

Western  Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

9-20-2016 12:00 AM


Essays in Market Structure and Liquidity

Adrian J. Walton
The University of Western Ontario

Supervisor
Adam Metzler
The University of Western Ontario

Graduate Program in Applied Mathematics
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of
Philosophy
© Adrian J. Walton 2016

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>

 Part of the [Dynamic Systems Commons](#), [Econometrics Commons](#), [Economic Theory Commons](#),
[Finance Commons](#), [Numerical Analysis and Computation Commons](#), and the [Other Applied Mathematics Commons](#)

Recommended Citation

Walton, Adrian J., "Essays in Market Structure and Liquidity" (2016). *Electronic Thesis and Dissertation Repository*. 4173.
<https://ir.lib.uwo.ca/etd/4173>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Market structure concerns the mechanisms for negotiating trades and the composition of trading participants, and can affect liquidity and price efficiency. More gains from trade can be realized from an asset that is more liquid, and a better allocation of risk and capital can be achieved when an asset's price is more efficient so it is important to understand market structure. This thesis uses theory and empirical methods to examine the effects of a few specific aspects of market structure.

In Chapter 1, we study a novel market structure on the New York Stock Exchange (NYSE), the Retail Liquidity Program (RLP), that allows liquidity providers to trade specifically with retail traders. We test whether it affected the quality of trading opportunities for retail and non-retail traders by measuring transaction costs before and after the RLP was launched. We find transaction costs are slightly lower for both retail and non-retail traders. We also find evidence of an improved price-discovery process from allowing market participants to distinguish between retail trades, which contribute little to price discovery, and non-retail trades, which contribute more so.

In Chapter 2, we extend a classic model of market microstructure to formalize the hypotheses and findings from Chapter 1 and to form new predictions. Under the models assumptions, prices are more efficient, and the effect on liquidity is ambiguous. We develop predictions on how informed traders adjust their trading strategies in the presence of the RLP.

In Chapter 3, we consider a market where a significant amount of trading is motivated by hedging. We use a classic microstructure model to examine how a market makers willingness to provide liquidity is affected by the need to learn about the underlying value of an asset as well as the inventory of a hedging trader. Under our models assumptions, a market maker provides more liquidity in the presence of hedging. We test our prediction empirically by studying the effect of predictable increases in trading volume that occur near the expiry of stock options. We find the evidence that hedging trades result in improved liquidity.

Keywords: Finance, market microstructure, liquidity, price efficiency, segmentation

Co-Authorship Statement

Chapter 1 is joint work with Corey Garriott and is adapted from our Bank of Canada Staff Working Paper with the permission of the Bank of Canada.

Acknowledgements

Thank you to my supervisor, Adam Metzler; to colleagues at the Bank of Canada, Nikil Chande, Corey Garriott, Sermin Gungor, Darcey McVanel and Joshua Slive; to my parents, John and Ruth; and to my wife, Jocelyn.

Contents

Abstract	ii
Co-Authorship Statement	iii
Acknowledgements	iii
List of Figures	viii
List of Tables	xi
List of Appendices	xviii
1 Retail Order Flow Segmentation: Empirics	1
1.1 Introduction	1
1.2 Hypotheses	5
1.3 Data	7
1.3.1 Treatment stocks	7
1.3.2 Control stocks	7
1.3.3 Data for computing information shares: returns and order flow	8
1.3.4 Data for conducting difference-in-differences: market quality measures	8
1.3.5 Summary statistics	9
1.4 Methodology	11
1.4.1 VAR and information shares	11
1.4.2 Difference-in-differences	14
1.5 Results	15
1.5.1 Conclusions	28
2 Retail Order Flow Segmentation: Theory	33
2.1 Introduction	33
2.2 Related Literature	35
2.3 Model with Market Orders	37
2.3.1 Informed Trader	37
2.3.2 Institutional Uninformed Trader	38
2.3.3 Retail Trader	38
2.3.4 Market Maker	38
2.3.5 Perfect Bayesian Nash Equilibrium	38
2.4 Model with Order Choice	45

2.4.1	Model Setup	46
2.4.2	Exchange Trading Mechanics	47
2.4.3	Market Maker	47
2.4.4	Active Noise Trader	48
2.4.5	Passive Noise Trader	48
2.4.6	Informed Trader	49
2.4.7	Market Maker's Strategy	49
2.4.8	Perfect Bayesian Nash Equilibrium	53
2.4.9	Informed Trader's Problem	55
2.4.10	Approximation: Large Switching Threshold	56
	Approximation: Large Switching Threshold and Small Probability of Informed Trading	58
2.4.11	Numerical Evidence	60
2.5	Conclusions	64
3	The Effect of Hedging on Liquidity	67
3.1	Introduction	67
3.1.1	Hedging and Program Trading	69
3.1.2	Transparency and Trading Platforms	70
3.1.3	Sunshine Trading vs. Front-Running	71
3.1.4	Equity Options and Information	72
3.2	Model	73
3.2.1	Single-Period Setting	73
	Noise Trader	74
	Hedger	74
	Informed Trader	74
	Market Maker	75
	Definition of Equilibrium	75
	Informed Trader's Strategy	76
	Market Maker's Strategy	76
3.2.2	Multiperiod Setting	77
	Noise Trader	77
	Hedger	77
	Market Maker	78
	Informed Trader	78
	Definition of Equilibrium	78
	Informed Trader's Strategy	79
	Market Maker's Strategy	79
3.2.3	Steady State Beliefs	80
3.2.4	Comparison to Other Models	81
3.2.5	Empirical Study	82
	Institutional Details	83
	Data	84
	Liquidity Measures	84
	Summary Statistics	84

3.2.6	Methodology	87
3.2.7	Results	89
3.3	Conclusions	91
References		94
A	Chapter 2 Proofs	99
A.1	Bid and Ask Prices	99
A.2	Proof of Proposition 2.3.1	99
A.3	Proof of Proposition 2.3.2	99
A.4	Alternate Proof of Proposition 2.3.2	102
A.5	Proof of Proposition 2.3.3	103
A.6	Proof of Lemma 2.4.1	104
A.7	Proof of Proposition 2.4.2	104
A.8	Proof of Proposition 2.4.5	111
B	Chapter 3 Proofs	112
B.1	Steady States	112
B.2	Example of Dynamic Optimization	115
Curriculum Vitae		117

List of Figures

1.1	Relative bid-ask spread for treatment and control stocks over the sample period. This figure shows the relative bid-ask spread for treatment and control stocks over the sample period. The blue line represents relative bid-ask spread for control stock; the red line represents relative bid-ask spread for treatment stocks. The vertical line indicates the launch of the RLP on 1 August 2012. . . .	12
1.2	Volume for treatment stocks over the sample period. This figure shows total volume and RLP volume.	13
1.3	Impulse-response functions for RLP and lit order flow. This figure plots orthogonalized impulse-response coefficients for each component of the order flow against their corresponding lags. The blue line represents the response of return to a one-standard deviation shock to the lit order flow; the red dashed line represents the response of return to a one-standard deviation shock to the RLP order flow.	17
2.1	Bid and ask prices. The figure shows the bid and ask prices as a function of the market maker’s prior about the true value of the asset. The figure was created with parameters $\alpha = \frac{1}{2}, \beta = \frac{1}{2}$ and $c = \frac{3}{10}$	40
2.2	Bid-ask spread. The figure shows the bid-ask spread versus the variance of the market maker’s belief about the true value of the asset. The parameter β is the degree of retail segmentation. The figure was created with parameters $\alpha = \frac{1}{2}, \beta \in \{0, \beta = \frac{1}{3}, \beta = \frac{2}{3}, \beta = 1\}$, and $c = \frac{3}{10}$	41
2.3	Numerical example. This figure shows the results of a numerical simulation of one possible path of the game for two values of β . The blue and green lines correspond to two simulations of the market maker’s prior as it evolves with one trade per unit time. The true value of the asset is shown with the red line. The figure was created with parameters $\alpha = \frac{1}{10}, \beta = \frac{1}{2}$ and $c = \frac{3}{10}$	43
2.4	Average liquidity. Each line on the graph shows the average bid-ask spread (on the y -axis) over periods 1 to N for several simulations of the game, where N is the total number of trades (on the x -axis). The solid blue line shows the result for measure zero of retail segmentation ($\beta = 0$) and the dashed green and red lines show the result for nonzero measure of retail segmentation ($\beta = \frac{1}{3}$ and $\beta = \frac{2}{3}$). The plot was created with parameters $\alpha = \frac{1}{5}, \beta = \{0, \frac{1}{3}, \frac{2}{3}\}$ and $c = \frac{3}{10}$ with 100 trials for each point on the x -axis.	44

2.5	Empirical trade and quote data. The figure shows empirical trading data aggregated across several trading venues in the US for roughly one minute of trading that occurred on 16 June 2011 for an S&P 500 exchange-traded fund. Grey bands indicate the bid-ask spread, coloured dots indicate trades, and volume is indicated by vertical bars at the bottom of the figure. This graph was reproduced with the permission of Nanex, LLC.	46
2.6	Model timing. A single period is represented by the actions shown from left to right between solid blue lines. In the first sub-period, limit orders are submitted, and in the second, market orders. The informed trader randomized between submitting a limit order or a market order, with probability (w.p.) ρ of submitting a market order. If the informed trader submits a limit order, he does so before the passive noise trader with probability α . If the informed trader submits a market order, he does so before the active noise trader with probability α	48
2.7	Market maker's expected profit. The figure shows the market maker's expected profit from quoting different bid-ask spreads: the red solid line represents expected profits from the inner pair $\{x_l, x_h\}$, the cyan dashed line the higher pair $\{x_h, x_H\}$, and the purple dash-dot line the lower pair $\{x_L, x_l\}$. The grey regions indicate where the market maker quotes either the upper or lower price pairs. The market maker prefers to quote the upper or lower price pair when they yield positive expected profit because the market maker acts competitively. The figure was created with model parameters: $x_h = \frac{4}{5}$, $x_l = \frac{1}{5}$, $x_L = \frac{1}{3}$, $x_H = \frac{2}{3}$ and $\alpha = \frac{1}{5}$	52
2.8	Numerical example of game dynamics. The figure shows the results of a simulation of one possible path of the game. The blue dashed line represents the market maker's belief about the asset's value which moves randomly according to the pattern of trades and quotes. The purple and cyan dashed lines represent the bid and ask prices quoted by the market maker. The solid red line represents the asset's true value. The grey area marks the end of the game when the market maker revises his quotes.	54
2.9	Optimal strategy. The figure shows the optimal probability of the informed trader submitting a market on the y -axis and the degree of segmentation on the x -axis. The prices used to generate the figure were $x_l = \frac{1}{3}$ and $x_h = \frac{2}{3}$. The blue line was created using a value of $\frac{1}{10}$ for α	57
2.10	Price efficiency. The y -axis is price inefficiency, defined as the reciprocal of the expected change in the market maker's prior evolves following a trade. On the x -axis is the degree of segmentation. The solid blue line represents price inefficiency when the informed trader acts strategically, while the dashed green line serves as a benchmark and represents the case when the informed trader submits limit orders and market orders with probabilities $\frac{1}{2}$. The figure shows that despite the informed trader's strategic order choice, price inefficiency decreases increases with the degree of segmentation. The figure was created using values of $\alpha = \frac{1}{10}$, $x_l = \frac{1}{3}$, and $x_h = \frac{2}{3}$	59

2.11	Simulated optimal strategy. The figure shows the numerical optimal strategy (solid blue) and the theoretical optimal strategy when H is large (dashed green) as a function of the degree of retail segmentation. Simulation parameters were $\alpha = \frac{1}{10}$, $x_l = \frac{1}{3}$, $\bar{\theta} = \frac{9}{10}$ and $\beta = \frac{1}{2}$. The probability of a market order is strictly decreasing and convex in the degree of segmentation for both the numerical and theoretical strategies.	63
2.12	Simulated optimal strategy. The figure shows the numerical optimal strategy as a function of the market maker's prior, for various degrees of retail segmentation. Simulation parameters were $\alpha = \frac{1}{10}$, $x_l = \frac{1}{3}$ and $\bar{\theta} = \frac{9}{10}$. The probability of a market order is strictly decreasing in each of the values for segmentation.	64
2.13	Simulated hitting time. The figure compares simulated expected hitting time (solid blue) to the theoretical approximation (dashed green) as the degree of retail segmentation increases. Simulation parameters were $\alpha = \frac{1}{10}$, $x_l = \frac{1}{3}$ and $\bar{\theta} = \frac{9}{10}$. The theoretical line was generated by plotting the reciprocal of the expected update given the parameters. In our approximation when H is large, hitting time is inversely proportional to the magnitude of the expected update. The dashed green line was scaled by a constant to be visually comparable with the solid blue line. Both decrease with the degree of retail segmentation at a comparable rate.	65
3.1	Price impact coefficient. This figure shows how the price impact coefficient varies with variance the hedging volume. The parameters chosen were $v_u = 0.2$, $v_v = 0.5$ and $h = 0.1$. The solid blue line shows price impact coefficient for the single period model from Equation 3.19 as a function of the variance of hedging trades. The green dashed line shows the price impact coefficient for the multiperiod case from Equation 3.31.	81
3.2	Average trading volume. This figure shows average daily trading volume for stocks the S&P 100 Index over the sample period. Red vertical lines mark options expiry dates.	87
3.3	Open interest. This figure shows average put open interest (the solid blue line) and average call open interest (the dashed red line) for stocks the S&P 100 Index over the sample period. Red vertical lines mark options expiry dates.	88
3.4	Causation diagram. This causation diagram shows the channels through which liquidity in the underlying stock and hedging volume are affected. Liquidity in the underlying stock and hedging volume may directly affect each other; if liquidity improves options hedging become less costly. This may lower the cost of options and lead to greater open interest which results in further hedging volume in the stock. This volume may again improve liquidity. Unobserved variables may affect both the activity in the options market and the liquidity of the underlying stock. Time to expiry qualifies as an instrument because it does not directly affect liquidity in the underlying stock, but only indirectly through the hedging channel.	89
3.5	The exogenous component of volume. The figure shows the exogenous component of volume $Volume_t$ versus time to expiry.	91

List of Tables

1.1	Execution fees for selected US trading venues, 2013-2014. This table shows the “maker” and “taker” prices charged to brokers and other financial intermediaries who trade at the venues given in column one. Fees are expressed as cents per 100 shares transacted, and a negative fee means a rebate. Venues are sorted by taker fee. <i>Venue</i> gives the name of the venue. <i>Transparency</i> is the nature of pre-trade price transparency on the venue, i.e., whether participants can observe the certain presence of a counterparty and the price at which it is willing to trade before a trade occurs. If this is unobservable, the transparency column reports dark. <i>Taker fee</i> is the fee (rebate) for trading immediately at the venue. <i>Maker fee</i> is the fee (rebate) for offering to trade on the venue with another counterparty. There is no maker fee at most dark venues because trades are executed immediately at the venue or not at all. <i>Time period</i> is the time period during which the price is representative of the venues pricing. <i>Source</i> is the data source used to obtain the price.	3
1.2	Market Quality Measures.	8
1.3	Summary statistics for treatment and control stocks. This table gives summary statistics on market quality and market cap for the 35 stocks identified as treatment stocks and the 35 matched control stocks. The columns of the table give the average, standard deviation, minimum, 25th percentile, 50th percentile, 75th percentile, and maximum for each measure. Panel A shows summary statistics for treatment stocks before the launch of the RLP, from April 2012 until July 2012, and Panel B shows summary statistics for treatment stocks after the launch of the RLP, August 2012 until August 2013. Panel C shows summary statistics for control stocks before the launch of the RLP, and Panel D shows summary statistics for control stocks after the launch of the RLP. Panel E shows the difference in means for each variable for both treatment and control stocks. <i>Volume</i> is the average number of shares traded per day in thousands of shares. <i>RLP Volume</i> is the average number of shares traded in the RLP per day in thousands of shares. <i>Relative Spread</i> is the average relative spread. <i>Effective Spread</i> is the average five-second effective spread. <i>Price Impact</i> is the average five-second price impact. <i>Autocorrelation</i> is the average daily absolute five-second autocorrelation of the midquote. <i>Market Cap</i> is average market capitalization over the period in billions	10

1.4 **Information shares.** This table gives summary statistics and results for a T-test on difference in means for information shares computed using a vector autoregression model. Information shares are computed monthly for each of the 35 treatment stocks. Panel A reports summary statistics. The columns of the table give the average, standard deviation, 25th percentile, 50th percentile and 75th percentile for each segment of the order flow. Lit is the information share of lit orders; RLP is the information share of RLP orders; Total order flow is the information share of all undifferentiated orders; RLP and lit is the sum of information shares for RLP and lit orders. Panel B reports the average difference between information shares for various segments of the order flow. Difference of lit, RLP is the difference between lit and RLP information shares; Difference of total and lit plus RLP is the difference between the sum of RLP and lit minus the aggregate information shares. N is the number of observations. The t-statistic is in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% level. 16

1.5 **Difference-in-differences event study of the impact of the launch of the RLP on relative bid-ask spreads.** The rows of the table give the regression coefficients and their associated t-statistics for specific variables across six different specifications of the event study on relative bid-ask spreads. A blank entry indicates exclusion from the regression. The columns of the table correspond to different specifications of the event study. In specification (1), only treatment stocks are included in the regression; in specification (2), treatment stocks and control stocks are included; in specification (3), *Market cap* and *Volume* are included; and so on. *Treatment* is a dummy variable that equals one during the period after the launch of the RLP for treatment stocks. *After* is a dummy variable that equals one during the period after the launch of the RLP for all stocks. *Market cap* is the daily market capitalization in billions. *Volume* is the number of shares traded per day in thousands of shares. *Market-wide liquidity* is the stock-specific factor score from principal component analysis. *10 – dayvolatility* is the 10-day rolling volatility of the midquote. *Lagged relative spread* is the relative bid-ask spread lagged by one day. *Constant* is the constant of regression. N is the number of observations. R^2 is the coefficient of determination. *, **, *** represent statistical significance at the 10%, 5%, and 1% level. Panel A shows results for the entire sample period, from April 2012 to August 2013. Panels B through E show results for a sample period limited to three months prior to the launch of the RLP and three months after. Panel B shows results when the sample period is limited to Q3 2012 and Q4 2012; Panel C shows to Q3 2012 and Q1 2013; Panel D shows Q3 2012 and Q2 2013; and Panel E shows Q3 2012 and Q3 2013. Each regression specification (1) to (6) in Panels B through E corresponds to those in Panel A, but we exclude reporting of variables other than Treatment and After for brevity. 19

1.6 **Difference-in-differences event study of the impact of the launch of the RLP on effective spreads.** The rows of the table give the regression coefficients and their associated t-statistics for specific variables across six different specifications of the event study on effective spreads. A blank entry indicates exclusion from the regression. The columns of the table correspond to different specifications of the event study. In specification (1), only treatment stocks are included in the regression; in specification (2), treatment stocks and control stocks are included; in specification (3), *Market cap* and *Volume* are included; and so on. *Treatment* is a dummy variable that equals one during the period after the launch of the RLP for treatment stocks. *After* is a dummy variable that equals one during the period after the launch of the RLP for all stocks. *Market cap* is the daily market capitalization in billions. *Volume* is the number of shares traded per day in thousands of shares. *Market-wide liquidity* is the stock-specific factor score from principal component analysis. *10-day volatility* is the 10-day rolling volatility of the midquote. *Lagged effective spread* is the five-second effective spread lagged by one day. *Constant* is the constant of regression. *N* is the number of observations. R^2 is the coefficient of determination. *, **, *** represent statistical significance at the 10%, 5%, and 1% level. Panel A shows results for the entire sample period, from April 2012 to August 2013. Panels B through E show results for a sample period limited to three months prior to the launch of the RLP and three months after. Panel B shows results when the sample period is limited to Q3 2012 and Q4 2012; Panel C shows Q3 2012 and Q1 2013; Panel D shows Q3 2012 and Q2 2013; and Panel E shows Q3 2012 and Q3 2013. Each regression specification (1) to (6) in Panels B through E corresponds to those in Panel A, but we exclude reporting of variables other than *Treatment* and *After* for brevity. 22

