission rate in this way. For example, in the second half of 1962, the dollar volume-weighted average one-way commission rate on NYSE common stocks is 0.82%.

Prior to July 1962, CRSP does not provide volume data. Thus, from 1925 to 1962 I calculate market value-weighted average commission rates instead. Of course, if dollar trading volume is proportional to market capitalization, then the two weighting schemes yield identical average commissions.

In December 1968, a modest volume discount was instituted on transactions over 1,000 shares. In 1971, the process of commission deregulation began, with commissions on the excess of any order over $500,000 determined "as mutually agreed." Deregulation gradually extended to smaller orders, until all commissions were deregulated on May 1, 1975, commonly known as May Day (see Jones and Seguin (1997), for example). The SEC (1977) reports that commissions on institutional trades fell rapidly; see also Stoll (1979), Melnik and Ofer (1978), and Jarrell (1984).

Given the discounts and deregulation that began in 1968, it is impossible to calculate average proportional commission rates for the more recent period. However, NYSE members annually report their income and expenses, and commission income is broken out separately. It is thus possible to use the time series of members' commission income as an indication of overall commissions on NYSE stocks. The problem is that this line item refers to all commission income, including stocks on other exchanges, as well as bonds and other securities. To get around this, annual commission income data back to 1966 can be compared to the estimated total commissions on NYSE stocks calculated using the fixed schedule. During the 1966-1968 fixed commission period, members' overall commission income averaged 2.75 times the amount calculated from the schedule, so total NYSE commissions are assumed to remain a constant 1 / 2.75 = 36.35% of members' commission income. This figure is then scaled by total dollar volume in NYSE stocks to arrive at a weighted average commission rate since 1968. These estimates appear to be quite accurate. For example, the estimated overall commission rate here

---

institution would direct the executing broker to give perhaps 60% of the commission to another firm that supplied research or other services to the institution. Stoll (1979) provides an excellent summary of such practices.

closely coincides with the institutional commissions reported in both Keim and Madhavan (1997) and Jones and Lipson (2000). Keim and Madhavan (1997) find an average commission rate of 0.20% for trades from Jan 1991 through Mar 1993 vs. an estimate in this paper of 0.24%. Jones and Lipson (2000) find an average 0.12% one-way commission for NYSE trades in 1997, compared to an estimate of 0.13% based on NYSE members' commission income.

Data since 1968 derived from members' commission income are spliced with the earlier data based on commission schedules, and Figure 3 displays the resulting time series. Note that the figure provides no estimates of proportional commissions before 1925. While NYSE commission schedules are available continuously back well into the 1800's, prior to the CRSP data it is more difficult to obtain volume or market cap information for all NYSE stocks, at least one of which is required to calculate a sensible weighted average commission rate.

Weighted average one-way proportional commissions begin at 0.27% in 1925, and are below 0.30% throughout the 1920's. As stock prices fall at the start of the Great Depression, proportional commissions rise to a local maximum of 0.67% in 1932. Commissions decline over the next five years but soon resume their gradual upward climb, reaching a fixed-commission era high of 0.88% in 1964 and 1965. Because of the fall in stock prices in 1973-1974, estimated average commissions reach 0.90% in 1974, despite partial commission deregulation. Commissions fall dramatically beginning in 1976, consistently declining by about half every seven or eight years since then.

2.3 Annualized trading costs

When summed together, bid-ask spreads and commissions represent an important and variable friction in trading US equities over the 20th century. Total one-way transaction costs (defined here as half the quoted spread on Dow Jones stocks plus average one-way commissions on NYSE stocks) are summarized in Figure 4. Total costs average 0.86% over the 1925-1998 period. Total costs have been below 0.50% since 1991 but were also below 0.50% from 1926 to 1928. Spreads and commissions together represented at least 1.00% of the dollar volume of trade for the entire period from 1953 to 1975.

This has interesting implications for transaction-intensive portfolio strategies. A number of researchers have identified such strategies that appear to provide returns in excess of their apparent risks (see Fama (1998) for a selective and skeptical summary). However, many such strategies (e.g., momentum strategies) require considerable turnover (see, for example, Moskowitz and Grinblatt (1999), especially p. 1269). Given the spread and commission levels measured here, it seems unlikely that a NYSE non-member would be able to realize profits on most transaction-intensive strategies. This is especially true for any trading strategy that requires trading in less liquid stocks, since spreads on the small-cap stocks involved in many of these candidate trading strategies are much wider than spreads on the large-cap stocks studied here.

As discussed in the introduction, spreads and commissions also eat into overall equity returns. Clearly, a buy-and-hold investor would incur very few such costs. So the relevant question is how often most investors trade and incur these costs. There is some evidence, such as Odean (1999), that investors trade too much, but for now consider investors' trading behavior as given and measure how much of the overall return on equities is lost to bid-ask spreads and commissions over the course of the 20th century.

Figure 5 displays the annual time series on share turnover in NYSE stocks. Annual share turnover is defined as annual share volume divided by total shares listed on the exchange. The time series is assembled from various issues of the annual NYSE Fact Book and extends from 1900 to the present. What is most notable is that annual turnover was often more than 200% in the first five years or so of the 1900's and regularly exceeded 100% up until about 1920. Turnover was back above 100% in 1928 and 1929 (132% and 119%, respectively), but following the stock market crash that began in 1929, turnover plummeted, reaching a low of only 9% in 1942. Though stock prices rebounded, volume remained at very low levels. During the 1940's and 1950's, for example, annual turnover averaged only 16%. Turnover has since climbed gradually, averaging 57% from 1983-1998, with local maxima in 1987 (73%) and in 1998 (76%).

When turnover is low, there is little room for spreads, commissions, and other transaction costs to whittle away aggregate investors' returns from holding common equity. For example, assume for simplicity that investors always trade with NYSE members and/or specialists, in the process incurring the half spread as well as one-way commissions.[4] Assume further that members and specialists hold a net zero position in equities. Then the aggregate net equity return in any period, R_{net}, is given by:

[4] In reality, NYSE members and specialists participate in approximately 20% of trading volume over the sample period.

$$R_{net} = R_m - V\,(½s + c), \tag{1}$$

where R_m is the gross equity return, V is the turnover per period, s is the proportional bid-ask spread, and c is the one-way proportional commission rate, expressed as a fraction of the beginning-of-period price.

Define $L \equiv V\,(½s + c)$. This variable measures aggregate proportional losses each period due to spread and commission trading frictions (in this case, aggregate annualized trading costs).[5] Figure 6 provides a time-series of this aggregate annualized trading cost measure over the 1900-1998 period.[6] Most notable is that, since about 1934, these trading costs average 0.21% annually, a figure that is almost negligible relative to aggregate investors' stock returns. In the first decade of this century, however, these trading costs were substantive, even on an aggregate basis, averaging 1.34% annually from 1900-1910. This high figure is clearly driven by the elevated turnover levels seen in these early years.

3. Implications for the equity premium

Could these trading costs account for some of the equity premium, and could the decline in trading costs account for some of the apparent decline in the equity premium in recent years? To address this question, consider the following description of a simple model, based loosely on Fisher (1994). Suppose that identical price-taking investors hold a riskless asset and a portfolio of stocks. The riskless asset can be purchased at zero cost. On the other hand, to buy or sell stocks, agents must pay an exogenously determined but random proportional transaction cost L_t that consists of a bid-ask spread plus a commission. By assumption, agents trade for exogenous rebalancing or liquidity reasons, so the size of the transaction cost does not affect the trading decision. Then, the trading costs can be considered exogenous, fixed taxes, and the usual stochastic discount factor methodology applies net of trading costs. That is, the portfolio of stocks has gross return R_t at time t, but in equilibrium it must be the case that:

$$E_t(m_{t+1}[R_{t+1} - L_{t+1}]) = 1. \tag{2}$$

Since $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$, this implies that:

$$\mathrm{Cov}_t(m_{t+1}, R_{t+1} - L_{t+1}) + E_t(m_{t+1})E_t(R_{t+1} - L_{t+1}) = 1 \tag{3}$$

[5] This is likely to be a conservative estimate of the annualized costs associated with aggregate equity investment, since it excludes all fees paid to investment managers.

Divide by $E_t(m_{t+1})$:

$$E_t(R_{t+1} - L_{t+1}) = E_t(m_{t+1})^{-1} + E_t(m_{t+1})^{-1} \mathrm{Cov}_t(m_{t+1}, R_{t+1} - L_{t+1}) \quad (4)$$

Rearranging yields the following expression for the expected gross return on the risky asset:

$$E_t(R_{t+1}) = E_t(m_{t+1})^{-1} + E_t(m_{t+1})^{-1} \mathrm{Cov}_t(m_{t+1},R_{t+1}) + E_t(L_{t+1}) + E_t(m_{t+1})^{-1} \mathrm{Cov}_t(m_{t+1},L_{t+1}) \quad (5)$$

The first two terms in this expression are the usual ones. The third term $E_t(L_{t+1})$ reflects the direct effect of transaction costs on gross expected returns. The last term is the indirect effect of transaction costs on expected returns if variation in transaction costs is related to the business cycle or is otherwise correlated with the stochastic discount factor.

