

Stock Price Clustering and Discreteness

Lawrence Harris

University of Southern California

Stock prices cluster on round fractions. Clustering increases with price level and volatility, and decreases with capitalization and transaction frequency. Clustering is pervasive. Price clustering will occur if traders use discrete price sets to simplify their negotiations. Exchange regulations require that most stocks be traded on eighths. Clustering on larger fractions will occur if traders choose to use discrete price sets based on quarters, halves, or whole numbers. An econometric model of clustering is derived and estimated. Projections from the results suggest that traders would frequently use odd sixteenths when trading low-price stocks, if exchange regulations permitted trading on sixteenths.

Stock prices cluster on round fractions. Integers are more common than halves; halves are more common than odd quarters; odd quarters are more common than odd eighths; other fractions are rarely observed. The phenomenon is remarkably persistent through time and across stocks.

An earlier version of this article was titled "Stock Price Clustering, Discreteness, and Bid/Ask Spreads." The author gratefully acknowledges support from the U.S. SEC and the NYSE. The comments and opinions in this article are those of the author only. In particular, the views expressed here do not necessarily reflect those of the U.S. SEC or its staff; or those of the directors, members, or officers of the NYSE. The author is grateful to seminar participants at Columbia, Hebrew, New York, Rutgers, Stanford, Tel-Aviv, and Yale Universities; the Universities of California at Berkeley, Minnesota Pennsylvania, Rochester, and Washington; and the SEC for helpful comments. The author is especially grateful to Steven Wheeler for providing the 1834 NYSE Board data, to Ken Lehn for providing the SEC data set, and to Jeremy Evnine of Wells Fargo Investment Advisers for providing market capitalization data. Address reprint requests to Lawrence Harris, Department of Finance and Business Economics, School of Business Administration, University of Southern California, Los Angeles, CA 90089-1421.

The Review of Financial Studies 1991 Volume 4, number 3, pp. 389-415
© 1991 The Review of Financial Studies 0893-9454/91/\$1.50

Price clustering occurs because traders use a discrete set of prices to specify the terms of their trades. Exchange regulations account for why fractions smaller than eighths are rarely used: Most stocks must be traded on eighths. Clustering on larger fractions can be explained if traders sometimes choose to restrict further their terms of trade to the sets of quarters, halves, or whole numbers. The use of these smaller sets may be customary or may be the result of explicit agreements among traders.

Discrete price sets are used to trade most assets. Usually, the sets are determined by custom or agreement rather than by regulation. Home prices, for example, are negotiated on 1000s or 5000s even though no authority requires the use of these discrete sets. Since negotiated home prices like \$273,221 are rare, there must be some reason why discrete prices are used.

In this article, it is assumed that traders use discrete price sets to lower the costs of negotiating. A small set limits the number of different bids and offers that can be made. Negotiations may therefore converge more rapidly since frivolous offers and counteroffers are restricted.¹ A small set also limits the amount of information that must be exchanged between negotiating traders. This reduces the time it takes to strike a bargain and it decreases the probability that two traders will believe that they have traded at different prices. These savings can be significant if trading is active.

Many exchanges require that quotes and transaction prices be stated as some multiple of a minimum price variation, or trading tick.² These regulations may simply ensure that all traders use the same discrete price set so that the above-mentioned benefits of discrete prices can be realized.³ Alternatively, exchanges may regulate minimum price variations to affect the provision of liquidity in their markets. The minimum price variation determines both the minimum quotable bid/ask spread and the maximum value of time precedence at a given price. If the bid/ask spread constraint is binding (as it usually is for low-price stocks and for actively traded stocks), dealer profits may be

¹A similar argument explains why bidden in English auctions are usually required to better the standing bid by some minimum price increment.

²The minimum price variation on the NYSE, (NYSE Rule 62) and on the AMEX (AMEX Rule 127) is $\$1/8$ for stocks priced at and above $\$1$, $\$1/16$ for stocks under $\$1$ and at or above $\$0.25$, and $\$1/32$ for stocks under $\$0.25$. The National Association of Securities Dealers permits trades on sixty-fourths for all stocks. Quotes in the NASDAQ system, however, must be a multiple of $\$1/8$ if the bid is above $\$10$ and $\$1/64$ if the bid is under $\$10$. These rules all permit exemptions. The AMEX allows foreign stocks trading under $\$5$ and some stocks formerly traded on NASDAQ to trade on sixteenths. The NYSE and the NASD have not granted any exemptions for common stocks.

³Minimum price variance regulations are probably not necessary among traders who know each other. Such traders will voluntarily use a discrete set to avoid a reputation for being difficult to negotiate with. In large public markets, public traders represented by brokers are generally anonymous. Minimum price variance regulations would then be necessary to ensure that none imposes costs on another.

supported.⁴ If the minimum price variation is small, time precedence is not valuable because dealers and other traders can cheaply acquire precedence through price priority by setting a quote or limit order with a slightly improved price.⁵

Interest in minimum price variation regulations is growing throughout the financial community. Traders are concerned about the effect of the regulations on transaction costs. They observe that a one-eighth spread on a \$2 stock is a relatively large 6.25 percent of price. Exchanges are considering smaller price variations to attract greater volume and market share. Proposals to switch to decimal pricing from fractional pricing would reopen the issue of minimum price variations. Finally, the regulation of minimum price variations must be considered whenever new contracts, trading instruments, and trading systems are introduced.

The availability of stock price data for many firms of varying characteristics provides an excellent opportunity to study discrete pricing. In this article, stock price clustering is examined to characterize the discrete price sets used by traders. The results have implications for how exchange minimum price variation regulations should be set.

The empirical analyses are based on the assumption that clustering for high-price stocks represents, at least partially, the use of discrete price sets that are coarser than the set determined by minimum price variation regulations. Under this assumption, data from high-price stocks are used to characterize cross-sectional determinants of these discrete sets. Stockprice clustering is found to increase with price level and volatility, and decrease with firm size and transaction frequency.

The discrete price sets that would be used to trade stocks if exchange discreteness regulations did not require a course grid are then estimated. The results suggest that stocks trading below \$20 would trade extensively on sixteenths, if sixteenths were permitted.

Other new results show that stock price clustering is remarkably persistent through time, across stocks, and across market structures. Clustering is found in quotes almost to the same extent as it is found in transaction prices; it is found in intradaily prices and in closing prices.

Writers of previous articles have recognized that stock price clustering is more than an amusing regularity. Osborne (1962)-who first described clustering in the academic literature-and Niederhoffer

⁴ Grossman and Miller (1988) argue that a minimum bid/ask spread may be necessary to ensure that dealers recover their fixed cost of making market.

⁵ Harris (1990) and Amihud and Mendelson (1990) discuss the importance of time precedence to traders who offer liquidity.

(1966) suggest that clustering may be inconsistent with efficient markets.⁶ Ball, Torous, and Tschoegl (1985) propose that clustering is related to price resolution. In this article, this hypothesis is further investigated. Harris (1990b) argues that clustering should be considered when analyzing the effects of price discreteness on estimators. In particular, if clustering results because traders use coarser discrete price sets than the set based on the regulated minimum price variation, the variance and serial covariance estimator biases identified in Gottlieb and Kalay (1985) and in Harris (1990b) will be more severe than previously thought. In general, all studies of empirical market microstructure must consider clustering because discreteness is a significant characteristic of transaction prices.

1. Initial Characterizations of Stock Price Clustering

Stock price clustering is pervasive. On December 31, 1987, 2431 of the 2510 closing prices reported in the CRSP Daily Stock Master Database are divisible by 1/8.⁷ Whole numbers (17.3 percent) are more common than halves (15.1), which are more common than odd quarters (12.8 and 14.1 percent) and odd eighths (10.1, 10.5, 9.5, and 10.5 percent). This frequency distribution is significantly different from the uniform distribution that would be expected if prices were randomly selected from the discrete set of eighths [$\chi^2(7) = 122.5$]. Although the last day of the year is often an unusual day for financial data, similar results can be found on virtually all other days.

Stock price clustering also appears in the earliest typeset journals of NYSE stock prices. Transaction price clustering distributions for four securities that traded on the New York Stock & Exchange Board (NYS&EB) between March 22 and April 15, 1854 (Table 1) are all qualitatively identical to CRSP clustering distributions. The highest priced security (New York Central Railroad) has the most clustering, while the lowest priced security (Parker Vein Coal Co.) has the least. There are no sixteenths in this sample.