1.7 **Difference-in-differences event study of the impact of the launch of the RLP on relative five-second price impacts.** The rows of the table give the regression coefficients and their associated t-statistics for specific variables across six different specifications of the event study on five-second price impacts. A blank entry indicates exclusion from the regression. The columns of the table correspond to different specifications of the event study. In specification (1), only treatment stocks are included in the regression; in specification (2), treatment stocks and control stocks are included; in specification (3), *Market cap* and *Volume* are included; and so on. *Treatment* is a dummy variable that equals one during the period after the launch of the RLP for treatment stocks. *After* is a dummy variable that equals one during the period after the launch of the RLP for all stocks. *Market cap* is the daily market capitalization in billions. *Volume* is the number of shares traded per day in thousands of shares. *Market-wide liquidity* is the stock-specific factor score from principal component analysis. *10-day volatility* is the 10-day rolling volatility of the midquote. *Lagged price impact* is the five-second relative price impact lagged by one day. *Constant* is the constant of regression. *N* is the number of observations. R^2 is the coefficient of determination. *, **, *** represent statistical significance at the 10%, 5%, and 1% level. Panel A shows results for the entire sample period, from April 2012 to August 2013. Panels B through E show results for a sample period limited to three months prior to the launch of the RLP and three months after. Panel B shows results when the sample period is limited to Q3 2012 and Q4 2012; Panel C shows Q3 2012 and Q1 2013; Panel D shows Q3 2012 and Q2 2013; and Panel E shows Q3 2012 and Q3 2013. Each regression specification (1) to (6) in Panels B through E corresponds to those in Panel A, but we exclude reporting of variables other than *Treatment* and *After* for brevity. 24

1.8 **Difference-in-differences event study of the impact of the launch of the RLP on the absolute value of five-second return autocorrelations.** The rows of the table give the regression coefficients and their associated t-statistics for specific variables across six different specifications of the event study on five-second return autocorrelations. A blank entry indicates exclusion from the regression. The columns of the table correspond to different specifications of the event study. In specification (1), only treatment stocks are included in the regression; in specification (2), treatment stocks and control stocks are included; in specification (3), Market cap and Volume are included; and so on. *Treatment* is a dummy variable that equals one during the period after the launch of the RLP for treatment stocks. *After* is a dummy variable that equals one during the period after the launch of the RLP for all stocks. *Market cap* is the daily market capitalization in billions. *Volume* is the number of shares traded per day in thousands of shares. *Market-wide liquidity* is the stock-specific factor score from principal component analysis. *10-day volatility* is the 10-day rolling volatility of the midquote. *Lagged autocorrelation* is the absolute five-second autocorrelation lagged by one day. *Constant* is the constant of regression. *N* is the number of observations. R^2 is the coefficient of determination. *, **, *** represent statistical significance at the 10%, 5%, and 1% level. Panel A shows results for the entire sample period, from April 2012 to August 2013. Panels B through E show results for a sample period limited to three months prior to the launch of the RLP and three months after. Panel B shows results when the sample period is limited to Q3 2012 and Q4 2012; Panel C shows Q3 2012 and Q1 2013; Panel D shows Q3 2012 and Q2 2013; and Panel E shows Q3 2012 and Q3 2013. Each regression specification (1) to (6) in Panels B through E corresponds to those in Panel A, but we exclude reporting of variables other than Treatment and After for brevity. 26

1.9	Difference-in-differences event study of the RLPs impact on market quality with a weighted panel of controls. The rows of the table give the regression coefficients and their associated t-statistics for specific variables across six different specifications of the event study for four market-quality measures. A blank entry indicates exclusion from the regression. The columns of the table correspond to different specifications of the event study. In specification (1), only treatment stocks are included in the regression; in specification (2), treatment stocks and control stocks are included; in specification (3), <i>Market cap</i> and <i>Volume</i> are included; and so on. Rather than using control stocks matched on-to-one with treatments (reported in Tables 3 through 6) a weighted panel of controls is used for each treatment stock. <i>Treatment</i> is a dummy variable that equals one during the period after the launch of the RLP for treatment stocks. <i>After</i> is a dummy variable that equals one during the period after the launch of the RLP for all stocks. <i>Market cap</i> is the daily market capitalization in billions. <i>Volume</i> is the number of shares traded per day in thousands of shares. <i>Market-wide liquidity</i> is the stock-specific factor score from principal component analysis. <i>10-day volatility</i> is the 10-day rolling volatility of the midquote. <i>Lagged autocorrelation</i> is the absolute five-second autocorrelation lagged by one day. <i>Constant</i> is the constant of regression. <i>N</i> is the number of observations. <i>R</i> ² is the coefficient of determination. *, **, *** represent statistical significance at the 10%, 5%, and 1% level. Panels A through D show results for relative spread, effective spread, price impact and autocorrelation over the entire sample period, from April 2012 to August 2013.	29
2.1	Trading outcomes, probabilities and market maker profits. This table shows combinations of the informed trader’s (Inf.) order type, whether the order is fast or slow, and the noise traders’ (Pass. N. and Act. N) actions, which limit orders are at the “top” of the order book (labelled Best bid and Best ask), each combination’s probability, and the market maker’s (MM) profit given an outcome. Bold entries in the Best bid and Best ask columns indicate the executed limit order. We have left out certain entries when they are irrelevant for computing expected profits. For example, when the informed trader submits a fast limit order, the passive noise trader’s order will never be executed, so the row combines the probabilities of the passive noise trader submitting a buy or a sell. The table is for outcomes given the true value of the asset is high and hence the high probability of a market order \bar{p} is used.	50
3.1	Summary Statistics. The rows of this table display summary statistics for the stocks in the sample period, stocks that were part of the S&P 100 for the 2014 calendar year.	85
3.2	Summary Statistics Expiration Days. The rows of the table shows summary statistics as in Table 1, only for trading days that were the last before an option expiry date. <i>Volatility</i> has been excluded since it is computed only using all observations in the sample period.	86

3.3	Summary Statistics Non-Expiration Days. This rows of this table shows summary statistics as in Table 1, excluding trading days that were the last before an option expiry date. <i>Volatility</i> has been excluded since it is computed only using all observations in the sample period.	86
3.4	Instrumental Variables Regression - First Stage. The columns of this table show the regression coefficients for the first stage of a two-stage instrumental variables methodology where volume is regressed on a set of covariates.	90
3.5	Instrumental Variables Regression. The columns of this table show the regression coefficients for the final stage of a two-stage instrumental variables methodology where four market quality measures are regressed on fitted values of trading volume and a set of covariates. \widehat{Volume} is the exogenous component of volume constucted from the first stage regression.	92
A.1	Conditional Probabilities. This table shows the probabilities of each of the probabilities $\kappa_{\bar{v}}$, which are conditional on the asset's true value.	100
B.1	Numerical simulation of convergence to the steady state. This table shows the mean and standard deviation of the impact coefficient for random initial conditions and after 50 Bayesian updates, as well as the model's predicted value for the steady state.	115

List of Appendices

Appendix A Chapter 2 Proofs	99
Appendix B Chapter 3 Proofs	112

Chapter 1

Retail Order Flow Segmentation: Empirics

1.1 Introduction

A major driver of market-structure innovation is the value of knowing whether a potential counterparty is a desirable trading partner. Counterparties such as hedge funds can be advantaged and tend to buy right before prices rise or sell right before prices fall. To limit the risk of losses, securities traders have incentives to do business on markets that restrict access to safer kinds of traders. In this paper, we study such a market that provides securities traders with access to a relatively safe class of counterparty, retail equity traders.¹ Markets that control who may participate are of interest because they are engaging in segmentation, and segmentation creates a tension. Although protected counterparties do receive improved trading opportunities, traders in the wider market may be left with poorer trading opportunities. Accordingly, regulators have expressed concerns that segmentation may be detrimental to overall market quality (International Organization of Securities Commissions (IOSCO) 2010 and Ontario Securities Commission (OSC) 2010).

We address questions about the market-quality effects of segmentation by studying the launch of a new trading facility that enables segmentation of retail trades. In August 2012, the New York Stock Exchange (NYSE) launched the Retail Liquidity Program (RLP). The RLP enables members to quote dark (non-displayed) limit orders that can be filled only by market orders that originate from retail traders. We study the RLP using Trade and Quote (TAQ) data from the NYSE and a difference-in-differences methodology. Our results show the RLP led to a mild but positive impact on overall market-quality measures. After launch, measures of the bid-ask spread, price impact and price efficiency improved in samples of four calendar quarters after its launch. We explain the market-quality results by showing the RLP improved the price-discovery process. A Hasbrouck (1991) vector autoregression (VAR) on the order flow shows the RLP improved market participants ability to forecast prices by distinguishing nonretail trades, which have greater predictive power.

The results are important because they provide a test case for recent theory on segmentation and market quality. We choose to test hypotheses drawn explicitly from the model in

¹Retail traders are individuals who trade for their personal account.

Zhu (2014). Our results confirm the model prediction that segmentation can improve price discovery, but they do not confirm the prediction that segmentation can hurt liquidity. One limitation of the model is that it is static and cannot study the potential for better price discovery to improve long-term liquidity. Our evidence is more consistent with papers that do allow an increase in information-based trade to have a dynamic effect on liquidity, such as new work in Rosu (2015).

The results are also important because they inform a current regulatory discussion. Data from Bloomberg shows that in 2016 so far 35% of US equity volumes have been executed outside of traditional exchanges. Concerns about the high share of trades executed in “dark pools” operated by dealers or networks of brokers have motivated securities regulators in the US, Canada and Australia to consider rules limiting the ability of traders to transact off-exchange. Our results are useful in regulatory discussions as they are part of a literature speaking to the positive aspects of off-exchange trade.

We test the effects of retail segmentation using data from the NYSE. The dataset is the Trade and Quote (TAQ) data from the NYSE in a span around the RLP launch date. The data contain information on all trades and the best bid and ask quotes and sizes on all stocks traded on the NYSE and NYSE Arca, an exchange owned and operated by the NYSE, with millisecond timestamps. RLP trades are identified by having subpenny prices, that is, prices that take values off the usual tick grid of one cent. At the time, subpenny trades were not otherwise possible on NYSE and NYSE Arca.

The data provide a good test of segmentation due to the fee level at the RLP. The fee is important because brokers search venues in a “pecking order” partly determined by trading fees (Menkveld, Yueshen, & Zhu, 2015). Table 1.1 gives a price schedule for US equity trading venues in 2013-2014. The source “Mechanical Markets” is an R script written by Kipp Rogers that searches US Securities and Exchange Commission (SEC) 606 reports and infers the volume-weighted average rebate paid by dark pools.² 2014 is the earliest date at which these reports exist with sufficient venue granularity.

The NYSE RLP is situated in the middle of the schedule, below dark pools in the “pecking order” but above traditional exchanges such as the NYSE. The RLP is competitive for order flow that would have been routed to traditional exchanges but not with off-exchange dark pools. As a consequence, retail liquidity programs have not captured a substantial share of the volumes executed in dark pools. The main reason is a trading fee of zero is not competitive with rebates. Another reason is that exchange-based trade carries legal risks. Regulation prevents exchanges from indemnifying brokers against operational errors such as those that occurred during the Facebook IPO, whereas dark pool operators are liable for trading errors.

We use the data to test four hypotheses drawn from Zhu (2014). The first and second hypotheses are that RLP trades contribute less to price discovery than non-RLP (hereafter referred to as “lit”) and that distinguishing between RLP trades and lit trades aids the price-discovery process. The first two hypotheses are tested to ensure trades on the NYSE are relatively more informed than trades on the RLP and that distinguishing the two improves price discovery. These are outcomes necessary to the market-quality prediction in Zhu (2014). Having verified the necessary conditions for the market-quality predictions, we then test the predictions. The third and fourth hypotheses are that price efficiency should improve while liquidity for lit trades

²Mechanical Markets is available online here: www.mechanicalmarkets.wordpress.com.

Table 1.1: **Execution fees for selected US trading venues, 2013-2014.** This table shows the “maker” and “taker” prices charged to brokers and other financial intermediaries who trade at the venues given in column one. Fees are expressed as cents per 100 shares transacted, and a negative fee means a rebate. Venues are sorted by taker fee. *Venue* gives the name of the venue. *Transparency* is the nature of pre-trade price transparency on the venue, i.e., whether participants can observe the certain presence of a counterparty and the price at which it is willing to trade before a trade occurs. If this is unobservable, the transparency column reports dark. *Taker fee* is the fee (rebate) for trading immediately at the venue. *Maker fee* is the fee (rebate) for offering to trade on the venue with another counterparty. There is no maker fee at most dark venues because trades are executed immediately at the venue or not at all. *Time period* is the time period during which the price is representative of the venues pricing. *Source* is the data source used to obtain the price.

Venue	Transparency	Taker fee	Maker fee	Time period	Source
Two Sigma	Dark	-18	-	2014	Mechanical Markets
UBS	Dark	-15	-	2014	Mechanical Markets
KCG	Dark	-11	-	2014	Mechanical Markets
Goldman Sachs	Dark	-10	-	2014	Mechanical Markets
Citadel	Dark	-9	-	2014	Mechanical Markets
BATS Y	Lit	-5	7	2013	Mechanical Markets
Direct Edge A	Lit	-4	6	2013	Mechanical Markets
NYSE RLP	Dark	0	3	2013	-
NYSE	Lit	25	-15	2013	O’Donoghue (2015)
BATS Z	Lit	29	-25	2013	O’Donoghue (2015)
NYSE Arca	Lit	30	-21	2013	O’Donoghue (2015)
Nasdaq	Lit	30	-20	2013	O’Donoghue (2015)
Direct Edge X	Lit	30	-21	2013	O’Donoghue (2015)

should deteriorate. We find price efficiency improves but liquidity for lit trades also slightly improves. The result is theoretically unexpected, so we discuss potential extensions to theory in the conclusions.

The methodology used to test the first and second hypotheses, which concern the informational characteristics of order flow, is structural VAR. We fit structural VAR models on returns and order flows and analyze how the RLP and lit components of the order flow contribute to price discovery. In the data, an impulse of lit trades causes a visibly larger response in the log return than does an impulse of RLP trades. We also compute the Hasbrouck (1991) information share of the RLP and lit order flows and show the sum is greater than that of the undifferentiated flows. Put differently, the segmented order flow is a better predictor of the price than the undifferentiated order flow. For both RLP and lit flow, the impact on return decays quickly on average after 10 minutes. Our interpretation is that the RLP is aiding price discovery at the 10-minute horizon.

The methodology used to test the third and fourth hypotheses, which concern market quality, is the difference-in-differences event study. During the sample period of our dataset, the RLP was launched on the NYSE's main exchange (simply referred to as the NYSE) but not NYSE Arca (simply referred to as Arca). We use stocks that traded only on Arca (and not on the NYSE) as a control group. Stocks that traded on Arca were eventually eligible for an RLP that launched in 2014. Overall, we find the RLP leads to a slight improvement in four standard market-quality measures: relative bid-ask spreads improve by around one basis point from an unconditional average of 12 basis points; effective spreads improve by around half a basis point from an unconditional average of 10 basis points; price impact decreases by half a basis point from an unconditional average of 3.5 basis points; and the return autocorrelation decreases by around 0.01 from an unconditional average of 0.06.

The results are economically small in size, likely because treatment stocks have an average of only 3.5% of trading volume in the RLP. To demonstrate the results are nevertheless robust, we use several event-study specifications. We examine the results using both the simple, single-difference event study and also the difference-in-differences event study. We estimate each difference-in-differences regression using six specifications that successively include more control variates. We fit the difference-in-differences model once over the entire sample and again over four within-period subsamples. Last, to ensure our selection of control stocks is robust, we construct a weighted panel of control stocks for each treatment stock in our sample and fit the event studies again. For each of the model specifications above, the general result of the paper persists: the RLP results in a slight improvement in market quality. The robustness exercises show that, in order to believe the impact is not present, one would have to believe another factor affected four market-quality measures on sets of stocks on the NYSE but not Arca around the launch of the RLP, a factor that is not explained by fixed effects, lags or common liquidity determinants, and a factor that persisted both throughout the sample and equally in each of the within-period subsamples.

In the market microstructure literature, order-flow segmentation is no longer novel and is studied in classic work by Easley, Kiefer, and O'Hara (1996), Battalio and Holden (2001) and Parlour and Rajan (2003). Recently, a substantial literature has arisen concerning dark pools that addresses segmentation directly and indirectly. Papers on the topic include Boni, Brown, and Leach (2013), Fleming and Nguyen (2013), Ready (2014), Buti, Rindi, and Werner (2010), Nimalendran and Ray (2014), and Degryse, Tombeur, Van Achter, and Wuyts (2013). Rela-

tive to the literature, our paper makes an incremental contribution by studying order-flow segmentation in a new market structure. Segmentation may have different costs and benefits on exchanges than it does among private broker networks or single-dealer platforms; these dark pools are market structures that are opaque and order-driven, whereas the RLP segments order flow within the competitive and quote-driven environment of the stock exchange. Battalio (1997) finds that bid-ask spreads tighten when a broker purchases order flow for execution off-exchange. Our findings are similar for exchange-based segmentation. A second incremental contribution relative to the literature is the study of a dataset in which it is possible to identify the trades of the segmented parties. Using a VAR model we can verify that segmentation increases the predictive power of the order flow, as is often supposed. Last, our third contribution to the literature is a result that is theoretically unexpected. Theory and intuition often predict that the outcome of segmentation should be to worsen market liquidity for a particular segment of the market, and we find the opposite effect.

Our paper is one of a number of papers that identify benefits to segmentation. Most papers focus on the context of darkness (lack of pre-trade transparency) rather than segmentation, but the two often go together. Foley and Putniņš (2016) is a related study that investigates restrictions on all dark price improvement in Canada. They find that dark trading benefits market quality by reducing quoted, effective and realized spreads and increasing informational efficiency. Using a regulatory dataset, Comerton-Forde, Malinova, and Park (2016) find the same restriction on dark trading drove liquidity from the dark market into the lit, resulting in higher fees for retail brokers and higher rebates for high-frequency market makers. Boni et al. (2013) study dark pools with participation constraints and find that stronger constraints lead to less serial correlation in returns, volume and volatility tend to lead other markets to a smaller degree, and more trade clustering occurs across days. Our study is distinguished from the above by studying the launch of a single trading facility rather than a rule change that indiscriminately affected multiple types of trading venue in different ways. We are also distinguished by studying a facility in the larger and more liquid US market.

The remainder of the paper is organized as follows: Section 1.2 describes our hypotheses in detail; Section 1.3 describes the data; Section 1.4 gives details on the methodology; Section 1.5 discusses the results; and Section 1.5.1 offers some conclusions.