This framework is quite different from much of the literature that relates transaction costs to the equity premium. Here, the amount of trading is exogenous. Most other papers in this area take the transaction cost as the primitive and solve for optimal trading, portfolio holdings, and/or consumption/investment decisions. Examples include Constantinides (1986), Aiyagari and Gertler (1991), He and Modest (1995), and Heaton and Lucas (1996). However, these other models do not explain trading volume very well, and researchers generally have very little idea why market participants trade the way they do. Hence, it may not be a bad idea to calibrate to observed trading behavior, leaving for another day the question of why people trade.

For the rest of this section, suppose that the indirect effect is zero, so that transaction costs drive a simple wedge between gross and net equity returns. Equivalently, suppose that variation in transaction costs over time is not a priced risk. Under these assumptions, the relevant equity premium is net of transaction costs. Recall that Figure 6 displays the time series of these aggregate realized trading costs. The (arithmetic) average annualized trading cost equals 0.39%. Using data from Shiller (1988) and updated through 1998 on his web page, the (arithmetic) average gross equity premium over the 1900-1998 period is 7.20%, which is slightly lower than other estimates mainly because he uses the prime commercial paper rate as the riskless rate. This implies that the average equity premium net of realized trading costs over the period is 6.81%.[7]

If aggregate realized trading costs vary over time but cannot be predicted and the required net equity premium remains constant at 6.81%, then the required gross equity premium is simply

---

[6] Recall that it is difficult to measure overall commissions in the pre-CRSP era. Henceforth, proportional one-way commissions are assumed to be a constant 0.27% before 1925, which is the weighted average commission rate in 1925.

[7] Even if the equity premium has remained constant over the 1900's, there is still considerable statistical uncertainty in estimates of the equity premium. For the Shiller (1988) data from 1900-1998, the standard error on the annual equity premium is 2.01%. See, for example, Ang, Bekaert, and Liu (2000).

equal to 6.81% plus the annualized trading cost for that year. This is displayed in Figure 7. As noted earlier, the annualized realized trading cost has been small since the late 1930's. Thus, for the last two-thirds of the sample period, the estimated gross equity premium is relatively constant at just above 7%. However, in the early 1900's, annualized trading costs are much higher, and the implied gross equity premium is also higher, ranging between 8% and 9%.

This exercise suggests that if agents simply add back the aggregate level of transaction costs in determining a gross required rate of return, there is a case to be made that the equity premium is perhaps 1% lower than it was early in the century. However, this exercise suggests that the equity premium has not declined substantially during the last quarter-century.

The decline in transaction costs has further implications for estimates of the equity premium based on historical data. If the decline in trading costs over the 20th century exceeded the expectations of an investor in 1900, then prices relative to dividends or earnings should be higher than they were in 1900, all else equal. Thus, some of the cumulative equity return over the past century is likely due to this expansion in multiples. To the extent that this expansion in multiples was unanticipated, the ex-ante equity premium would be lower than the ex-post excess return on equity.

To see this, consider a simple Gordon growth model of equity returns.[8] Net required returns on equity are r, investors receive dividends D_t and pay trading costs L_tP_t each period, and trading costs and dividends are both expected to grow at a rate g, so that dividend-price ratios and the ratio of trading cost to price (that is, L_t) are expected to remain constant. Then

$$P_t = E_t(D_{t+1} - L_{t+1}P_{t+1}) \,/\, (r - g) \tag{6}$$

$$\quad = (1 + g)(D_t - L_tP_t) \,/\, (r - g) \tag{7}$$

Rearranging yields:

$$D_t \,/\, P_t = L_t + (r - g) \,/\, (1 + g) \tag{8}$$

That is, if net required returns and expected dividend growth rates do not change, then the last term is constant, and dividend-price ratios move *one-for-one* with the incidence of trading costs. For example, in 1900, dividend yields were 4.37%, and I estimate that aggregate trading costs were 1.10% of total stock market wealth. In 1998, I estimate aggregate trading costs of 0.13%, which implies a decline in the dividend yield to 3.40%, all else equal. The actual dividend yield

[8] This does not hinge on assumptions about dividend or trading cost growth rates. A similar result can be derived using arbitrary growth rates in the more general log-linear framework of Campbell and Shiller (1988).

in 1998 was 1.36%, so multiples have expanded even further, but the point is that some of the increase in price-dividend ratios can be attributed directly to the decline in trading costs.

To put it another way, the average annual log gross excess return over commercial paper from 1901-1998 was 4.97%. Of that annualized excess return, 25 basis points can be attributed to the decline in proportional trading costs if the decline was not anticipated. In that case, the estimated gross equity premium is a slightly lower 4.72%, and after subtracting off the average annualized trading costs of 39 basis points, the estimated net equity premium would be 4.33%. As before, including trading costs brings down the estimate of the equity premium based on historical data, but of course these estimates remain large relative to ex post variability in aggregate measured consumption.

Given the linear relationship implied by equation (8), it is interesting to observe that the dividend yield and average bid-ask spread are fairly highly correlated, with a correlation coefficient of 0.634 (see Table 1). Figure 8 overlays the two time series on top of each other. While there appears to be considerable noise in both measures during the early part of the century, the two series coincide quite closely during the last quarter of the sample period. In particular, both dividend yields and bid-ask spreads have fallen substantially, especially during the last 20 years. While not a formal test, this provides further support for the hypothesis that liquidity and stock price levels are closely related.

However, an annual time-series regression of dividend yields on contemporaneous proportional bid-ask spreads yields a slope coefficient of 4.897. That is, small changes in bid-ask spreads imply big changes in dividend-price ratios. This suggests that bid-ask spreads and other transaction cost variables are broader measures of a time-varying, priced factor called liquidity (for lack of a better term). The next section explores this possibility further.

4. Bid-ask spread and turnover as predictive variables

The calibration exercise of the previous section approaches transaction costs rather mechanistically. It is certainly true that the size of the bid-ask spread and the fraction of shares traded in a period are important determinants of the aggregate transaction costs paid. But the evidence at the end of the last section suggests that prices are much more sensitive to changes in trading cost variables than a literal transaction cost model would imply. Therefore, in this section, I consider the possibility that the bid-ask spread, commissions and/or the volume of trade proxy for a priced liquidity factor.

Why might liquidity be priced? There are a number of possible explanations. For example, as noted by Glosten and Milgrom (1985), bid-ask spreads may reflect the degree of information asymmetry. If the marginal investor is uninformed, she may demand higher rates of return when the adverse selection problem is more severe. Empirical tests of this hypothesis have focused on expected returns in the cross-section. Amihud and Mendelson (1986) and Brennan and Subrahmanyam (1996) find cross-sectional evidence that is consistent with this hypothesis. Wider bid-ask spreads are associated with higher expected returns.

There could be a similar relationship in the time series of equity returns. If adverse selection varies over time and expected returns are increasing in the amount of adverse selection present, then bid-ask spreads should be positively associated with equity returns. The long time series of bid-ask spreads assembled here provides an opportunity to test this hypothesis.

Of course, adverse selection may not be the only mechanism linking expected returns to bid-ask spreads. For example, suppose that professional market-makers (including members on the floor and specialists in the case of the NYSE) are the main sources of bid and ask quotes. Then the overall spread level could mainly reflect the financial condition of brokerage houses and related financial institutions. Specifically, if market-makers face binding capital constraints, they are likely to quote less aggressively, and overall spreads are likely to rise. If these financial constraints appear when the overall economy is at a cyclical low, it seems natural that expected returns would also be high at that time, and spreads would act as a cyclical marker for high expected returns.

Another possibility is that investors are prone to waves of excessive optimism and pessimism. When investors are over-optimistic, perhaps they are willing to trade actively, providing liquidity to the market and reducing spreads. If investors are excessively pessimistic, they avoid holding and trading stocks, reducing liquidity and increasing spreads. Whatever the direction of excess, it is eventually reversed, and stock prices return to “normal” levels. This behavioral explanation also implies that turnover would be negatively associated with future stock returns. Mean reversion in stock prices would also be consistent with this story.

In the context of the stochastic discount factor formulation of equation (5), these explanations all imply that both the indirect and direct effects contribute to expected gross equity returns. Equivalently, variable transaction costs are correlated with some source of priced risk.

