A CENTURY OF STOCK MARKET LIQUIDITY AND TRADING COSTS

Charles M. Jones
Graduate School of Business
Columbia University

First version: May 1, 2000
This version: May 22, 2002

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 the Federal Reserve Bank of New York, New York University, 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 1900-2000, along with an annual estimate of the weighted-average commission rate for trading NYSE stocks since 1925. Spreads are cyclical, especially during periods of market turmoil. 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 the transaction cost measures that also proxy for liquidity – spreads and turnover – predict stock returns one year or more ahead. High spreads predict high stock returns; high turnover predicts low stock returns. These liquidity variables dominate traditional predictor variables, such as the dividend yield. The evidence suggests that time-series variation in aggregate liquidity is an important determinant of conditional expected stock market 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. Equilibrium 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 1900 through 2000,

(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 (2001), 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), in a related paper, 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." A number of authors have explored this question in the cross-section of stocks, including Amihud and Mendelson (1986), Easley, Hvidkjaer, and O'Hara (2002), Brennan and Subrahmanyam (1996), and Pastor and Stambaugh (2001), to name a few. This paper is concerned with the link between asset pricing and variation in aggregate liquidity, but over time rather than in the cross-section. 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, also independently suggested in Amihud (2002), 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 three years ahead. Over the entire 20th 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 shows that variation in trading costs can account for variation in price-dividend ratios, and provides suggestive evidence that trading cost variables contain information about future stock returns. 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

[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.

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 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, the New York Times, and the New York Herald, up until around 1950. Closing bid-ask data have been used in other contexts by Arnold, Hersch, Mulherin, and Netter (1999), Calomiris and Wilson (1999), and Fisher and Weaver (1999), but their existence is not well known.

I use all of these printed sources to collect monthly bid-ask spread data on a subset of stocks in the Dow Jones averages. Each monthly spread observation for each stock is entered at least twice using two different sources, one of which is always the Wall Street Journal. When the data from the initial two sources do not match, a third source is used to break the tie.

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 1900-1928 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 sources were 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, spreads on nearby days 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 original recording of bid and ask prices that might have escaped the filters.

From 1987 to 2000, I use intraday data and standard filters to calculate the time-weighted average proportional quoted spread for each DJIA stock on each trading day.[3] 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 1900-2000 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

[3] Quoted spreads are based on national best bid-offered prices and exclude all quotes that are not eligible for inclusion in the national BBO. Filters also exclude all quotes outside of regular trading hours (9:30am to 4:00pm), as well as all quotes with a dollar spread of more than $4.

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 depend on contemporaneous stock market movements. Spreads skyrocketed during the depths of the Great Depression, especially in 1932, 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). This is confirmed by simple correlations; the contemporaneous correlation between annual excess stock returns and proportional spreads is –0.236 (Table 1 Panel C).

It is not surprising that spreads rise during market downturns, since price is in the denominator of the proportional spread measure, and dollar spreads were limited to a discrete grid of eighths for most of the sample. To demonstrate this point, Figure 2 displays average dollar spreads per share on the Dow Jones sample stocks. At least for Dow stocks, dollar spreads display a steady downward trend from 1960 on. Dollar spreads are also punctuated by a number of sharp spikes, but these do not necessarily correlate with the spikes in proportional spreads. There is no obvious cyclicality in dollar spreads. Of course, proportional spreads are the relevant measure in a return measurement context or for cross-sectional comparisons, and so the rest of the paper uses proportional spreads exclusively.

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.[4]

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 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.

[4] 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

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), Ofer and Melnik (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 closely coincides with the institutional commissions reported in both Keim and Madhavan (1997) and Jones and Lipson (2001). 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 (2001) 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

---

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.

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 and the equity premium

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 0.84% over the 1925-2000 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. For now, I consider investors' trading behavior as given and measure how much of the aggregate 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. 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 to 88% in 2000, with a local maximum in 1987 (73%).

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 outside investors always trade with a specialist or other liquidity provider (such as a floor trader), and in the process all investors incur the half spread as well as one-way commissions.[5] Assume further that liquidity providers hold a net zero position in equities. Then the aggregate net equity return in any period, R_{nt}, is given by:

$$R_{nt} = R_t - V_t\,(½S_t + C_t), \tag{1}$$

where R_t is the usual (gross) aggregate equity return, V_t is the turnover per period, S_t is the proportional bid-ask spread, and C_t is the one-way proportional commission rate, both expressed as a fraction of the beginning-of-period price.

Define $K_t \equiv V_t\,(½S_t + C_t)$. This variable measures aggregate proportional losses each period due to spread and commission trading frictions (in this case, aggregate annualized trading costs).[6] Figure 6 provides a time-series of this aggregate annualized trading cost measure over the 1900-2000 period.[7] Most notable is that, since 1934, these trading costs average 0.21% annually, a figure that is almost negligible relative to aggregate investors' stock returns. In the

[5] Fisher (1994) develops an equilibrium model under essentially these assumptions. In reality, NYSE members and specialists participate as principals in approximately 20% of trading volume, and this fraction has remained remarkably constant over the sample period.

[6] 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.

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 turnover levels seen in these early years.

Table 1 Panel A shows how these costs are related to aggregate equity returns. Over the whole 1900-2000 sample period, the (arithmetic) average trading cost equals 0.38% per year. Using data from Shiller (1988) and updated through 2000 on his web page, the (arithmetic) average gross equity premium over the 1900-2000 period is 7.06%, which is slightly lower than other estimates mainly because he uses the prime commercial paper rate rather than a Treasury yield as the riskless rate. This implies that the average equity premium net of realized trading costs over the period is 6.69%.[8]

If aggregate realized trading costs vary over time but cannot be predicted and the required net equity premium remains constant at 6.69%, then the required gross equity premium is simply equal to 6.69% 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%.

Overall, 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 annual equity premium is perhaps 1% lower than it was early in the century. However, aggregate transaction costs do not suggest that the equity premium has declined substantially during the last quarter-century. I return to this analysis in the next section.

2.4 The elasticity of trading volume

As pointed out above, NYSE turnover is very low from the 1930's through the 1950's. It is interesting to note that total NYSE one-way transaction costs climb substantially during this period. In fact, over the 1926-2000 sample period where more complete transaction cost data are available, the correlation between annual NYSE turnover and one-way transaction costs is –0.80.

---

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

Though the magnitude of this correlation is substantial, the qualitative result is not surprising, since transaction costs should be of first-order importance in many trading decisions. Note that causality may also run in the other direction, since NYSE members may have responded to an exogenous decline in trading volume by hiking minimum commission rates. In fact, this was usually the tack taken by the NYSE in publicly justifying higher commission levels.

Another way to quantify the relationship here between price and quantity is to calculate the elasticity between trading volume and proportional one-way transaction costs. During the same sample period 1926-2000, the elasticity between turnover and one-way transaction costs is –1.13. This elasticity is substantial and at the high end of the range of previous estimates. For example, in evaluating the likely effect of a transaction tax, Schwert and Seguin (1993) cite several studies with elasticities ranging from –0.25 to –1.00. In short, this long time series confirms earlier findings that trading activity is quite sensitive to the level of trading costs.

3. Frameworks for analysis

It is possible to develop equilibrium models given transaction costs, which can be used to generate implications for asset price and trading behavior. Examples include Constantinides (1986), Aiyagari and Gertler (1991), and He and Modest (1995). However, these other models do not explain trading volume very well, and transaction costs typically have only a second-order effect on expected returns.

In this paper, I take a different approach. I take transaction costs and trading volume as given, and apply an accounting identity that equates prices with future discounted net cash flows. This is not an equilibrium model, but is instead designed to illuminate the empirical link between trading costs, trading volume, and returns. This may provide a foundation for future theoretical work.[9]

---

[8] Note that, even if the equity premium has remained constant over the entire century, there is still considerable statistical uncertainty in estimates of the equity premium. For the gross return data from 1900-2000, the standard error on the annual equity premium is 2.02%. See, for example, Ang, Bekaert, and Liu (2002).

