Applications of the High-Low Spread Estimator to Historical CRSP and Datastream Data

Shane A. Corwin and Paul Schultz
January 2010

Corwin and Schultz (2010) derive an estimator for the bid-ask spread based on daily high and low prices. To demonstrate the applicability of the high-low spread estimator, we provide two illustrative analyses. The first is an analysis of historical spreads on NYSE stocks from 1926 through 2006, based on daily CRSP data. The second is an analysis of historical spreads on two non-U.S. markets, based on daily data from Datastream.

In most cases, intraday data is preferable for calculating trading costs. However, this data is only available for recent years and selected markets. For example, TAQ data is only available from 1993 on, while the harder to use ISSM data is unavailable for NYSE or Amex stocks before 1983 and for Nasdaq stocks before 1987. Researchers who wish to examine the impact of trading costs on asset pricing need a much longer time-series of data. Likewise, researchers who examine the profitability of trading strategies or the robustness of anomalies may need trading cost estimates for years or markets for which intraday data do not exist. In these cases, as illustrated below, the high-low spread estimator should prove valuable to researchers.

1. Estimating Historical Spreads for U.S. Stocks using Daily CRSP Data

Using high and low price data from CRSP, we calculate bid-ask spreads for each NYSE stock each month from 1926 to 2006. As in the Corwin and Schultz (2010), monthly spreads are defined as the average of all two-day spreads within the calendar month, negative two-day spread estimates are set to zero, and we require a minimum of 12 daily price ranges to calculate a monthly spread. The results are illustrated in Figures 1 and 2.

Panel A of Figure 1 plots the cross-sectional average of high-low spread estimates for NYSE stocks each month from 1926 through 2006. Results are shown for the full sample of NYSE stocks and for the smallest and largest market capitalization deciles. Examining the market-wide average, we see that spreads display considerable variation over time. They were very high in the early years of the depression, with mean spreads exceeding 10% for several months in 1932 and 1933. Spreads declined in 1935 and 1936 but increased sharply as the market performed poorly in 1937 and 1938. Spreads declined steadily until the early 1950's and remained relatively low through the early 1970's. The recession of 1974-1975 is clearly visible in the figure as a period of increased spreads. Spreads are also relatively high in the early 1990's and during the tech bubble of the late 1990's. As expected, the results show that small

stocks tend to have higher execution costs than large stocks. However, the graph also illustrates that the difference between these groups is highly variable. For most months, spreads are 1% to 2% higher for small stocks than large stocks. During the depression, on the other hand, small stock spreads sometimes exceeded large stock spreads by 50%. So, at the time that spreads were 8% or 9% for large stocks, they were around 60% for small stocks. This shows that trading strategies involving small stocks were extremely expensive during the depression. It also indicates that if the returns to small stocks contain a premium to compensate for trading costs, that premium would have been especially high in the 1930's.

Panel B of Figure 1 provides a similar graph for the 1950-2006 subperiod. By omitting the depression and altering the scale of the graph, we get a clearer picture of the intertemporal variation in spreads over the last 50 years. Here, the impact of recessions and stock market declines in 1974-1975 and 1991-1992, the 1987 crash, and the "technology bubble" are clearly visible. The difference between spreads of small and large stocks was relatively large in the mid-1970's and also in the early 1990's. However, in recent years, the difference in spreads between small and large stocks has shrunk to almost nothing. Thus, while trading strategies involving small stocks may have been prohibitively expensive during the mid-1970s and early-1990s, these trading strategies may be more profitable today.

The point of this exercise is to illustrate how the high-low estimator can be used in practice. However, there is also a lesson in the analysis: trading costs prior to the early 1940s are too large to be ignored. The high-low estimator allows researchers who are studying this period to incorporate bid-ask spreads.

2. Estimating Historical Spreads for International Stocks Using Datastream Data

To demonstrate the applicability of the high-low estimator to non-U.S. markets, we estimate high-low spreads for individual stocks in Hong Kong and India using daily high and low prices from Datastream.[1] As discussed below, each of these markets provides a specific event around which we expect execution costs to change. Results for nine additional countries covered by Datastream are provided in a separate application. Again, we include only those stock-months with at least 12 daily spread observations and we set all negative estimates to zero before taking the monthly average.

Hong Kong was significantly affected by the Asian Currency Crisis beginning in October 1997, when its currency came under pressure. During this period, the equity market in Hong Kong became more

[1] Reuters provides intraday data for many international markets starting in 1996, but earlier intraday data is limited. Lesmond (2005) studies the ability of the Roll (1984), Amihud (2002), and Lesmond, Ogden, and Trzcinka (1999) measures to explain differences in bid-ask spreads within and across emerging markets.