How could one determine the importance of the direct and indirect effects? The direct effect implies that average future stock returns should be increasing in the bid-ask spread at time *t*. If required net equity returns are unrelated to the level of transaction costs, then an increase in bid-ask spreads implies an increase in the wedge between gross and net required returns, and thus an increase in next period's expected gross equity return. The indirect effect goes in the same direction, as long as greater liquidity (as proxied by bid-ask spreads) is associated with lower expected returns. So it is impossible to distinguish between the pure transaction cost hypothesis and the broader liquidity story based purely on the sign of the correlation between bid-ask spreads and future returns. Note, however, that both models imply that the spread should have some forecasting ability, in contrast to the implicit random walk null that excess returns are unforecastable.

It is less clear how turnover might be related to expected returns in a rational model. There is a substantial literature that finds a positive contemporaneous association between volume and volatility (the classic survey is Karpoff (1987)). If volatility persists through time and is an undesirable, priced factor, then high turnover could be associated with high volatility and thus high expected returns. Alternatively, trading volume could simply be procyclical based on some sort of wealth effect, in which case high turnover might be associated with low expected returns. Another possibility is a behavioral story similar to that discussed above, where high trading activity is a marker for excessive investor optimism, and thus is linked to low future expected returns.

If these explanations are true, then the indirect effect is important, and in particular high turnover is associated with low future stock returns. In contrast, the direct effect goes the other direction. If only the direct effect is important, then high turnover predicts high future stock returns, since an increase in turnover results in an increased incidence of transaction costs. Thus, it might be possible to distinguish the broader liquidity hypothesis from the pure transaction cost story. Both explanations can also be tested against the null of constant gross expected excess returns.

Turning to the data, Table 1 provides some summary statistics on the relationship between asset returns and the various measures of trading costs and trading activity used in this paper. The dividend-price ratio and other variables are included for comparison to some of the other work on predictability.[9] The earnings-price ratio is included because Lamont (1998) finds that it has incremental ability to predict stock returns, with higher earnings forecasting lower stock returns, holding dividends constant. Stock returns, dividend yields, and earnings yields are all from Shiller (2000).

Panel A contains means, standard deviations, and first-order autocorrelations. Note that all variables except turnover are continuously compounded or in logs. In contrast to the arithmetic averages reported earlier, the average annual (log) equity premium over the 1900-1998 period, compared to the return on prime commercial paper, is 4.97%. The dividend-price ratio and earnings-price ratio are both fairly persistent, with first-order autocorrelations of 0.719 and 0.702, respectively. Note that these are annual observations, in contrast to the quarterly or monthly observations that are more common in the predictability literature, so it is less surprising that annual autocorrelations should be further from unity. However, the long sample period is part of the explanation as well. Over the second half of the 1900's, which corresponds more closely to most work on the predictive power of the dividend yield, the first-order autocorrelation in that variable is a much higher 0.857.

Panel B contains the contemporaneous correlation matrix between stock returns, the riskless rate, the trading cost and activity measures introduced in this paper (along with their first differences), and dividend and earnings yields. The correlations confirm the earlier graphical analysis that excess stock returns are negatively associated with changes to spreads and commission rates (ρ = –0.604 and –0.343, respectively). On the other hand, excess stock returns are positively associated with changes in turnover, with a correlation coefficient of 0.448. Commissions and turnover have a strong negative association (ρ = –0.623). As in Lamont (1998), the dividend-price ratio is closely associated with the earnings-price ratio; the correlation coefficient between the two is 0.697.

One other notable finding is that the dividend yield and average bid-ask spread are fairly highly correlated, with a correlation coefficient of 0.640. Figure 8 overlays the two time series

[9] Campbell and Shiller (1988) and Hodrick (1992) are good early examples. However, recent work, including Ang and Bekaert (2000) and Goyal and Welch (1999), among others, questions the predictive ability of dividend yield.

on top of each other. While there appears to be considerable noise in both measures during the early part of the century, the two series coincide quite closely during the last quarter of the sample period. In particular, both dividend yields and bid-ask spreads have fallen substantially, especially during the last 20 years.

If liquidity is a time-varying, priced factor, then these trading cost and activity measures should signal changes in expected equity returns. To investigate this, I estimate simple and multiple regressions using the trading variables as well as the traditional predictor variables.

There are a number of well-known econometric problems associated with inference in forecasting regressions. There are no long-horizon forecasting regressions, so there are no problems with overlapping data. But there are finite sample problems associated with autocorrelation in the predictor variable. When the forecasting variable is highly autocorrelated, and innovations to the dependent variables and independent variable are contemporaneously correlated, OLS is still consistent, and OLS standard errors are correct in large samples. However, Stambaugh (1999) shows that the small-sample distribution of the coefficient estimates may depart substantially from the T-distribution implied by the standard ideal conditions. In general, T-statistics and hence R^2 measures tend to be biased away from zero. He recommends simulations to assess the bias in various estimated moments of interest.

In this paper, statistical inference is conducted using a bootstrap variant of the simulations in Stambaugh (1999). To be more precise, assume that x_{t-1} predicts y_t, and x_t follows an AR(1) process:

$$y_t = a + bx_{t-1} + e_t$$

$$x_t = q + fx_{t-1} + u_t,$$

where e_t and u_t may be correlated. For each iteration, Stambaugh simulates from $N(0,\Sigma)$, where Σ is the estimated covariance matrix between e_t and u_t, and uses these random variables to construct x_t and y_t under the null hypothesis of $\beta = 0$. He then regresses y_t on x_{t-1}, and across all iterations conducts inference by comparing the distribution of b under the null to the original OLS estimate of b. The same general approach is used here, except that I resample with replacement from the bivariate distribution (e_t, u_t) rather than generate normal random vectors. This preserves the distributional structure of the innovations rather than imposing normality on them. Reported slope coefficients are bias-adjusted, and hypothesis tests are conducted by comparing the simulated distribution of b under the null of no predictability to the OLS estimate.

The results are reported in Table 2. On a univariate basis, both bid-ask spreads and turnover are able to predict the next year's aggregate stock return. Predicted stock returns are increasing in the prior year's bid-ask spread, which is consistent with the hypothesis that liquidity is desirable and is a priced factor in stock returns. Though there is considerable uncertainty in the estimated slope, the bias-adjusted slope coefficient of 20.957 implies that expected returns are fairly sensitive to spread levels. For example, a 10 basis point decrease in the bid-ask spread implies a decline of about 210 bps in expected returns over the following year.

During 1998, bid-ask spreads on DJIA stocks average 0.175%. At this spread level, the univariate regression forecasts expected excess returns for 1999 of –6.53%. Even without the benefit of hindsight, this is not a very reasonable forecast. Applying this regression using current spread levels is problematic because of two recent changes in the structure of the trading market. First, in June 1997 the NYSE reduced its minimum price increment from eighths ($0.125) to sixteenths ($0.0625). Spreads on NYSE stocks fell immediately and substantially as a result of the tick size reduction. For example, proportional spreads on DJIA stocks averaged 0.224% in May 1997 and 0.152% in July 1997. Second, technological changes have made it easier for investors away from the floor to submit limit orders, thereby competing with the specialist and floor-based liquidity providers. It seems likely that the result of this increased competition is a narrower spread. This narrower spread may or may not affect expected returns, however. In any case, it seems unlikely that expected excess returns on stocks are actually negative at present.

Turnover also predicts stock returns the next year, with an OLS R^2 of 4.86% and a p-value of 0.096. High turnover predicts low stock returns in the future. Again, there is considerable uncertainty in the parameter estimates, but 1998's turnover of 76% implies a more reasonable (log) excess return of 3.9% in the following year. There do not appear to be any changes to the structure of the market that might render this relationship nonstationary. The data are consistent with an exuberance explanation that results in swings in the volume of trade. The data could also be consistent with a rational model in which trading volume is procyclical, as long as the business cycle accounts for changes in expected equity returns. But the data are not consistent with the simple, mechanistic explanation that higher turnover means a greater incidence of transactions costs and thus higher required gross returns.

Putting the trading cost and activity measures together, spreads, commissions, and turnover account on an unadjusted basis for about 9% of the variance of excess returns.

Commissions do not seem to provide any incremental explanatory power. It is interesting to note, however, that consistent with recent research, neither dividend yield nor earnings yield appears to add much predictive power at an annual predictive horizon over this long sample period. Neither is there any evidence that the risk-free rate can predict excess stock returns at this horizon and over this sample period in the US, in contrast to the international evidence at shorter forecast horizons reported in Ang and Bekaert (2000).

Next I investigate whether these liquidity variables are able to forecast cross-sectional differences in equity returns. Specifically, these variables are used to predict next year's zero-investment portfolio returns in the Fama and French (1993) three-factor model. Neither the liquidity nor the traditional predictor variables have any reliable ability to predict HML, the relative book-to-market value factor, so the focus hereafter is on the size factor portfolio SMB, the return on small-cap stocks minus the return on large-cap stocks.