[9] It is also possible to develop an equilibrium model taking trading costs and volume as given. This is the path pursued by Fisher (1994), and it implies a stochastic discount factor formulation that incorporates per-period realized trading costs K_t as an exogenous, fixed tax:

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

This can be rearranged to express the gross equity return as:

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

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

The framework employed here simply considers realized trading costs as a negative dividend. Using this notion, it is possible to extend the constant dividend growth model to account for trading costs, and this is done in Section 3.1. Section 3.2 relaxes the constant dividend growth and transaction cost assumptions using the Campbell-Shiller (1988) log-linear framework. This framework motivates the empirical work reported in Section 4.

3.1. A constant dividend growth model with trading costs

Consider first a simple variation on the constant dividend growth model. Suppose that in period *t*, each investor receives a dividend D_t and pays transaction costs κ_t, both of which are expressed in dollars. If the discount rate *r* is constant, and both dividends and transaction costs are expected to grow at a constant rate *g*, then the (ex-dividend and ex-transaction cost) security price at time *t* is given by:

$$P_t = \frac{E_t\left(D_{t+1} - \kappa_{t+1}\right)}{r-g}. \tag{2}$$

Rearranging, we can write the dividend yield as:

$$\frac{D_t}{P_t} = \frac{r-g}{1+g} + K_t, \tag{3}$$

where $K_t = \kappa_t / P_t$ is the proportional realized transaction cost. This makes clear that the dividend-price ratio and proportional transaction costs are directly related. In fact, in this model they are linearly related with a slope of one, all else equal. That is, if net required returns and expected dividend growth rates do not change, then dividend-price ratios move *one-for-one* with proportional trading costs incurred each period. For example, if there is an unanticipated permanent decrease of 100 basis points in realized annual transaction costs, and there are no related changes in *r* or *g*, then dividend yields should also fall by 100 basis points.[10]

There have been substantial declines in transaction costs over the past century. Can these declines account for the observed expansion of valuation multiples? In 1900, dividend yields were 4.37%, and I estimate that aggregate trading costs were 1.05% of total stock market wealth. In 2000, estimated aggregate trading costs are 0.17% per year. This implies a decline in the

---

returns if variation in transaction costs is related to the business cycle or is otherwise correlated with the stochastic discount factor.

[10] This result does not require constant expected returns and dividend growth. It can also be derived using the approximate log-linear model of Campbell and Shiller (1988) and Campbell (1991).

dividend yield to 3.49% if expected (net) returns and growth rates were the same at the end of both centuries. The actual dividend yield in 2000 was 1.22%, so valuation multiples have expanded even further. Nevertheless, some of the increase in price-dividend ratios can be attributed directly to the decline in trading costs.

Fama and French (2002) and others point out that, at least in the United States, part of the historical equity premium is associated with the expansion of valuation multiples. If this expansion of multiples was not anticipated, then the *ex post* equity premium is an overestimate of the *ex ante* equity premium. Based on the evidence in the previous paragraph, 88 basis points of the decline in dividend yields can be attributed to the decline in trading costs. If this decline had not occurred, then this model implies that dividend yields would have been 88 basis points higher in 2000, which implies that *prices would have been 42% lower*.

While declining trading costs appear to have large price effects, these declines have occurred over the course of 100 years, so their effect on average returns is smaller. If this century's trading cost declines had not been anticipated, the 42% price differential translates into an average (log) return differential of 53 basis points per year. That is, the estimated *ex ante* equity premium would be 53 basis points lower than *ex post* average returns. Using the geometric averages in Table 1 Panel A, the estimate of the *ex ante* equity premium becomes 4.05%, 53 basis points below the historical net equity premium of 4.58%.

Including trading cost declines in this way brings down the estimate of the equity premium based on historical data, but the reductions are small relative to the overall historical equity premium. Overall, the decline in transaction costs can directly explain only a small part of the equity premium puzzle.

In the data, there is evidence that dividend-price ratios and transaction costs are positively related, consistent with the model. Table 1 Panel C lists the correlation between the dividend-price ratio and K_t as 0.201. A linear regression of D_t / P_t on K_t gives a slope coefficient of 0.84, which is statistically indistinguishable from unity. However, shocks to K_t are not permanent. The first-order autocorrelation for K_t in Table 1 Panel B is 0.864, which is statistically different from one. The model in the next section accommodates temporary shocks to trading costs, as well as shocks that are correlated with shocks to expected returns or dividend growth.

3.2. A log-linear model with variation in trading costs

The model in the previous subsection is somewhat unsatisfactory, in that it allows only one-time, unexpected shifts in proportional transaction costs. However, the evidence in the figures indicates that aggregate transaction costs contain a cyclical component, so it might be useful to generalize the model to allow temporary changes in proportional transaction costs.

This can be accomplished using the log-linearized present value identity introduced by Campbell and Shiller (1988) and Campbell (1991). As before, in period t, each investor receives a dividend D_t and pays transaction costs κ_t, both of which are expressed in dollars. Begin with the following basic identity:

$$1 = R_{t+1}^{-1}\left(\frac{P_{t+1} + D_{t+1} - \kappa_{t+1}}{P_t}\right). \tag{4}$$

Note that in this derivation, returns R are always net returns after including the effect of trading costs. Multiply both sides by P_t / D_t, multiply and divide the right-hand side by D_{t+1}, and rearrange slightly to obtain:

$$\frac{P_t}{D_t} = R_{t+1}^{-1}\left(\frac{D_{t+1}}{D_t}\right)\left(1 + \frac{P_{t+1}}{D_{t+1}}\left[1 - \frac{\kappa_{t+1}}{P_{t+1}}\right]\right). \tag{5}$$

Take logs of both sides:

$$p_t - d_t = -r_{t+1} + \Delta d_{t+1} + \ln\left(1 + \exp\left[p_{t+1} - d_{t+1} + \ln(1 - \exp\{k_{t+1} - p_{t+1}\})\right]\right). \tag{6}$$

Assume that the log price-dividend ratio and proportional transaction costs are stationary around long-run means $p - d$ and $k - p$, respectively. Define:

$$h\left(p_{t+1} - d_{t+1}, k_{t+1} - p_{t+1}\right) = \exp\left[p_{t+1} - d_{t+1} + \ln(1 - \exp\{k_{t+1} - p_{t+1}\})\right] \tag{7}$$

and take a Taylor expansion of the last term on the RHS. The linearized log price-dividend ratio can be written as:

$$p_t - d_t \doteq -r_{t+1} + \Delta d_{t+1} + \rho\left[p_{t+1} - d_{t+1} + \delta\left(k_{t+1} - p_{t+1}\right)\right] + c\,, \tag{8}$$

where c is a constant that depends on the long-run means,

$$\rho = \bar{h}\left(1 + \bar{h}\right)^{-1}, \quad \bar{h} = h(p - d, k - p) \tag{9}$$

is a constant close to but less than one, and

$$\delta = \frac{e^{k-p}}{1 - e^{k-p}} \tag{10}$$

is a small positive constant for reasonable transaction costs. Iterate forward, take expectations, and impose a transversality condition to arrive at:

$$p_t - d_t \doteq c + E_t \sum_{j=1}^{\infty} \rho^{j-1} \left(-r_{t+j} + \Delta d_{t+j} - \rho\delta \left[k_{t+j} - p_{t+j} \right] \right). \tag{11}$$

This identity shows clearly that prices are increasing in future dividend growth, decreasing in future discount rates, and decreasing in future proportional trading costs.

As emphasized by Cochrane (2001), in a stationary world absent frictions, price-dividend ratios can only vary if expected returns change or if expected dividend growth changes. However, when transaction costs are introduced into the model, scaled prices depend on the magnitude of those costs. Since prices are reduced by the capitalized value of future transaction costs, any change in the expected path of those costs will affect dividend yields. While the earlier model could only accommodate a one-time permanent change in transaction costs, this identity is better suited for assessing the impact of temporary changes in trading costs. Permanent or temporary, the effect is in the same direction. When future costs go down, current prices must rise.