Clustering also increases with price level in the CRSP sample. A very high-resolution characterization of the relation appears in Figure 1. This figure plots 200 price-clustering frequency distributions—one for every integer price level between 1 and 200. All CRSP daily closing stock prices divisible by 1/8 in the 25 years between January 1963 and

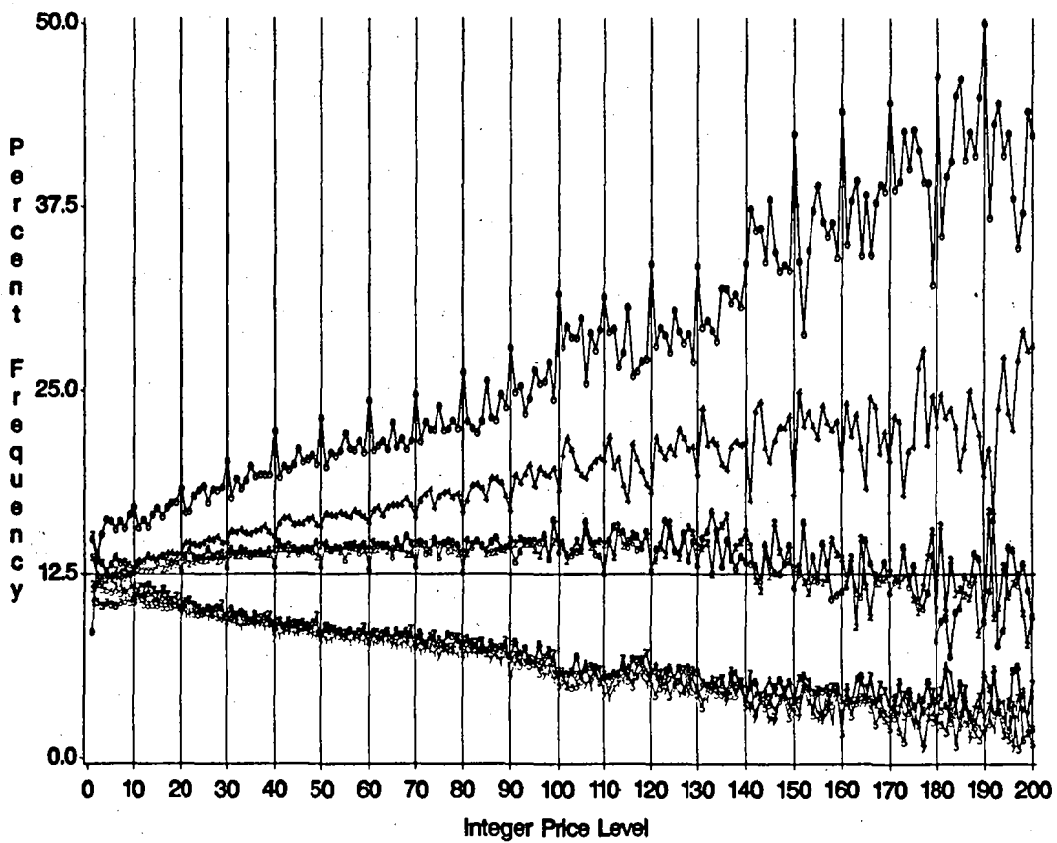
⁶See also Osborne (1962) and Niederhoffer (1965, 1966). Their concern can be summarized as follows: If stock prices reveal information about underlying values and if the fractional portions of underlying values are uniformly distributed, then the fractional portions of stock prices should be uniformly distributed.

⁷Of the remainder, 54 reported prices are the average of bid and ask quotes, 23 are closing prices reported as odd sixteenths, and 2 are closing prices reported as odd thirty-seconds.

Table 1
Historic price clustering frequency distributions for four stocks that frequently traded on the NYSE Board between March 22 and April 15, 1854

Security	Price range	N	Price fraction frequencies							
			0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1	$\frac{5}{4}$	$\frac{3}{2}$	2
Parker Vein Coal Co.	7-8	159	18.9%	11.3	16.4	6.3	13.8	13.8	11.3	8.2
Cumberland Coal Co.	26-31	235	23.4	4.3	14.9	11.1	22.1	6.8	13.6	3.8
Erie Railroad	69-80	403	20.1	6.5	21.6	7.2	18.6	7.7	13.4	5.0
New York Central Railroad	108-109	126	30.2	0.8	13.5	2.4	31.7	2.4	15.9	3.2

This transaction data sample starts with the first available NYS&EB typeset price records. These four stocks account for more than half of the trading activity of the NYS&EB at that time. Source: NYSE archives.



December 1987 are used.⁸ Clustering appears at all price levels except the lowest.⁹ As expected, clustering is greatest in the highest price levels.

These frequency distributions also provide indirect evidence of clustering on whole tens and fives. The line connecting the zero eighth frequencies across the price level clustering distributions spikes up at every five integers starting at 5. The spikes on the tens are higher than the spikes on the fives, which suggests round integer clustering. Further 'investigation of this remarkably regular, but relatively small, phenomenon is beyond the scope of this article.

The distributions plotted in Figure 1 allow us to dismiss one possible explanation for clustering. Suppose that stock prices are obtained by rounding continuously and uniformly distributed underlying stock values to the nearest eighth, as in Gottlieb and Kalay (1985). Assume, however, that "nearest" does not only depend on linear distance. Let it also depend on the attraction of round fractions. If whole numbers have the greatest attraction, followed in order by halves, odd quarters, and odd eighths, these assumptions imply stock price clustering. They also imply that the 1/8 and 7/8 price frequencies, which are adjacent to the attractive 0/8, should be less than the 3/8 and 5/8 price frequencies, which are adjacent to the less attractive 4/8. The data provide no support for these additional implications. There are no systematic frequency differences among the four odd eighths.

Since high prices reveal the most information about coarse discrete price sets, consider how price levels in this CRSP sample are distributed. The frequency distribution of the integer portion of stock prices

⁸ Subsamples spanning shorter time intervals produce similar (but noisier) results.

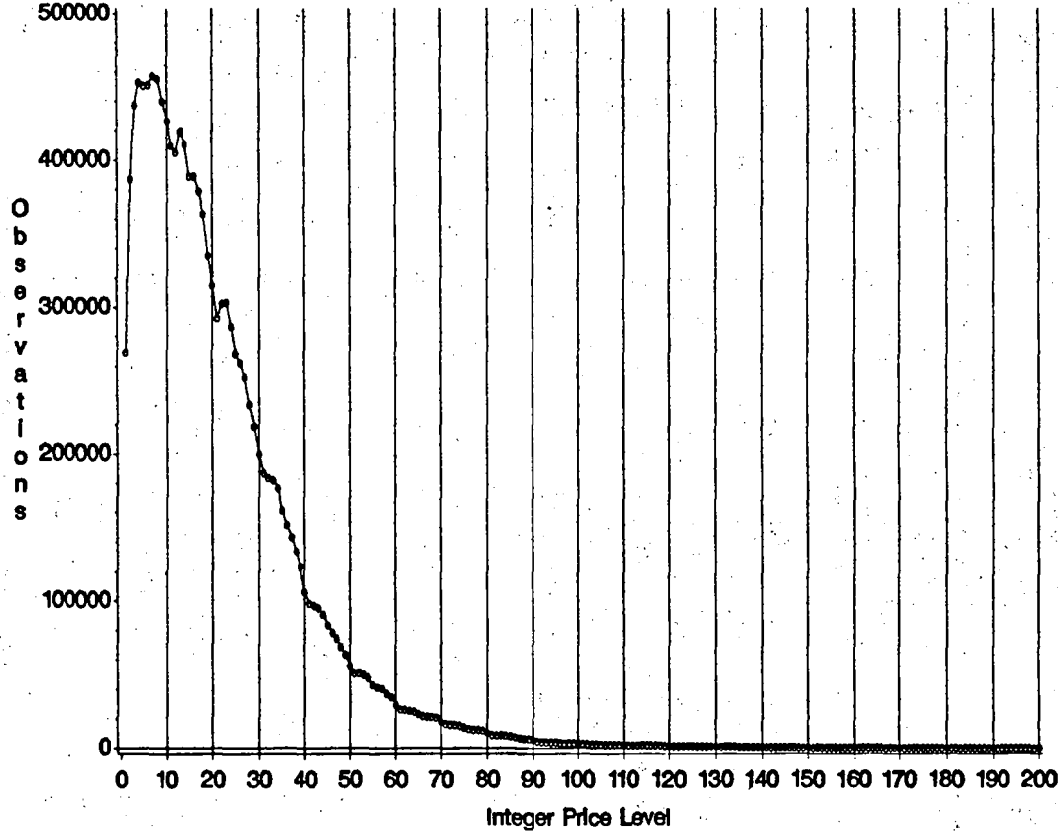
⁹ The ranking of the eighths for integer price level 1, from least common to most common, is 0,1, 3,2,4,5,6,7. This near monotonic ranking (the three-eighth frequency is only slightly greater than the two-eighth frequency) arises because fewer stocks trade near \$1 than near \$2. In this interval, the minimum price variation causes bid/ask spreads to be a very high fraction of price. Firms whose stocks trade in this interval therefore have strong incentives to reverse split. The incentive is greatest for stocks trading at the lower end of this interval. To a much lesser extent, a similar pattern can be identified for integer price level 2.