The same forecasting variables are used, but now the dependent variable is the annual SMB factor return. CRSP data are needed to classify firms and calculate portfolio returns, so the sample period for this exercise begins in 1927. The results are in Table 3. As before, spreads and turnover have considerable univariate explanatory power. Interestingly, commissions also reliably predict small stock excess returns. Together the three transaction cost or liquidity variables explain over 22% of the variance in next year's SMB realization. Dividend yields also have some ability to predict SMB one year ahead, with an unadjusted R^2 of 8.27%, but the liquidity variables have much more forecasting ability. Coefficients on the earnings yield and the risk-free rate are indistinguishable from zero.

Casual examination of the time series of the liquidity variables in Figures 1 and 5 suggests that there is considerably more variation in the first third of the 1900's than in the period since then. To check whether the associations found above are stationary over the sample, we next examine subsamples of the data. Because of the small amount of data (98 annual observations in the prediction regressions) and the limited power of these regressions to identify changes in expected returns, it is not practical to consider more than two subsamples. Thus, the sample is arbitrarily divided into two halves. The latter half (1950-1998) corresponds fairly closely to the sample used in a number of papers on dividend yield and excludes the period around the Great Depression. In both subsamples, excess market returns and excess small stock returns are forecast using spread, turnover, and dividend yield as predictors. Note that because

SMB data is not available before 1926, the first subperiod comprises 1900-1949 for the excess market returns regressions and 1927-1949 for the SMB regressions.

The results are in Table 4. Liquidity variables reliably forecast stock returns in both subsamples, though it is not too surprising that the associations are stronger during the first half of the century. After 1950, spreads and turnover do not reliably predict aggregate stock returns, but over the same interval they continue to reliably predict the next year's excess return on small stocks. In contrast, one cannot reject the hypothesis that, at this annual horizon, dividend yield has no forecasting power over either stock return measure in either subperiod. Tests for subsample stability never reject the hypothesis that the two subperiods have identical coefficients, so there is no evidence of a structural break or nonstationarity.

Up to now, all the tests have been conducted at an annual frequency. It is natural to ask whether the predictive relationship is driven by short-run associations. If so, spreads and turnover might be expected to do a good job of predicting returns one month or one quarter ahead, with little incremental forecasting power later on. To investigate this, I run forecast regressions at various horizons. Specifically, the same annual forecasting variables are used, but now the dependent variable is the excess stock market return over the next k months, where k = 1, 2, 3, 6, 9, or 12.

Table 5 contains the results. When k = 3, neither spreads nor turnover exhibit forecasting power. At k = 6 and k = 9, spreads predict well, but there is no reliable relationship with turnover. Thus, spreads appear to have the most predictive power two to three quarters ahead, while turnover has its biggest impact four quarters ahead.

Based on the findings from the predictive regressions, the last estimates consist of a vector autoregression using excess stock returns, bid-ask spreads, and turnover. Results are reported in Table 6. The coefficients are very similar to those from the predictive regressions. In particular, lagged stock returns do not seem to have any predictive power, indicating that spreads and turnover are capturing something more than momentum or reversals in stock prices.

5. Conclusions

This paper provides the first comprehensive look at some of the frictions faced by equity investors over the past 100 years. The main results are as follows. Bid-ask spreads on Dow Jones stocks gradually declined over the course of the century but are punctuated by sharp rises during periods of market turmoil. Proportional one-way commissions rise dramatically to a peak of nearly 1% in the late 1960's and early 1970's, and fall sharply following commission deregulation in 1975. Turnover is extremely high in the first decade of the 1900's, and plunges in the wake of the Great Depression, remaining low for several decades thereafter.

The sum of half-spreads and one-way commissions, multiplied by annual turnover, is an estimate of the annual proportional cost of aggregate equity trading. This cost drives a wedge between gross equity returns and net equity returns. This wedge can account for a small part of the observed equity premium, but suggests that the gross equity premium is perhaps 1% lower today than it was early in the 1900's. Finally, and perhaps most importantly, the paper presents evidence that these measures of liquidity – spreads and turnover – predict stock returns one year ahead. High spreads predict high stock returns; high turnover predicts low stock returns. This suggests that liquidity is an important determinant of conditional expected returns.

There are many possible directions for future work. As suggested by Chordia, Roll, and Subrahmanyam (2000), it would be interesting to explore why aggregate liquidity varies over time. Specifically, it should be possible to identify specific macroeconomic influences, including measures derived from the bond market, that correlate with the measures of liquidity introduced here. While this paper has focused on the time-series behavior of aggregate stock returns, these liquidity variables may be able to explain the cross-section of returns, and/or these liquidity variables may be related to other factors identified in the literature, including book-to-market and momentum factors.

References

Aiyagari, S. Rao and Mark Gertler (1991), "Asset returns with transactions costs and uninsured individual risk," Journal of Monetary Economics 27:311-331.

Amihud, Yakov and Haim Mendelson (1986), "Asset pricing and the bid-ask spread," Journal of Financial Economics 17(2):223-249.

Ang, Andrew and Geert Bekaert (2000), "Stock return predictability: is it there?" working paper, Columbia University.

Ang, Andrew, Geert Bekaert, and Jun Liu (2000), "Why stocks may disappoint," working paper, Columbia University.

Arnold, Tom, Philip Hersch, J. Harold Mulherin, and Jeffry Netter (1999), "Merging markets," Journal of Finance 54(3):1083-1107.

Brennan, Michael J. and Avanidhar Subrahmanyam (1996), "Market microstructure and asset pricing: on the compensation for illiquidity in stock returns," Journal of Financial Economics 41(3):441-464.

Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam (1998), "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns," Journal of Financial Economics 49(3):345-373.

Calomiris, Charles W. and Berry Wilson (1999), "Bank capital and portfolio management: the 1930's capital crunch and scramble to shed risk," NBER working paper 6649.

Campbell, John Y. and Robert J. Shiller (1988), "The dividend-price ratio and expectations of future dividends and discount factors," Review of Financial Studies 1:195-228.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam (2000), "Commonality in liquidity," Journal of Financial Economics 56(1):3-28.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam (2001), "Market liquidity and trading activity," Journal of Finance 56(2):501-530.

Conrad, Jennifer S., Kevin M. Johnson, and Sunil Wahal (1997), "Institutional trading and soft dollars," Journal of Finance 56(1):397-416.

Constantinides, George M. (1986), "Capital market equilibrium with transaction costs," Journal of Political Economy 94:842-862.

Eleswarapu, Venkat R. and Marc R. Reinganum (1993), "The seasonal behavior of the liquidity premium in asset pricing," Journal of Financial Economics 34(3):373-386.

Fama, Eugene F. (1998), “Market efficiency, long-term returns, and behavioral finance,” Journal of Financial Economics 49(3):283-306.

Fisher, Lawrence and Daniel G. Weaver (1999), "Unbiased equally weighted return indexes and new estimates of the size of short-term anomalies in the relative prices of common stocks listed on the New York and American Stock Exchanges,” working paper, Baruch College.

Fisher, Stephen J. (1994), “Asset trading, transaction costs and the equity risk premium,” Journal of Applied Econometrics 9:S71-S94.

Froot, Kenneth A, Michael Kim, and Kenneth Rogoff (1995), “The law of one price over 700 years,” NBER working paper W5132.

Glosten, Lawrence and Paul Milgrom (1985), “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders,” Journal of Financial Economics 14(1):71-100.

Goyal, Amit and Ivo Welch (1999), “Predicting the equity premium,” working paper, University of California at Los Angeles.

Hasbrouck, Joel and Duane J. Seppi (2001), “Common factors in prices, order flows, and liquidity,” Journal of Financial Economics 59(3):383-411.

He, Hua and David M. Modest (1995), “Market frictions and consumption-based asset pricing,” Journal of Political Economy 103:94-117.

Heaton, John and Deborah J. Lucas (1996), “Evaluating the effects of incomplete markets on risk sharing and asset pricing,” Journal of Political Economy 104(3):443-487.

Huberman, Gur and Dominika Halka (1999), “Systematic liquidity,” working paper, Columbia University.

Jarrell, Gregg A. (1984), “Change at the Exchange: the causes and effects of deregulation,” Journal of Law and Economics 27(2):273-312.

Jones, Charles M. and Marc L. Lipson (2001), “Sixteenths: direct evidence on institutional trading costs,” Journal of Financial Economics 59(2):253-278.

Jones, Charles M. and Paul J. Seguin (1997), “Transaction costs and price volatility: evidence from commission deregulation,” American Economic Review 87(4):728-737.

Jorion, Philippe and William N. Goetzmann (1999), “Global stock markets in the twentieth century,” Journal of Finance 54(3):953-980.

Karpoff, Jonathan M. (1987), “The relation between price changes and trading volume: a survey,” Journal of Financial and Quantitative Analysis 22(1):109-126.

Keim, Donald B. and Ananth Madhavan (1997), "Transactions costs and investment style: an inter-exchange analysis of institutional equity trades," Journal of Financial Economics 46(3):265-292.

Lamont, Owen (1998), "Earnings and expected returns," Journal of Finance 53(5):1563-1587.