Thus, changes in transaction costs could account for some of the observed variation in dividend yields. Cochrane (2001) argues that dividend growth is nearly unforecastable, so nearly all the variation in the price-dividend ratio must be due to changes in expected returns. In a model with trading frictions, this need not be so. Both changes in expected returns and changes in expected trading costs can cause variation in price-dividend ratios. To see this more concretely, rewrite equation (11) as:

$$p_t - d_t \doteq c + E_t \left(-\eta_r + \eta_d - \eta_k \right)$$

where η_r is the discounted sum of future expected returns, η_d is the discounted sum of future dividend growth rates, and η_k is the discounted sum of future proportional transaction costs. Using this notation, the variance of dividend-price ratios can be decomposed into three variance terms and three covariance terms:

$$\operatorname{var}(p_t - d_t) = \operatorname{var}(\eta_r) + \operatorname{var}(\eta_d) + \operatorname{var}(\eta_k) - 2\operatorname{cov}(\eta_r, \eta_d) + 2\operatorname{cov}(\eta_r, \eta_k) - 2\operatorname{cov}(\eta_d, \eta_k)$$

There are two terms of interest here, var(η_k) and cov(η_r,η_k). If cov(η_r,η_k) = 0, then changes in transaction costs are simply changes in a proportional wealth tax. These costs affect prices either via cash flows or via the direct wedge between gross and net returns. Said another way, if this covariance is zero, then variation in dividend-price ratios contains no information about changes in required (net) returns. The equity premium calibration exercise in Section 2.3 implicitly makes this zero covariance assumption and also that changes in transaction costs are uncorrelated with changes in dividend growth rates, $\text{cov}(\eta_d,\eta_k) = 0$.

However, if $\text{cov}(\eta_r,\eta_k) \neq 0$, then innovations in expected future transaction costs contain information about future required returns. A nonzero covariance, combined with persistence in transaction cost shocks, would imply in general that current transaction cost levels would forecast future stock returns. Also, since K_t is composed of spreads, commissions, and turnover, it should be possible to look at these components of realized transaction costs to see if any of the components have a disproportionate impact.

The preliminary evidence suggests that the components of transaction costs may signal changes in required returns. In Table 1 Panel C, correlations between the dividend-price ratio and the components of K_t vary widely. It was mentioned earlier that the correlation between dividend yields and K_t is 0.201 over the full sample period. Dividend yields are not significantly correlated with commissions ($\rho = 0.081$), and they are not significantly correlated with turnover ($\rho = -0.001$). This is surprising, since an increase in either commissions or turnover represents an increase in the incidence of transaction costs, either of which should be associated with a decline in prices and an increase in dividend yield if the covariance terms are unimportant in the dividend-price ratio variance decomposition.

On the other hand, the dividend yield and average bid-ask spread are strongly correlated, with a correlation coefficient of 0.709. Figure 8 overlays the last 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.

However, an annual time-series regression of dividend yields on contemporaneous proportional bid-ask spreads yields a slope coefficient of 5.335. That is, small changes in bid-ask spreads imply big changes in dividend-price ratios. If all other covariance terms were zero, proportional spreads and dividend yields would move in tandem, and the slope coefficient would be one. This means that bid-ask spreads must also covary with some other element of the accounting identity – changes in expected returns, changes in expected dividends, or other parts of K_t.

While not a formal test, the variation in these component correlations suggests that there is more to the story than a simple transaction cost mechanism. The next section introduces the alternative that liquidity is a time-varying, priced factor for aggregate stock returns, and a forecasting regression framework is used to distinguish between competing explanations of the link between stock prices and the components of realized transaction costs.

4. Liquidity and time-varying expected returns

The evidence at the end of the last section suggests that scaled stock prices covary with components of realized transaction costs. These valuation multiples are much more sensitive to changes in trading cost variables than a mechanistic 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, and the discussion here is by no means exhaustive. 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 tend to 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.

Turnover is negatively correlated with spread, so explanations concerning spreads may also apply to turnover. In addition, turnover appears to be procyclical, and it may proxy separately for state variables related to the business cycle. For example, higher turnover may reflect some sort of wealth effect, or it may be the result of some individuals liquidating their publicly traded assets in order to pursue nontraded investment opportunities, such as startup ventures, which might be relatively more attractive during booms.

Another possibility is a behavioral one. Perhaps 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.

The log-linear decomposition suggests that forecasting regressions can be used to gauge the link between future stock returns and the variables introduced in this paper. Formally, there are three hypotheses of interest that can be distinguished using forecasting regressions:

The pure transaction cost hypothesis. If $\text{cov}(\eta_r,\eta_k) = 0$, then there is only a mechanical link between transaction costs and stock returns, as discussed in the previous section. Future net stock returns should be unrelated to transaction costs. In a forecasting regression using gross stock returns, K_t should carry a slope of one, and future stock returns should be increasing in all three components of K_t – spreads, commissions, and turnover – since an increase in any one of these variables increases the level of transaction costs borne by investors in the aggregate.

The priced liquidity hypothesis. If $\text{cov}(\eta_r,\eta_k) \neq 0$, then K_t and/or its components may be state variables that are correlated with expected stock returns. The components of K_t could be proxies for time-varying liquidity. If expected returns are decreasing in liquidity, more liquidity should predict low expected returns. Spreads should have a positive slope, turnover should be negatively related to future returns, and, if commission levels are unrelated to liquidity, commissions should not predict future returns in either direction.[11]

The priced liquidity hypothesis and the pure transaction cost hypothesis make the same prediction on the relationship between spreads and future returns. So it is impossible to distinguish between the two based purely on the sign of the correlation between bid-ask spreads and future returns. In contrast, turnover can distinguish the two hypotheses. The priced liquidity hypothesis implies that high turnover is associated with high liquidity, which should predict *low* future stock returns. The pure transaction cost hypothesis implies that higher turnover increases the incidence of transaction costs, and investors will require higher gross returns to compensate. Under the pure transaction cost hypothesis, high turnover should predict *high* future returns.

Random walk null. Alternatively, it may be the case that gross expected returns are constant or at least unrelated to the variables identified here. In that case, no variable should have any ability to predict future gross stock returns.

4.1 Summary statistics

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.[12] 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).

Table 1 Panel B contains means, standard deviations, and first-order autocorrelations. The dividend-price ratio and earnings-price ratio are both fairly persistent, with first-order autocorrelations of 0.840 and 0.707, respectively. Spreads and turnover exhibit similar persistence. 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 not surprising that annual autocorrelations should be fairly distant from unity. Tests fail to reject a unit root only

[11] In the stochastic discount factor formulation discussed in a previous footnote, the pure transaction cost hypothesis corresponds to the hypothesis that $\text{cov}_t(m_{t+1}, K_{t+1}) = 0$, and the priced liquidity hypothesis corresponds to a nonzero conditional correlation between m_{t+1} and K_{t+1}.

for weighted average commissions, which have an annual autocorrelation of 0.962. This makes sense, since commissions are basically characterized by a secular upward trend during the cartel years and a steady decline since commission deregulation.

Panel C contains the contemporaneous correlation matrix between stock returns, the riskless rate, the trading cost and activity measures introduced in this paper, and dividend and earnings yields. The correlations confirm the earlier graphical analysis that excess stock returns are negatively associated with contemporaneous 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.

4.2 Forecasting regressions

To distinguish between the various hypotheses, I estimate simple and multiple forecasting 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:

---

[12] Campbell and Shiller (1988) and Hodrick (1992) are good early examples. However, recent work, including Ang and Bekaert (2001) and Goyal and Welch (2002), among others, questions the predictive ability of dividend yield.

$$y_t = \alpha + \beta x_{t-1} + e_t$$

$$x_t = \theta + \phi x_{t-1} + u_t,$$

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 β under the null to the original OLS estimate of β. The same general approach is used here, except that I bootstrap by resampling 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 and intercepts are bias-adjusted, and hypothesis tests are conducted by comparing the simulated distribution of β 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 22.191 implies that expected returns are fairly sensitive to spread levels. For example, based on the point estimates, a 10 basis point decrease in the bid-ask spread implies a decline of about 222 bps in expected returns over the following year.