←

Figure 1

Frequency distributions of the fractional portion of CRSP daily stock prices by integer price level

Two hundred price clustering frequency distributions—one for every integer price level between 1 and 200—are plotted. Each distribution is represented by eight points (labeled 0 through 7 for each of the various eighths) located vertically over the appropriate integer price level. Each point represents the percentage of prices observed within the associated integer price level on that eighth. For example, the height of point 4 over price level 50 represents the percentage of prices observed at $50\frac{1}{2}$ in the interval from 50 to $50\frac{7}{8}$. The horizontal reference line at 12.5 percent is the frequency that would be observed for all eighths if there were no clustering. The sample includes all CRSP daily closing prices for the years 1963 to 1987 that are divisible by 1/8 and are greater than or equal to \$1. There are a total of 14,064,759 such prices, 14,912 of which are greater than or equal to \$201.



(Figure 2) shows that most prices are less than \$50 (92.4%). A significant fraction is less than \$10 (27.1%). These are stocks for which the minimum price variation is a relatively large fraction of price. Although stock prices of \$100 and greater are relatively uncommon (0.68%), this huge sample of more than 14 million observations contains 94,949 such prices. The relative paucity of data at these price levels explains why the clustering distributions in Figure 1 are noisier for high price levels than for low ones.¹⁰

An unexpected regularity appears in this price level frequency distribution. Inflections in the distribution function occur at approximately every 10 integers starting at 12 and are discernible through 80. Further investigation of this unusual characteristic is beyond the scope of this article.

Stock price clustering is present in intraday as well as closing prices. As we have already seen, clustering is present in the 1854 NYS&EB transactions data set. Using a 1982 Fitch sample, Harris (1989a) shows that the clustering in NYSE closing prices is virtually identical to clustering in earlier intraday transaction prices. The former is greater than the latter, but the differences are trivial compared to the frequency variation among the eighths.

Clustering also characterizes bid-and-ask quote distributions. Table 2 shows average frequencies at which bids, asks, and trade prices are reported on whole numbers. The whole number frequency summarizes clustering. It will be larger than 12.5 percent if whole numbers are more common than halves, if halves are more common than odd quarters, or if quarters are more common than odd eighths. All primary market trade prices and the quotes that stood prior to those trades

¹⁰Note that any nonuniform stock price distribution that is smooth over all eighth-divisible prices implies stock price clustering. Prices will cluster within any integer price level over which the slope of the density function k is not zero. For smooth price distributions, the frequency of the zero eighth, when compared to the seven odd eighths immediately above it, will be higher when the slope of the density distribution is negative and lower when it is positive.

This observation may partly explain the increased clustering observed in Figure 1 for high-priced stocks (for which the integer price density distribution in Figure 2 is declining), but it cannot explain the clustering of the low-priced stocks (for which the integer price density distribution is increasing). In either event, the effect must be of secondary importance since the distribution of stock prices over all eighth-divisible prices (not presented) is not smooth. It is very jagged with local peaks at almost every zero eighth.



Figure 2

Frequency distribution of the integer portion of CRSP daily stock prices

The sample includes all CRSP daily closing prices for the years 1963 to 1987 that divisible by 1/8 and are greater than or equal to \$1. There are a total of 14,064,759 such prices, 14,912 of which are greater than or equal to \$201.

Table 2

Mean whole number frequencies for trade prices and bid and ask quotes in transactions data during the week of September 28, 1987

Price level subsample	Mean whole no. frequency			Average price	No. of stocks
	Trades	Bids	Asks		
NYSE and AMEX stocks					
All stocks	14.9%	14.2%	15.9%	\$25.9	2411
Over \$40	18.4	18.8	19.4	65.2	424
\$20-\$40	15.3	14.0	16.8	29.0	732
\$10-\$20	14.6	13.7	15.4	15.0	646
Under \$10	12.8	11.8	13.0	6.2	609
NASD over-the-counter stocks					
All stocks	18.9	17.8	19.3	15.3	2711
Over \$40	25.0	26.5	25.2	56.3	107
\$20-\$40	23.8	22.9	23.8	27.0	559
\$10-\$20	18.7	18.6	18.9	14.4	980
Under \$10	15.9	13.4	16.8	5.9	1065

The data were obtained from a consolidated transaction and quote data set maintained by the SEC. The quote sample consists of the last quotes available before each trade. Only data from the primary market for each stock are used. The reported means are computed from the sample of whole number price and whole number quote frequencies observed in the time series for each stock.

for the week of September 28 to October 2, 1987 are examined.¹¹ For each stock, the various whole number frequencies are computed. These frequencies are then averaged across all stocks. The results show that the whole number bid, ask, and trade price frequencies all are about equal, regardless of price level. Whole number asks are slightly more common than whole prices and whole prices are slightly more common than whole bids. These differences are small and are unlikely to provide much insight into the causes of clustering.

The results in Table 2 also show that stock price clustering appears in dealer markets as well as exchange markets. The NASD stocks display more clustering than do NYSE and AMEX stocks at all price levels.

The final characterization of stock price clustering in this section shows that stock price clustering increases when prices are volatile. Average whole number price frequencies before and after the October 1987 stock market crash are presented in Table 3. The sample includes all primary market trade prices for all common equities for the weeks of October 12-16 and 19-23. Whole number frequencies are com-

¹¹ The data come from a stock transaction data base assembled by the SEC, which was originally obtained to study the October 19, 1987 stock crash. Only one week of data is used because there are holes in the SEC quote data base during the test of the October sample. Fortunately, a week of transaction data is more than sufficient to address this question. Quotes other than the last bid and ask quotes before a transaction are excluded to ensure that differences in the rates at which quotations and transactions occur will not affect the results. Such differences could conceivably depend on the eighths.

Table 3
Mean whole number price frequencies before and after the October 19, 1987 stock market crash

	October 12-16		October 19-23		
Price level subsample	Mean frequency	Average price	Mean frequency	Average price	No. of stocks
NYSE and AMEX stocks					
All stocks	16.8%	\$24.2	21.1%	\$19.5	2401
Over \$40	22.2	69.8	33.7	56.5	308
\$20-\$40	17.1	31.2	23.1	25.2	678
\$10-\$20	16.1	16.4	19.4	13.1	642
Under \$10	14.9	6.4	15.9	5.1	773
NASDAQ over-the-counter stocks					
All stocks	21.9	14.1	23.9	11.7	2641
Over \$40	35.0	62.7	43.4	56.2	73
\$20-\$40	30.0	28.9	31.8	24.3	393
\$10-\$20	23.8	15.5	25.6	12.7	847
Under \$10	17.5	6.1	19.3	4.9	1328

The data were obtained from a consolidated transaction and quote data set maintained by the SEC. The sample includes each common stock that traded both before and after the stock market crash of October 19, 1987. The reported means are computed from the sample of whole number price frequencies observed in the time series for each stock. Only data from the primary market for each stock are used.

puted for each stock-during both weeks. These frequencies are then averaged across all stocks that traded during both weeks.¹² The results show that price clustering increased after the crash at all price levels, and in both exchange and dealer markets. The before-and after-crash frequency differences are all overwhelmingly statistically significant. These results are especially noteworthy because less clustering would normally be expected after prices fall.