Mehra, Rajnish and Edward Prescott (1985), "The equity premium: a puzzle," Journal of Monetary Economics 15:145-161.

Moskowitz, Tobias J. and Mark Grinblatt (1999), "Do industries explain momentum?" Journal of Finance 54(4):1249-1290.

New York Stock Exchange, *NYSE Fact Book*, published annually since 1956.

Odean, Terrance (1999), "Do investors trade too much?" American Economic Review 89(5):1279-1298.

Ofer, Aharon R. and Arie Melnik (1978), "Price deregulation in the brokerage industry: an empirical analysis," Bell Journal of Economics 9(2):633-641.

Schwert, G. William (1990), "Indexes of US stock prices from 1802 to 1987," Journal of Business 63(3):399-426.

Schwert, G. William and Paul J. Seguin (1993), "Securities transaction taxes: an overview of costs, benefits and unresolved questions," Financial Analysts Journal 49(5):27-35.

Securities and Exchange Commission, United States (1977), "The effect of the absence of fixed rates of commissions," 5th report to Congress, May 26, 1977.

Shiller, Robert J. (1988), *Market volatility*, MIT Press, Cambridge MA.

Shiller, Robert J. (2000), *Irrational exuberance*, Princeton University Press, Princeton NJ.

Siegel, Jeremy J. (1992a), "The equity premium: stock and bond returns since 1802," Financial Analysts Journal, January/February, 28-38,46.

Siegel, Jeremy J. (1992b), "The real rate of interest from 1800-1900: a study of the U.S. and U.K.," Journal of Monetary Economics 29:227-252.

Stambaugh, Robert F. (1999), "Predictive regressions," Journal of Financial Economics 54(3):375-421.

Stoll, Hans R. (1979), "Regulation of securities markets: an examination of the effects of increased competition," Monograph Series in Finance and Economics, 1979-2, New York University Salomon Center.

Stoll, Hans R. and Robert E. Whaley (1983), “Transaction costs and the small firm effect,” Journal of Financial Economics 12(1):57-79.

Appendix A
NYSE minimum commissions prior to May 1975 for all stocks with a share price at least $1

19th century to at least 1902

Minimum commission was 0.125% of the par value of the shares

Pre-War to May 7, 1919

Share price	Minimum commission per 100 shares
$1 to $9.875	$6.25
$10 and over	$12.50

May 8, 1919 to October 30, 1924

Share price	Minimum commission per 100 shares
$1 to $9.875	$7.50
$10 to $124.875	$15
$125 and over	$20

October 30, 1924 to January 3, 1938

Share price	Minimum commission per 100 shares
$1 to $9.875	$7.50
$10 to $99.875	$12.50 + 0.1% of amount traded, rounded down to the nearest $2.50
$100 to $199.875	$25
$200 and over	$10 + 0.1% of amount traded, rounded down to the nearest $5

January 3, 1938 to March 16, 1942

Share price	Minimum commission per 100 shares
$1 to $10	$4 + 1% of amount traded
Over $10	$13 + 0.1% of amount traded, rounded down to the nearest $1.00

March 16, 1942 to November 3, 1947

Share price	Minimum commission per 100 shares
$1 to $10	$5 + 1% of amount traded
$10 to $90	$12.75 + 0.25% of amt. traded, rounded down to the nearest $0.25
Over $90	$35

November 3, 1947 to November 9, 1953

Share price	Minimum commission per 100 shares
$1 to $10	$5 + 1% of amount traded
$10 to $40	$10 + 0.5% of amount traded
Over $40	$26 + 0.1% of amount traded (with a $50 maximum)

November 9, 1953 to May 1, 1958

Share price	Minimum commission per 100 shares
$1 to $20	$5 + 1% of amount traded
$20 to $50	$15 + 0.5% of amount traded
Over $50	$35 + 0.1% of amount traded (with a $50 maximum)

May 1, 1958 to March 30, 1959

Share price	Minimum commission per 100 shares
$1 to $4	$4 + 2% of amount traded
$4 to $22	$8 + 1% of amount traded
$22 to $50	$19 + 0.5% of amount traded
Over $50	$39 + 0.1% of amount traded (with a $75 maximum)

March 30, 1959 to December 5, 1968

Share price	Minimum commission per 100 shares
$1 to $4	$3 + 2% of amount traded (with a $6 minimum)
$4 to $24	$7 + 1% of amount traded
$24 to $50	$19 + 0.5% of amount traded
Over $50	$39 + 0.1% of amount traded (with a $75 maximum)

December 5, 1968 to April 6, 1970

On the first 1,000 shares of an order, the minimum commission remains unchanged.

Share price	Minimum commission per 100 shares on shares in excess of 1,000
$1 to $28	$4 + 0.5% of amount traded
$28 to $30	$18
$30 to $90	$3 + 0.5% of amount traded
Over $90	$39 + 0.1% of amount traded

except that the minimum commission for an order is never to exceed $100,000.

April 6, 1970 to March 24, 1972

Share price	Minimum per 100 shares on the first 1,000 shares of an order
$1 to $4	$4.50 + 3% of amount traded
$4 to $23	$10.50 + 1.5% of amount traded
$23 to $24	$22 + 1% of amount traded
$24 to $50	$34 + 0.5% of amount traded
Over $50	$54 + 0.1% of amount traded (with a $75 maximum)

On shares in excess of 1,000 per order, the minimum commission remains unchanged.
Effective April 5, 1971, the commission on the portion of an order exceeding $500,000 may be negotiated between broker and customer.

March 24, 1972 to September 25, 1973

Share price	Minimum for orders of 100 shares
$1 to $8	$6.40 + 2% of amount traded
$8 to $25	$12 + 1.3% of amount traded
Over $25	$22 + 0.9% of amount traded (with a $65 maximum)

Money involved	Minimum for multiple round-lot orders
$100 to $2,500	$12 + 1.3% of amount traded
$2,500 to $20,000	$22 + 0.9% of amount traded
$20,000 to $30,000	$82 + 0.6% of amount traded
$30,000 to $300,000	$142 + 0.4% of amount traded

On multiple round-lot orders, there is an additional charge for each round lot of 100 shares:
- $6.00 per round lot for the first to tenth round lot
- $4.00 per round lot for the eleventh round lot and above

The minimum commission on each round lot in a multiple round-lot order cannot exceed the minimum commission for a 100-share order.

Effective April 5, 1971, the commission on the portion of an order exceeding $500,000 may be negotiated between broker and customer.
Effective April 24, 1972, the commission on the portion of an order exceeding $300,000 may be negotiated between broker and customer.

September 25, 1973 to November 19, 1974

The minimum increased by 10% on orders up to $5,000 and by 15% on orders between $5,000.01 and $300,000.
Effective April 1, 1974, on orders of $2,000 or less, the commission was as mutually agreed.

November 19, 1974 to May 1, 1975

The minimum increased by an additional 8% on orders over $5,000.

Table 1
Summary statistics

All variables except turnover are in logs. The sample consists of annual observations 1900-1998. Aggregate stock returns r_{mt}, risk-free rates r_{ft}, dividend-price ratio $d_t - p_t$, and earnings-price ratio $e_t - p_t$ are all from Shiller (2000). ρ is the first-order autocorrelation.

	Mean	Std. dev.	ρ
r_{mt}	0.0960	0.1916	0.018
r_{ft}	0.0463	0.0291	0.878
$r_{mt} - r_{ft}$	0.0497	0.1959	0.028
$d_t - p_t$ (dividend yield)	0.0451	0.0150	0.722
$e_t - p_t$ (earnings yield)	0.0770	0.0269	0.699
s_t (proportional spread)	0.0065	0.0021	0.666
c_t (proportional commission)	0.0047	0.0023	0.980
V_t (turnover)	0.6048	0.6018	0.866
$L_t = V_t (½s_t + c_t)$	0.0039	0.0033	0.842

Correlations	$r_{mt} - r_{ft}$	r_{ft}	s_t	c_t	V_t
$r_{mt} - r_{ft}$	1.000				
r_{ft}	-0.220	1.000			
s_t	-0.268	-0.271	1.000		
c_t	0.005	-0.122	0.168	1.000	
V_t	0.017	0.098	-0.135	-0.623	1.000
Δs_t	-0.604	0.118	0.446	0.020	-0.117
Δc_t	-0.343	-0.185	0.404	0.165	-0.050
ΔV_t	0.448	-0.046	-0.164	0.022	0.210
$d_t - p_t$	-0.454	-0.168	0.640	-0.064	0.011
$e_t - p_t$	-0.283	0.097	0.233	0.071	-0.021

Correlations	Δs_t	Δc_t	ΔV_t	$d_t - p_t$	$e_t - p_t$
Δs_t	1.000				
Δc_t	0.353	1.000			
ΔV_t	-0.400	-0.119	1.000		
$d_t - p_t$	0.335	0.315	-0.277	1.000	
$e_t - p_t$	0.132	0.106	-0.081	0.697	1.000

Table 2
Annual excess return forecast regressions, 1900-1998

The dependent variable is the excess stock market return over the next year. Forecasting variables include proportional bid-ask spreads s_t, weighted-average commission rates c_t, turnover V_t, risk-free rates r_{ft}, the dividend-price ratio $d_t - p_t$, and the earnings-price ratio $e_t - p_t$; the last three are from Shiller (2000). Lower-case letters denote logs. Bias-adjusted slope coefficients and hypothesis tests are based on a bootstrap variant of simulations in Stambaugh (1999); p-values are in italics below the coefficient estimates and reflect one-sided tests of the univariate null that the coefficient is zero.