During 2000, bid-ask spreads on DJIA stocks average 0.181%. At this spread level, the univariate regression forecasts expected excess returns for 2001 of –10.29%. Despite the *ex post* accuracy of this forecast (the CRSP value-weighted excess return in 2001 was –15.83%), this is probably 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).[13] Spreads on NYSE stocks fell immediately and substantially as a result of the tick

[13] On January 29, 2001, after the end of the sample period for this paper, the NYSE began trading all its listed issues in decimal increments and reduced the minimum tick size from one-sixteenth to one cent per share. Penny increments caused a further decline in quoted spreads (see, for example, Bacidore, Battalio, Jennings, and Farkas (2001)), especially for small share amounts. The NYSE phased in decimalization beginning on August 28, 2000, but none of the DJIA stocks in this sample began to trade in decimals before January 29, 2001.

size reduction. For example, proportional spreads on DJIA stocks averaged 0.224% immediately before the change in May 1997 and 0.152% in July 1997, immediately after the change. Second, technological improvements 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 4.33% and a p-value of 0.089. High turnover predicts low stock returns in the future. Again, there is considerable uncertainty in the parameter estimates, but the year 2000 turnover of 88% implies a still negative (log) excess return of –1.71% 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. The pure transaction cost hypothesis is rejected in favor of the priced liquidity hypothesis.

This conclusion is confirmed by looking at the forecasting ability of K_t, the aggregate realized transaction cost measure. This variable has no ability to forecast future gross returns, and in fact the estimated coefficient is insignificantly negative, implying that higher transaction costs today are associated with lower gross stock returns in the future. But since K_t is constructed by multiplying per-trade costs by the volume of trade, the result is consistent with the other univariate results, since spreads are positively related to future returns, while turnover is negatively related to future returns.

Putting the trading cost and activity measures together, spreads, commissions, and turnover account on an unadjusted basis for just over 9% of the variance of excess returns. Commissions do not seem to provide any incremental explanatory power, which is also consistent with the priced liquidity hypothesis rather than the pure transaction cost hypothesis. It is interesting to note, however, that consistent with some other recent research, neither dividend yield nor earnings yield appears to provide reliable 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 (2001).

One might worry that spreads and/or turnover are proxying for conditional volatility, which can be related to expected returns in a variety of models. However, there is no evidence that conditional volatility predicts future stock returns in this model. A GARCH(1,1)-in-mean model, estimated on the monthly time series of stock returns beginning in 1900, generates only an insignificant relationship between conditional volatility and next-period returns. Since volatility does not predict future returns, it appears that these liquidity variables contain information orthogonal to volatility measures. The liquidity variables are not simply volatility measures renamed.

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 1926. The factor return is defined as the value-weighted return on stocks in the smallest five market cap deciles less the value-weighted return on all stocks in the largest five market cap deciles. All NYSE, AMEX, and Nasdaq stocks are included in the portfolio returns as long as data are available from CRSP, but decile breakpoints are calculated using only NYSE stocks.

Small stocks have higher betas than large stocks, so the size factor is positively correlated with the excess return on the market ($\rho = 0.445$). As a result, one might expect these liquidity variables to also forecast SMB, given their ability to forecast the excess market return. In order to account for this, I conduct two different kinds of forecasting regressions for SMB. In addition to the standard forecasting regressions, I also orthogonalize SMB by projecting the factor return on the contemporaneous excess market return.

The results of this exercise are in Table 3. Panel A contains slope coefficients and related statistics for various univariate forecasting regressions. As before, spreads and turnover individually display considerable explanatory power, with R^2's of 13.83% and 11.57%, respectively. The estimated bias-adjusted slope coefficient on spread is 25.630, which is even larger than the analogous number from Table 2. Dividend yields also have some ability to predict SMB one year ahead (p-value = 0.061), but the liquidity variables have more forecasting ability. None of the other variables exhibit reliable forecasting power. When these predictor variables are used to forecast the orthogonal component of SMB, the results are little changed. This suggests that these liquidity variables contain separate information about expected returns on small stocks.

Panel B summarizes a number of multivariate forecasting regressions. Together the three transaction cost or liquidity variables explain almost 18% of the variance in next year's SMB realization, and 13.50% of the variance in next year's SMB that is orthogonal to the excess market return. Most of this is driven by the predictive power of the bid-ask spread, which has p-values near 0.04 in every specification. Dividend yields, earnings yields, and riskless rates generally have little forecasting ability, and they are always swamped by the forecasting ability of the liquidity variables.

4.3 Subsample and robustness checks

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 (101 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 (1951-2000) 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 1900-1950 for the excess market returns regressions and 1926-1950 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 these variables continue to have some ability to predict the next year's excess return on small stocks. Dividend yield can forecast SMB in both subsamples, but its coefficient in predicting annual market returns is indistinguishable from zero. 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 high or low frequency variation in expected returns. For example, if there is short-run variation in expected returns, then 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 j months, where j extends from one month to three years. For horizons longer than one year, there is some overlap between successive observations. The bootstrapped hypothesis tests account for any autocorrelation induced by this overlap.

Table 5 contains the results. At forecast horizons of one month and one quarter, neither spreads nor turnover exhibit forecasting power. At horizons of two and three quarters ($j = 6$ and $j = 9$ months), spreads predict well, but there is no reliable relationship with turnover. At horizons over one year, spreads and turnover both have strong predictive ability, with p-values less than 0.02 in some cases. At horizons of two or three years, it is also interesting to note that both dividend yields and earnings yields exhibit similarly strong forecasting power. Thus, it appears that all of these forecasting variables – both the traditional predictors and liquidity variables – are able to pick up low-frequency variation in expected returns.

4.4 A vector autoregression

Proportional spreads have price in the denominator, and the numerator is confined to a discrete grid of eighths of a dollar for most of the sample period. If stock prices fall but the dollar spread remains unchanged, then proportional spreads must rise. Thus, one might worry that the predictive power of spreads is simply due to mean reversion in stock prices.

A vector autoregression using excess stock returns, bid-ask spreads, and turnover can address this issue. Results are reported in Table 6. Negative excess stock returns in period t do in fact cause a reliable rise in proportional spreads in period $t + 1$, though it is impossible to know based on this specification whether discreteness is the cause. So it is possible that the predictive regression is just picking up mean reversion along the lines identified by Fama and French (1988). However, the return prediction equation puts these concerns to rest. Lagged stock returns do not drive out the predictive power of spreads. This indicates that spreads in particular are capturing something more than pure momentum or reversals in stock returns.

The vector autoregression can also be used to characterize the effect of transaction cost shocks on returns over time (see also Campbell and Ammer, 1993). Figure 9 shows the impulse response functions for the response of stock returns to a one standard deviation orthogonalized shock to either spreads or turnover. A one-standard deviation shock to spreads is a change of about 14 basis points, and a positive shock to spreads raises expected returns by 3.04%, 2.24%, and 1.24% in the first three years after the shock. Thereafter, the shock has a half-life on the order of one year. A similar calculation can be performed for a turnover shock. A unit shock to turnover is 27%, and a positive shock to turnover leads to lower returns in succeeding years. Such a turnover shock has a fairly large effect on expected returns, with a one-year ahead effect of –5.64% on expected stock returns. Note, however, that in recent years turnover shocks have been much smaller than 27%, and of course smaller shocks imply smaller effects on expected returns. A turnover shock lasts longer than a shock to bid-ask spreads, which is a function of turnover's strong persistence (its autoregressive coefficient in the VAR is 0.916).

Overall, the VAR and forecasting regression evidence indicates that spreads and turnover should be thought of as liquidity variables rather than transaction cost variables. Both variables forecast returns out to three years, and both variables are able to forecast small firm returns even better than they forecast aggregate market returns. Finally, these liquidity variables operate at fairly low frequencies, with modest but persistent effects on future expected returns. Because of this persistence, liquidity can have large effects on share 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 (2001), 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.

Amihud, Yakov (2002), "Illiquidity and stock returns: cross-section and time-series effects," Journal of Financial Markets 5(1):31-56.

Ang, Andrew and Geert Bekaert (2001), "Stock return predictability: is it there?" working paper, Columbia University.

Ang, Andrew, Geert Bekaert, and Jun Liu (2002), "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.

Bacidore, Jeff, Robert Battalio, Robert Jennings, and Susan Farkas (2001), "Changes in order characteristics, displayed liquidity, and execution quality on the New York Stock Exchange around the switch to decimal pricing," New York Stock Exchange working paper 2001-02.

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. (1991), "A variance decomposition for stock returns," Economic Journal 101:157-179.

Campbell, John Y. and John Ammer (1993), "What moves the stock and bond markets? A variance decomposition for long-term asset returns," Journal of Finance 48(1):3-37.

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.