2. Cross-sectional Characterizations of Stock Price Clustering

This section examines individual stock price clustering distributions to more completely characterize the clustering phenomenon. The above results show that clustering varies with price level, volatility, and exchange listing. Other variables such as firm size and transaction frequency may also be related to clustering. Since these variables are all correlated, the cross-sectional characterizations presented in this section are based on multivariate regression analyses.

¹²Equal-weighted averaging of the stock whole number frequencies ensures that any changes in the frequency of trade after the crash that might be correlated with price level do not affect the results. To ensure that the price level classified means are not affected by reclassifications caused by postcrash changes in price levels, a single price level is assigned to each stock based on the average price for that stock in the sample.

2.1 The dependent variables

These regression analyses examine two univariate measures that summarize clustering in the price fraction frequency distribution for a given stock. The first is the summed difference between the even and odd eighth sample frequencies. This difference measures the extent to which quarters are more common than odd eighths. The second is the sample frequency of whole prices. This frequency is a more general measure of clustering because it is large when whole numbers are more common than halves, halves, are more common than odd quarters, or quarters are more common than odd eighths.¹³ It also has an attractive model-based interpretation: The whole number frequency estimates the average price variation used by traders in a simple model in which traders use various discrete price sets based on different minimum price variations.¹⁴

Clustering frequency distributions for individual stocks are estimated from time-series data. The estimates can be noisy because prices are not serially dependent even if returns are. For example, a low-price stock that last traded at 1% will more likely next trade on the third, fourth, or fifth eighth than on any other eighth. In a one-year price time-series for this stock, no whole numbers maybe observed simply because the closest whole numbers are both 33 percent away from 1%. The whole number frequency is therefore a noisier measure of clustering in a time-series sample than in a cross-sectional sample of equal size.¹⁵ The difference is greatest for low-price stocks.

Estimator noise in the univariate clustering measures fortunately does not affect the consistency of regression coefficient estimators: Under reasonable assumptions, this noise has zero mean and is uncorrelated with the firm characteristics that are used as regressors. The noise, however, does increase the standard error of the regression.

Better estimates of the clustering frequencies can be obtained by taking the time dependence among the eighths into account. A full

¹³ The whole number frequency is the weighted sum of independent measures of three different kinds of clustering. The first measure, $d_0 = f_0 - f_1 + f_2 - f_3 + f_4 - f_5 + f_6 - f_7$, is the summed difference between even and odd eighths. The second and third measures, $d_1 = f_0 - f_2 + f_4 - f_6$ and $d_2 = f_0 - f_4$, are the summed differences between even and odd quarters, and even and odd halves, respectively. The whole number frequency, f_0 is exactly equal to $\frac{1}{8}d_0 + \frac{1}{4}d_1 + \frac{1}{2}d_2$.

¹⁴ Assume that a fraction α_8 of prices arise from traders that use a discrete price set based on a one-eighth price variation. Further, assume that these traders are equally likely to trade on any odd or even eighth. Likewise, let fractions α_4 , α_2 , and α_1 of prices arise from traders that use discrete price sets based on quarters, halves, and wholes. Since some discrete price set must be used, the various fractions sum to 1.

The average price variation used by traders in this model is a measure of clustering. It is equal to $\alpha_8/8 + \alpha_4/4 + \alpha_2/2 + \alpha_1$. This expression is identically equal to the whole number (zero eighth) price frequency implied by this model. The sample whole number frequency therefore estimates the average price variation used by traders in this model. A similar model appears in Ball, Torous, and Tschoegl (1985).

¹⁵ To a much lesser extent, this problem also affects the summed difference between even and odd eighths.

information estimator would be based on the cyclic Markov chain that characterizes this process. In this article, a less expensive ad hoc method is used to adjust the standard frequency estimators to reflect the domain over which prices are observed to wander. If prices do not often visit the region near a given eighth, the frequency for that eighth is adjusted upward. If prices dwell in that region, the frequency is adjusted downward. The two summary statistics for clustering described above are computed from the adjusted frequencies.

The adjustments to the standard sample frequencies are based on estimates of the domain of the observed price path over the various eighths. The domain is defined by assuming that a domain event over a given eighth occurs whenever prices change so that the price path passes over or arrives on that eighth. For example, a price change from 1/4 to 5/8 creates three domain events. (A passage over 3/8, a passage over 1/2, and an arrival at 5/8.) The domain is then estimated by the frequency distribution of these domain events. The adjusted eighth frequency estimates are $\hat{f}_i^* = \hat{f}_i + \frac{1}{8} - \hat{d}_i$ for $i = 0, \dots, 7$, where $\hat{f}_i$ is the standard sample clustering frequency, and $\hat{d}_i$ is the sample frequency of domain events occurring on the i th eighth.

This estimator has several attractive properties. First, the adjustments are very small for high-price stocks. Their large absolute price changes cause their domain to be evenly distributed over the various eighths. Second, the adjusted frequencies always sum to 1. Third, if the price domain does not span the A_h eighth, little can be learned about that eighth. Its adjusted frequency will be 12.5 percent. Fourth, clustering can be observed in the adjusted frequencies even if all price changes for a stock are 1/8. This is because zero price changes are not counted as domain events. If they were counted, the estimated domain distribution and the raw sample clustering distribution would be identical and no clustering would ever be indicated. Prices can cluster when no price changes are greater than 1/8 if zero price changes are more likely on even eighths than on odd eighths. A final property of the adjusted estimator is illustrated with an example: Suppose that price bounces between 3/8, and 5/8 stopping at 4/8 on each passage. The unadjusted fraction frequencies for this stock are $\hat{f}_3 = \hat{f}_5 = \frac{1}{4}$ and $\hat{f}_4 = \frac{1}{2}$ and the domain frequencies are likewise $\hat{d}_3 = \hat{d}_5 = \frac{1}{4}$ and $\hat{d}_4 = \frac{1}{2}$, so that the adjusted frequencies are all equal to 1/8: With no further information about the price process, it is impossible to determine whether prices cluster for this stock. The unadjusted frequencies indicate clustering but the adjusted ones do not.

2.2 The independent variables

The regressors used in the multivariate regression analyses are suggested primarily by the price resolution hypothesis advanced by Ball,

Torous, and Tschoegl (1985) in their study of gold price clustering. They hypothesize that price clustering depends on how well known is the underlying value of the security. If value is not well known, prices will cluster.

The price resolution hypothesis can be motivated by assuming that traders use discrete price sets to lower their costs of negotiating. The size of the discrete set, and therefore the extent of the clustering, depends on a balance between lower negotiation costs and lost gains-from-trade. Negotiating costs will be low if traders use a coarse set, but gains-from-trade may be lost if the set does not include a price that is acceptable to both parties. Lost gains-from-trade are likely if little dispersion exists among trader reservation prices, such as when underlying security values are well known. Traders will therefore use a fine set of prices (and little price clustering will be observed) when underlying security values are well known.

Several variables can proxy for the unobserved degree of reservation price dispersion. Time-series volatility should be correlated with reservation price dispersion because information is not uniformly distributed and interpreted when events cause values to change quickly. Firm size should be inversely correlated with reservation price dispersion because more information is produced and distributed about large firms than about small firms (fewer analysts follow small firms) and because large firms are generally better diversified than small firms (and therefore easier to value). Transaction frequency should be inversely correlated with reservation price dispersion because trading tends to reveal stock values by aggregating the information possessed by different traders.

Two other variables are used as regressors. Price level is included because an eighth represents a smaller fraction of price for high-price stocks than for low-price stocks. As a first approximation, traders are assumed to use discrete price sets based on minimum price variations that are constant fractions of price. This implies larger price variations (and hence more clustering) for high-price stocks than for low-price stocks.

Finally, a dummy variable that indicates whether a stock primarily trades in a dealer market is included as a regressor to determine whether clustering is related to market structure. Niederhoffer (1965) suggests that clustering is caused by the submission of public limit orders at round prices. If so, more clustering should appear in public auction exchange markets than in dealer markets because limit orders are represented in the former but not in the latter. Only dealers set quotes in dealer markets. In contrast, the negotiating cost arguments presented above suggest that prices should cluster more in dealer markets than in public auction exchange markets. Dealers can profit

by effectively establishing a reputation as low-cost negotiators because they play a repeated game. They are thus less likely to use a fine discrete set than would an anonymous trader in a large public auction market.