1.2 Hypotheses

There are four hypotheses tested in the paper. The first two hypotheses are about how the RLP alters the informational character of the order flow. The second two hypotheses are about how the RLP impacts market quality.

The motivation for the hypotheses derives from Zhu (2014), which models traders choices to use either a dark midquote crossing facility or a traditional exchange.³ As is common in microstructure, there are three types of agent: informed traders, uninformed traders and exchange-based market makers. The dark crossing facility matches buy and sell market orders at the exchanges midquote. Traders strategically choose venues in equilibrium. In the crossing facility, execution is not certain, since there can be more buy orders than sell orders

³A midquote crossing facility matches non-displayed buy and sell orders at the midpoint of the best displayed bid and offer prices posted on other exchanges.

or vice versa. The uncertainty of execution in the crossing facility discourages the participation of informed agents more than it does the uninformed, because the information possessed by informed agents is short-lived. Thus the dark crossing facility endogenously segments the market, concentrating informed activity in the public exchange. Due to the concentration of informed activity, price efficiency is better, and liquidity on the exchange is worse.

We motivate hypotheses using Zhu (2014) because the NYSE RLP is much like the dark crossing network in the model. It guarantees price improvement versus the main exchange, and it presents execution risk since limit orders are not displayed. Unlike the model crossing network, the NYSE RLP segments the market exogenously; liquidity is only accessible by brokers executing retail market orders. Nevertheless, we believe the theory is a good match because the active mechanism in Zhu (2014) is segmentation. The role of darkness in the model is to create the execution risk that incentivizes the segmentation. In a sense, the data are ideal, since it is already true by assumption that segmentation occurs, so it is possible to test the predicted impact of the segmentation directly. This leads us to the following two hypotheses:

Hypothesis 1.2.1 *The RLP order flow is less informed than the non-RLP order flow.*

Hypothesis 1.2.2 *Segmentation improves the informativeness of the total order flow.*

To ensure the RLP segmentation is between the more- and less-informed components of the order flow, we first test the hypothesis that RLP flow is indeed less informative than lit. Then we test whether the facility increases the informativeness of the order flow overall. If so, the segmentation does offer a superior way to discover prices from the order flow, as in the model. We follow with the hypotheses on the impact:

Hypothesis 1.2.3 *Participation in the RLP affects a stocks liquidity.*

Hypothesis 1.2.4 *Participation in the RLP affects a stocks price efficiency.*

The removal of the retail order flow to the RLP concentrates the more informed order flow on the main exchange, which should improve price efficiency. However, more informed order flow is more costly to fill. Market makers could compensate by widening bid-ask spreads on the main exchange. The third and fourth hypotheses ask whether these two impacts result from the first and second hypotheses.

It is not clear that the outcome will be as in Zhu (2014). One limitation of the model is that it is static. The model shows the option to trade in the dark concentrates informed agents on the exchange, which otherwise resembles the classic limit-order market modelled in Glosten and Milgrom (1985). The impact of concentrating informed agents on an exchange is given in Glosten and Milgrom (1985) Proposition 5, which also points to dynamic effects. Although an increase in informed activity has the immediate impact of increasing the bid-ask spread, future spreads are tighter as informational differences between the informed agents and the market maker decrease more quickly. This intuition is formalized in Rosu (2015), who predicts that an increase in informed traders information results in an immediate increase in bid-ask spreads followed by a decrease in bid-ask spreads, which occurs at a speed proportional to the degree of informed trading, as in Glosten and Milgrom (1985). It is possible the same economic mechanism could be active on the RLP, resulting in superior price efficiency as well as superior liquidity.

1.3 Data

Our dataset contains information on trades and best bid and ask quotes on all stocks traded on the NYSE and Arca for 333 trading days from 1 April 2012 to 1 August 2013. The data consist of time-stamped reports of all trade prices and quantities and time-stamped reports of all best bid and best ask prices and quantities, for each stock and exchange. The trades are not marked by the sign of trade (buyer- or seller-initiated), so we impute the sign of trade using the Lee and Ready (1991) algorithm.⁴

We mark trades on the NYSE that have subpenny prices after 1 August 2012 as RLP and all other trades are marked lit. The NYSE reports that no trades can take a subpenny price on the NYSE's main venue except via the RLP. Indeed, there were no trades on the NYSE before 1 August 2012 in our sample that had subpenny prices. This approach may slightly underestimate the total activity in the RLP. For example, for stocks with bid-ask spreads greater than one cent, RLP trades could occur at prices on the regular tick grid of one cent.

1.3.1 Treatment stocks

Treatment stocks are defined as stocks that had at least 1% RLP volume share. We choose 35 treatment stocks using the following criteria. Before choosing treatment stocks, we sample the data. We drop all small-cap stocks (stocks with a market capitalization under US\$2 billion), all exchange-traded funds and all share classes other than common equity. Stocks that were cross-listed in Canada were removed from the sample, since they were eligible for a similar program to the RLP, Alpha IntraSpread. We drop stocks that had a minimum price below \$2.00 at any time during the sample period, since they may have been eligible for subpenny pricing on the NYSE due to their low price. We drop stocks that are eligible for a separate RLP on another NYSE-operated exchange, NYSE MKT. Before sampling, the data included trades and quotes for 3,993 stocks that trade on the NYSE. After removing small-cap stocks, exchange-traded funds, cross-listed stocks, non-common equity share classes and stocks with a low price, 2,265 stocks remained.

Of the 2,265 sample stocks, 49 had an average of over 1% of total volume on the NYSE that traded in the RLP. Of these 49 stocks with relatively heavy usage of the RLP, 14 had sparsely populated data, that is, fewer than 10 days of complete data before and after the launch of the RLP. We designate the remaining 35 stocks the sample of treatment stocks.

1.3.2 Control stocks

We create a pool of control stocks from the set of all stocks that were traded on Arca and not on the NYSE and therefore were ineligible for the RLP. From the set of stocks that traded on Arca and not on the NYSE, we sample using the same criteria used to define treatment stocks (except for the threshold for RLP activity). There are 184 candidate control stocks that fit the criteria. We create a matched sample by pairing treatment stocks one-to-one with control stocks. For each treatment stock, we select the nearest neighbour by average sample market capitalization

⁴The Lee and Ready (1991) algorithm assigns buyer- and seller-initiated designations to trade and quote data using trading prices, prevailing quotes, and the sequence of previous trade initiations.

Table 1.2: **Market Quality Measures.**

Relative bid-ask spread	$\frac{ask-bid}{midquote}$
Effective spread	$2sign\ of\ trade\ \frac{price-midquote}{midquote}$
Five-second price impact	$sign\ of\ trade\ \frac{midquote_{t+5seconds}-midquote_t}{midquote_t}$
Five-second return autocorrelation	$corr(return_{t+5seconds}, return_t)$

without replacement. Measures of liquidity for certain control stocks on 26 December 2012 were extreme, so we drop the day from the sample.

1.3.3 Data for computing information shares: returns and order flow

For the purposes of computing information shares, we compute a return variable and three order-flow variables: five-minute log midquote returns, five-minute net RLP order flow, five-minute net lit order flow, and the five-minute net undifferentiated order flow (RLP and lit together), for days after the launch of the RLP on 1 August 2012. Five-minute log midquote returns are computed for every five-minute time increment by taking the log of the ratio of the midquote to the five-minute-lagged midquote. Order flow variables are computed for each five-minute time increment by summing trading volumes within the period, where buyer-initiated trades and seller-initiated trades are signed positive and negative, respectively. All order flow variables are signed positive for net buying and negative for net selling. The choice of five minutes results in a granular set of observations while ensuring the RLP net order flow is non-zero for most time intervals.

1.3.4 Data for conducting difference-in-differences: market quality measures

For the purposes of running the event studies, we compute daily averages of four standard market-quality measures: the relative bid-ask spread, the effective spread, the five-second price impact and the five-second return autocorrelation. The first three measure liquidity, while the return autocorrelation is used to measure price efficiency. The choice of five seconds is standard for price impacts and autocorrelations. For these measures we use the standard formulae given in Table 1.2. The subscripts t and $t + 5$ seconds denote observations at a particular time t and five seconds later; bid , ask and $midquote$ denote the best bid and ask prices, and their mean, the midquote; $sign\ of\ trade$ denotes whether a trade was buy- or sell-initiated as computed using the Lee and Ready (1991) algorithm; $return$ denotes the five-second log midquote return; and $corr$ denotes the correlation operator.

We eliminate trades from the sample that are flagged as occurring during the opening or closing auctions. We take daily averages of the above measures over standard trading hours, from 9:30 a.m. to 4:00 p.m.

1.3.5 Summary statistics

Table 1.3 provides summary statistics on market quality and market capitalization for the 35 stocks identified as treatment stocks and the 35 matched control stocks. Panel A shows summary statistics for treatment stocks before the launch of the RLP, from April 2012 until July 2012, and Panel B shows summary statistics for treatment stocks after the launch of the RLP, from August 2012 until August 2013. Panel C shows summary statistics for control stocks before the launch of the RLP, and Panel D shows summary statistics for control stocks after the launch of the RLP. The columns of the table give the average, standard deviation, minimum, 25th percentile, 50th percentile, 75th percentile and maximum for each market-quality measure and for market capitalization.

Volume is the average number of shares traded per day in thousands of shares. RLP Volume is the average number of shares traded in the RLP per day in thousands of shares. Relative Spread is the average daily relative spread. Effective Spread is the average daily effective spread. Price Impact is the average daily five-second price impact. Autocorrelation is the average daily absolute five-second autocorrelation of the midquote.

For treatment stocks, average volume decreased from 2,541K shares per day to 2,492K after the launch of the RLP. RLP volume after launch was 89K shares per day or roughly 3.5% of total volume. Since overall volume decreased, it is unlikely that the RLP attracted new order flow that was previously traded off the NYSE. Each of the liquidity measures improved for treatment stocks after the launch of the RLP. The average relative bid-ask spread decreased from 12.1 to 10.5 basis points. Similarly, the average effective spread decreased from 9.9 basis points to 8.7 basis points, while the average price impact decreased from 3.6 basis points to 3.2 basis points. Average absolute autocorrelation remained at 0.06. Average market capitalization increased after the launch of the RLP from \$188.24 to \$237.71 billion.

For control stocks, average volume decreased from 311K to 243K shares per day after the launch of the RLP. Control stocks had less volume than treatment stocks across the sample period. This is a weakness of the control group and one reason we also perform the regression using a weighted panel of control stocks (Table 1.9). The weighted panel ensures the result is not spuriously driven by a particular selection of the 35 control stocks from the 184 candidates.

Liquidity measures for control stocks were relatively unchanged when compared to the liquidity measures of the treatment stocks. The relative average bid-ask spread increased from 12.8 to 12.9 basis points. The average effective spread decreased from 7.2 to 6.8 basis points. Average price impact increased from 2.5 to 2.8 basis points. Average absolute autocorrelation increased from 0.09 to 0.10. Market capitalization for control stocks increased after the launch of the RLP from \$182.44 to \$215.06 billion.

Panel E shows the difference in the means of market-quality measures and market-quality factors before and after the launch of the RLP for treatment and control stocks. Volume decreased for both treatment and control stocks over the sample, while market capitalization increased. Each of the liquidity measures for treatment stocks decreased after the launch of the RLP: the relative spread by 1.5 basis points, effective spread by 1.2 basis points and price impact by 0.5 basis points. For control stocks, these liquidity measures had no consistent pattern: the effective spread decreased slightly, by 0.37 basis points, while the relative spread and price impact increased slightly, by 0.11 and 0.24 basis points, respectively. There was no average change in autocorrelation for treatment stocks, while the autocorrelation for control stocks

Table 1.3: **Summary statistics for treatment and control stocks.** This table gives summary statistics on market quality and market cap for the 35 stocks identified as treatment stocks and the 35 matched control stocks. The columns of the table give the average, standard deviation, minimum, 25th percentile, 50th percentile, 75th percentile, and maximum for each measure. Panel A shows summary statistics for treatment stocks before the launch of the RLP, from April 2012 until July 2012, and Panel B shows summary statistics for treatment stocks after the launch of the RLP, August 2012 until August 2013. Panel C shows summary statistics for control stocks before the launch of the RLP, and Panel D shows summary statistics for control stocks after the launch of the RLP. Panel E shows the difference in means for each variable for both treatment and control stocks. *Volume* is the average number of shares traded per day in thousands of shares. *RLP Volume* is the average number of shares traded in the RLP per day in thousands of shares. *Relative Spread* is the average relative spread. *Effective Spread* is the average five-second effective spread. *Price Impact* is the average five-second price impact. *Autocorrelation* is the average daily absolute five-second autocorrelation of the midquote. *Market Cap* is average market capitalization over the period in billions

Panel A: Summary statistics for treatment stocks before the launch of the RLP.

	Mean	Std. Dev.	Min	P25	P50	P75	Max
Volume	2540.88	4039.83	29.75	507.82	1320.61	2537.54	38129.25
Relative Spread	12.05	7.20	2.59	8.35	10.40	13.47	48.63
Effective Spread	9.91	6.59	2.18	6.61	8.49	11.41	53.90
Price Impact	3.64	1.65	-0.21	2.43	3.48	4.54	26.57
Autocorrelation	0.06	0.04	0.00	0.03	0.05	0.08	0.30
Market Cap	188.24	376.56	16.33	28.46	64.45	157.35	2216.54

Panel B: Summary statistics for treatment stocks after the launch of the RLP.

	Mean	Std. Dev.	Min	P25	P50	P75	Max
Volume	2491.85	4896.33	22.69	398.95	1156.48	2351.92	207285.34
RLP volume	89.38	402.89	0.00	1.80	11.15	38.35	9202.98
Relative spread	10.54	5.97	2.05	7.13	9.14	12.15	48.36
Effective spread	8.73	5.48	1.84	5.68	7.46	10.31	41.91
Price impact	3.19	1.43	-3.14	2.20	3.02	3.90	26.08
Autocorrelation	0.06	0.05	0.00	0.03	0.05	0.09	0.50
Market cap	237.71	459.43	12.11	36.30	81.11	167.06	2570.55

Panel C: Summary statistics for control stocks before the launch of the RLP.

	Mean	Std. Dev.	Min	P25	P50	P75	Max
Volume	310.83	446.25	10.73	88.03	169.45	349.81	5431.86
Relative spread	12.83	8.55	2.31	6.80	10.54	16.49	60.56
Effective spread	7.15	4.52	1.81	3.88	5.88	9.07	41.25
Price impact	2.52	1.65	-0.08	1.37	2.07	3.16	13.92
Autocorrelation	0.09	0.06	0.00	0.04	0.08	0.12	0.59
Market cap	182.44	326.78	10.92	28.99	63.32	152.18	1663.42

Panel D: **Summary statistics for control stocks after the launch of the RLP.**

	Mean	Std. Dev.	Min	P25	P50	P75	Max
Volume	243.09	383.16	6.00	68.91	127.13	262.45	7072.84
Relative spread	12.94	11.99	1.69	6.22	10.09	16.47	457.72
Effective spread	6.78	5.33	1.08	3.52	5.32	8.42	108.73
Price impact	2.76	2.09	-0.70	1.48	2.25	3.40	42.24
Autocorrelation	0.10	0.09	0.00	0.04	0.09	0.14	0.79
Market cap	215.06	408.38	13.94	32.99	72.52	159.60	2533.08

Panel E: **Difference in means before and after the launch of the RLP for treatment and control stocks.**

	Treatment stocks	Control stocks
Volume	-49.03	-67.74
Relative spread	-1.51	0.11
Effective spread	-1.18	-0.37
Price impact	-0.45	0.24
Autocorrelation	0	0.01
Market cap	49.47	32.62

increased by 0.01 after the launch of the RLP.

Figures 1.1 and 1.2 show the liquidity history for treatment stocks over the sample period. 1.1 shows the relative bid-ask spread, while 1.2 shows the total and RLP volume in thousands of shares per day.

The relative bid-ask spread is roughly 11 basis points on average in the beginning of the sample and falls to roughly 10 basis points from September 2012 until January 2013. Total volume is roughly constant over the sample period. Volume in the RLP increased steadily from after the RLPs launch in August 2012 until September 2012. RLP volume remains roughly constant at 3.5% of total volume thereafter.

1.4 Methodology

We use two statistical methodologies: structural VAR for our first two hypotheses on the informational characteristics of the order flow, and the difference-in-differences event study to test our second two hypotheses on the effects of the RLP on market quality.

1.4.1 VAR and information shares

We fit the structural VAR model using the five-minute order-flow and return data described above for each month and stock in our sample, starting with August 2012 (the first treatment month) and every subsequent month. We fit two specifications. First, we fit a structural VAR

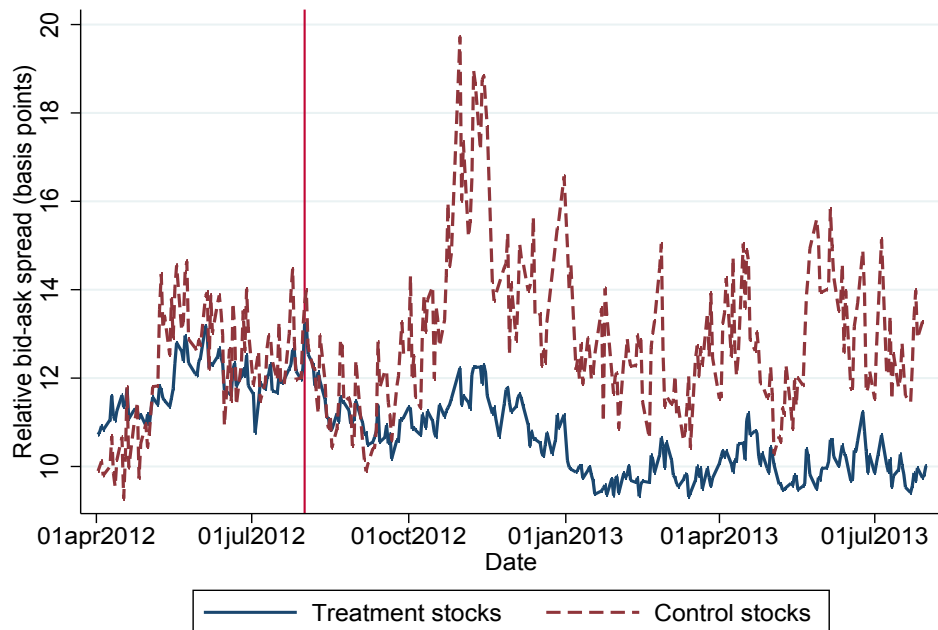


Figure 1.1: **Relative bid-ask spread for treatment and control stocks over the sample period.** This figure shows the relative bid-ask spread for treatment and control stocks over the sample period. The blue line represents relative bid-ask spread for control stock; the red line represents relative bid-ask spread for treatment stocks. The vertical line indicates the launch of the RLP on 1 August 2012.

on return and the total order flow:

$$\begin{aligned}
 r_t &= \alpha_1 + \sum_{\tau=1}^6 \beta_{r,r}^{\tau} r_{t-\tau(5 \text{ min})} + \sum_{\tau=0}^6 \beta_{r,flow}^{\tau} flow_{t-\tau(5 \text{ min})} + \varepsilon_{1,t} \\
 flow_t &= \alpha_2 + \sum_{\tau=1}^6 \beta_{flow,r}^{\tau} r_{t-\tau(5 \text{ min})} + \sum_{\tau=1}^6 \beta_{flow,flow}^{\tau} flow_{t-\tau(5 \text{ min})} + \varepsilon_{2,t}
 \end{aligned} \tag{1.1}$$

where r_t is the five-minute log-return; $flow_t$ is the signed total five-minute net order flow; β are coefficients indexed by variables with subscripts and by time with superscripts; α are constants; and ε are error terms. The variable τ indexes the lag terms, for example, and $t_{\tau}(5 \text{ min})$ means a lag of τ times 5 minutes for $\tau = 0, 1, 2$ etc. We fit the model using six lag terms, hence the upper limit on the summands is 6. Following Hasbrouck (1991), the limit of summation for $flow_t$ in the return process starts from 0 to allow order flow to have a contemporaneous effect on return. The assumption is that prices are driven by order flow and not the reverse. The model imposes that there are no contemporaneous effects between return and itself, flow and itself, and no contemporaneous effect of return on flow.