Int.	s_t	c_t	V_t	$d_t - p_t$	$e_t - p_t$	r_{ft}	OLS R^2
-0.102	20.957						5.55%
0.025	*0.022*						
0.010		2.329					0.88%
0.338		*0.386*					
0.090			-0.055				4.86%
0.009			*0.096*				
-0.043				1.535			2.29%
0.080				*0.139*			
-0.018					0.699		1.35%
0.176					*0.169*		
0.067						-0.576	0.32%
0.278						*0.235*	
-0.013	18.942	-13.074	-0.062				8.97%
0.365	*0.041*	*0.828*	*0.114*				
-0.032				1.558	0.299	-0.280	2.47%
0.248				*0.366*	*0.325*	*0.346*	
-0.036	22.556	-14.651	-0.067	-2.185	1.169	-0.299	9.86%
0.360	*0.061*	*0.839*	*0.118*	*0.778*	*0.183*	*0.364*	

Table 3
Annual forecast regressions of SMB, 1927-1998

The dependent variable is SMB, the excess return over the next year on small stocks vs. large stocks, as defined by Fama and French (1993). Forecasting variables include proportional bid-ask spreads s_t, weighted-average commission rates c_t, turnover V_t, risk-free rates r_{ft}, the dividend-price ratio $d_t - p_t$, and the earnings-price ratio $e_t - p_t$, the last three are from Shiller (2000). Lower-case letters denote logs. Bias-adjusted slope coefficients and hypothesis tests are based on a bootstrap variant of simulations in Stambaugh (1999); p-values are in italics below the coefficient estimates and reflect one-sided tests of the univariate null that the coefficient is zero.

Int.	s_t	c_t	V_t	$d_t - p_t$	$e_t - p_t$	r_{ft}	OLS R^2
-0.168	29.361						18.56%
0.002	*0.003*						
-0.076		16.768					8.56%
0.052		*0.069*					
0.088			-0.162				8.93%
0.066			*0.036*				
-0.096				2.776			8.27%
0.016				*0.019*			
-0.029					0.697		1.90%
0.158					*0.159*		
0.034						-0.227	0.05%
0.394						*0.373*	
-0.176	25.557	6.116	-0.039				22.39%
0.059	*0.008*	*0.305*	*0.339*				
-0.107				3.839	-0.775	0.387	9.56%
0.036				*0.032*	*0.770*	*0.272*	
-0.203	22.961	6.394	-0.089	1.533	-0.735	0.780	25.46%
0.070	*0.045*	*0.323*	*0.228*	*0.274*	*0.723*	*0.127*	

Table 4
Forecast regression subperiod analysis

A subset of the regressions in Tables 2 and 3, with the full 1900-1998 sample divided into two halves. The dependent variable is the excess stock market return or SMB over the next year. Forecasting variables include proportional bid-ask spreads s_t, turnover V_t, and the dividend-price ratio $d_t - p_t$. Lower-case letters denote logs. Bias-adjusted slope coefficients and hypothesis tests are based on a bootstrap variant of simulations in Stambaugh (1999); p-values are in italics below the coefficient estimates and reflect one-sided tests of the univariate null that the coefficient is zero.

Subperiod 1: prior to 1950							
Dep. variable: excess market return 1900-1949				Dep. variable: SMB 1927-1949			
Int.	s_t	V_t	$d_t - p_t$	Int.	s_t	V_t	$d_t - p_t$
-0.269	37.762			-0.325	44.956		
0.010	*0.009*			*0.004*	*0.006*		
0.108		-0.059		0.138		-0.175	
0.134		*0.138*		*0.119*		*0.056*	
-0.120			2.254	-0.122			2.898
0.177			*0.165*	*0.134*			*0.129*
-0.189	33.992	-0.023	-1.044	-0.244	42.643	-0.053	-1.022
0.158	*0.027*	*0.334*	*0.655*	*0.054*	*0.004*	*0.234*	*0.666*

Subperiod 2: 1950-1998							
Dependent variable: excess market return				Dependent variable: SMB			
Int.	s_t	V_t	$d_t - p_t$	Int.	s_t	V_t	$d_t - p_t$
0.009	0.877			-0.101	18.108		
0.613	*0.433*			*0.078*	*0.088*		
0.054		0.077		0.077		-0.190	
0.708		*0.762*		*0.193*		*0.020*	
-0.089			2.329	-0.082			2.459
0.133			*0.179*	*0.073*			*0.094*
-0.130	-13.236	0.055	2.679	0.037	-11.428	-0.185	2.281
0.362	*0.662*	*0.608*	*0.220*	*0.427*	*0.699*	*0.083*	*0.196*

Table 5
Univariate forecast regressions at varying horizons, 1900-1998

The dependent variable is the excess stock market return over the next *k* months. Forecasting variables include proportional bid-ask spreads s_t, turnover V_t, and the dividend-price ratio $d_t - p_t$. Lower-case letters denote logs. Intercepts are omitted. Bias adjustments and hypothesis tests are based on a bootstrap variant of simulations in Stambaugh (1999); p-values are in italics below the coefficient estimates and reflect one-sided tests of the univariate null that the coefficient is zero.

Independent variable	Bias-adjusted slope coefficients					
	$k = 1$	$k = 2$	$k = 3$	$k = 6$	$k = 9$	$k = 12$
s_t	1.486	-1.070	-3.009	12.762	19.862	20.957
	0.236	*0.633*	*0.781*	*0.032*	*0.009*	*0.022*
V_t	0.004	-0.002	-0.002	-0.015	-0.017	-0.055
	0.667	*0.434*	*0.440*	*0.272*	*0.277*	*0.096*
$d_t - p_t$	0.196	0.094	-0.500	0.057	1.688	1.535
	0.245	*0.401*	*0.829*	*0.459*	*0.064*	*0.139*
OLS R^2 with all 3 ind. variables	1.02%	0.69%	0.92%	6.51%	8.25%	8.72%

Table 6
Annual VAR of excess returns, spreads, and turnover

Annual data from 1900-1998. Excess stock returns $r_{mt} - r_{ft}$ are from Shiller (2000). Proportional bid-ask spreads s_t are scaled up by 100 relative to earlier tables and can be thought of as percentage spread. Annual share turnover is given by V_t. Bias-adjusted slope coefficients and hypothesis tests are based on a bootstrap variant of simulations in Stambaugh (1999); p-values are in italics below the coefficient estimates and reflect one-sided tests of the univariate null that the coefficient is zero.

Dep. Variable	Constant	$r_{mt} - r_{ft}$	s_t (in %)	V_t	R^2
$r_{m,t+1} - r_{f,t+1}$	-0.072	0.101	0.204	-0.052	9.43%
	0.089	*0.160*	*0.029*	*0.119*	
s_{t+1} (in %)	0.270	-0.260	0.635	0.010	47.87%
	0.000	*0.001*	*0.000*	*0.432*	
V_{t+1}	-0.120	0.246	0.264	0.915	78.79%
	0.000	*0.047*	*0.046*	*0.000*	

Figure 1. Bid-ask spreads on Dow Jones stocks
(all DJ stocks 1898-1928, DJIA stocks 1929-present)