2.3 Results

Several samples of data are employed. A long CRSP sample is used to demonstrate robustness of results concerning price level, firm size, and volatility; a Fitch sample, is used to examine the relation between transaction frequency and clustering; and an SEC consolidated transaction data set is used to contrast clustering in dealer and exchange markets.

The CRSP sample consists of all NYSE and AMEX stocks for which there are 200 daily price observations on eighths during any year between 1963 and 1987. Prices recorded as the average of a bid and an ask are omitted where identified.¹⁶ Stocks that split more than 2 for 1 or reverse split less than 1 for 2 are not included because splits artificially change price levels. Stocks for which capitalization could not be determined are also excluded. The 1987 subset of this sample will be referred to as the 1987 CRSP sample.

Four dependent variables and four independent variables are computed for each stock year in the sample. The dependent variables are $\hat{f}_0$, the sample whole number (zero eighth) frequency; $\hat{f}_0^*$, the adjusted whole number frequency; EvenOdd, the summed difference between the even and odd eighth frequencies; and EvenOddAdj, the summed difference between the even and odd eighth adjusted frequencies.

The independent variables are AvePrice, the average price for the given year; **STDAP**, the standard deviation of overlapping five-day price changes over the year;¹⁷ LogMkVal, the log of the average of the beginning- and end-of-year total stock capitalizations; and CloseEnd, a dummy for closed-end funds, which are identified by SIC code 6723. The closed-end dummy is placed in the regression model to provide additional evidence that clustering is related to price resolution. Price resolution for closed-end funds should be finer than for other firms because closed-end funds are portfolios of marketable (and therefore easily valued) assets.

The regressions are estimated separately for each year in the sample, for five-year subperiods, and for the whole sample. Since the results are similar across sample periods, estimates for all four dependent

¹⁶CRSP fails to identify these prices before 1971. See Foerster and Kcim (1991).

¹⁷ Five-day price changes are used to minimize the effects of clustering, bid/ask spreads, and price discreteness on measured volatility.

Table 4

Cross-sectional ordinary least squares regressions of stock price clustering measures on various stock characteristics for a 1987 CRSP stock sample

Dependent variable	Independent variables estimates and <i>t</i> -statistics					<i>R</i> ²	Root mean-squared error
	Intercept	AvePrice	STDAP	LogMkVal	CloseEnd		
$\hat{f}_0$	22.0% (30.8)	0.0840 (6.2)	1.24 (3.7)	-0.755 (-11.6)	-0.632 (-1.14)	0.248	4.05%
$\hat{f}_0^*$	22.2% (35.6)	0.0870 (7.4)	1.18 (4.1)	-0.758 (-13.4)	-0.768 (-1.59)	0.306	3.53%
EvenOdd	20.9% (10.9)	0.198 (5.5)	5.60 (6.3)	-1.48 (-8.5)	-2.10 (-1.41)	0.3544	10.8%
EvenOddAdj	21.0% (11.0)	0.199 (5.5)	5.60 (6.3)	-1.49 (-8.6)	-2.09 (-1.41)	0.3545	10.8%

The dependent variables are $\hat{f}_0$, the sample frequency of whole number (zero eighth) prices; $\hat{f}_0^*$, the whole number price frequency adjusted for the observed domain of prices near the zero eighth; EvenOdd, the summed difference between the even and the odd sample eighth frequencies; EvenOddAdj, the Summed difference between the even and the odd sample eighth frequencies, computed from adjusted frequencies. The independent variables are AvePrice, the average stock price; STDAP, the standard deviation of overlapping five-day price changes; LogMkVal, the log of common stock market capitalization, and CloseEnd, a dummy that indicates closed-end funds. The sample consists of all NYSE and AMEX stocks for which there are 200 daily closing trade prices divisible by an eighth during 1987. There are 1979 stocks in the sample.

variables are reported only for 1987 (Table 4). All parameter estimates have their expected signs and all but CloseEnd are statistically significant. The adjusted frequency models fit better than the unadjusted frequency models. Residual scatter plots (not shown) show that the reduction in residual error is greatest for low-price stocks, as expected. Parameter estimates for the adjusted and unadjusted models are all quite close. Estimation noise in the dependent variables apparently is not correlated with the regressors.

Yearly regression estimates are reported in Table 5 for the adjusted whole number frequency model. The results are generally consistent across time. The only notable exceptions are that STDAP is significantly negative in 1963, 1964, 1966, and 1967. It may be consequential that these are all years for which CRSP fails to indicate which reported prices are averages of bid and ask quotes. They are also years for which market values computed from CRSP data are least reliable.¹⁸

Better fitting models and stronger results (not reported) are obtained when the time-series sample for computing the clustering measures is lengthened to five years from one year.¹⁹ These results show that the clustering measures are estimated with less error in longer samples.

¹⁸Market values often have to be computed from later reports of shares outstanding projected backward after adjusting for intervening splits.

¹⁹The goodness of fit measure, R^2 , ranges between 38 and 48 percent for the five-year subsample regression models that use the adjusted zero eighth frequency as the dependent variable.

Table 5

Cross-sectional ordinary least squares regressions of the adjusted whole number frequency measure of stock price clustering on various stock characteristics

Year	Independent variables					R^2	N
	Intercept	Ave-Price	STDAP	LogMkVal	CloseEnd		
63	22.2	0.141	-1.78	-0.561	-1.799*	0.330	1412
64	19.9	0.131	-1.83	-0.380	-1.537*	0.369	1664
65	21.6	0.132	0.24*	-0.727	-0.234*	0.323	1758
66	23.8	0.209	-1.19	-0.889	0.658*	0.379	1813
67	20.5	0.147	-0.54	-0.550	-0.494*	0.347	1883
68	22.8	0.153	0.19*	-0.797	-0.840*	0.401	1969
69	22.1	0.151	0.40*	-0.682	-1.110*	0.333	1986
70	21.0	0.162	0.29*	-0.553	-0.660*	0.309	2139
71	22.5	0.162	0.37*	-0.800	-0.452*	0.356	2304
72	19.5	0.088	1.35	-0.511	0.914*	0.321	2447
73	22.1	0.037	2.22	-0.618	-1.891	0.257	2246
74	22.3	0.091	0.92	-0.604	-1.085	0.170	1965
75	21.6	0.092	1.37	-0.656	-0.160*	0.205	1966
76	21.1	0.060	2.92	-0.709	-0.461*	0.268	2081
77	21.2	0.062	2.94	-0.716	-0.274*	0.230	2053
78	21.6	0.096	1.53	-0.725	-0.661*	0.233	2125
79	22.9	0.099	1.72	-0.841	-0.903*	0.198	2071
80	22.7	0.140	0.47*	-0.756	-1.796	0.261	2048
81	23.0	0.148	0.75	-0.821	-1.396	0.262	2015
82	23.3	0.117	1.06	-0.807	-1.624	0.233	1973
83	22.0	0.127	0.90	-0.757	-1.180	0.296	2062
84	20.9	0.083	1.29	-0.610	-1.361	0.193	1963
85	20.0	0.039	2.09	-0.548	-0.472*	0.212	1935
86	20.8	0.045	2.01	-0.626	-0.720*	0.285	1901
87	22.2	0.087	1.18	-0.758	-0.768*	0.306	1979
Mean	21.7	0.112	0.834	-0.680	-0.812	0.283	25

All estimates have t -statistics of greater than 2 in absolute value, unless marked by an asterisk. The dependent variable is $\hat{f}_n$, the sample frequency of whole number (zero eighth) prices, adjusted for the observed domain of prices near the zero eighth. The independent variables are AvePrice, the average stock price; STDAP, the standard deviation of overlapping five-day price changes; LogMkVal, the log of common stock market capitalization; and CloseEnd, a dummy that indicates closed-end funds. The annual samples consist of all NYSE and AMEX stocks for which there are 200 daily closing trade prices divisible by an eighth.

A sample of Fitch data from the NYSE is examined to characterize the relation between trading activity and clustering. The same methods described above are used to construct the sample and the analysis variables.²⁰ An additional variable, InvSQRTans, and inverse of the square root of the average number of transactions per day, is also computed. Since the clustering should decrease with transaction frequency, the estimated coefficient for InvSQRTans should be positive. The inverse square root transformation is based on some information-

²⁰ Since the sample time period is slightly larger, a stock is included only if it had 220 daily price observations on eighths, instead of the 200 required for each CRSP year. Also, a small number of stocks for which the Pitch ticket symbol could not be matched to a CRSP CUSIP number were dropped. The match is necessary to obtain capitalizations and SIC codes.

theoretic considerations.²¹ Similar results are obtained using other transformations. The sample period covers the 14 months from December 1981 to January 1983. This sample will be referred to as the Fitch 1982 sample.