We also fit a second structural VAR, now on return and the segregated lit flow and RLP

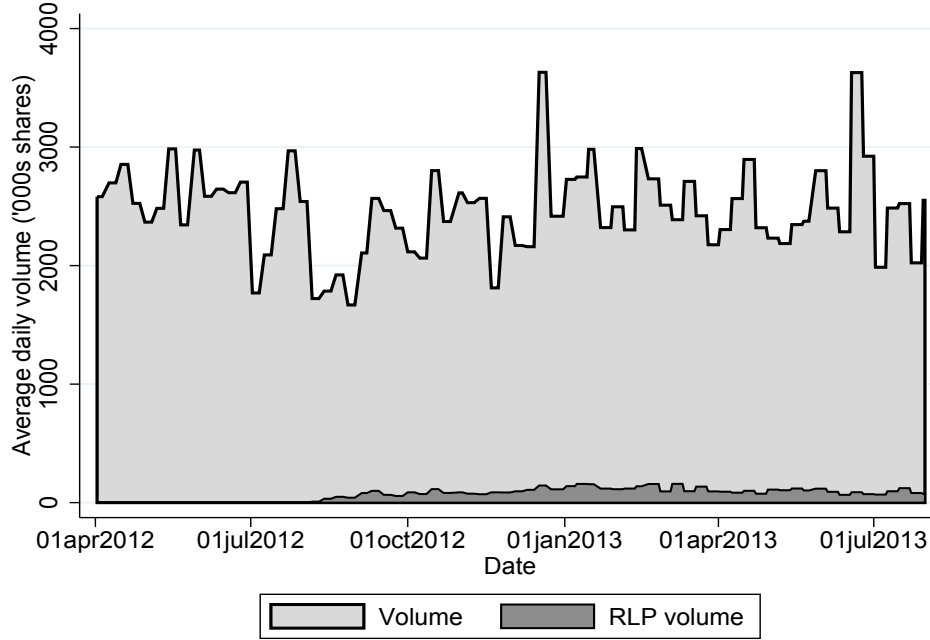


Figure 1.2: **Volume for treatment stocks over the sample period.** This figure shows total volume and RLP volume.

flow:

$$\begin{aligned}
 r_t &= \alpha_1 + \sum_{\tau=1}^6 \beta_{r,r}^\tau r_{t-\tau(5 \text{ min})} + \sum_{\tau=0}^6 \beta_{r,RLP}^\tau RLP_{t-\tau(5 \text{ min})} + \sum_{\tau=0}^6 \beta_{r,lit}^\tau lit_{t-\tau(5 \text{ min})} + \varepsilon_{1,t} \\
 RLP_t &= \alpha_2 + \sum_{\tau=1}^6 \beta_{RLP,r}^\tau r_{t-\tau(5 \text{ min})} + \sum_{\tau=1}^6 \beta_{RLP,RLP}^\tau RLP_{t-\tau(5 \text{ min})} + \sum_{\tau=1}^6 \beta_{RLP,lit}^\tau lit_{t-\tau(5 \text{ min})} + \varepsilon_{2,t} \quad (1.2) \\
 lit_t &= \alpha_3 + \sum_{\tau=1}^6 \beta_{lit,r}^\tau r_{t-\tau(5 \text{ min})} + \sum_{\tau=1}^6 \beta_{lit,RLP}^\tau RLP_{t-\tau(5 \text{ min})} + \sum_{\tau=1}^6 \beta_{lit,lit}^\tau lit_{t-\tau(5 \text{ min})} + \varepsilon_{3,t}
 \end{aligned}$$

where r_t is the five-minute log-return; RLP_t is the net five-minute order flow specifically in the RLP, and lit_t is the net five-minute order flow not in the RLP; and β , α and ε are as above. As in the VAR described above, the limits of summation for RLP_t and lit_t in the return process start from 0 to allow the flows to have a contemporaneous effect on return.

From the VAR models we compute two sets of results: the orthogonalized impulse response functions and the corresponding information shares. The information shares are those used by Hasbrouck (1991) and have been used to assess the contribution of the order flow to price in a variety of settings. Both the impulse response functions and the information shares derive

from the moving-average representations of the VAR models:

$$\begin{pmatrix} r_t \\ flow_t \end{pmatrix} = \sum_{\tau=0}^{\infty} \begin{pmatrix} a_{\tau} & b_{\tau} \\ c_{\tau} & d_{\tau} \end{pmatrix} \begin{pmatrix} \varepsilon_{r,t-\tau} \\ \varepsilon_{flow,t-\tau} \end{pmatrix} \quad (1.3)$$

$$\begin{pmatrix} r_t \\ RLP_t \\ lit_t \end{pmatrix} = \sum_{\tau=0}^{\infty} \begin{pmatrix} e_{\tau} & f_{\tau} & g_{\tau} \\ h_{\tau} & i_{\tau} & j_{\tau} \\ k_{\tau} & l_{\tau} & m_{\tau} \end{pmatrix} \begin{pmatrix} \varepsilon_{r,t-\tau} \\ \varepsilon_{RLP,t-\tau} \\ \varepsilon_{lit,t-\tau} \end{pmatrix} \quad (1.4)$$

where the terms a_T through m_T are the coefficients of the orthogonalized impulse-response functions for step τ , and the ε are serially uncorrelated innovations. The information content of a time series as an explainer of the return process is the cumulation of its associated impulse-response coefficients in the moving-average representation (Hasbrouck, 1991). For each component of the order flow we compute information shares, the proportion of variance of return attributable to the component. For the above models the information shares of the total undifferentiated order flow, RLP order flow, lit order flow and the aggregate segmented (the sum of shares of RLP and lit) order flows are:

$$share_{flow} = \frac{\hat{\sigma}_{flow}^2 (\sum_{\tau=0}^{\infty} b_{\tau})^2}{\hat{\sigma}_r^2 (\sum_{\tau=0}^{\infty} a_{\tau})^2 + \hat{\sigma}_{flow}^2 (\sum_{\tau=0}^{\infty} b_{\tau})^2} \quad (1.5)$$

$$share_{RLP} = \frac{\hat{\sigma}_{RLP}^2 (\sum_{\tau=0}^{\infty} f_{\tau})^2}{\hat{\sigma}_r^2 (\sum_{\tau=0}^{\infty} e_{\tau})^2 + \hat{\sigma}_{RLP}^2 (\sum_{\tau=0}^{\infty} f_{\tau})^2 + \hat{\sigma}_{lit}^2 (\sum_{\tau=0}^{\infty} g_{\tau})^2} \quad (1.6)$$

$$share_{lit} = \frac{\hat{\sigma}_{lit}^2 (\sum_{\tau=0}^{\infty} g_{\tau})^2}{\hat{\sigma}_r^2 (\sum_{\tau=0}^{\infty} e_{\tau})^2 + \hat{\sigma}_{RLP}^2 (\sum_{\tau=0}^{\infty} f_{\tau})^2 + \hat{\sigma}_{lit}^2 (\sum_{\tau=0}^{\infty} g_{\tau})^2} \quad (1.7)$$

$$share_{segmented} = \frac{\hat{\sigma}_{RLP}^2 (\sum_{\tau=0}^{\infty} f_{\tau})^2 + \hat{\sigma}_{lit}^2 (\sum_{\tau=0}^{\infty} g_{\tau})^2}{\hat{\sigma}_r^2 (\sum_{\tau=0}^{\infty} e_{\tau})^2 + \hat{\sigma}_{RLP}^2 (\sum_{\tau=0}^{\infty} f_{\tau})^2 + \hat{\sigma}_{lit}^2 (\sum_{\tau=0}^{\infty} g_{\tau})^2} \quad (1.8)$$

where $\hat{\sigma}_{flow}^2$, $\hat{\sigma}_{RLP}^2$ and $\hat{\sigma}_{lit}^2$ are the estimated variances of the VAR innovations for the respective error terms (using root mean-squared error) for the respective model.

1.4.2 Difference-in-differences

We use the differences-in differences event study to assess the impact of the RLP launch on market quality. The methodology compares the change in market quality for treatment stocks to the change in market quality for control stocks. We used daily averages of the four market-quality measures described above. Specifically, we regress each market-quality measure on a treatment dummy equalling one during the period in which a stock was eligible for the RLP, on an after-period dummy equalling one during the period following the launch of the RLP

for all stocks, on stock and control fixed effects, and on control variates. For all measures, the difference-in-differences model specification is:

$$measure_{i,t} = \beta treatment_{i,t} + \gamma after_t + \delta X_i + FE_i + \varepsilon_{i,t} \quad (1.9)$$

where i is the index for the stock (including both treatments and controls), t is the day index, $measure$ is the metric of interest (e.g. relative bid-ask spread), $treatment$ is a treatment dummy equalling one if stock i is a treatment stock and the date is after 1 August 2012, $after$ is an after-period dummy equalling one if the date is after 1 August 2012, FE is a fixed effect for each treatment stock and control stock, X is a vector of control variates for the stock, and ε is the error term. The fixed effects span the sample, so there is no constant of regression. The difference-in-differences impact coefficient is β .

The model is fit over six specifications. Specification 1 excludes the control-stock observations and all control variates. The regression coefficient β is then the treatment effect from a simple or “single-difference” event study that compares market quality only for treatment stocks before and after the launch of the RLP. Specification 2 includes the control-stock observations and continues to exclude control variates. Specifications 3 through 6 introduce a steadily greater number of control variates. Specification 3 includes the common equity-market liquidity determinants, the stocks log market capitalization and log daily volume. Specification 4 includes a market-wide liquidity factor, the stock-specific factor score from principal component analysis. The factor score equals the first principal component of the daily observations of the treatment and control market-quality measure multiplied by the stock-specific eigenvalue. Specification 5 includes the rolling 10-day volatility for each stock. Last, Specification 6 includes the previous day’s value of the market-quality metric for the stock. For each specification, standard errors are clustered by stock and date.

The model is fit both over the entire post-period of data and also over four subsamples of data spanning four different time periods. All model fits use observations from the three-month pre-period sample, April to July 2012. They differ by the post-period data. The post-period in the first fit, reported in each table’s Panel A, is the entire sample after 1 August 2012. The second through fifth fits compare the pre-period to four three-month post-periods: Q4 2012, Q1 2013, Q2 2013, and Q3 2013. The second through fifth fits are reported in each tables Panels B to E.

1.5 Results

Figure 1.3 shows the impulse-response function for the monthly VARs fit on the log return and the two order flows (RLP and lit) averaged over all months and all 35 treatment stocks.

The contemporaneous impact on the return of a one standard-deviation shock to the lit order flow is around 5 basis points. The lit flow still impacts the return after five minutes by around 2 basis points. For the RLP flow, the contemporaneous and first-lag return impacts of a one standard-deviation shock are both less than one basis point. Both flows appear to have no appreciable impact on the return after 10 minutes (two lags).

Table 1.4 Panel A shows summary statistics for the information shares of segments of the order flow. The share for Total order flow is the information share of the order flow for the VAR model fit by stock by month on the log return and total order flow. The share for Lit is the

Table 1.4: **Information shares.** This table gives summary statistics and results for a T-test on difference in means for information shares computed using a vector autoregression model. Information shares are computed monthly for each of the 35 treatment stocks. Panel A reports summary statistics. The columns of the table give the average, standard deviation, 25th percentile, 50th percentile and 75th percentile for each segment of the order flow. Lit is the information share of lit orders; RLP is the information share of RLP orders; Total order flow is the information share of all undifferentiated orders; RLP and lit is the sum of information shares for RLP and lit orders. Panel B reports the average difference between information shares for various segments of the order flow. Difference of lit, RLP is the difference between lit and RLP information shares; Difference of total and lit plus RLP is the difference between the sum of RLP and lit minus the aggregate information shares. N is the number of observations. The t-statistic is in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% level.

Panel A: **Summary statistics for information shares**

	Mean	Std. Dev.	P25	P50	P75
Total order flow	26.27	19.56	12.12	21.98	36.52
Lit	25.40	19.42	10.85	20.56	37.05
RLP	2.37	8.25	0.17	0.66	1.94

Panel B: **T-test for difference in means**

Difference of lit share and RLP share	23.03*** (16.25)
Difference of total and lit plus RLP shares	1.49*** (4.21)
N	419

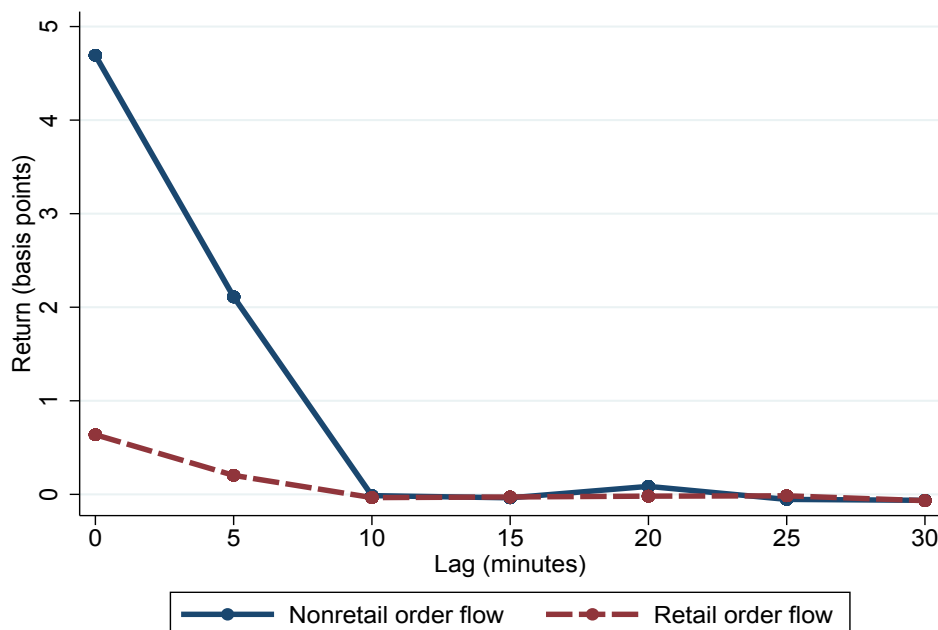


Figure 1.3: **Impulse-response functions for RLP and lit order flow.** This figure plots orthogonalized impulse-response coefficients for each component of the order flow against their corresponding lags. The blue line represents the response of return to a one-standard deviation shock to the lit order flow; the red dashed line represents the response of return to a one-standard deviation shock to the RLP order flow.

information share of the lit order flow for the VAR model fit by stock by month on log return, lit order flow and RLP order flow. The share for RLP is the information share of the RLP order flow for the VAR model fit by stock by month on log return, lit order flow and RLP order flow.

The total order flow had an information share of 26.3% on average. For lit order flow, the information share was 25.4% on average. For RLP order flow the information share was 10 times lower, 2.4% on average.

Hypotheses 1.2.1 and 1.2.2 in this paper ask whether the RLP order flow is demonstrably less informed and whether segmentation adds information to the order flow. Table 1.4 Panel B reports the differences of means between order flows and t-statistics on the differences. Difference of lit and RLP is the difference between the lit and RLP information shares. Difference of total and lit plus RLP is the difference between the information share of the total order flow with the information shares of the sum of the segmented lit and RLP order flows. The difference between lit and RLP was on average 23.0% with statistical significance. This result is to be expected and demonstrates the desirability of RLP order flow for intermediaries. The sum of lit and RLP information shares was on average 1.5% more than the information share of the total order flow with statistical significance. The increase in the information earned by differentiating RLP and lit order flow shows that using the RLP marker to distinguish RLP trades from lit trades can increase the explanatory power of the order flow. Segmentation does appear to remove noise from the signal, as hypothesized. We next measure the impact on market quality.

Hypotheses 1.2.3 and 1.2.4 in this paper concern whether participation in the RLP affects stocks liquidity and price efficiency. Figure 1.1, Panel A shows the relative bid-ask spread of treatment stocks and control stocks over the sample period. Before the launch of the RLP, the average bid-ask spreads for treatment and control stocks were close on average and co-moved. In September 2012 the level of the control series shifts upward and the level of the treatment series shifts downward. The spike in the relative bid-ask spread of treatment stocks on 26 December 2012 is an outlier and was dropped from the sample. The results were stronger with the outlier.

Tables 1.5 to 1.9 give the results of difference-in-differences event studies on four market-quality measures. The rows of the tables give the regression coefficients and their associated t-statistics for specific variables across six different specifications of the event study on relative bid-ask spreads, effective spreads, price impacts and absolute return autocorrelations. A blank entry indicates a variate is not included in the regression.

The columns of Tables 1.5 to 1.9 correspond to the different specifications of the event study. Specification 1 excludes the control-stock observations and gives the treatment effect of a classic, single-difference event study. Specifications 2 to 6 estimate the differenced differences. Specifications 3 to 6 include a progressively larger set of control variates, starting with log market capitalization and log daily volume, then adding the market-wide liquidity factor, then the 10-day moving average of the closing price, and last the lagged value of the dependent variable. Treatment is a dummy variable that indicates the period after the launch of the RLP for treatment stocks. After is a dummy variable that indicates the period after the launch of the RLP for all stocks. Market cap is the daily market capitalization in billions. Volume is the number of shares traded per day. Market-wide liquidity is the stock-specific factor score from principal component analysis. 10-day volatility is the 10-day rolling volatility of the close price. The variable Lagged market quality is the market-quality measure for the stock lagged by one day. *, **, *** represent statistical significance at the 10%, 5%, and 1% level.

Each Panel A for Tables 1.5 to 1.8 shows results for the entire sample period, from April 2012 to August 2013. We then test subsamples to ensure our result holds throughout the sample period. Panels B through E show results for a sample period limited to three months prior to the launch of the RLP and three months after. Panel B shows results when the sample period is limited to Q3 2012 and Q4 2012; Panel C shows results when the sample period is limited to Q3 2012 and Q1 2013; Panel D shows results when the sample period is limited to Q3 2012 and Q2 2013; and Panel E shows results when the sample period is limited to Q3 2012 and Q3 2013. Each regression specification 1 to 6 in Panels B through E corresponds to those in Panel A. For Panels B through E we exclude reporting of variables other than Treatment and After for brevity.