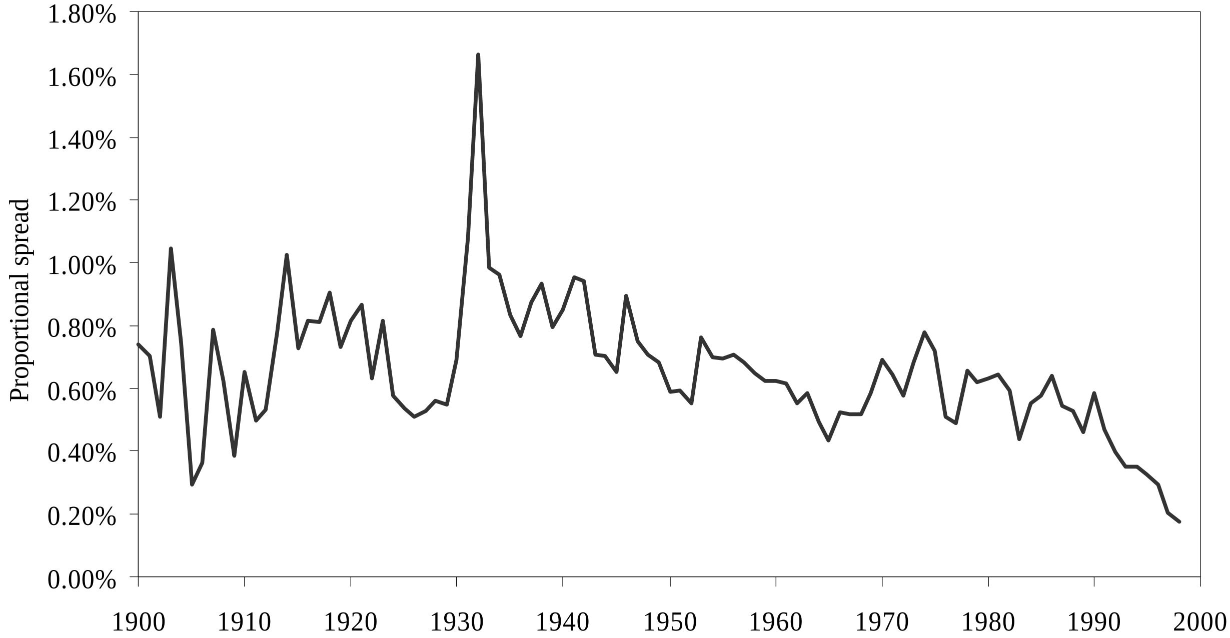

Figure 2. S&P Composite Index (12/31/1899 = 100)

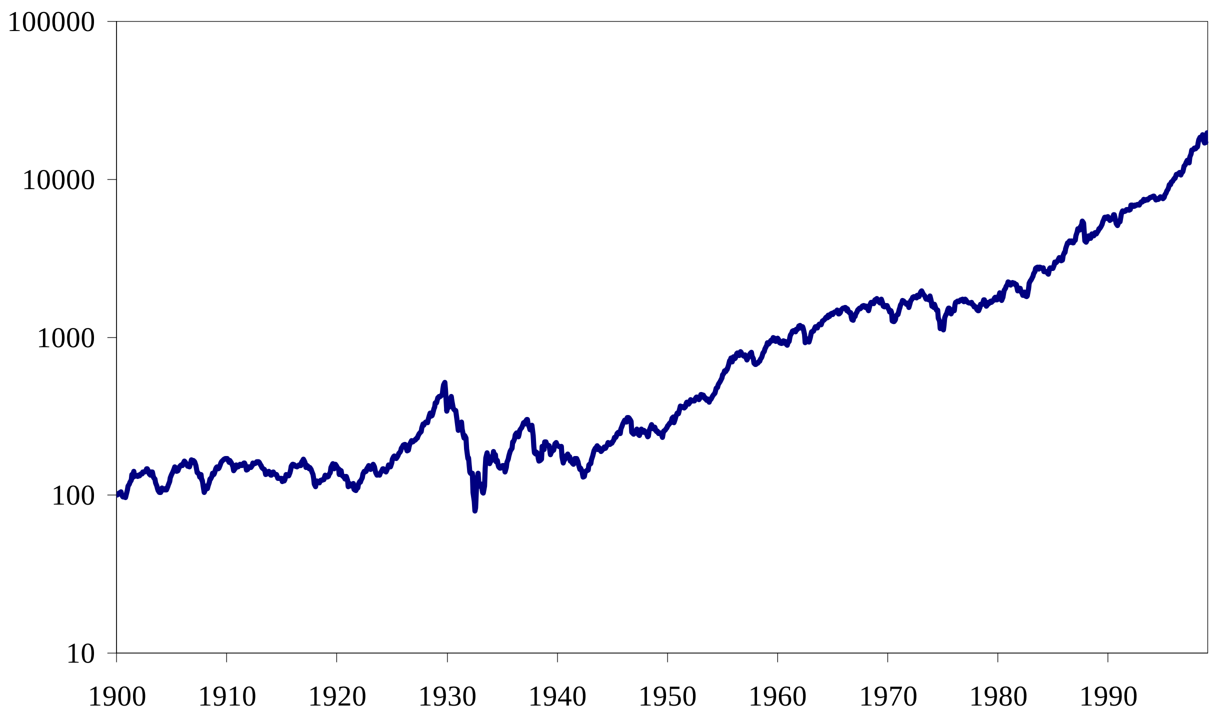

Figure 3. Average commissions on round-lot transactions in NYSE stocks
(based on fixed schedule pre-1968 and member commission revenue thereafter)

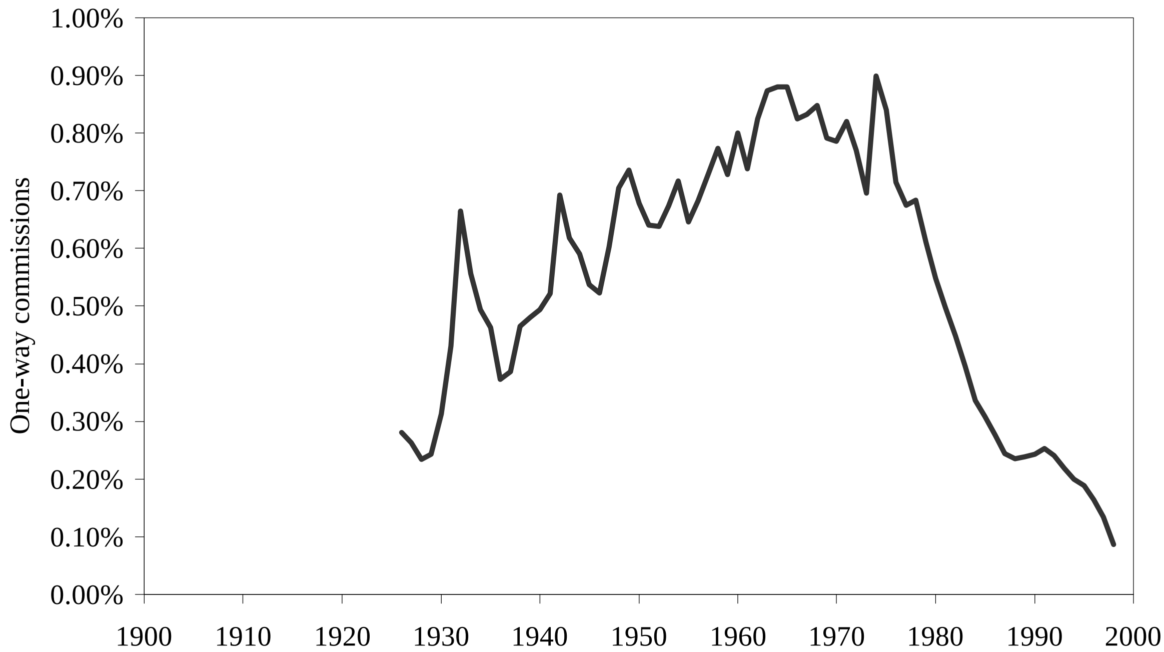

Figure 4. Average one-way transaction costs (half-spread + NYSE commission)

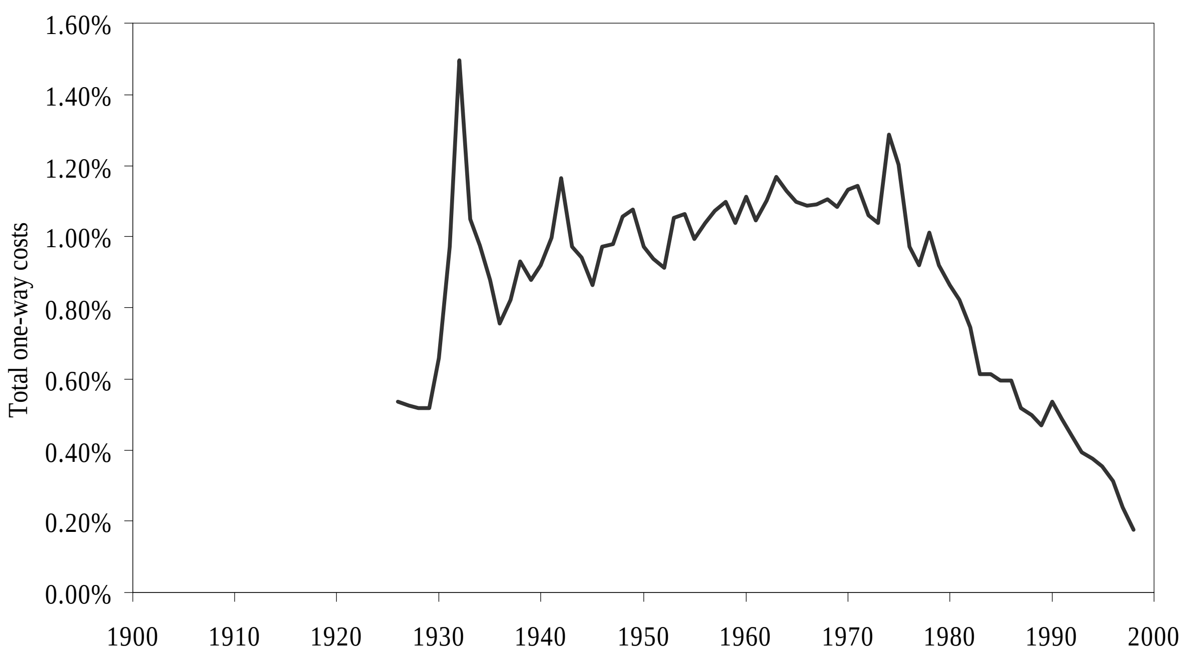

Figure 5. Annual share turnover on NYSE stocks
(Source: NYSE Fact Books)