Cochrane, John (2001), *Asset Pricing*, Princeton University Press, Princeton NJ.

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.

Easley, David, Soeren Hvidkjaer, and Maureen O’Hara (2002), “Is information risk a determinant of asset returns?” Journal of Finance, forthcoming.

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.

Fama, Eugene F. and Kenneth R. French (1988), “Permanent and temporary components of stocks prices,” Journal of Political Economy 96(2):246-273.

Fama, Eugene F. and Kenneth R. French (1993), “Common risk factors in the returns on stocks and bonds”, Journal of Financial Economics 33(1), 3-56.

Fama, Eugene F. and Kenneth R. French (2002), “The equity premium,” Journal of Finance, forthcoming.

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 (2002), “Predicting the equity premium with dividend ratios,” Yale ICF Working Paper 02-04.

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.

Hodrick, Robert J. (1992), "Dividend yields and expected stock returns: alternative procedures for inference and measurement," Review of Financial Studies 5(3):357-386.

Huberman, Gur and Dominika Halka (2001), "Systematic liquidity," Journal of Financial Research 24(2):161-178.

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.

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.

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

Pastor, Lubos and Robert F. Stambaugh (2001), "Liquidity risk and expected stock returns," working paper, University of Chicago.

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

The sample consists of annual observations 1900-2001 for returns, 1900-2000 for all other variables. Lower-case letters denote logs. 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).

Panel A: Gross vs. net average equity returns

	Raw returns		Excess returns	
	Mean	Std. dev.	Mean	Std. dev.
Arithmetic averages				
Gross equity return	11.88%	19.81%	7.06%	20.34%
Aggregate transaction costs	0.38%	0.32%	0.38%	0.32%
Net equity return	11.50%	19.82%	6.69%	20.34%
Geometric averages				
Gross equity return	9.98%	19.14%	4.98%	20.54%
Aggregate transaction costs	0.39%	0.32%	0.39%	0.32%
Net equity return	9.59%	19.24%	4.58%	20.65%

Panel B: Other summary statistics

	Mean	Std. dev.	First-order autocorrelation
$r_{mt} - r_{ft}$	0.0464	0.2056	0.023
S_t (avg. proportional spread)	0.0064	0.0021	0.706
C_t (avg. one-way commission rate)	0.0046	0.0023	0.962
V_t (turnover)	0.6093	0.5967	0.881
$K_t = V_t(\frac{1}{2}S_t + C_t)$ (annual costs)	0.0038	0.0032	0.864
D_t / P_t (dividend yield)	0.0426	0.0164	0.840
E_t / P_t (earnings yield)	0.0745	0.0298	0.707
r_{ft}	0.0466	0.0289	0.877

Panel C: Contemporaneous correlations

	$r_{mt} - r_{ft}$	S_t	C_t	V_t	K_t	$d_t - p_t$	$e_t - p_t$	r_{ft}
$r_{mt} - r_{ft}$	1.000							
S_t	-0.236	1.000						
C_t	0.008	0.243	1.000					
V_t	0.016	-0.191	-0.620	1.000				
K_t	-0.033	0.079	-0.460	0.939	1.000			
$d_t - p_t$	-0.335	0.709	0.081	-0.001	0.201	1.000		
$e_t - p_t$	-0.276	0.323	0.134	-0.017	0.100	0.747	1.000	
r_{ft}	-0.226	-0.283	-0.134	0.102	0.080	-0.165	0.076	1.000

Table 2
Annual excess return forecast regressions, 1900-2001

The dependent variable is the log excess stock market return for the next calendar year. Annual forecasting variables include proportional bid-ask spreads S_t, weighted-average commission rates C_t, turnover V_t, aggregate trading costs $K_t = (½S_t + C_t)V_t$, 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 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	K_t	$d_t – p_t$	$e_t – p_t$	r_{ft}	OLS R^2
-0.143	22.191							6.08%
0.023	*0.021*							
-0.023		3.829						1.16%
0.288		*0.344*						
0.035			-0.059					4.33%
0.129			*0.089*					
0.029				-8.339				2.23%
0.168				*0.864*				
0.175					0.057			2.71%
0.149					*0.139*			
0.150						0.059		1.54%
0.164						*0.158*		
0.027							-0.645	0.36%
0.282							*0.211*	
-0.028	19.837	-13.888	-0.065					9.16%
0.367	*0.042*	*0.834*	*0.118*					
0.148					0.025	0.026	-0.212	2.84%
0.215					*0.359*	*0.398*	*0.404*	
-0.143	23.975	-13.485	-0.066		-0.099	0.086	-0.238	9.82%
0.356	*0.089*	*0.836*	*0.130*		*0.770*	*0.213*	*0.391*	

Table 3
Annual forecast regressions of SMB, 1926-2001

The dependent variable is either SMB_{t+1}, the excess return over the next year on small stocks vs. large stocks, or the part of SMB_{t+1} orthogonal to the contemporaneous excess market return $r_{m,t+1}$. Annual forecasting variables include proportional bid-ask spreads S_t, weighted-average commission rates C_t, turnover V_t, aggregate trading costs $K_t = (½S_t + C_t)V_t$, 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 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. Panel A summarizes 14 univariate regressions; multivariate forecast regressions are reported one per line in Panel B. In Panel A, OLS R^2's are below the p-values. In both panels, R^2's exclude the explanatory power of $r_{m,t+1}$.

Panel A: Slope coefficients for univariate regressions

$r_{m,t+1}$	S_t	C_t	V_t	K_t	$d_t – p_t$	$e_t – p_t$	r_{ft}
No	25.630	16.852	-0.184	-21.505	0.073	0.047	-0.384
	0.007	*0.070*	*0.030*	*0.861*	*0.061*	*0.181*	*0.286*
	13.83%	8.36%	11.57%	3.53%	5.05%	1.58%	0.27%
Yes	21.699	15.883	-0.145	-10.787	0.058	0.017	-0.121
	0.006	*0.052*	*0.033*	*0.743*	*0.068*	*0.352*	*0.438*
	10.97%	7.19%	7.31%	1.00%	2.76%	0.17%	0.02%

Panel B: Multivariate regressions

$r_{m,t+1}$	Intercept	S_t	C_t	V_t	$d_t – p_t$	$e_t – p_t$	r_{ft}	OLS R^2
No	-0.070	19.442	-1.619	-0.108				17.89%
	0.297	*0.039*	*0.545*	*0.195*				
No	0.198				0.105	-0.049	0.180	5.68%
	0.131				*0.101*	*0.715*	*0.377*	
No	-0.383	29.818	-2.100	-0.140	-0.076	0.005	0.496	20.10%
	0.130	*0.038*	*0.557*	*0.176*	*0.729*	*0.473*	*0.239*	
Yes	-0.118	16.790	5.990	-0.040				13.50%
	0.158	*0.037*	*0.307*	*0.356*				
Yes	0.157				0.137	-0.103	0.431	5.34%
	0.139				*0.028*	*0.913*	*0.214*	
Yes	-0.374	24.904	4.948	-0.080	-0.013	-0.056	0.759	17.29%
	0.103	*0.044*	*0.344*	*0.266*	*0.552*	*0.736*	*0.116*	

Table 4
Forecast regression subperiod analysis

A subset of the regressions in Tables 2 and 3, with the full sample divided into two halves. The dependent variable is the log excess stock market return or SMB over the next year. Annual 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: up to 1950							
Dep. variable: excess market return 1901-1950				Dep. variable: SMB 1926-1950			
Int.	S_t	V_t	$d_t - p_t$	Int.	S_t	V_t	$d_t - p_t$
-0.290	39.880			-0.463	58.787		
0.013	*0.009*			*0.001*	*0.001*		
0.064		-0.067		0.105		-0.240	
0.117		*0.115*		*0.046*		*0.067*	
0.457			0.154	0.755			0.258
0.124			*0.124*	*0.044*			*0.043*
-0.298	33.609	-0.025	-0.026	-0.224	45.008	-0.094	0.030
0.290	*0.050*	*0.335*	*0.556*	*0.300*	*0.007*	*0.233*	*0.384*

Subperiod 2: 1951-2000							
Dependent variable: excess market return				Dependent variable: SMB			
Int.	S_t	V_t	$d_t - p_t$	Int.	S_t	V_t	$d_t - p_t$
-0.044	5.591			-0.097	17.420		
0.274	*0.333*			*0.093*	*0.117*		
-0.003		-0.012		0.057		-0.172	
0.477		*0.271*		*0.120*		*0.049*	
0.106			0.042	0.304			0.091
0.304			*0.255*	*0.088*			*0.070*
0.294	-14.250	-0.036	0.065	0.469	-19.086	-0.188	0.090
0.317	*0.652*	*0.449*	*0.317*	*0.179*	*0.751*	*0.167*	*0.219*

Table 5
Various forecast horizons

Slope coefficients for univariate forecasting regressions over the 1900-2001 sample, where the dependent variable is the excess stock market return over the next *j* months. Forecasting variables are all measured annually, and 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 $e_t - p_t$. 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 slope coefficient is zero. R^2's are given below the p-values.