The estimated regression for the adjusted whole number frequency is

$$\begin{aligned} \hat{f}_i^* = & 13.8\% + 0.0489 \text{ AvePrice} + 1.92 \text{ STD}\Delta P \\ & (9.1) \quad (2.74) \quad (5.3) \\ & - 0.219 \text{ LogMkVal} - 2.37 \text{ CloseEnd} \\ & (-2.01) \quad (-4.2) \\ & + 11.5 \text{ InvSQRTans}, \quad R^2 = .329, \quad N = 1153. \\ & (9.3) \quad \text{RMSE} = 2.95\% \end{aligned} \quad (1)$$

The estimates all have their expected signs and are statistically significant. The results for the other dependent variable regressions (not reported) are similar.

Finally, to determine whether prices cluster more in dealer markets than in exchange markets, a consolidated transaction data sample obtained from the SEC is examined. These data include all equity transactions at all U.S. exchanges. The sample spans September 28 to October 16, 1987. Only data from before the stock market crash are used to ensure that the results are not influenced by the crash. A new indicator variable, Dealer, takes a value of 1 if the primary market for the stock is the over-the-counter market, and 0, otherwise. Otherwise, the experimental design is nearly identical to that described above.²²

The estimated regression for the adjusted whole number frequency is

²¹ Suppose that each transaction conveys information about underlying value plus some noise. Underlying value estimators computed from the transaction sequence would then have standard errors that are proportional to the square root of the number of transactions observed. Price uncertainty is therefore proportional to the inverse square root of the number of transactions.

The inverse transformation has a second attractive property. Clustering should not decrease without limit as transaction frequency increases because other factors also determine minimum price variations. Since the inverse of the transaction frequency is bounded below at zero, negative predicted values for the dependent variables are less likely for frequently traded stocks than if a linear specification were used.

²² Only data from the primary market for each security are used. The primary market is assumed to be that market in which the stock most frequently traded. A stock appears in the sample only if it traded on eighth more than 20 times during the sample period, if it traded on at least 12 of the 15 days in the sample, and if more than 90 percent of the total number of trades at prices divisible by 1/8. The standard deviation of price changes is computed from daily closing prices, and the market values are computed from a list of shares outstanding on September 30, provided by Wells Fargo Investment Advisers.

$$\begin{aligned} \hat{f}_0^* = & 7.02\% + 0.105 \text{ AvePrice} + 1.33 \text{ STD}\Delta P \\ & (2.7) \quad (8.2) \quad (6.8) \\ & + 0.234 \text{ LogMkVal} + 11.97 \text{ InvSQRTans} \\ & (1.20) \quad (7.3) \\ & + 4.98 \text{ Dealer,} \quad R^2 = .118, \quad N = 3355. \\ & (13.5) \quad \text{RMSE} = 10.0\% \end{aligned} \quad (2)$$

Prices apparently cluster more in dealer markets. All other estimates but LogMkVal have their expected signs and are statistically significant. The positive estimated LogMkVal coefficient sign may be due to the cross-sectional correlation of LogMkVal with InvSQRTans. The fit of the model is relatively poor because the time-series sample used to compute the clustering frequencies is very short. The results for the other dependent variable regressions (not reported) are similar.²³

3. Odd Sixteenth Usage Frequency Projections

In this section, an econometric model for stock price clustering is derived and estimated. In the model, it is assumed that traders sometimes use discrete price sets based on fractions larger than the 1/8 dollar minimum price variation permitted by exchange regulations. Maximum likelihood estimates of the model are used to project how often traders would use odd sixteenths to trade low-price stocks if exchanges permitted trading on sixteenths. Since high- and low-price stocks differ by size, volatility, and transaction frequency, these cross-sectional differences are explicitly modeled.

3.1 The discrete price model

Suppose that all price negotiations have two parts. Traders first explicitly or implicitly decide to limit their terms of trade to some particular discrete price set. They then negotiate a specific price from within that set. The discrete price sets may differ from trade to trade.

Denote by α_8 the fraction of prices that arise from traders who use a discrete price set based on a one-eighth minimum price variation. Assume that these traders are equally likely to trade on any odd or even eighth. Likewise, let a_8 , a_4 , and a_2 denote the fractions of prices that arise from traders who use discrete price sets based on quarters,

²³ This test may unfairly represent clustering in the OTC markets because many OTC stocks above a dollar and under \$10 trade on sixteenths. Those that frequently did are excluded from the sample. The remaining sample may therefore be biased toward stocks for which traders use a large minimum price variation. To determine whether this bias is significant, the analysis was repeated, assuming that all stocks could trade on sixteenths. Although the new sample includes 102 additional (generally small) stocks, the results are virtually identical to those reported in the text.

halves, and wholes: Since some discrete price set must be used, these fractions sum to 1. The clustering distribution implied by this model is

$$f_i = \begin{cases} \alpha_8/8 + \alpha_4/4 + \alpha_2/2 + \alpha_1, & i = 0, \\ \alpha_8/8 + \alpha_4/4 + \alpha_2/2, & i = 4, \\ \alpha_8/8 + \alpha_4/4, & i = 2, 6, \\ \alpha_8/8, & i = 1, 3, 5, 7, \end{cases} \quad (3)$$

where f_i is the implied frequency of prices on the i th, eighth.

Now suppose that if there were no price discreteness regulations, traders would use a discrete set based on a minimum price variation that is some fraction R of price. Assume that this desired relative minimum price variation, R , varies from trade to trade, and let $F(R)$ denote the cumulative distribution function of R over all positive numbers.

Each pair of traders must translate their value of R into a particular discrete price set based on eighths, quarters, halves, or whole numbers. Assume that they multiply R by the price level P to express their desired minimum price variation in terms of price. They then round this product to the nearest minimum price variation in the set of $(1/8, 1/4, 1/2, 1)$, where nearness is measured by geometric distance. These assumptions imply the following rules for obtaining the basis, d , of their discrete price set:

$$d = \begin{cases} 1/8, & \text{if } P \times R \leq k_8, \\ 1/4, & \text{if } k_8 < P \times R \leq k_4, \\ 1/2, & \text{if } k_4 < P \times R \leq k_2, \\ 1, & \text{if } k_2 < P \times R, \end{cases} \quad (4)$$

where

$$k_8 = \sqrt{(1/8 \times 1/4)}, \quad (5a)$$

$$k_4 = \sqrt{(1/4 \times 1/2)}, \quad (5b)$$

$$k_2 = \sqrt{(1/2 \times 1/1)}. \quad (5c)$$

are the geometric midpoints. These rules imply the following expressions for the fractions α_8 , α_4 , α_2 , and α_1 :

$$\alpha_8 = F(k_8/P), \quad (6a)$$

$$\alpha_4 = F(k_4/P) - F(k_8/P), \quad (6b)$$

$$\alpha_2 = F(k_2/P) - F(k_4/P), \quad (6c)$$

$$\alpha_1 = 1 - F(k_2/P). \quad (6d)$$

Parametrize the distribution function $F(R)$ as $F(R; \mu_i, v)$, where μ_i is a location parameter for stock i and v is a common shape parameter. Let the location measure vary by stock according to

$$\mu_i = \gamma_0 + \gamma_1 \text{AvePrice}_i + \gamma_2 \text{STDRet}_i + \gamma_3 \text{LogMkVal}_i + \gamma_4 \text{InvSQRTTrans}_i, \quad (7)$$

where STDRet is the five-day standard deviation of percentage returns.²⁴ The results of the previous section suggest that γ_2 should be positive, γ_3 should be negative, and γ_4 should be positive. The parameter γ_1 , should be zero since the desired relative minimum price variation is expressed as a fraction of price.

This model can be estimated in a cross-sectional sample of stocks if the functional form of F is specified. This analysis uses the gamma distribution. This distribution ranges over all positive numbers and is relatively easy to use. The parametrization is chosen so that m_i is the mean of the distribution. The resulting gamma distribution function is

$$F(R; \mu_i, v) = \int_0^R \frac{\lambda(\lambda r)^{v-1} e^{-\lambda r}}{\Gamma(v)} dr, \quad (8)$$

where $\lambda = v/\mu_i$.