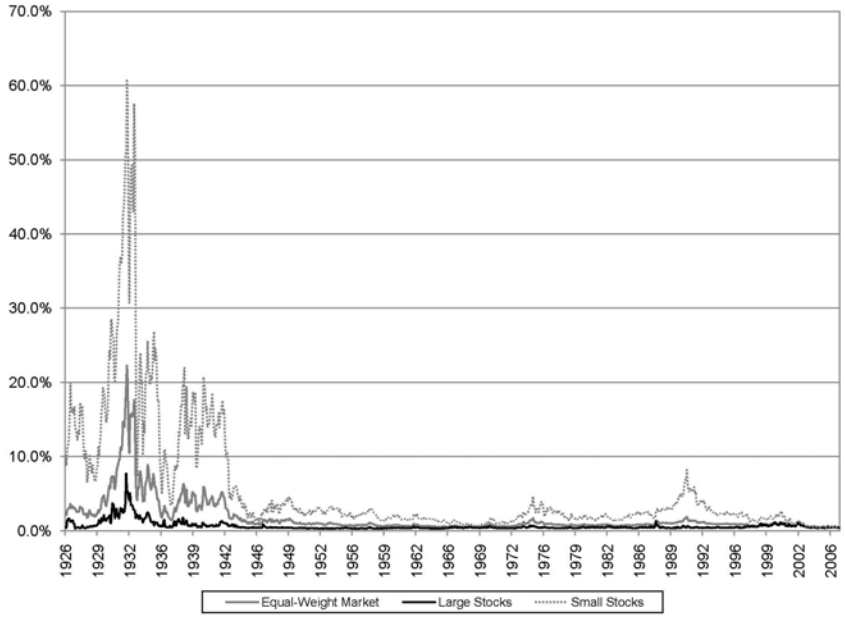
volatile, with the Hang Sang index falling 23% between October 20 and 23, 1997. We expect a significant increase in execution costs in the Hong Kong market during this period.

The cross-sectional average of high-low spread estimates for stocks in Hong Kong is plotted by month in Panel A of Figure 2. Because data coverage in Datastream increases over time, the graph also plots the number of firms used to compute the market-wide average in each month. As expected, average bid-ask spreads in Hong Kong increased sharply starting in October 1997. Average spreads increased from approximately 0.75% prior to 1997 to over 1.5% in late 1997, peaking at 2.3% in February 2000. This shift in spreads coincides with the Asian Currency Crisis and related turmoil in Hong Kong's equity markets in 1997-1998.

As of 1994, the Bombay Stock Exchange (BSE) was India's dominant market, accounting for 75% of equity volume. In November 1994, the National Stock Exchange (NSE) opened, providing Indian investors with an order-driven electronic limit order book, reduced tick sizes, satellite technology with links to sites all over India, and improved settlement and clearing standards (see Shaw and Thomas (2000)). By October 1995, NSE had surpassed the BSE, becoming the dominant equities market in India. We expect execution costs to decrease with the introduction of this new market structure.

Monthly high-low spread estimates for India are plotted in Panel B of Figure 2. Again, the graph shows the cross-sectional average across all stocks with available data in a given month, along with the number of firms used to compute the market-wide average each month. As we predicted, the average bid-ask spread across stocks in India decreased sharply in early 1995. Bid-ask spreads dropped from an average of approximately 4.5% in early 1994 to approximately 1.5% in early 1995. Spreads remain low after the introduction of the NSE, ranging from one to two percent from 1995 through 2006. This shift in spreads is consistent with the hypothesis that the change in market structure brought about by the introduction of the NSE led to a significant and permanent decrease in execution costs in India.