Forecast horizon	Univariate forecast variable					
in months	S_t	C_t	V_t	$d_t - p_t$	$e_t - p_t$	r_{ft}
j = 1	1.766	-1.180	0.004	0.007	0.006	0.046
	0.191	*0.652*	*0.338*	*0.235*	*0.306*	*0.636*
	0.85%	0.03%	0.20%	0.71%	0.39%	0.33%
j = 3	-0.468	-0.563	-0.004	-0.004	-0.004	0.109
	0.541	*0.525*	*0.398*	*0.556*	*0.559*	*0.661*
	0.00%	0.06%	0.17%	0.00%	0.00%	0.42%
j = 6	11.286	-3.158	-0.016	0.010	0.020	-0.071
	0.053	*0.593*	*0.269*	*0.356*	*0.291*	*0.474*
	3.64%	0.08%	0.93%	0.46%	0.46%	0.03%
j = 9	21.491	-2.118	-0.020	0.074	0.050	-0.766
	0.009	*0.542*	*0.751*	*0.053*	*0.141*	*0.101*
	8.89%	0.16%	1.08%	5.19%	1.59%	1.15%
j = 12	22.191	3.829	-0.059	0.057	0.059	-0.645
	0.021	*0.344*	*0.089*	*0.139*	*0.158*	*0.211*
	6.08%	1.16%	4.33%	2.71%	1.54%	0.36%
j = 18	28.425	4.065	-0.086	0.103	0.131	-0.515
	0.019	*0.371*	*0.064*	*0.077*	*0.046*	*0.309*
	5.74%	0.37%	4.88%	3.95%	4.00%	0.14%
j = 24	36.682	12.137	-0.124	0.173	0.174	-1.097
	0.010	*0.239*	*0.031*	*0.024*	*0.026*	*0.169*
	6.92%	1.16%	7.15%	6.63%	4.97%	0.75%
j = 36	35.772	14.149	-0.153	0.208	0.207	-1.919
	0.023	*0.240*	*0.019*	*0.018*	*0.023*	*0.069*
	4.70%	1.09%	7.53%	5.56%	4.89%	2.37%

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

Annual data from 1900-2000. 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
$r_{m,t+1} - r_{f,t+1}$	-0.101	0.096	0.196	-0.051	8.71%
	0.115	*0.169*	*0.048*	*0.134*	
S_{t+1} (in %)	-0.440	-0.230	0.698	0.010	55.06%
	0.000	*0.001*	*0.000*	*0.411*	
V_{t+1}	-0.703	0.232	0.221	0.916	78.58%
	0.000	*0.042*	*0.078*	*0.000*	

Figure 1. Bid-ask spreads on Dow Jones stocks
(all DJ stocks 1900-1928, DJIA stocks 1929-present)

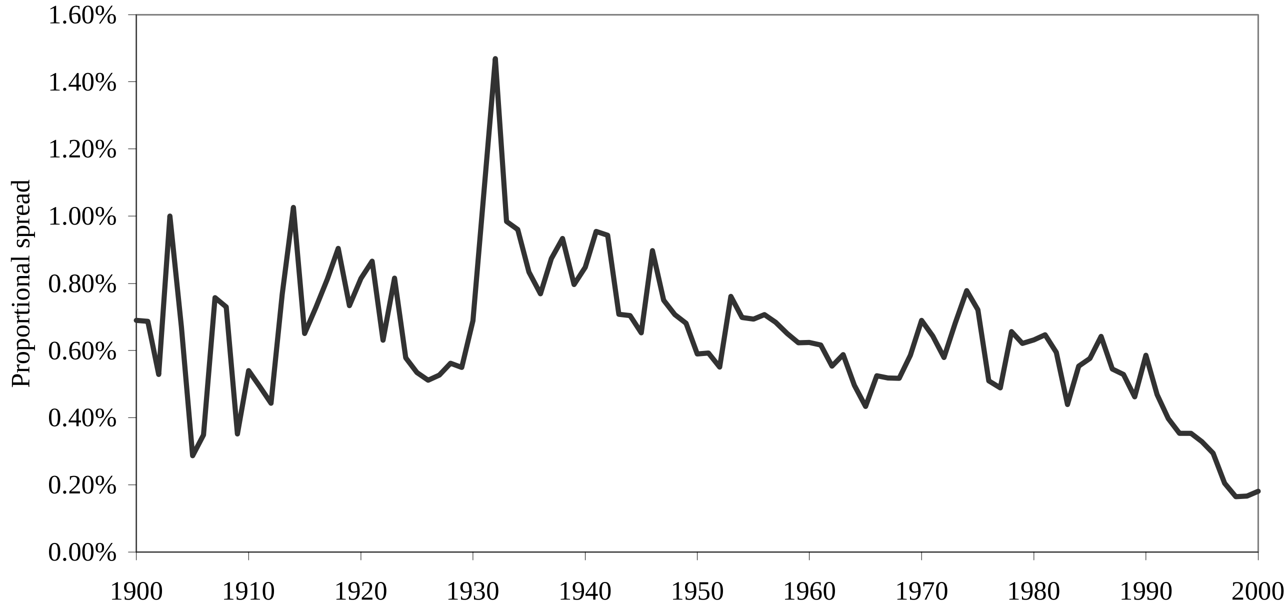

Figure 2. Average bid-ask spread (in $) on Dow Jones stocks
(all DJ stocks 1900-1928, DJIA stocks 1929-present)

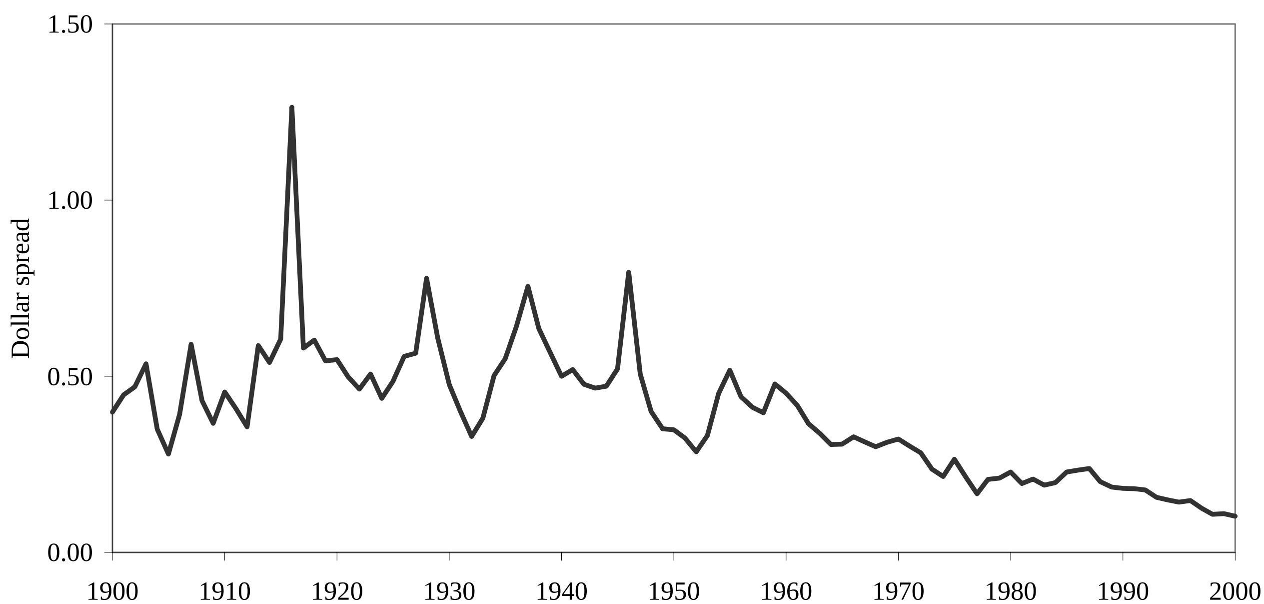

Figure 3. Average commissions on round-lot transactions in NYSE stocks
(based on fixed schedule pre-1968 and member commission revenue thereafter)