The model is estimated using the full information maximum likelihood estimation method. The data consist of stock sample frequency distributions of the various denominator fractions used to express prices. These frequencies are computed from adjusted clustering frequencies as follows:

$$\hat{\beta}_1 = \hat{f}_0 \quad (\text{wholes}), \quad (9a)$$

$$\hat{\beta}_2 = \hat{f}_1 \quad (\text{halves}), \quad (9b)$$

$$\hat{\beta}_4 = \hat{f}_2 + \hat{f}_6 \quad (\text{odd quarters}), \quad (9c)$$

$$\hat{\beta}_8 = \hat{f}_1 + \hat{f}_3 + \hat{f}_5 + \hat{f}_7 \quad (\text{odd eighths}), \quad (9d)$$

Model implications for these frequencies are

$$\beta_1 = \alpha_8/8 + \alpha_4/4 + \alpha_2/2 + \alpha_1, \quad (10a)$$

$$\beta_2 = \alpha_8/8 + \alpha_4/4 + \alpha_2/2, \quad (10b)$$

$$\beta_4 = 2(\alpha_8/8 + \alpha_4/4), \quad (10c)$$

$$\beta_8 = 4\alpha_8/8, \quad (10d)$$

with the α 's given by Equation (6). The log likelihood for a given

²⁴ The return standard deviation appears in this specification instead of the price change standard deviation because R is the desired relative minimum price variation.

stock data vector is therefore

$$\text{Log } \mathcal{L} = T(\hat{\beta}_1 \log \beta_1 + \hat{\beta}_2 \log \beta_2 + \hat{\beta}_4 \log \beta_4 + \hat{\beta}_8 \log \beta_8), \quad (11)$$

where T is the number of time-series observations.

3.2 Empirical results

The model is estimated for the 1987 CRSP sample and for the 1982 Fitch sample. To equally weight all stocks in the analysis, the time-series sample size T is taken to be the same for all stocks. So that the statistical inferences will err on the conservative side, T is assumed to be equal to the minimum number of observations necessary for inclusion in the sample.

The parameter estimates (and asymptotic t-statistics) for the CRSP sample are

$$\begin{aligned} \hat{\mu}_i = & 1.03\% + 0.0000256 \text{ AvePrice}_i + 0.0293 \text{ STDRet}_i \\ & (46.4) \quad (0.43) \quad (12.6) \\ & - 0.0632 \text{ LogMkVal}_i \\ & (-42.3) \\ \hat{\rho} = & 0.638, \quad T = 200, \quad N = 1979, \quad \text{Log } \mathcal{L} = -507,500. \quad (12) \\ & (24.9) \end{aligned}$$

The coefficient estimates are all as expected.²⁵ Most notably, the estimate for AvePrice is not significantly different from zero. This suggests that the structural form of the model is sufficient to account for the relation between price level and clustering in this sample.

The cross-sectional mean of μ_i , the estimated mean of the desired relative minimum price variation, is 0.40 percent for the entire sample (Table 6). The mean estimate declines with price level: It is 0.54 percent for stocks priced under \$10 and 0.23 percent for stocks priced above \$40. The mean desired relative minimum price variation declines with price level because high-price stocks have low return volatilities and high market values (Table 6).

The estimated model implies denominator fraction frequencies that closely fit the various observed minimum denominator frequencies (Table 7). For the whole sample, the difference between the mean implied (fitted) frequency and the mean observed frequency ranges from a high of 1.1 percentage points for the whole number to a low of -0.8 percentage points for the odd eighths. The fit is also good for stock subsamples classified by price level. Root mean squared

²⁵ The t-statistics may be overstated because price fractions are not serially independent. The overstatement is probably more than offset by the conservative treatment of the time-series sample size and by using the adjusted frequencies.

Table 6

Cross-sectional means of estimated desired relative minimum price variations, $\hat{\mu}_i$, and independent variables

Variable	All stocks	Price level subsamples			
		Under \$10	\$10-20	\$20-40	Over \$40
1987 CRSP sample ¹					
Mean $\hat{\mu}_i$	0.398%	0.544	0.437	0.337	0.233
AvePrice (mean price)	\$24.7	6.0	14.8	28.8	61.7
STDRet (5-day return STD)	4.81%	5.78	4.79	4.37	4.19
LogMkVal (log market value)	12.2	10.4	11.6	13.0	14.6
No. of stocks	1979	501	507	655	316
No. of negative estimates of $\hat{\mu}_i$	4	0	0	0	4
1982 Fitch sample ²					
Mean $\hat{\mu}_i$	0.442%	0.587	0.476	0.403	0.293
AvePrice (mean price)	\$23.4	7.0	15.2	28.0	55.4
STDRet (5-day return STD)	4.87%	5.62	4.85	4.73	4.47
LogMkVal (log market value)	12.5	10.8	12.0	13.1	14.1
InvSQRTtrans (1/ $\sqrt{\text{mean trades/day}}$)	0.24	0.30	0.26	0.23	0.18
No. of stocks	1153	155	392	497	109
No. of negative estimates of $\hat{\mu}_i$	0	0	0	0	0

The desired relative minimum price variation is the minimum price variation that traders would base their discrete price sets upon if there were no minimum price variation regulations, expressed as a fraction of price. The desired relative minimum price variation estimates are computed from maximum likelihood estimates of the discrete price model presented in Section 3.1.

$$\hat{\mu}_i = 1.03\% + 0.0000256 \text{ AvePrice}_i + 0.0293 \text{ STDRet}_i - 0.0632 \text{ LogMkVal}_i$$

$$\hat{\mu}_i = 0.251\% - 0.00113 \text{ AvePrice}_i + 0.0493 \text{ STDRet}_i - 0.0211 \text{ LogMkVal}_i + 0.991 \text{ InvSQRTtrans}_i$$

errors of the fitted frequencies range between the 2.76 and 5.08 percentage points, and are consistent across price levels.

Goodness of fit across the various frequencies is important because the estimated model is used to project sixteenth usage frequencies. These projections are out of sample projections because the data on eighths, quarters, halves, and wholes do not span sixteenths. Therefore, they crucially depend on the shape of the distribution of R , the underlying desired relative minimum price variation. These goodness of fit measures suggest that the gamma distribution provides a useful description of this distribution.

Goodness of fit across price levels is important. The sixteenth usage frequency projections for low-price stocks depend on information about discrete price set usage learned from high-price stocks for which the current one-eighth minimum price variation constraint is often not binding. These projections will not be reliable if the estimated model does not adequately describe both high- and low-price stocks.

The projected usage frequency for discrete price sets based on Sixteenths is $\alpha_{16} = F(k_{16}/P)$, where $k_{16} = \sqrt{1/16} \times 1/8$. The projected odd sixteenth usage frequency is half of this frequency.

Table 7
Cross-sectional means of observed and fitted price fraction frequencies and their root mean squared differences for the estimated discrete price model

Price fraction	Variable type	All stocks	Price level subsamples			
			Under \$10	\$10-20	\$20-40	Over \$40
1987 CRSP sample						
Whole nos.	Observed	16.3%	14.7	15.8	16.7	19.1
	Fitted	15.2	12.7	14.0	16.0	19.1
	RMSE	3.65	3.41	3.29	3.22	5.11
Halves	Observed	14.0	12.8	13.4	14.5	15.8
	Fitted	14.6	12.7	13.9	15.5	16.9
	RMSE	2.76	2.53	2.55	2.85	3.21
Odd quarters	Observed	26.5	25.1	26.1	27.2	28.2
	Fitted	26.3	25.4	26.5	26.9	26.3
	RMSE	3.62	3.11	2.97	3.55	5.12
Odd eighths	Observed	43.1	47.5	44.7	41.6	36.9
	Fitted	43.9	49.2	45.6	41.6	37.8
	RMSE	5.08	3.99	4.18	5.42	6.89
	N	1979	501	507	655	316
1982 Fitch sample						
Whole nos.	Observed	17.1%	15.3	16.2	17.7	19.7
	Fitted	16.0	13.0	14.5	17.2	20.5
	RMSE	3.24	3.73	3.27	2.95	3.73
Halves	Observed	14.4	12.3	13.8	15.0	16.6
	Fitted	15.3	13.0	14.3	16.2	17.6
	RMSE	2.63	2.47	2.43	2.77	2.85
Odd quarters	Observed	27.0	25.5	26.1	27.7	28.8
	Fitted	26.5	25.6	26.6	26.7	26.0
	RMSE	3.67	3.32	3.02	3.87	5.12
Odd eighths	Observed	41.6	46.9	43.9	39.5	35.0
	Fitted	42.2	48.5	44.6	39.9	35.9
	RMSE	4.77	3.59	4.10	5.35	5.67
	N	1153	155	392	497	109

The fitted frequencies are implied from maximum likelihood estimates of the discrete price model presented in Section 3.1.