Table 1.5 Panel A shows that the relative spread decreased for treatment over the sample period. The result is consistently statistically significant for each of the regression specifications. For treatment stocks (specification 1), the average relative spread decreased by 1.4 basis points over the sample period. When control stocks are included in the regression (specification 2), the difference is more pronounced; relative spreads decrease by 2.0 basis points. This is because control stocks experience a widening of relative spreads over the sample period, resulting in a negative estimate of the treatment variable that was larger in magnitude. The R^2 drops from 0.839 to 0.608 because control stocks are added to the regression as dependent variables going from specification 1 to specification 2. In the remaining specifications, 3 to 6,

Table 1.5: Difference-in-differences event study of the impact of the launch of the RLP on relative bid-ask spreads. The rows of the table give the regression coefficients and their associated t-statistics for specific variables across six different specifications of the event study on relative bid-ask spreads. A blank entry indicates exclusion from the regression. The columns of the table correspond to different specifications of the event study. In specification (1), only treatment stocks are included in the regression; in specification (2), treatment stocks and control stocks are included; in specification (3), *Market cap* and *Volume* are included; and so on. *Treatment* is a dummy variable that equals one during the period after the launch of the RLP for treatment stocks. *After* is a dummy variable that equals one during the period after the launch of the RLP for all stocks. *Market cap* is the daily market capitalization in billions. *Volume* is the number of shares traded per day in thousands of shares. *Market-wide liquidity* is the stock-specific factor score from principal component analysis. *10 – dayvolatility* is the 10-day rolling volatility of the midquote. *Lagged relative spread* is the relative bid-ask spread lagged by one day. *Constant* is the constant of regression. N is the number of observations. R^2 is the coefficient of determination. *, **, *** represent statistical significance at the 10%, 5%, and 1% level. Panel A shows results for the entire sample period, from April 2012 to August 2013. Panels B through E show results for a sample period limited to three months prior to the launch of the RLP and three months after. Panel B shows results when the sample period is limited to Q3 2012 and Q4 2012; Panel C shows to Q3 2012 and Q1 2013; Panel D shows Q3 2012 and Q2 2013; and Panel E shows Q3 2012 and Q3 2013. Each regression specification (1) to (6) in Panels B through E corresponds to those in Panel A, but we exclude reporting of variables other than Treatment and After for brevity.

Panel A: Impact on relative spread, entire sample.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-1.389** (-2.43)	-1.987*** (-3.11)	-1.262** (-2.20)	-0.931** (-2.03)	-0.932** (-2.08)	-0.653* (-1.87)
After		0.661 (1.62)	0.832* (1.93)	0.991** (2.54)	1.066*** (2.74)	0.785** (2.38)
Market cap			-5.302*** (-3.79)	-3.189** (-2.42)	-3.126** (-2.38)	-2.488** (-2.22)
Volume			-1.797*** (-5.18)	-1.690*** (-5.39)	-1.808*** (-5.73)	-1.512*** (-5.11)
Market-wide liquidity				0.904*** (15.03)	0.891*** (15.08)	0.716*** (7.16)
10-day volatility					4.602*** (2.98)	3.820*** (3.12)
Lagged relative spread						0.225** (2.13)
Constant	11.81*** (29.21)	12.09*** (42.83)	119.7*** (5.12)	84.20*** (3.66)	84.24*** (3.67)	67.73*** (3.28)
N	11640	23090	23090	20650	20650	20580
R^2	0.839	0.608	0.626	0.827	0.828	0.854

Panel B: Impact on relative spread, Q2 2012 and Q1 2013.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.857 (-1.58)	-2.779*** (-2.87)	-2.301*** (-2.63)	-1.980** (-2.35)	-1.990** (-2.36)	-1.263** (-2.25)
After		1.953** (2.32)	2.089** (2.54)	1.936** (2.43)	1.977** (2.46)	1.277** (2.35)
N	4366	8639	8639	8190	8190	8120
R ²	0.880	0.749	0.763	0.856	0.856	0.887

Panel C: Impact on relative spread, Q2 2012 and Q2 2013.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-2.222*** (-3.92)	-1.787** (-2.42)	-0.724 (-1.13)	-0.947** (-2.07)	-0.949** (-2.07)	-0.315 (-1.40)
After		-0.437 (-0.85)	-0.201 (-0.34)	0.899* (1.89)	0.917* (1.88)	0.368 (1.52)
N	4339	8639	8639	8190	8190	8190
R ²	0.879	0.815	0.832	0.850	0.850	0.900

Panel D: Impact on relative spread, Q2 2012 and Q3 2013.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-1.927** (-2.54)	-2.033** (-2.01)	-1.058 (-1.18)	-1.350*** (-2.64)	-1.342*** (-2.64)	-0.489** (-1.97)
After		0.108 (0.15)	0.535 (0.76)	1.737*** (3.65)	1.752*** (3.69)	0.703*** (3.13)
N	4474	8918	8918	8330	8330	8330
R ²	0.826	0.807	0.830	0.880	0.881	0.921

Panel E: Impact on relative spread, Q2 2012 and Q4 2012.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-2.171*** (-2.94)	-3.365** (-2.15)	-2.515* (-1.69)	-1.270** (-2.06)	-1.300** (-2.13)	-1.137* (-1.95)
After		1.203 (0.84)	2.167 (1.41)	1.771*** (3.07)	1.808*** (3.11)	1.629*** (2.93)
N	3008	5921	5921	5460	5460	5460
R ²	0.867	0.456	0.467	0.900	0.900	0.904

the magnitude of the regression coefficient for relative spread on the treatment dummy attenuates as more and more covariates are added to the regression. The sign remains negative and statistically significant, dropping to 0.7 basis points for specification 6.

In Table 1.5, Panels B through E show that participation in the RLP leads to lower relative spreads in four subsample time periods. The treatment effect in specification 2 is negative and statistically significant in all panels. For specification 6, it misses significance in Panel C, which studies the post-period Q1 2013.

Table 1.6 shows a milder result for effective spreads. Panel A shows the results for the regression over the entire sample. The regression coefficient for the treatment dummy is negative but is statistically significant only for specifications 1 and 2. In specification 2, the treatment effect is 0.9 basis points, but in specification 6 the effect is 0.2 basis points and has a t-statistic of only 1.38.

In Table 1.6, Panels B through E show how the results for effective spread vary over time. The results are hit and miss. The weakest period is Panel B, comparing Q2 2012 to Q4 2012. While each coefficient of the treatment dummy is negative, the only one with statistical significance is the simple difference-in-differences, specification 2. Panels C through E do show evidence of a decrease in effective spreads when comparing Q2 2012 to the remaining periods. The regression coefficients in specification 6 for panels C through E range from 0.3 basis points to 0.6 basis points and are nearly significant, with t-statistics all greater than 1.63.

Table 1.7 gives the effect of the RLP on price impact. Panel A shows that price impact decreased for treatment stocks over the sample period. The decrease resulting from the RLP ranges between 0.4 and 0.6 basis points depending on the specification. In specification 6, the RLP leads to a decrease in price impact of 0.5 basis points with a t-statistic of 4.38.

In Table 1.7, Panels B through E show how the effect of the RLP on price impact varies over time. All regression specifications except for Panel A specification 1 show negative and statistically significant coefficients for the treatment effect. In specification 6 for Panels B through E, the treatment effect ranges from 0.6 basis points to 0.9 basis points.

Table 1.8 gives the effect of the RLP on price efficiency as measured by the absolute autocorrelation of the return of the midquote. Panel A shows the RLP increased price efficiency over the sample period. The simple treatment effect (specification 1) is positive, small (0.003) and statistically insignificant, indicating that absolute autocorrelation was generally unchanged for treatment stocks over the sample period. When control stocks are added to the regression (specification 2), the regression coefficient for the treatment dummy becomes negative and statistically significant, -0.01. Control stocks experienced an increase in absolute autocorrelation over the sample, while treatment stocks did not. Specifications 3 through 6 continue to show the RLP decreased the absolute autocorrelation of the midquote. The impact in these specifications ranges from -0.008 to -0.009 indicating an improvement in price efficiency.

Table 1.8 Panels B through E show how the result on autocorrelation varies by the time period. The impact misses significance in Panel B for specifications 3 to 6, meaning we fail to find good evidence the RLP had an impact on price efficiency in Q4 2012. In fact, the simple treatment effect (Panel B specification 1) is positive and significant, and the treatment effect becomes negative and significant when control stocks are included in the regression (specification 2). For Panels C through E, the treatment effect ranges from -0.01 to -0.04 and is significant in specifications 3 through 6. For these panels, the simple treatment effect (specification 1) is small and insignificant, and the addition of control stocks and control variates makes the

Table 1.6: Difference-in-differences event study of the impact of the launch of the RLP on effective spreads. The rows of the table give the regression coefficients and their associated t-statistics for specific variables across six different specifications of the event study on effective spreads. A blank entry indicates exclusion from the regression. The columns of the table correspond to different specifications of the event study. In specification (1), only treatment stocks are included in the regression; in specification (2), treatment stocks and control stocks are included; in specification (3), *Market cap* and *Volume* are included; and so on. *Treatment* is a dummy variable that equals one during the period after the launch of the RLP for treatment stocks. *After* is a dummy variable that equals one during the period after the launch of the RLP for all stocks. *Market cap* is the daily market capitalization in billions. *Volume* is the number of shares traded per day in thousands of shares. *Market-wide liquidity* is the stock-specific factor score from principal component analysis. *10-day volatility* is the 10-day rolling volatility of the midquote. *Lagged effective spread* is the five-second effective spread lagged by one day. *Constant* is the constant of regression. *N* is the number of observations. R^2 is the coefficient of determination. *, **, *** represent statistical significance at the 10%, 5%, and 1% level. Panel A shows results for the entire sample period, from April 2012 to August 2013. Panels B through E show results for a sample period limited to three months prior to the launch of the RLP and three months after. Panel B shows results when the sample period is limited to Q3 2012 and Q4 2012; Panel C shows Q3 2012 and Q1 2013; Panel D shows Q3 2012 and Q2 2013; and Panel E shows Q3 2012 and Q3 2013. Each regression specification (1) to (6) in Panels B through E corresponds to those in Panel A, but we exclude reporting of variables other than Treatment and After for brevity.

Panel A: Impact on effective spread, entire sample.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-1.114** (-2.41)	-0.938** (-2.01)	-0.430 (-1.08)	-0.440 (-1.42)	-0.441 (-1.46)	-0.244 (-1.38)
After		-0.194 (-0.86)	0.264 (0.99)	0.518** (2.29)	0.577** (2.57)	0.327** (2.29)
Market cap			-4.909*** (-4.30)	-3.344*** (-3.37)	-3.295*** (-3.32)	-1.878*** (-3.98)
Volume			-0.493** (-2.43)	-0.405** (-2.50)	-0.497*** (-3.04)	-0.294** (-2.50)
Market-wide liquidity				0.563*** (8.73)	0.553*** (8.79)	0.340*** (4.65)
10-day volatility					3.607*** (3.88)	1.989*** (2.95)
Lagged eff. spread						0.420*** (5.45)
Constant	9.751*** (29.78)	8.392*** (41.02)	92.73*** (4.87)	66.39*** (3.97)	66.42*** (3.98)	38.06*** (4.74)
N	11640	23090	23090	20650	20650	20580
R^2	0.867	0.761	0.783	0.804	0.804	0.841

Panel B: Impact on effective spread, Q2 2012 and Q4 2012.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.515 (-1.18)	-0.959* (-1.69)	-0.648 (-1.39)	-0.661 (-1.41)	-0.672 (-1.44)	-0.354 (-1.47)
After		0.452 (1.14)	0.808* (1.92)	0.827** (2.00)	0.872** (2.10)	0.477** (2.11)
N	4401	8709	8709	8260	8260	8260
R ²	0.906	0.750	0.762	0.757	0.757	0.780

Panel C: Impact on effective spread, Q2 2012 and Q1 2013.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-1.686*** (-3.61)	-1.109** (-2.13)	-0.456 (-1.08)	-0.608* (-1.89)	-0.611* (-1.91)	-0.311* (-1.95)
After		-0.580** (-2.16)	-0.0778 (-0.21)	0.606* (1.92)	0.639** (2.01)	0.338** (2.13)
N	4339	8639	8639	8190	8190	8190
R ²	0.896	0.819	0.831	0.843	0.843	0.882

Panel D: Impact on effective spread, Q2 2012 and Q2 2013.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-1.759*** (-2.87)	-1.216 (-1.61)	-0.633 (-0.96)	-0.838** (-2.03)	-0.829** (-2.02)	-0.579 (-1.63)
After		-0.550 (-1.16)	0.121 (0.23)	0.992** (2.28)	1.010** (2.34)	0.711* (1.83)
N	4474	8918	8918	8330	8330	8330
R ²	0.857	0.756	0.771	0.811	0.812	0.828

Panel E: Impact on effective spread, Q2 2012 and Q4 2012.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-1.831*** (-3.02)	-1.055 (-1.49)	-0.523 (-0.83)	-0.689* (-1.76)	-0.727* (-1.87)	-0.481 (-1.63)
After		-0.782* (-1.88)	0.452 (0.84)	1.259*** (3.19)	1.305*** (3.33)	0.885*** (2.96)
N	3008	5921	5921	5460	5460	5460
R ²	0.892	0.837	0.855	0.881	0.882	0.894

Table 1.7: Difference-in-differences event study of the impact of the launch of the RLP on relative five-second price impacts. The rows of the table give the regression coefficients and their associated t-statistics for specific variables across six different specifications of the event study on five-second price impacts. A blank entry indicates exclusion from the regression. The columns of the table correspond to different specifications of the event study. In specification (1), only treatment stocks are included in the regression; in specification (2), treatment stocks and control stocks are included; in specification (3), *Market cap* and *Volume* are included; and so on. *Treatment* is a dummy variable that equals one during the period after the launch of the RLP for treatment stocks. *After* is a dummy variable that equals one during the period after the launch of the RLP for all stocks. *Market cap* is the daily market capitalization in billions. *Volume* is the number of shares traded per day in thousands of shares. *Market-wide liquidity* is the stock-specific factor score from principal component analysis. *10-day volatility* is the 10-day rolling volatility of the midquote. *Lagged price impact* is the five-second relative price impact lagged by one day. *Constant* is the constant of regression. *N* is the number of observations. R^2 is the coefficient of determination. *, **, *** represent statistical significance at the 10%, 5%, and 1% level. Panel A shows results for the entire sample period, from April 2012 to August 2013. Panels B through E show results for a sample period limited to three months prior to the launch of the RLP and three months after. Panel B shows results when the sample period is limited to Q3 2012 and Q4 2012; Panel C shows Q3 2012 and Q1 2013; Panel D shows Q3 2012 and Q2 2013; and Panel E shows Q3 2012 and Q3 2013. Each regression specification (1) to (6) in Panels B through E corresponds to those in Panel A, but we exclude reporting of variables other than Treatment and After for brevity.

Panel A: Impact on five-second price impact, entire sample.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.356*** (-2.62)	-0.627*** (-4.25)	-0.486*** (-3.51)	-0.510*** (-4.14)	-0.510*** (-4.27)	-0.456*** (-4.38)
After		0.299*** (3.53)	0.469*** (4.81)	0.540*** (5.82)	0.574*** (6.29)	0.503*** (6.77)
Market cap			-1.516*** (-4.97)	-0.976*** (-3.46)	-0.946*** (-3.41)	-0.553*** (-3.27)
Volume			-0.0414 (-0.43)	-0.0183 (-0.21)	-0.0730 (-0.84)	-0.0145 (-0.20)
Market-wide liquidity				0.190*** (9.36)	0.184*** (9.38)	0.125*** (5.08)
10-day volatility					2.129*** (3.96)	1.700*** (3.62)
Lagged price impact						0.116*** (4.32)
Constant	3.529*** (36.30)	3.001*** (46.58)	27.61*** (5.21)	18.62*** (3.76)	18.64*** (3.80)	10.74*** (3.58)
N	11640	23090	23090	20650	20650	20580
R^2	0.411	0.494	0.513	0.537	0.540	0.567

Panel B: Impact on five-second price impact, Q2 2012 and Q4 2012.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.144 (-1.11)	-0.832*** (-3.97)	-0.745*** (-3.86)	-0.769*** (-3.93)	-0.776*** (-4.03)	-0.671*** (-4.77)
After		0.699*** (4.01)	0.823*** (4.70)	0.831*** (4.85)	0.861*** (5.03)	0.733*** (6.35)
N	4366	8639	8639	8190	8190	8120
R ²	0.434	0.585	0.596	0.597	0.599	0.633

Panel C: Impact on five-second price impact, Q2 2012 and Q1 2013.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.738*** (-5.10)	-1.066*** (-6.28)	-0.909*** (-5.70)	-0.954*** (-6.97)	-0.957*** (-7.11)	-0.892*** (-7.15)
After		0.330*** (3.22)	0.516*** (3.99)	0.699*** (6.44)	0.723*** (6.74)	0.658*** (6.98)
N	4339	8639	8639	8190	8190	8190
R ²	0.482	0.559	0.570	0.579	0.582	0.600

Panel D: Impact on five-second price impact, Q2 2012 and Q2 2013.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.501*** (-2.79)	-0.725*** (-3.13)	-0.583*** (-2.64)	-0.654*** (-3.85)	-0.649*** (-3.91)	-0.602*** (-3.60)
After		0.227 (1.43)	0.458*** (2.68)	0.687*** (5.06)	0.697*** (5.21)	0.641*** (4.64)
N	4474	8918	8918	8330	8330	8330
R ²	0.408	0.509	0.520	0.556	0.559	0.564

Panel E: Impact on five-second price impact, Q2 2012 and Q3 2013.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.722*** (-3.77)	-0.814*** (-3.40)	-0.684*** (-2.85)	-0.737*** (-4.05)	-0.767*** (-4.41)	-0.720*** (-4.21)
After		0.0929 (0.58)	0.496*** (2.92)	0.749*** (5.74)	0.785*** (6.18)	0.706*** (5.39)
N	3008	5921	5921	5460	5460	5460
R ²	0.452	0.573	0.590	0.625	0.629	0.634

Table 1.8: **Difference-in-differences event study of the impact of the launch of the RLP on the absolute value of five-second return autocorrelations.** The rows of the table give the regression coefficients and their associated t-statistics for specific variables across six different specifications of the event study on five-second return autocorrelations. A blank entry indicates exclusion from the regression. The columns of the table correspond to different specifications of the event study. In specification (1), only treatment stocks are included in the regression; in specification (2), treatment stocks and control stocks are included; in specification (3), Market cap and Volume are included; and so on. *Treatment* is a dummy variable that equals one during the period after the launch of the RLP for treatment stocks. *After* is a dummy variable that equals one during the period after the launch of the RLP for all stocks. *Market cap* is the daily market capitalization in billions. *Volume* is the number of shares traded per day in thousands of shares. *Market-wide liquidity* is the stock-specific factor score from principal component analysis. *10-day volatility* is the 10-day rolling volatility of the midquote. *Lagged autocorrelation* is the absolute five-second autocorrelation lagged by one day. *Constant* is the constant of regression. *N* is the number of observations. R^2 is the coefficient of determination. *, **, *** represent statistical significance at the 10%, 5%, and 1% level. Panel A shows results for the entire sample period, from April 2012 to August 2013. Panels B through E show results for a sample period limited to three months prior to the launch of the RLP and three months after. Panel B shows results when the sample period is limited to Q3 2012 and Q4 2012; Panel C shows Q3 2012 and Q1 2013; Panel D shows Q3 2012 and Q2 2013; and Panel E shows Q3 2012 and Q3 2013. Each regression specification (1) to (6) in Panels B through E corresponds to those in Panel A, but we exclude reporting of variables other than Treatment and After for brevity.