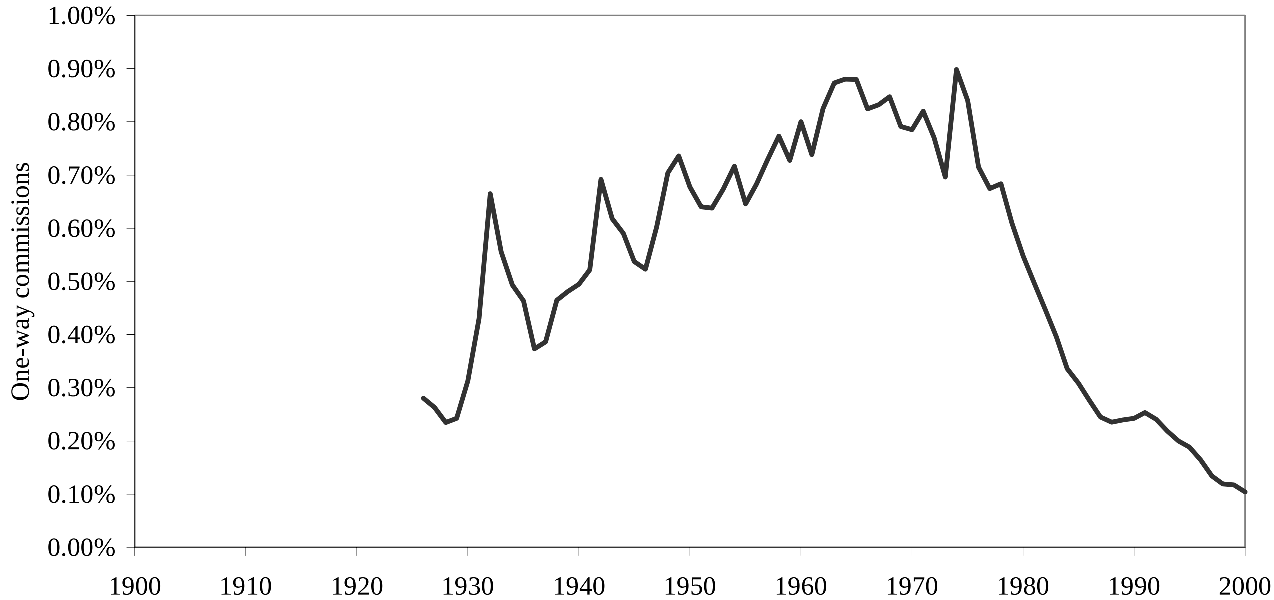

Figure 4. Average one-way transaction costs (half-spread + NYSE commission)

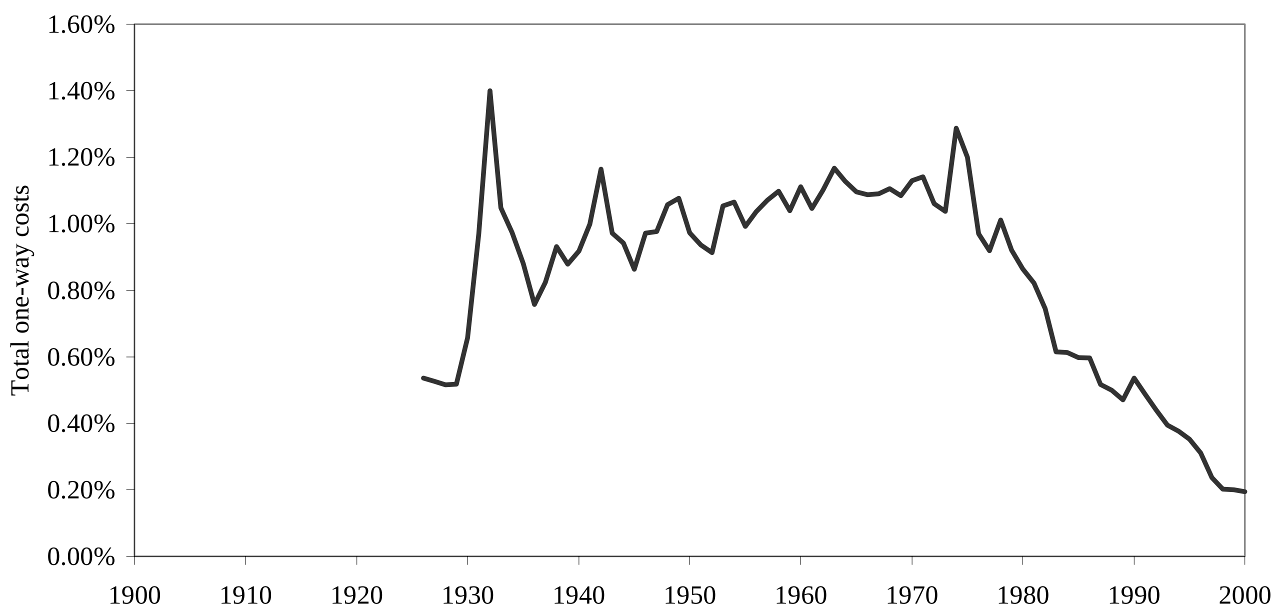

Figure 5. Annual share turnover on NYSE stocks
(Source: NYSE Fact Books)

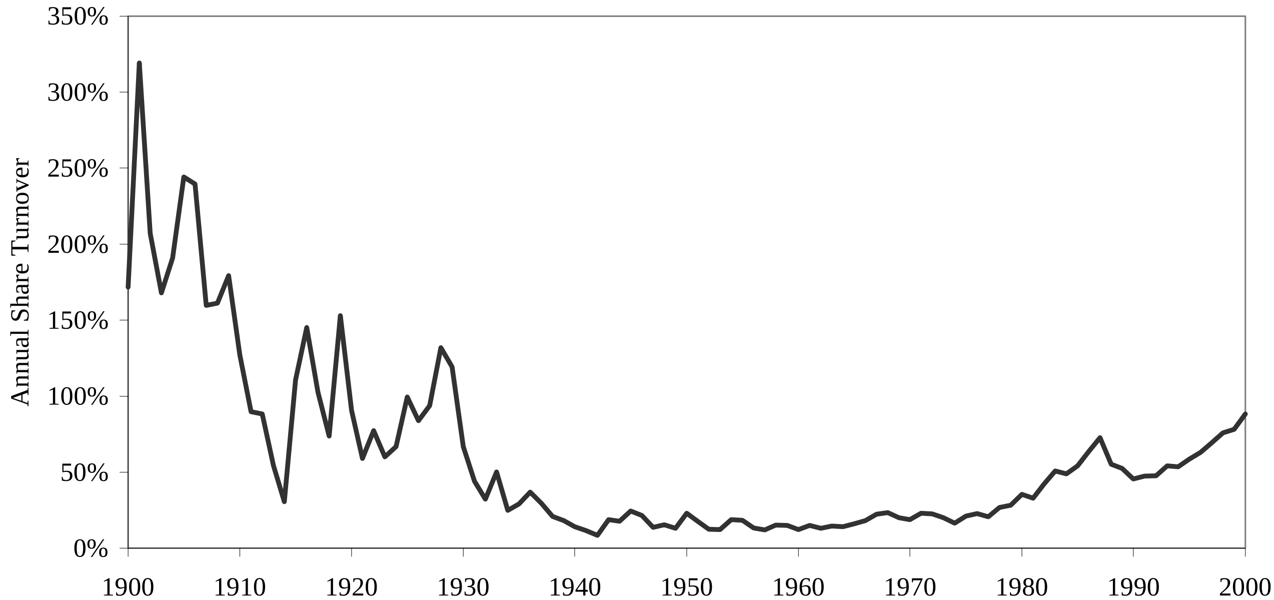

Figure 6. Estimated annualized trading costs on NYSE stocks 1900-2000
= turnover * [bid-ask half-spread + one-way commission]