The cross-security mean projected odd sixteenth usage frequency is 36.7 percent for the entire sample (Table 8). This mean declines with price level. For stocks under \$10, the mean projected usage frequency is 45.7 percent; for stocks over \$40, it is 28.8 percent.

Projections of odd thirty-second usage frequencies are less reliable than projections of odd sixteenth frequencies because they depend on the shape of the gamma distribution over a region that is further from observed data. The average projected odd thirty-second usage frequency is 28.1 percent for the entire sample (Table 8).²⁶

²⁶There is a small upward bias in the sixteenth and thirty-second usage frequency projections. The mean μ_i must be positive but the linear specification does ensure that it is. The estimation procedure assigns zero values if negative values would otherwise be estimated. This occurs for four of the 1979 firms in the CRSP sample (Table 6). When the μ_i is equal to zero, all of the mass of the gamma distribution is at zero. This implies no price clustering, regardless of the regulated minimum price variation. For these stock, the odd sixteenth and the odd thirty-second price frequency projections are both equal to 50 percent.

Table 8
Cross-sectional means of projected odd sixteenth and odd thirty-second price usage frequencies

		Price level subsamples			
Projected usage frequency	All stocks	Under \$10	\$10-20	\$20-40	Over \$40
1987 CRSP sample					
Odd sixteenths	36.7%	45.7	37.9	32.6	28.8
Odd thirty-seconds	28.1%	38.6	28.4	23.5	20.5
No. of stocks	1979	501	507	655	316
1982 Fitch sample					
Odd sixteenths	34.1%	43.6	36.6	30.7	26.8
Odd thirty-seconds	25.2%	35.2	27.4	21.9	18.7
No. of stocks	1153	155	392	497	109

The projected odd sixteenth and odd thirty-second price usage frequencies are implied from maximum likelihood estimates of the discrete price model presented in Section 3.1.

The parameter estimates for the Fitch sample are

$$\begin{aligned} \hat{\mu}_t = & 0.251\% - 0.00113 \text{ AvePrice}_t + 0.0493 \text{ STDRet}_t \\ & (3.7) \quad (-11.4) \quad (14.5) \\ & - 0.0211 \text{ LogMkVal}_t + 0.991 \text{ InvSQRTTrans}_{t-1} \\ & (-5.5) \quad (19.4) \\ \hat{\rho} = & 0.640, \quad T = 220, \quad N = 1153, \quad \text{Log } \mathcal{L} = -328,526. \quad (13) \\ & (25.0) \end{aligned}$$

The estimated coefficients for STDRet, LogMkVal, and InvSQRTTrans are all of the expected sign and are statistically significant. The small AvePrice coefficient, however, is significantly different from zero. This may indicate some minor model misspecification. Perhaps the mean number of transactions per day, which is correlated with AvePrice, is improperly transformed.

Goodness-of fit statistics and odd sixteenth and odd thirty-second price frequency projections are presented in Tables 7 and 8 for the Fitch model. The results are similar to those obtained for the CRSP data. The root mean squared error statistics suggest that the Fitch sample model fits slightly better than the CRSP sample model. This may be due to the additional information in the transaction frequency variable, or because the samples differ.

4. Conclusion

4.1 Summary

Stock price clustering is pervasive. Clustering distributions from the mid-nineteenth century appear no different from those observed in the late twentieth century.

This article characterizes stock price clustering and its relation to several observable characteristics of stock prices. Clustering increases with price level and volatility. Clustering decreases with market value and transaction frequency: Clustering is greater in dealer markets than in public auction exchange markets.

An econometric model of stock price clustering is estimated. The model assumes that traders sometimes use discrete price sets based on fractions larger than the 1/8 dollar minimum price variation permitted by exchange regulations. Projections from the estimates suggest that traders would frequently use odd sixteenths to trade low-price stocks if exchanges permitted trading on sixteenths.

4.2 Additional perspectives and directions for future research

This empirical clustering analysis is based on an informal model in which traders choose discrete price sets to minimize negotiation costs. Formalization of this model may produce new testable hypotheses. Game-theoretic methods should be used to study discreteness in price negotiations.

In this article, only unconditional clustering distributions are examined. Clustering must also affect price change distributions. An empirical study of clustering transition frequencies (clustering distributions conditional on previous prices) could reveal more about why traders use discrete prices.

A joint implication of the cost-of-negotiation clustering model and of the efficient market hypothesis is that there should be no profitable trading rule based on clustering transition frequencies (after properly considering and modeling the bid/ask spread). Niederhoffer and Osborne (1966) consider this question and find some profitable trading rules. Their study should be updated. A test of the weak-form efficient market hypothesis should also be undertaken based on implied symmetries in the cyclic Markov chain transition matrix that characterizes the conditional stock price clustering distribution.

Clustering appears in the prices of many other assets besides stocks. Goodhart and Curcio (1990) examine decimal price clustering in foreign exchange markets and Ball, Torous, and Tschoegl (1985) examine the same in the London gold market. No studies have been made of clustering in bond markets, option markets, or futures markets. A study of the determinants of clustering across various asset classes would be especially interesting.

Stock splits change price levels with generally little effect on other firm attributes. A study of clustering surrounding splits would allow much stronger inferences about its relation with price level.

The empirical results suggest that traders would frequently use odd sixteenths to trade low-price stocks, if they could. These results do

not imply, however, that' bid/ask spreads would decrease. Using a discrete bid/ask spread model similar to the econometric discrete price model used in this Article, Harris (1989b) projects that average spreads for stocks trading under \$10 would decrease by about 20 percent. The potential cost savings to liquidity demanding investors could be quite significant.

References

- Amihud, Y., and H. Mendelson, 1990, "Option Market Integration: An Evaluation," working paper, New York University.
- Ball, C. A., W. A. Torous, and A. E. Tschoegl, 1985, "The Degree of Price Resolution: The Case of the Gold Market," *Journal of Futures Markets*, 5,29-43.
- Brennan, M., and T. Copeland, 1988, "Stock Splits, Stock Prices, and Transaction Costs," *Journal of Financial Economics*, 22,83-101.
- Foerster, S. R., and D. B. Keim, 1991, "Direct Evidence of Non-trading and Implications for Daily Return Autocorrelations," working paper, University of Pennsylvania.
- Goodhart, C., and R. Curcio, 1990, "Asset Price Discovery and Price Clustering in the Foreign Exchange Market," working paper, London School of Economics.
- Gottlieb, G., and A. Kalay, 1985, "Implications of the Discreteness of Observed Stock Price," *Journal of Finance*, 40,135-153.
- Grossman, S. J., and M. H. Millett, 1988, "Liquidity and Market Structure." *Journal of Finance* ,43 617-633.
- Harris, L., 1989a, "A Day-end Transaction Price Anomaly," *Journal of Financial and Quantitative Analysis*, 24, 24 -45.
- Harris, L., 1989b. "Stock Price Clustering. Discreteness, and Bid/Ask Spreads," working paper, University of Southern California.
- Harris, L., 1990a, "Liquidity, Trading Rules, and Electronic Trading Systems," New York University Salomon Brothers Center Monograph Series 90-4.
- Harris, L., 1990b. "Estimation of Stock Variance and Serial Covariances from Discrete Observations," *Journal of Financial and Quantitative Analysis*, 25,291-306.
- Niederhoffer, V., 1965, "Clustering of Stock Prices," *Operations Research* 13,258-265.
- Niederhoffer, V., 1966, 'A New Look at Clustering of Stock Prices." *Journal of Business*, 39, 309-313.
- Niederhoffer, V., and M.F.M. Osborne, 1966, "Market Making and Reversals on the Stock Exchange," *Journal of the American Statistical Association* 61,697-916.
- Osborne, M. F. M., 1962, "Periodic Structure in the Brownian Motion of Stock Prices," *Operations Research*, 10,345-379.