Panel A: **Impact on return autocorrelation, entire sample.**

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.00338 (1.20)	-0.0115** (-2.53)	-0.00975** (-2.15)	-0.00938** (-2.13)	-0.00937** (-2.13)	-0.00805** (-2.04)
After		0.0164*** (4.49)	0.0144*** (4.07)	0.0144*** (4.20)	0.0143*** (4.11)	0.0124*** (3.95)
Market cap			-0.00434 (-0.77)	0.00277 (0.44)	0.00269 (0.42)	0.00255 (0.45)
Volume			-0.00978*** (-6.86)	-0.00973*** (-6.54)	-0.00958*** (-6.35)	-0.00872*** (-6.10)
Mkwd. liquidity				0.00165*** (2.76)	0.00167*** (2.78)	0.00147*** (2.75)
10-day volatility					-0.00584 (-0.46)	-0.00473 (-0.42)
Lagged acorr.						0.123*** (9.08)
Constant	0.0583*** (29.32)	0.0724*** (39.31)	0.269*** (2.80)	0.154 (1.44)	0.154 (1.44)	0.136 (1.44)
N	11640	23090	23090	20650	20650	20580
R^2	0.163	0.210	0.215	0.218	0.218	0.229

Panel B: Impact on return autocorrelation, Q2 2012 and Q4 2012.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.00629** (2.21)	-0.00921* (-1.83)	-0.00763 (-1.61)	-0.00705 (-1.51)	-0.00702 (-1.51)	-0.00582 (-1.41)
After		0.0158*** (3.65)	0.0143*** (3.20)	0.0132*** (2.99)	0.0131*** (2.95)	0.0111*** (2.84)
N	4366	8639	8639	8190	8190	8120
R ²	0.163	0.254	0.262	0.268	0.268	0.279

Panel C: Impact on return autocorrelation, Q2 2012 and Q1 2013.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.00283 (1.06)	-0.0148** (-2.49)	-0.0113* (-1.92)	-0.0115** (-2.10)	-0.0115** (-2.10)	-0.00995** (-2.06)
After		0.0177*** (3.30)	0.0166*** (3.02)	0.0206*** (3.90)	0.0205*** (3.86)	0.0178*** (3.78)
N	4339	8639	8639	8190	8190	8190
R ²	0.178	0.242	0.247	0.250	0.250	0.261

Panel D: Impact on return autocorrelation, Q2 2012 and Q2 2013.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.00379 (0.93)	-0.0177*** (-2.63)	-0.0145** (-2.23)	-0.0161*** (-2.59)	-0.0162*** (-2.61)	-0.0145** (-2.52)
After		0.0217*** (3.98)	0.0206*** (3.92)	0.0250*** (5.12)	0.0250*** (5.06)	0.0225*** (4.98)
N	4474	8918	8918	8330	8330	8330
R ²	0.194	0.238	0.244	0.248	0.248	0.255

Panel E: Impact on return autocorrelation, Q2 2012 and Q3 2013.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.00443 (-1.03)	-0.0416*** (-5.19)	-0.0390*** (-5.04)	-0.0425*** (-5.38)	-0.0424*** (-5.32)	-0.0376*** (-5.10)
After		0.0375*** (5.62)	0.0368*** (5.54)	0.0403*** (5.54)	0.0401*** (5.42)	0.0353*** (5.09)
N	3008	5921	5921	5460	5460	5460
R ²	0.172	0.270	0.275	0.281	0.281	0.292

treatment coefficients negative and significant.

The general takeaway from the difference-in-differences regression results in Tables 1.5 through 1.8 is that the RLP led to slightly higher liquidity as well as slightly greater price efficiency. Except for the effective spread, the result is robust across regression specifications and over time in our sample period. For the effective spread and price impact, the result was weak in Q4 2012. This may be due to a gradual adoption of the RLP by liquidity providers and retail brokers.

To demonstrate that our results are robust to the selection of control stocks, we repeat the regressions with an alternative methodology for the choice of control observations. Rather than matching treatments and controls one-to-one, we use a weighted average of all stocks in our pools of candidate control stocks. Once for each of the 35 treatment stocks, for each of the 184 candidate control stocks, weights were generated equal to the squared difference between a treatment stock and the candidate control stocks market capitalization divided by the sum of such differences over the candidate control stocks. Hence for each treatment stock, the weights on the candidate control stocks add to one. The weights were then used to generate a weighted set of 184 control stock observations for each daily observation of a treatment stock.

Table 1.9 Panel A shows the results for relative bid-ask spread using the weighted panel of controls. Each regression coefficient for the treatment dummy is negative with statistical significance except for specification 3. The treatment effect in specification 6 is 0.7 basis points.

Table 1.9 Panel B shows the results for effective spread using the weighted panel of controls. The results are not statistically significant, and we fail to conclude from the difference-in-differences event study that the RLP had an effect on the effective spread. The treatment effect ranges from 1.1 to 0.3 basis points.

Table 1.9 Panel C shows the results for price impact using the weighted panel of controls. Each regression coefficient for the treatment dummy is negative with significance. The treatment effect ranges from 0.4 to 0.6 basis points.

Table 1.9 Panel D shows the results for absolute autocorrelation of returns using the weighted panel of controls. The simple treatment effect (specification 1) is small, positive and insignificant, and the addition of control stocks and control variates to the regression produces negative and significant regression coefficients for the treatment dummy in specifications 4 through 6.

In general, the results of the difference-in-differences regressions using a weighted panel of control stocks are similar to those produced by the one-to-one matched sample presented in Tables 1.5 through 1.8. All market-quality measures tend to improve slightly following the launch of the RLP. However, results for effective spread were statistically insignificant when a weighted panel of controls was used.

1.5.1 Conclusions

We find the launch of the NYSE's Retail Liquidity Program resulted in a small positive impact on market quality. While it is not surprising that retail traders might benefit from RLP due to its mandated price improvement, the overall effects of segmentation are more challenging to predict. Our results indicate that other classes of traders are not worse off when retail traders are segmented.

Table 1.9: **Difference-in-differences event study of the RLPs impact on market quality with a weighted panel of controls.** The rows of the table give the regression coefficients and their associated t-statistics for specific variables across six different specifications of the event study for four market-quality measures. A blank entry indicates exclusion from the regression. The columns of the table correspond to different specifications of the event study. In specification (1), only treatment stocks are included in the regression; in specification (2), treatment stocks and control stocks are included; in specification (3), *Market cap* and *Volume* are included; and so on. Rather than using control stocks matched on-to-one with treatments (reported in Tables 3 through 6) a weighted panel of controls is used for each treatment stock. *Treatment* is a dummy variable that equals one during the period after the launch of the RLP for treatment stocks. *After* is a dummy variable that equals one during the period after the launch of the RLP for all stocks. *Market cap* is the daily market capitalization in billions. *Volume* is the number of shares traded per day in thousands of shares. *Market-wide liquidity* is the stock-specific factor score from principal component analysis. *10-day volatility* is the 10-day rolling volatility of the midquote. *Lagged autocorrelation* is the absolute five-second autocorrelation lagged by one day. *Constant* is the constant of regression. *N* is the number of observations. R^2 is the coefficient of determination. *, **, *** represent statistical significance at the 10%, 5%, and 1% level. Panels A through D show results for relative spread, effective spread, price impact and autocorrelation over the entire sample period, from April 2012 to August 2013.

Panel A: **Impact on relative spread, entire sample, weighted panel of controls.**

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-1.389** (-2.43)	-1.201* (-1.86)	-0.932 (-1.47)	-0.978** (-2.22)	-1.040** (-2.51)	-0.738** (-2.24)
After		-0.208 (-0.54)	-0.348 (-0.98)	0.337 (1.14)	0.497** (2.09)	0.355* (1.72)
Market cap			-1.859* (-1.91)	0.288 (0.48)	0.303 (0.51)	0.126 (0.27)
Volume			-2.437*** (-8.68)	-2.068*** (-10.18)	-2.130*** (-11.65)	-1.774*** (-9.20)
Market-wide liquidity				1.000*** (31.63)	1.004*** (38.57)	0.779*** (7.12)
10-day volatility					3.871*** (2.59)	3.165*** (2.59)
Lagged relative spread						0.244** (2.17)
Constant	11.81*** (29.21)	12.63*** (45.24)	70.98*** (4.41)	32.24*** (3.12)	32.26*** (3.17)	27.92*** (3.16)
N	11640	1718905	1718905	1548750	1548400	1543150
R^2	0.839	0.568	0.592	0.786	0.801	0.831

Panel B: Impact on effective spread, entire sample, weighted panel of controls.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-1.114**	-0.520	-0.346	-0.512	-0.547	-0.334
	(-2.41)	(-1.08)	(-0.76)	(-1.44)	(-1.60)	(-1.59)
After		-0.656***	-0.498***	-0.0769	0.0206	0.0158
		(-3.32)	(-2.62)	(-0.46)	(0.16)	(0.18)
Market cap			-1.822***	-0.502	-0.494	-0.281
			(-3.58)	(-1.45)	(-1.45)	(-1.39)
Volume			-0.462***	-0.328***	-0.381***	-0.238***
			(-3.24)	(-2.89)	(-3.74)	(-3.36)
Market-wide liquidity				0.524***	0.525***	0.337***
				(20.92)	(23.83)	(11.25)
10-day volatility					2.604***	1.515**
					(2.74)	(2.53)
Lagged relative spread						0.388***
						(10.54)
Constant	9.751***	7.146***	41.21***	18.55***	18.74***	11.19***
	(29.78)	(48.33)	(4.65)	(3.13)	(3.20)	(3.27)
N	11640	1718905	1718905	1548750	1548400	1543150
R ²	0.867	0.618	0.627	0.676	0.690	0.738

Panel C: Impact on five-second price impact, entire sample, weighted panel of controls.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.356***	-0.577***	-0.513***	-0.588***	-0.597***	-0.523***
	(-2.62)	(-3.96)	(-3.70)	(-4.89)	(-5.19)	(-5.28)
After		0.244***	0.319***	0.445***	0.483***	0.479***
		(3.44)	(4.48)	(6.62)	(7.94)	(9.39)
Market cap			-0.718***	-0.271*	-0.267*	-0.189*
			(-3.63)	(-1.79)	(-1.80)	(-1.68)
Volume			-0.0983*	-0.0541	-0.0914**	-0.0407
			(-1.84)	(-1.18)	(-2.14)	(-1.25)
Market-wide liquidity				0.173***	0.171***	0.106***
				(18.68)	(19.94)	(9.16)
10-day volatility					1.412***	1.042***
					(2.82)	(2.70)
Lagged price impact						1336.4***
						(8.96)
Constant	3.529***	2.532***	14.95***	7.293***	7.512***	4.823***
	(36.30)	(50.01)	(4.42)	(2.88)	(3.01)	(2.62)
N	11640	1718905	1718905	1548750	1548400	1543150
R ²	0.411	0.476	0.481	0.510	0.517	0.553

Panel D: **Impact on five-second return autocorrelation, entire sample, weighted panel of controls.**

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.00338 (1.20)	-0.00562 (-1.57)	-0.00587 (-1.57)	-0.00678* (-1.89)	-0.00679* (-1.89)	-0.00611* (-1.87)
After		0.00993*** (4.16)	0.00688*** (2.93)	0.00924*** (3.95)	0.00940*** (4.01)	0.00848*** (3.91)
Market cap			0.00969* (1.79)	0.0164*** (2.90)	0.0165*** (2.90)	0.0150*** (2.91)
Volume			-0.0119*** (-11.82)	-0.0112*** (-11.23)	-0.0114*** (-11.23)	-0.0106*** (-11.25)
Mktwd. liquidity				0.00245*** (3.62)	0.00243*** (3.62)	0.00222*** (3.63)
10-day volatility					0.00821 (1.16)	0.00762 (1.21)
Lagged acorr.						0.0979*** (10.85)
Constant	0.0583*** (29.32)	0.0942*** (51.85)	0.0859 (0.99)	-0.0314 (-0.35)	-0.0297 (-0.33)	-0.0246 (-0.30)
N	11640	1718905	1718905	1548750	1548400	1543150
R ²	0.163	0.104	0.111	0.116	0.116	0.124

We analyze the mechanism by which segmentation affects market quality by computing the information share of each component of the order flow using the techniques of Hasbrouck (1991). The analysis shows that RLP order flow impounds significantly less information into prices than does lit order flow. The result demonstrates the economics of the program: intermediaries are exposed to lower adverse selection risk when offering liquidity to retail traders. We find the sum of the information shares of RLP trades and lit trades is larger than the information share of the total order flow, indicating there is more information available from the order flow when it is segmented.

We measure the effect of the RLP by testing measures of market quality before and after the launch of the NYSE's RLP using a difference-in-differences regression. Bid-ask spread, effective spread, price impact and autocorrelation decrease for stocks that saw relatively heavy use of the RLP. Our result is robust to the time period and the choice of control stocks. First, we match controls to treatments one-to-one from a pool of control stocks that were ineligible for the RLP. Second, we run the regression using a weighted panel of all candidate control stocks. Our weakest results are those for effective spreads. The weighted panel of control stocks eliminated statistical significance for effective spreads. Both model specifications return the same overall conclusion: market quality mildly increased for stocks that had relatively high volumes in the RLP for the sample as a whole and throughout four subperiods.

Our analysis is consistent with other empirical evidence that segmentation of retail traders, either through broker internalization, payment for order flow, or other programs similar to the NYSE's RLP, have beneficial or innocuous effects on market quality. Although many retail segmentation programs operate as dark pools, we demonstrate a mechanism that is consistent with explicit segmentation and unrelated to pre-trade transparency. We argue the effect on market quality is due more to segmentation than to pre-trade opacity. In the NYSE's case, pre-trade opacity was introduced mostly to enable market makers to offer better pricing to retail

traders than is allowed by regulation for stocks constrained by the regulatory minimum tick size.

We frame our results as tests of hypotheses generated by the static theory model of Zhu (2014). Segmentation may worsen liquidity for lit trades, all else equal, but since the order flow contributes more to price discovery, all else is not equal. Illiquidity due to informational differences between market order submitters and limit order submitters may diminish more quickly, resulting in greater average liquidity. Our results show the worsening in liquidity is either economically insignificant or dominated by the effects of greater price efficiency. A possible extension of Zhu (2014) or of other papers on dark markets and segmentation might explore the dynamics to show whether concentrating informed agents on the exchange may improve liquidity in a multi-period setting. The dynamic effect of concentrating informed agents is noted briefly in Glosten and Milgrom (1985) and more thoroughly studied in Rosu (2015). An extension to these models could include the interaction of segmentation with dark liquidity, and investigate how the effect varies if information is long-lived or short-lived.

Chapter 2

Retail Order Flow Segmentation: Theory

2.1 Introduction

Segmentation of retail orders is a feature of certain trading venues that allows liquidity providers to offer trades specifically to retail traders, individuals who trade using their own accounts and on behalf of themselves. As shown in Chapter 1, retail traders are desirable counterparties because they are “uninformed”, that is, they are unlikely to have private information about the fundamental value of an asset or about future trading patterns. Many stock exchanges offer retail segmentation facilities that allow market participants to submit limit orders that are only made available for execution against market orders submitted by retail brokers. These are called retail price improvement programs (RPIs) or retail liquidity programs (RLPs). Examples include Alpha Exchange in 2011, NYSE in 2012, NASDAQ and BATS in 2013, and Aequitas in 2015. These market structures have implications for liquidity and price efficiency, and may also have effects on trading costs that vary across market participants. In Chapter 1, we motivated hypotheses about market quality using the theoretical framework of Zhu (2014). In this chapter, we develop two models that focus on specifics of RLPs that can offer a richer set of predictions about RLPs.

RLPs only allow limit orders to be submitted at prices that are strictly better than the best bid and offer prices available to non-retail traders. Retail clients enjoy the slightly improved prices while retail brokers typically pay a lower fee to the exchange when submitting their clients’ orders to the RLP. Limit orders in RLPs tend to lack pre-trade transparency (they are “dark” orders), that is, their prices or quantities are not displayed publically before a trade occurs. One could imagine an RLP with pre-trade transparency, but one reason they tend to be dark is because of complications that arise from regulations around publically displayed orders. Specifically, prices for publically displayed orders are confined to a grid defined by the tick size, that is, the minimum price increment for displayed quotes. The tick size is most often one cent. Dark orders are allowed to take prices off the tick grid and thereby offer a mechanism to offer slightly better prices to retail traders when bid-ask spreads are one cent wide.

We model an exchange as a limit order book for non-retail traders, and an RLP with all retail traders submitting market orders to the RLP. After a trade occurs in the RLP, market participants will know the trade was executed in the RLP. In equity markets, trades are announced publically in real-time and one can infer trades that are executed in an RLP either because

trades are explicitly identified or by virtue of their prices being within the best bid and offer prices for non-retail traders. Since retail trades are identified after they occur, non-retail trades (we will use “institutional” and “non-retail” interchangeably) are also identified by exclusion. Institutional traders include classes of “informed traders” such as arbitrageurs, high-frequency traders, hedge funds, etc. We focus on the effect of RLPs segmenting and identifying uninformed and informed trades.

We develop two extensions of Glosten and Milgrom (1985) that include segmentation of retail orders. The model has a market maker who maintains a bid and an ask price at which a risky asset is traded anonymously with informed and uninformed traders. Anonymity drives the central tension of the model; the market maker does not know if he is trading with an informed or uninformed trader. Since the asset’s payout is uncertain, there is adverse selection; the market maker risks posting prices that are too low or too high. The bid and ask price therefore must balance losses incurred when trading with informed traders and profits made from trading with uninformed traders. The market maker is a Bayesian agent, that is, after each trade he revises his belief about the asset’s true value. As trades occur, more information about the asset’s true value is reflected in prices. The bid-ask spread and the rate at which the market maker learns about the true value of the asset are endogenous.

In our first model, we deviate from the classic model by changing the anonymity of trading. While the classic model features purely anonymous trading, we assume that certain trades can be associated with a particular class of counterparty, namely, retail traders. We assume a fraction of uninformed traders are retail and, following trades, their identity is revealed through the mechanism described above. In the first extension, we study how liquidity, as measured by the bid-ask spread, and how price efficiency are affected by the segmentation of retail trades. We find that all else equal, bid-ask spreads for non-retail traders increase, while price efficiency increases. However, if there are sufficiently many trades, then over time the increased certainty that arises from the gain in price efficiency results in improved average liquidity.

In the second model, we study how a long-lived informed trader reacts to segmentation strategically and the equilibrium effects on price efficiency. We generalize the first model to allow traders to submit either limit orders or market orders. Order choice is an important feature of markets as central limit order books are a predominant system of trade in equities, foreign exchange, options, and some highly liquid fixed-income markets. All else equal, identification of retail market orders amplifies the information conveyed by non-retail market orders. To counteract the effect, the informed trader uses fewer market orders and more limit orders to complete trades. Although the informed trader strategically mitigates the increase in price efficiency to preserve future profits, price efficiency is still greater in equilibrium.

Electronic trading venues have proliferated in the past two decades, as has competition to provide liquidity from market participants who act as intermediaries. For liquidity providers, retail order flow is the most desirable component of total order flow due to its balance between buy and sell orders, the tendency for retail orders to be small, and lack of correlation of retail trades with future price movements. Retail trades are ideal for liquidity providers to earn profits from the bid-ask spread. One of the original mechanisms to segment retail is called payment for order flow. Large securities dealers are willing to pay retail brokers for execution of their clients’ trades. Doing so guarantees them a desirable component of the order flow, and they profit through bid-ask spreads. A similar practice is called internalization. A broker-dealer may fill a retail client order against its own inventory rather than sending the order to an exchange

and paying a fee. This practice is not allowed in Canada for small orders but is common in the US.