Panel A - Average High-Low Spreads for NYSE Stocks by Month, 1926-2006



Panel B - Average High-Low Spreads for NYSE Stocks by Month, 1950-2006

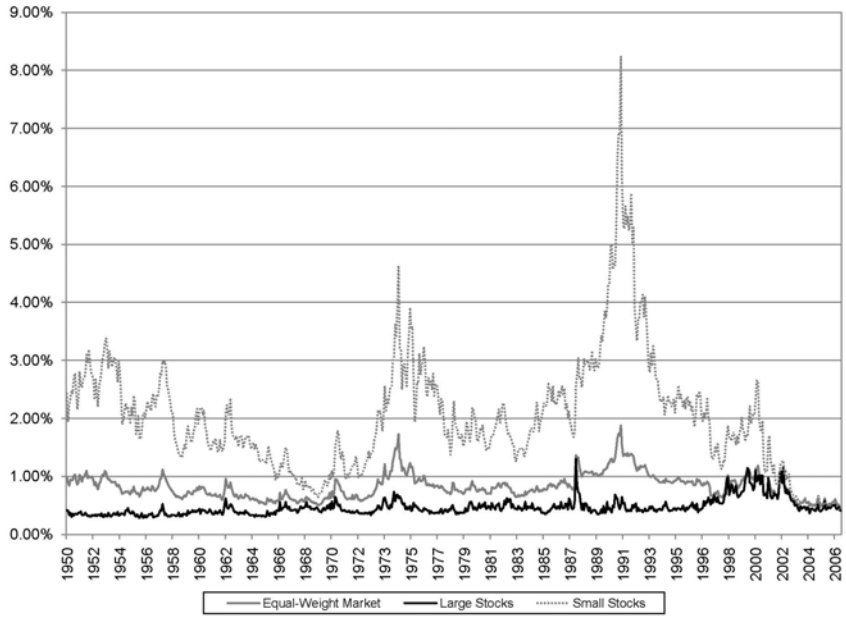
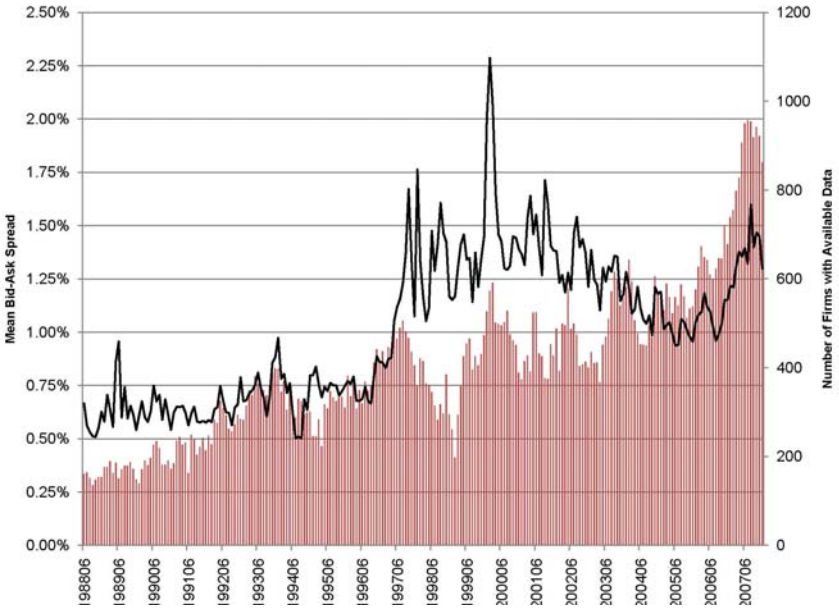


Figure 1 - Historical High-Low Spread Estimates based on CRSP Data

High-low spreads are estimated for each stock each month by averaging two-day spread estimates within the month. The graph plots the equally weighted average spread by month across all stocks with at least 12 daily spread observations within the month. Results are shown for the full sample of NYSE stocks, and for the smallest and largest deciles by market capitalization. The graph also shows the number of firms included in the average each month. Panel A shows results from 1926-2006 and while Panel B shows results from 1950-2006. All data are from CRSP.

Panel A - Average High-Low Spreads for Stocks in Hong Kong by Month, 1988-2007



Panel B - Average High-Low Spreads for Stocks in India by Month, 1990-2007

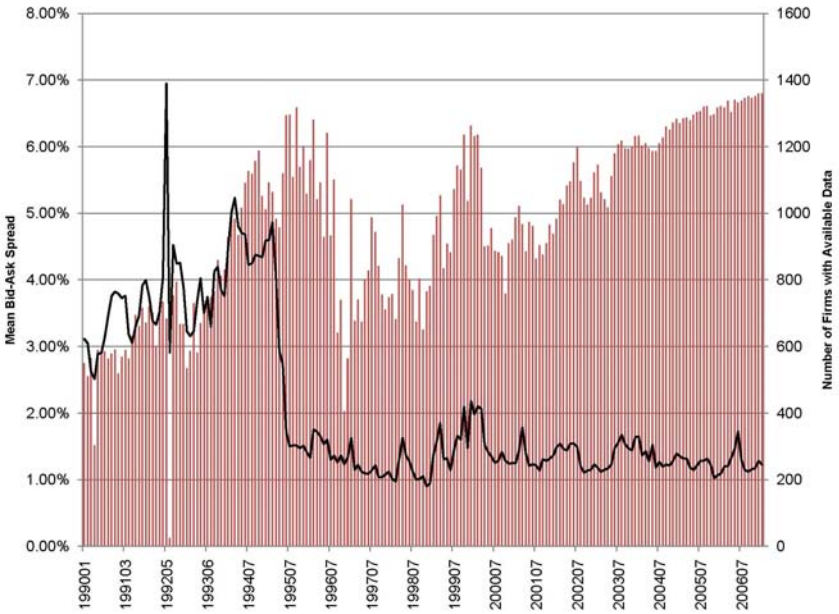


Figure 2 - Historical High-Low Spread Estimates based on Datastream Data

High-low spreads are estimated for each stock each month by averaging two-day spread estimates within the month. The graph plots the equally weighted average spread by month across all stocks with at least 12 daily spread observations within the month. The graph also shows the number of firms included in the average each month. Panel A shows results for stocks in Hong Kong and Panel B shows results for stocks in India. All data are from Datastream.

Corwin, Shane A., and Paul Schultz, 2010, A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices, working paper, University of Notre Dame.

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1106193