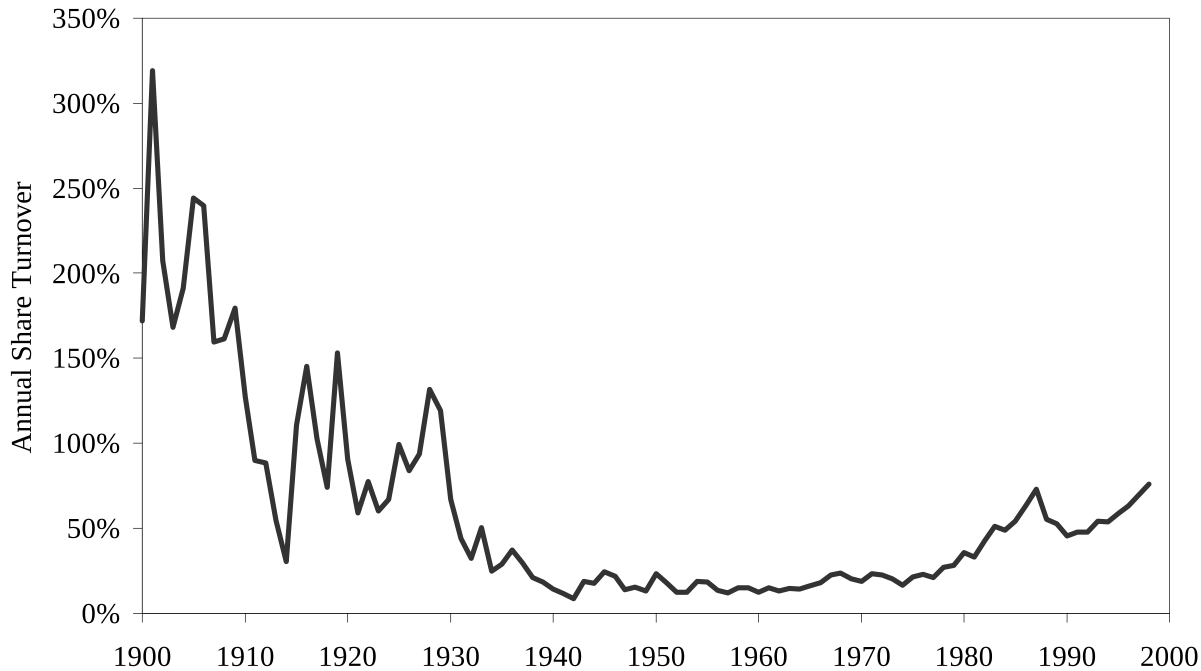

Figure 6. Estimated annualized trading costs on NYSE stocks 1900-1998
= turnover * [bid-ask half-spread + one-way commission]

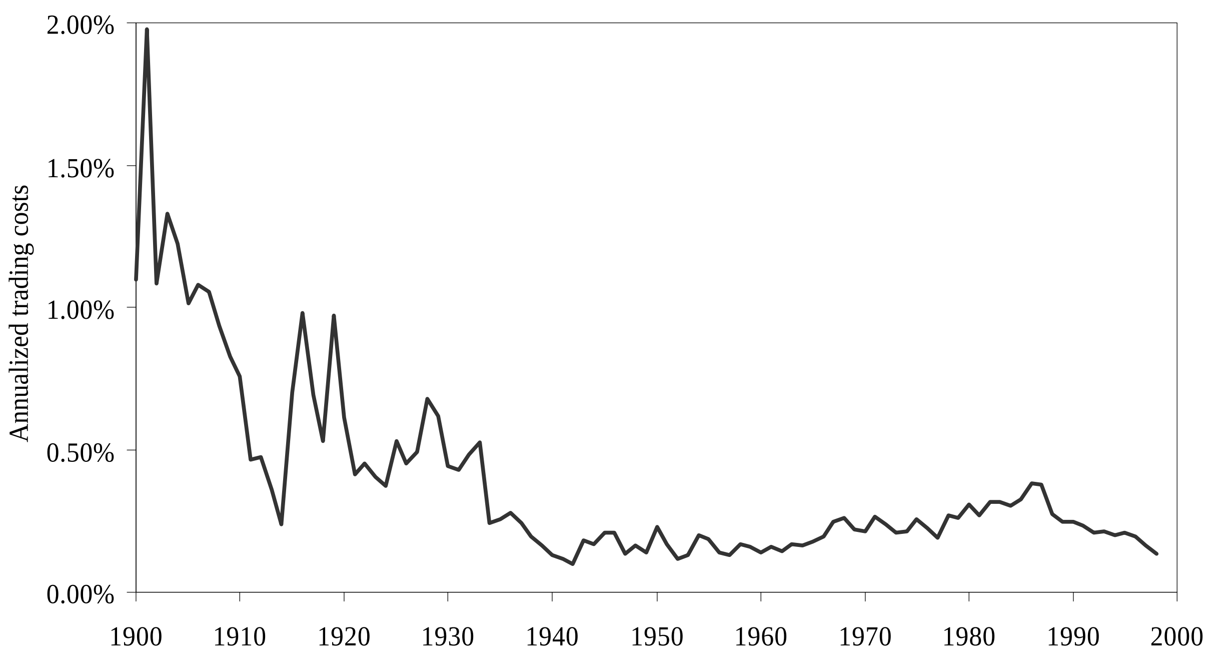

Figure 7. Gross vs. net US equity premium
(based on estimated annualized trading costs)

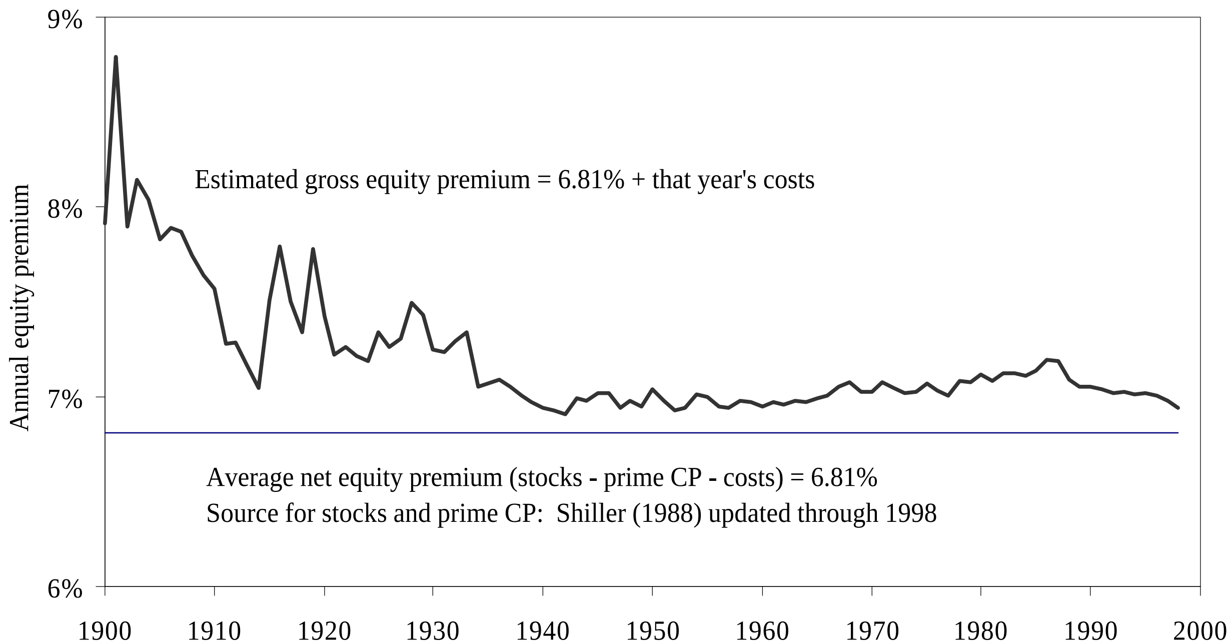

Figure 8. Average bid-ask spreads vs. dividend yield