Stock exchanges and other trading venues profit from trading fees charged to traders or their brokers. As broker internalization and payment for order flow schemes became popular and institutionalized, fewer trades were executed on exchanges. Exchanges sought ways to recover lost order flow by offering services that are economically similar. Easley et al. (1996) note that the New York Stock Exchange (NYSE) has estimated that 35% of small orders are executed off-exchange. In Canada, regulation of trading and exchanges has historically been stricter than in the US. However, the first RLP opened in Canada, on Alpha Exchange in 2011. In Canada, RLPs are an alternative to payment for order flow and broker internalization that comply with the Ontario Securities Commission's Order Exposure Rule. It dictates that small orders be routed through a fully accessible exchange unless they offer price improvement.

2.2 Related Literature

Our paper adds to the body of microstructure papers that study game-theoretic equilibria with adverse selection. Different characteristics of retail and institutional trading have been shown empirically in a stream of literature on internalization and payment for order flow, including in Chapter 1 which is based on Garriott and Walton (2016).

Boehmer and Kelley (2009) study the role of institutional ownership of stocks over two decades. They find that stocks with greater institutional ownership tend to be more informationally efficient. We read this as evidence for the important role institutional trading plays in price discovery. Easley et al. (1996) empirically study stocks known to be used in payment for order flow schemes on the NYSE. They find significantly more information content in trades executed on the NYSE versus those executed on the Cincinnati Stock Exchange, which they argue is from retail-like order flow being internalized rather than being directed to the NYSE. Battalio (1997) empirically studies a payment for order flow scheme on NYSE-listed securities. The study finds that bid-ask spreads tighten after the entry of a large firm that purchased retail order flow and argues that any harm to liquidity is therefore economically insignificant. Our predictions on liquidity are ambiguous.

Our second model examines the choice between limit and market orders. Our result that informed traders submit more limit orders than market with the degree of segmentation is consistent with the empirical literature that shows limit orders are an important source of price discovery. Jovanovic and Menkveld (2011) model limit order submitters with private information and find that they may increase welfare by benefiting other traders. Uninformed traders react in equilibrium by using a greater number of market orders. The result is complementary to ours; their paper suggests the presence of sophisticated traders results in unsophisticated traders using a greater share of market orders, while our model predicts that in response to retail segmentation, sophisticated traders use a greater share of limit orders. Brogaard, Hendershott, and Riordan (2015) show that limit orders play a larger role than market orders in price discovery. Using Canadian equity data, they find that limit orders in general, but especially those that come from high-frequency traders command a significant information share using the techniques of Hasbrouck (1995). Goettler, Parlour, and Rajan (2005) simulate a dynamic limit order market with rational traders and find a Markov-perfect equilibrium. They

find that the midpoint of the best limit buy and limit sell orders is not a good proxy for the true value of an asset because of the information content of limit orders. Lo and Sapp (2003) empirically study submission of limit and market orders in a foreign exchange market. They find evidence that the balance of limit and market orders is affected by observable variables suggesting that traders choose order types strategically.

This paper is related to work on choice of trading venue and the interaction of a main exchange and a dark pool or crossing network. While we do not model traders' choice of venue, the connection is relevant because venue-choice results in endogenous segmentation across venues. Degryse, Van Achter, and Wuyts (2009) model the interaction of a dealer market and a crossing network. They find that a crossing network alongside a dealer market produces greater welfare for traders. Fleming and Nguyen (2013) study dark liquidity in the US treasuries market. They find greater use of dark liquidity at volatile times and that its informational role becomes relatively less important during those volatile times. Higher usage of dark liquidity is associated with higher market depth, lower bid-ask spreads and higher trading intensity.

Much of the literature on venue choice focuses on dark pools. Our study is related since RLPs are a specific type of dark pool. In general, the themes are similar: dark pools endogenously segment order flow, changing the relative concentration of information across trading venues. Boulatov and George (2013) model a market with informed traders who trade using limit orders and find that if limit orders are hidden (dark), prices become more efficient because competition between traders intensifies. Buti, Rindi, and Werner (2015) model a dark pool alongside a limit order book and show that a dark pool can have adverse effects on market quality. Boni et al. (2013) empirically study dark pools with participation constraints and find that those with constraints that are amenable to buy-side traders provide better execution quality to traders. Comerton-Forde and Putniņš (2014) study the impact of dark trading using Australian equity data and find that large dark trades are not associated with lower levels of liquidity, but that dark smaller trades are. They argue that uninformed traders are more likely to execute in the dark pool increasing adverse selection in the main exchange.

Two studies use data on a Canadian RLP. Comerton-Forde et al. (2016) study a 2012 rule change in Canada that resulted in a reduction in levels of trading in Canadian dark pools. They find the rule change resulted in greater quoted depth in the lit market, indicating that dark segmented trading had a negative effect on the lit market. Foley and Putniņš (2016) study the same restrictions on dark trading in Canada in 2012. They find that dark limit order markets are beneficial to market quality, reducing quoted, effective and realized spreads and increasing informational efficiency. Our model predicts that retail segmentation has an ambiguous effect on liquidity which may explain why conclusive effects are difficult to detect.

Our models are also related to work on optimal execution strategies. We derive an optimal execution strategy for a simplified case in order to incorporate game-theoretic elements into our model. A line of literature exists that focuses on more realistic models of execution that might be implemented in practice. This paper informs how the overall quality of those strategies might change when changes in market structure occur that involve segmentation. Cartea and Jaimungal (2015) develop an optimal execution strategy that balances the use of limit and market orders, as do traders in our model. Almgren (2003) derives optimal liquidation strategies for an asset while taking into account volatility risk and price impact. The more rapidly the trade execution must take place, the more uncertainty about the price of execution. Alfonsi and Schied (2010) model execution strategies in a limit order book and find that placing determin-

istic quantities at homogeneous time intervals is optimal, as opposed to randomizing traders. They find under general assumptions about a limit order book, price manipulation strategies are not effective. Obizhaeva and Wang (2013) model the effects of dynamic supply and demand on trading strategies within a limit order book setting. They show that the resilience of the order book is the key determinant of a strategy. Traders mix their strategies between large and small trades. Forsyth, Kennedy, Tse, and Windcliff (2012) derive optimal trading strategies under a mean-quadratic-variation criteria when the asset follows a geometric Brownian motion with price impact.

There are two main contributions of our paper: first, we give results on segmentation of retail traders through retail price improvement that give rise to testable empirical predictions; second, our second model contributes to microstructure theory as a novel version of the Glosten and Milgrom (1985) with dynamic order choice.

2.3 Model with Market Orders

Time is discrete with N periods denoted by t_1, t_2, \dots, t_N over which trading for a single risky asset takes place. After the final trading period at t_N the asset pays a random dividend v that we normalize to be 1 and 0 with ex-ante probabilities $\frac{1}{2}$. Denote by \tilde{v} the realization of the random variable v . We normalize the risk-free rate to be 0.

There are four classes of trader: a market maker who intermediates all trades, a measure α of informed institutional traders, a measure $(1 - \alpha)(1 - \beta)$ of uninformed institutional traders and a measure $(1 - \alpha)\beta$ of retail traders. The total measure is 1.

The market maker sets bid and ask prices b_i and a_i , each for a single unit of the asset, before trading at each time t_i when a single trader arrives at the market anonymously. The arriving trader is chosen uniformly at random from the total trading population independently of time t_i or the history of previous arrivals. The trader observes the market maker's quotes and chooses either to buy or sell. An institutional trader arriving at time t_i may buy or sell a single unit of the asset at prices a_i or b_i ; a retail trader arriving at time t_i may buy or sell a single unit of the asset at improved prices $a_i - c(a_i - b_i)$ or $b_i + c(a_i - b_i)$ for fixed $c \in [0, \frac{1}{2}]$. The prices available to the retail trader are within the bid-ask spread by some fraction c to capture the effect of an RLP's price improvement.¹ Each individual trader who arrives at the market trades only once. The market maker may revise his quotes at the beginning of each trading period. Our equilibrium concept is Perfect Bayesian. The classes of trader and the market maker are discussed in greater detail below, as well as further details on the game's equilibrium.

2.3.1 Informed Trader

An informed trader who arrives at the market observes a perfect signal of the asset's value \tilde{v} . He is risk-neutral and maximizes profit, that is, if he arrives at t_i and observes $\tilde{v} = 1$ ($\tilde{v} = 0$) then he buys (sells) if $a_i < 1$ ($b_i > 0$).

¹Prices in RLPs are typically set this way, allowing prices at specific fractions of the institutional best bid and offer.

2.3.2 Institutional Uninformed Trader

An institutional uninformed trader who arrives at the market realizes a random exogenous liquidity demand to either buy or sell one unit of the asset, each with equal probability. His demand is perfectly inelastic so he either buys or sells with probabilities $\frac{1}{2}$.

2.3.3 Retail Trader

A retail trader is identical to an institutional uninformed trader with the exception that he may trade with the market maker at the improved buy and sell prices $a_i - c(a_i - b_i)$ and $b_i + c(a_i - b_i)$ for fixed $c \in [0, \frac{1}{2}]$.² The parameter β represents the proportion of uninformed traders that are identified as retail and able to use the RLP. We will also refer to β as the *degree of retail segmentation* and investigate its role in the models throughout the paper. The parameter β has two interpretations which are indistinguishable in this model: the measure of retail traders present in the market, all of whom use the RLP; or the measure of uninformed traders who are retail and use the RLP.

2.3.4 Market Maker

The market maker is risk-neutral and acts perfectly competitively, making zero profit from each trade in expectation. Since trade is anonymous, the market maker cannot infer whether an institutional trade was informed or uninformed. However, the market maker can observe the prices of each trade and infer when a trade was retail as well as if a trade was a buy or a sell. The market maker does not observe \tilde{v} but gradually learns the value of \tilde{v} from the pattern of trades.

The market maker maintains a belief about the true value of the asset, that is, a probability distribution that is revised as more information is learned from trading. Since the asset's payoff is Bernoulli, the market maker's belief is characterized by a single number, the probability of the asset paying 1 after all trading takes place. Let p_i denote the market maker's Bayesian prior about \tilde{v} before trading occurs at time t_i and denote the set of actions that the market maker can observe by $\mathcal{A} = \{I_{buy}, I_{sell}, R_{buy}, R_{sell}\}$ where the letters I and R refer to institutional and retail. Let $A_i \in \mathcal{A}$ be the action observed at t_i .

2.3.5 Perfect Bayesian Nash Equilibrium

A Perfect Bayesian Nash equilibrium (Fudenberg & Tirole, 1991) in this context is a pair of sequences of bid and ask prices $\{a_i\}$ and $\{b_i\}$ for $i = 1, \dots, N$ that satisfy the zero expected profit conditions in Equations 2.1 and 2.2 for all i . Each period there are multiple pairs a_i and b_i that lead to zero profit in expectation, so we refine the equilibrium to one where both the bid and ask prices independently result in zero expected profit, given the history of trade A_{i-1}, A_{i-2}, \dots . Equation 2.1 represents expected profit from posting an ask a_i and Equation 2.2 represents expected profits from posting a bid b_i . The equations are linked through the term that represents price improvement for retail trades $c(a_i - b_i)$ which depends on the bid-ask

²For a detailed study on the level of price improvement in dark markets, see Brolley (2016).

spread. Let the \mathbb{E}_m^i be the expectation operator for the market maker's information set at time i .

$$(1 - \beta(1 - \alpha))\mathbb{E}_m^i[a_i - v \mid A_i = I_{buy}; A_{i-1}, A_{i-2}\dots] + \beta(1 - \alpha)\mathbb{E}_m^i[a_i - c(a_i - b_i) - v \mid A_i = R_{buy}; A_{i-1}, A_{i-2}\dots] = 0 \quad (2.1)$$

$$(1 - \beta(1 - \alpha))\mathbb{E}_m^i[v - b_i \mid A_i = I_{sell}; A_{i-1}, A_{i-2}\dots] + \beta(1 - \alpha)\mathbb{E}_m^i[v - b_i - c(a_i - b_i) \mid A_i = R_{sell}; A_{i-1}, A_{i-2}\dots] = 0. \quad (2.2)$$

The terms $\beta(1 - \alpha)$ and $(1 - \beta(1 - \alpha))$ are the probabilities of trading with a retail trader and with a non-retail trader respectively. Solving Equations 2.1 and 2.2 simultaneously yields

$$a_i = \frac{4\alpha p_i(p_i - 1) + p_i(1 - \alpha)(\alpha + 2\beta c - 2\alpha p_i - 1)}{4\alpha^2 p_i(p_i - 1) + (\alpha^2 \beta c - \alpha - 1)(1 - \alpha)} \quad (2.3)$$

$$b_i = \frac{p_i(1 - \alpha)(\alpha + 2\beta c - 2\alpha p_i - 1)}{4\alpha^2 p_i(p_i - 1) + (\alpha^2 \beta c - \alpha - 1)(1 - \alpha)}. \quad (2.4)$$

Details of the calculation are shown in Appendix A.1.

Figure 2.1 shows the bid and ask prices versus the market maker's Bayesian prior on the true value of the asset. The bid and ask prices both increase with the market maker's Bayesian prior. The bid price is always less than the ask price to ensure the market maker profits when trading with uninformed traders. Both prices are also always between the two possible values of the asset, 1 and 0. Although this means the market maker will lose when trading with an informed trader, he can recover these losses through the bid-ask spread.

The bid-ask spread s_i is given by

$$s_i := a_i - b_i = \frac{4\alpha p_i(p_i - 1)}{4\alpha^2 p_i(p_i - 1) + (\alpha^2 \beta c - \alpha - 1)(1 - \alpha)}. \quad (2.5)$$

Notice the bid-ask spread is increasing in the variance of the market maker's belief $p_i(1 - p_i)$; the bid-ask spread reflects the market maker's uncertainty about the asset because it corresponds to adverse selection risk from informed traders.

Figure 2.2 shows the bid-ask spread versus the market maker's prior variance for various parameter values. The bid-ask spread increases with the variance of the market maker's belief about the true value of the asset $p_i(1 - p_i)$. The market maker begins the game with a prior of $\frac{1}{2}$ which corresponds to when his uncertainty about the true value of the asset is highest. As the market maker learns about the true value and the variance decreases, he is willing to quote a tighter bid-ask spread.

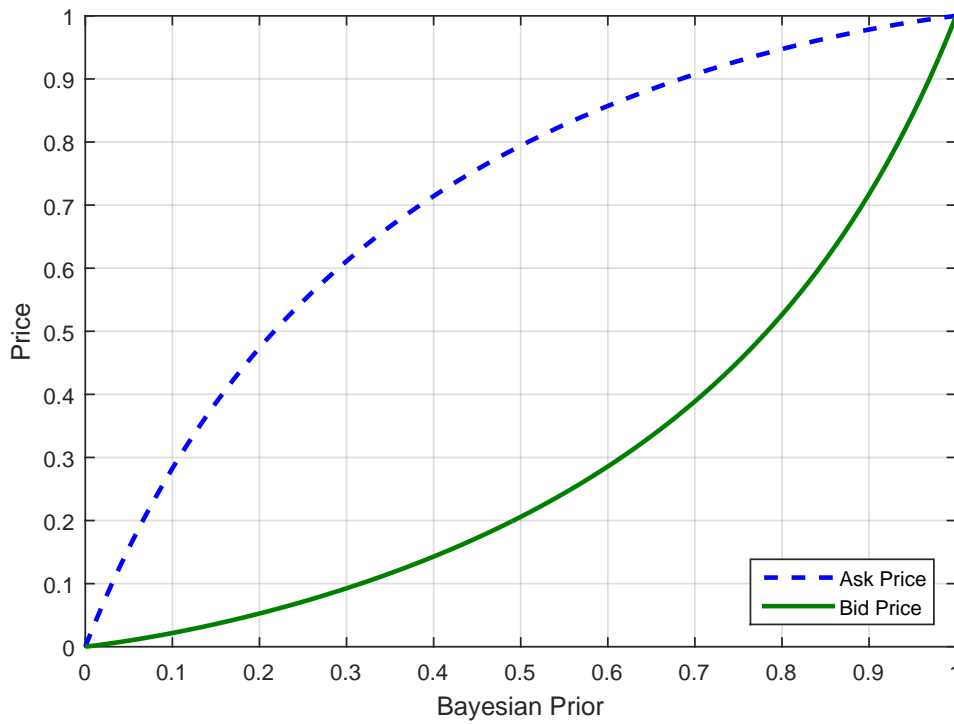
Proposition 2.3.1 *All else equal, the bid-ask spread for institutional investors widens as a larger proportion of retail traders are segregated from the total trading population. That is, $s_i(\beta)$ is increasing in β . A proof is found in Appendix A.2.*

The greater the measure of retail traders β , the more profits the market maker forgoes because of the better prices offered to retail traders. To compensate, the market maker offers a wider spread to institutional traders.

After each trade, the market maker updates his belief about the asset's payoff. The market maker's Bayesian update for p_i is

$$\frac{p_{i+1}}{1 - p_{i+1}} = \frac{p_i}{1 - p_i} \xi_i \quad (2.6)$$

Figure 2.1: **Bid and ask prices.** The figure shows the bid and ask prices as a function of the market maker's prior about the true value of the asset. The figure was created with parameters $\alpha = \frac{1}{2}$, $\beta = \frac{1}{2}$ and $c = \frac{3}{10}$.

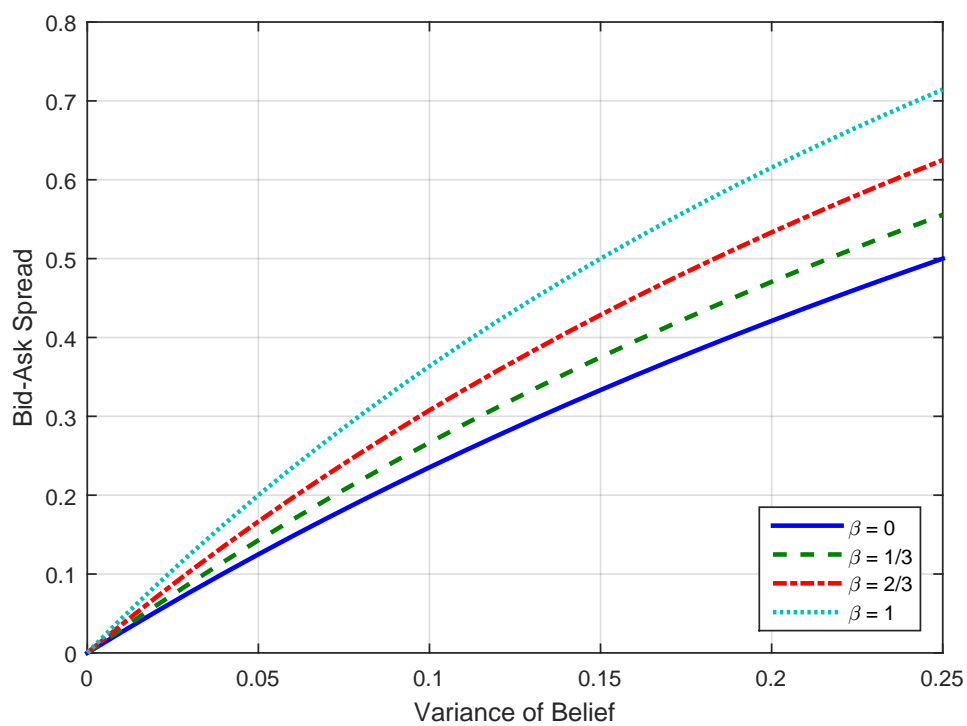


$$\xi_i = \begin{cases} \frac{2\alpha+(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)} & \text{if } A_i = I_{buy} \\ \frac{(1-\alpha)(1-\beta)}{2\alpha+(1-\alpha)(1-\beta)} & \text{if } A_i = I_{sell} \\ 1 & \text{if } A_i = R_{buy} \text{ or } R_{sell}. \end{cases} \quad (2.7)$$

which comes from Bayes Rule for Bernoulli random variables. The factors ξ_i are ratios of the probability of an observed action given the asset's true value is high, over the probability of the action given the asset's true value is low. A table of the conditional probabilities is given in Table A.1.