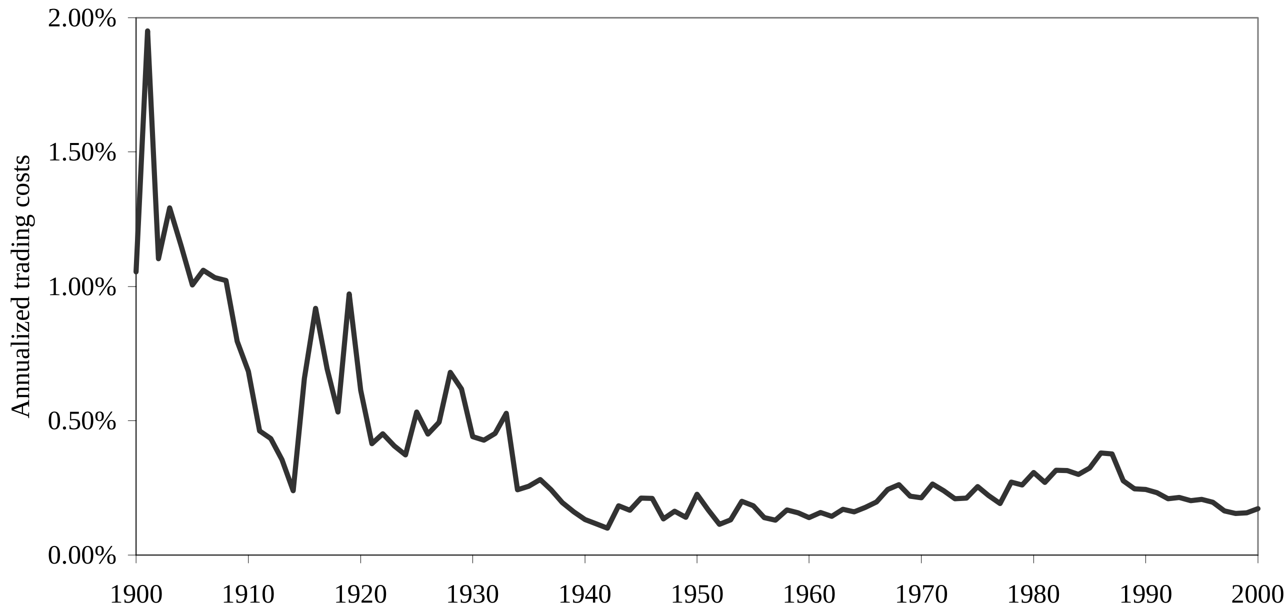

Figure 7. The declining gross equity premium
(based on estimated annualized trading costs)

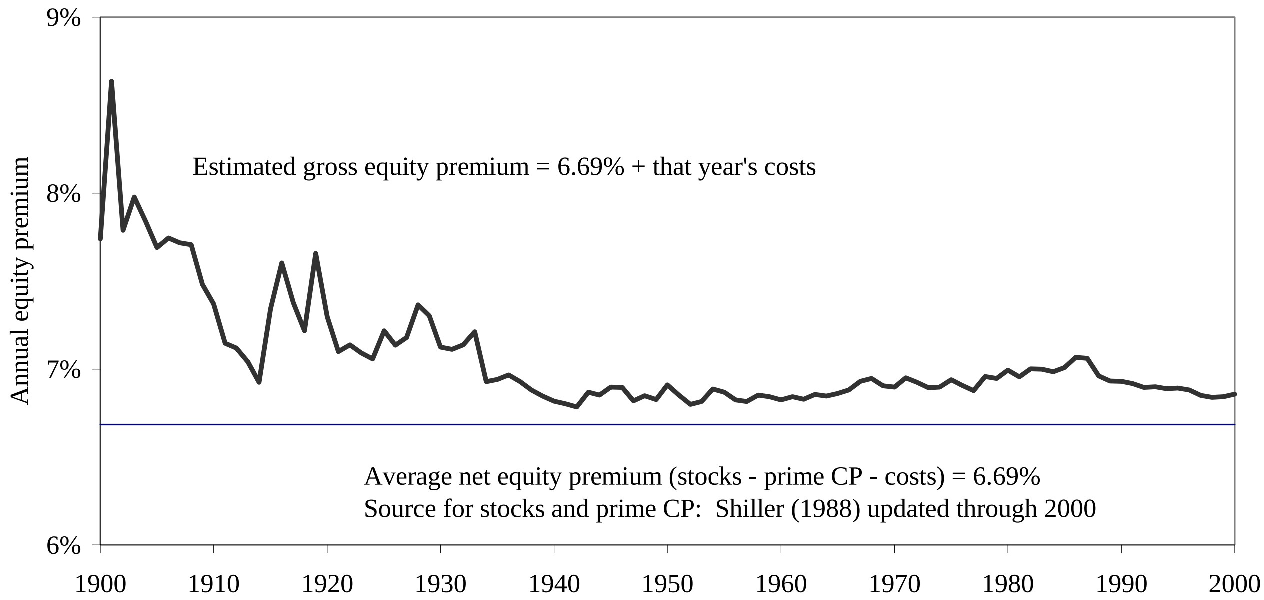

Figure 8. Average bid-ask spreads vs. dividend yield

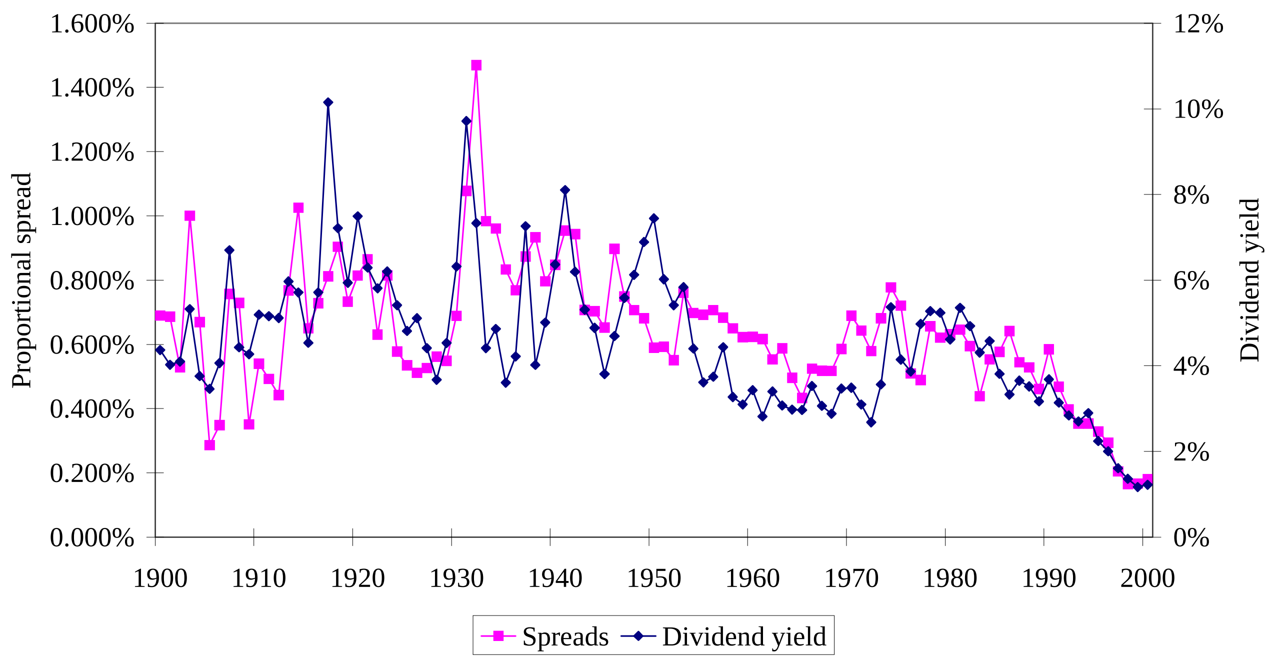

Figure 9. Response of Excess Stock Return to Unit Impulse