After finitely many trades, the market maker will never be certain of the asset's true value. Certainty corresponds to the case of the ratio $\frac{p_i}{1-p_i}$ being either zero or infinity, neither of which are attainable by sequential multiplication by a finite number of factors ξ_i . However, it is interesting to consider the case when a given amount of information has been impounded into the market maker's belief. Consider a joint hypothesis test to accept the hypothesis that $\nu = 1$ and simultaneously reject the hypothesis that $\nu = 0$. This can be done by defining an upper boundary H for the likelihood ratio $\lambda_i := \frac{p_i}{1-p_i}$. If λ_i crosses the boundary, the market maker is able to accept that $\nu = 1$ with some level of confidence. It is shown in the proof of Proposition 2.3.2 that the corresponding lower boundary to simultaneously reject $\nu = 0$ is $\frac{1}{H}$. Note that H must be chosen from one of the possible values taken by λ_i . λ_i takes values on the lattice $(\frac{2\alpha+(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)})^k$ for $k \in \mathbb{Q}$ since each time the market maker observes a trade and updates his belief about the asset's value, he either divides or multiplies by either 1 or the factor $\frac{2\alpha+(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)}$. If H is not chosen on this lattice, the process λ_i may overshoot H leading to any analysis being

Figure 2.2: **Bid-ask spread.** The figure shows the bid-ask spread versus the variance of the market maker's belief about the true value of the asset. The parameter β is the degree of retail segmentation. The figure was created with parameters $\alpha = \frac{1}{2}$, $\beta \in \{0, \beta = \frac{1}{3}, \beta = \frac{2}{3}, \beta = 1\}$, and $c = \frac{3}{10}$.



imprecise. This problem of the process overshooting the boundary will be relevant in Section 2.4.

Let the game end when the likelihood ratio $\lambda_i := \frac{p_i}{1-p_i}$ reaches the upper boundary H or the lower boundary $\frac{1}{H}$, $\lambda_i = H$ or $\lambda_i = \frac{1}{H}$ for $H \in [1, \infty)$. This means the number of periods in the game N is a random variable. A priori, it is not clear if segmentation will result in the market maker learning more efficiently. Price impact for institutional traders increases, while price impact for retail traders becomes zero. While more is learned on average from an institutional trade, the market maker's belief is static after retail trades since they convey no information. We show below that segmentation always results in faster learning in the sense of a decrease in the expected number of trades $\mathbb{E}[N]$ before either the upper or lower boundary is hit.

Proposition 2.3.2 *As the measure of retail traders increases, on average, the market maker is able to learn about the asset's value faster. That is, the expected number of trades $\mathbb{E}[N]$ needed for the likelihood ratio λ_i to reach the upper boundary H or lower boundary $\frac{1}{H}$ is*

$$\mathbb{E}[N] = \frac{\frac{H-1}{H+1} \log(H)}{\alpha \log\left(1 + \frac{2\alpha}{(1-\alpha)(1-\beta)}\right)}, \quad (2.8)$$

and $\mathbb{E}[N]$ is decreasing with the measure of retail traders β . A proof is found in Appendix A.3.

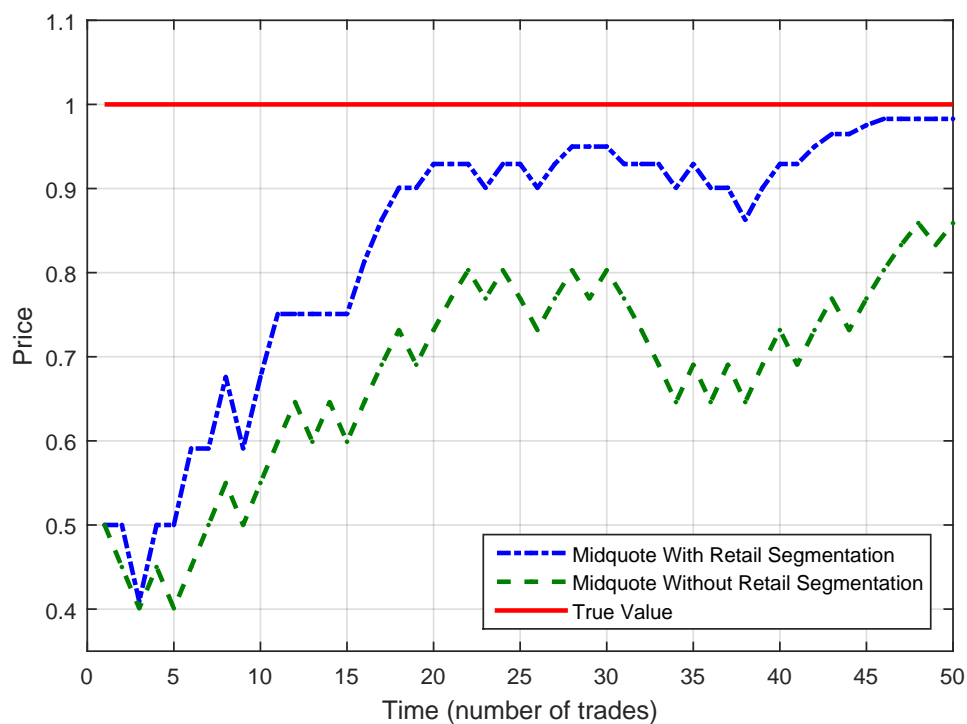
Figure 2.3 shows the results of a numerical simulation of a path of the game. The market maker learns about the asset's true value faster when retail trades are segmented. Both processes were generated using the same random draws, but the process represented by the blue line was created by identifying certain random trades as retail and taking into account changes to the market maker's Bayesian update. The green line serves as a benchmark where no retail segmentation occurs. The correlation between the processes is clear. This comes from the pattern of institutional buys and sells being common to both processes. However, the blue line occasionally moves horizontally; these movements correspond to retail trades where the market maker's prior remains constant. Each move taken the blue line is greater in magnitude than moves for the green. This is because the market maker is relatively more certain that institutional trades come from informed traders when retail trades are segmented. It is apparent in the figure that the process with retail segmentation has a stronger tendency toward the true value of the asset, that is, prices are more efficient.

The proposition above shows that information is impounded into prices faster when retail orders are segmented from the total order flow. The components of the formula have a natural interpretation analogous to "time equals distance over speed". The numerator is the expected value of the log-likelihood ratio when it hits either boundary, that is, the expected distance the log-likelihood ratio moves, while the denominator is the expected move at each step, that is, the average speed of the log-likelihood ratio per trade where there is one trade per unit of time.

Proposition 2.3.2 implies the likelihood ratio evolves more quickly, but does not say anything about the possibility of an increase in errors in the hypothesis test. It is possible to calculate the probability that the hypothesis test defined by H results in an error.

Proposition 2.3.3 *The probability of the market maker making a error in inference, Type 1 or Type 2, either accepting the hypothesis that the asset's value is high when it is low or vice versa*

Figure 2.3: **Numerical example.** This figure shows the results of a numerical simulation of one possible path of the game for two values of β . The blue and green lines correspond to two simulations of the market maker's prior as it evolves with one trade per unit time. The true value of the asset is shown with the red line. The figure was created with parameters $\alpha = \frac{1}{10}$, $\beta = \frac{1}{2}$ and $c = \frac{3}{10}$.



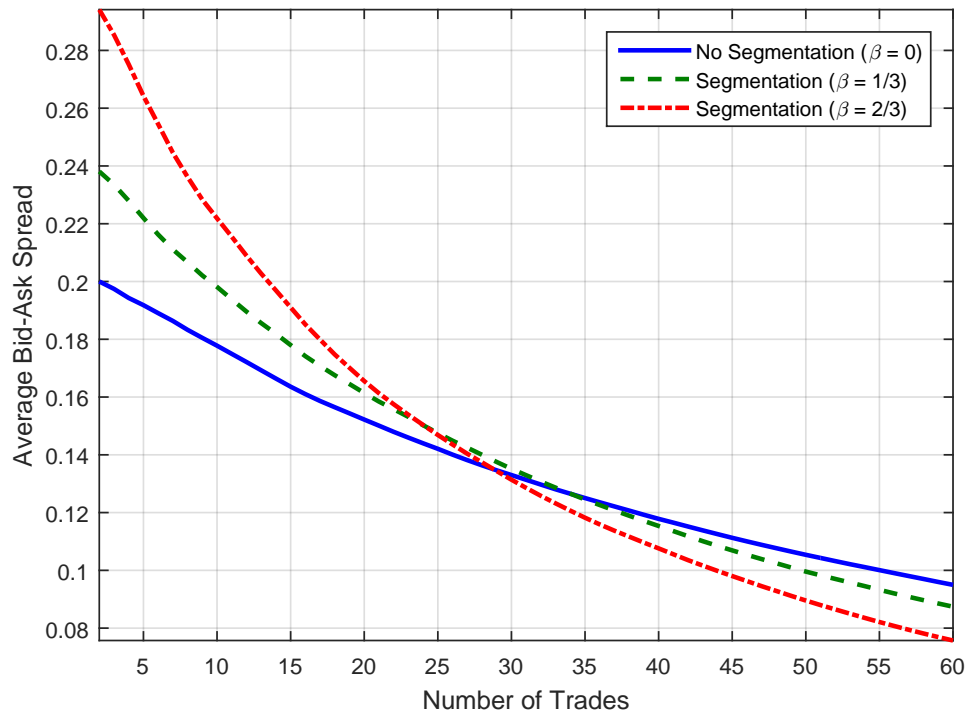
is given by

$$\frac{1}{1 + H}$$

The probability of the test resulting in error in inference is independent of the degree of retail segmentation β . A proof is found in Appendix A.4 and A.5.

The probability of error in the hypothesis test is important to combine with Proposition 2.3.2 because, a priori, it is possible that while the average number of trades required to gain a given level of confidence about the asset's true value may decrease, the probability of error may increase. If this were true, it would be difficult to come to a conclusion about price efficiency. However, Proposition 2.3.3 shows that this is not the case; the probability of the test resulting in error is independent of the measure of retail traders β while the expected duration of the test is decreasing in β .

Figure 2.4: **Average liquidity.** Each line on the graph shows the average bid-ask spread (on the y-axis) over periods 1 to N for several simulations of the game, where N is the total number of trades (on the x-axis). The solid blue line shows the result for measure zero of retail segmentation ($\beta = 0$) and the dashed green and red lines show the result for nonzero measure of retail segmentation ($\beta = \frac{1}{3}$ and $\beta = \frac{2}{3}$). The plot was created with parameters $\alpha = \frac{1}{5}$, $\beta = \{0, \frac{1}{3}, \frac{2}{3}\}$ and $c = \frac{3}{10}$ with 100 trials for each point on the x-axis.



While Proposition 2.3.1 predicts that, all else equal, liquidity should worsen with segmentation, Proposition 2.3.2 predicts that the market maker will learn faster about the asset's true value leading to less uncertainty about the asset and therefore tighter bid-ask spreads. It is not clear which one of these opposing forces dominates on average. While liquidity may initially be worse with segmentation, it will improve faster due to the increase in price efficiency.

To understand the effect of segmentation on liquidity over time, we consider the bid-ask spread averaged over all trading periods 1 through N . It is not clear how to prove a result about average liquidity in general. The formula for the bid-ask spread becomes intractable under expectation over multiple periods, so we proceed with numerical simulation. Figure 2.4 shows that average liquidity improves if sufficiently many trades occur. The result suggests that segmentation may have benefits for uninformed institutional traders as well as retail traders. The green dashed line and red dash-dot line are initially higher than the solid blue line, indicating lower liquidity because, as in Proposition 2.3.1, all else equal, the bid-ask spread is larger when retail trades are segmented. However, the dashed lines decrease at a faster rate than the solid blue line. This is because price efficiency is greater when retail trades are segmented. Depending on how many trades occur in the game N , the dashed lines may or may not cross the solid line, indicating that the effect of retail segmentation on average liquidity is ambiguous.

2.4 Model with Order Choice

In the section above, we developed a model where informed traders and uninformed traders use market orders. In practice, central limit order books allow traders to use both limit orders and market orders. A market order has the advantage of certainty of execution, while a limit order allows a trader to capture the bid-ask spread but with no guarantee of execution. Traders may minimize price impact by using a combination of order-types since it may be more difficult for a market maker to learn from the pattern of trade and quotes than if trades were conducted using only a single order type. As discussed in Section 2.2, the choice between market orders and limit orders has been shown to be an important aspect of trading behaviour. We are interested in how informed traders mitigate the informational impact of their trades when retail orders are segmented, and if the improvement in price efficiency predicted in Section 2.3 is still present when traders use market orders and limit orders strategically.

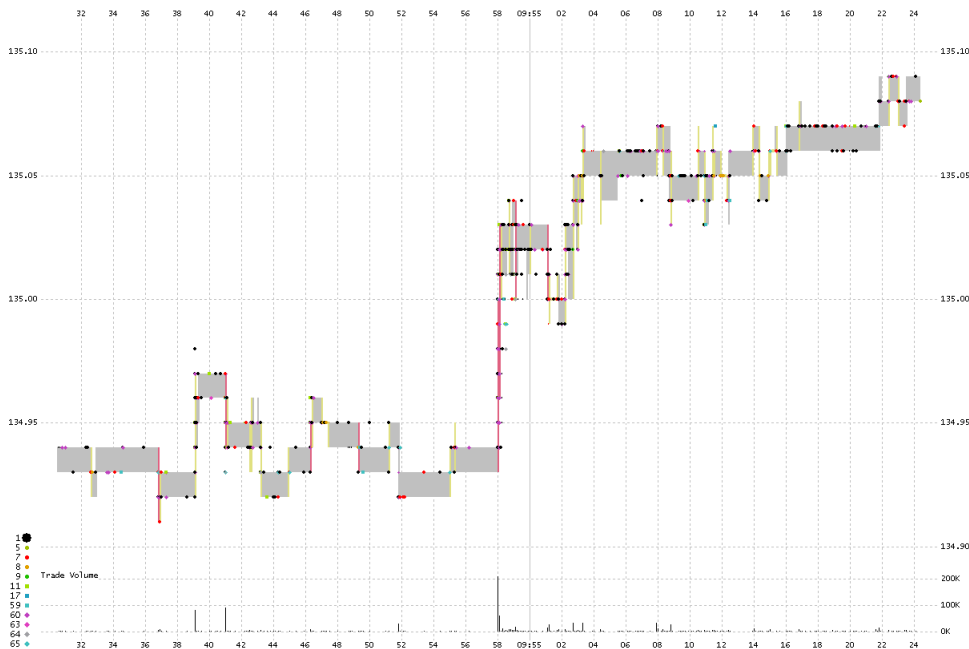
In this section we develop a model in the spirit of the Glosten and Milgrom (1985) model above that is simplified in certain dimensions in order to accommodate the possibility for traders to choose which types of orders they use. We examine the effect of segmentation of retail traders on price-efficiency and the optimal strategy of the informed trader. Our model is related to Liu and Kaniel (2004) which also extends Glosten and Milgrom (1985) to include the choice between market and limit orders. Our paper is distinguished by having a different dynamic setup: Liu and Kaniel (2004) is a three-period model while ours has potentially infinitely many trading periods. Another difference in our model is that we impose a tick constraint on prices. While we focus on segmentation of retail market orders, Liu and Kaniel (2004) focuses on the total number of uninformed traders. The results are generally compatible with ours; that informed traders strategically submit limit orders to avoid moving the market maker's prior, and that the informed trader's strategy will affect the informativeness of limit orders.

Our aim is to design a model with the following features: traders can submit either limit orders or market orders; the market maker can observe the state of the limit order book and therefore learn from that information; and retail market orders are identifiable as such, as is the case with RLPs. In the previous section, the informed trader's strategy was trivial while in this section, the informed trader will be long-lived and balance per-period profits with the informational impact of his trades. We will focus on a case where the market maker begins

the game willing to quote a bid-ask spread that is constrained by a minimum price increment. After learning sufficiently about the asset's value, the market maker revises his quotes either upward or downward. The informed trader will attempt to delay this revision since the market maker's quote revision will tend to diminish his potential to earn future trading profits.

Figure 2.5 shows that empirically, at high frequencies, prices and quotes move discretely on the grid defined by the minimum price increment which is commonly one cent. The figure shows trades and quotes for roughly one minute of trading that occurred on 16 June 2011 for an S&P 500 exchange-traded fund, SPY. Our model has certain features in common with the figure; that most trades occur at either the best bid or ask price, and that these prices often remain constant for several trades.

Figure 2.5: Empirical trade and quote data. The figure shows empirical trading data aggregated across several trading venues in the US for roughly one minute of trading that occurred on 16 June 2011 for an S&P 500 exchange-traded fund. Grey bands indicate the bid-ask spread, coloured dots indicate trades, and volume is indicated by vertical bars at the bottom of the figure. This graph was reproduced with the permission of Nanex, LLC.



2.4.1 Model Setup

Time is discrete and infinite denoted by t_0, \dots, t_i, \dots over which trading for the risky asset takes place. There is one trade in each period for an asset that pays a random dividend v of value 1 or 0 at $i = \infty^3$ with ex-ante probabilities $\frac{1}{2}$. Let the realized draw of the asset's value be \tilde{v} . The risk-free rate is normalized to zero.

³The interpretation of the asset's value being realized when time runs to infinity is that there are many opportunities to trade the asset before the public revelation of information. This applies well to electronic markets where high-frequency trading algorithms operate at the millisecond frequency.

There are four classes of trader: a market maker, a single informed trader, passive noise traders, and active noise traders which we describe in more detail below. Trading takes place on an exchange as described below.

2.4.2 Exchange Trading Mechanics

Trading takes place sequentially with one trade every period. The exchange makes four possible prices available, a “tick grid”: $0 < x_L < x_l < x_h < x_H < 1$. The prices are symmetric around one-half; $x_h = 1 - x_l$ and $x_L = 1 - x_H$. Each period is divided into two sub-periods. In the first, traders submit limit orders and in the second, market orders. The exchange operates as a limit order book that holds three orders: it allows the market maker to quote a bid and ask price simultaneously as well as a single bid or ask order from one other trader. Each order is for a single unit of the risky asset.⁴

Limit orders are firm commitments to trade that last until the end of a period and are cancelled thereafter regardless of whether they were executed. If there are two limit orders at a given price, the order submitted first is accepted into the limit order book while the other fails to be accepted into the limit order book.⁵

We leave aside price improvement for retail traders (the parameter c in the previous model) as a simplification. The model setup captures the effects the change to the anonymity brought about by segmentation, which is the focus of this section.

In the second sub-period, traders may each submit a market order for one unit of the asset. The first market order submitted is executed against the corresponding limit order and the period ends. Each period, one trader’s limit order will be accepted. After that, the market maker submits his bid and ask limit orders. The state of the limit order book is publically observed after each period. Figure 2.6 shows the model timing described above graphically. We describe each class of trader below.

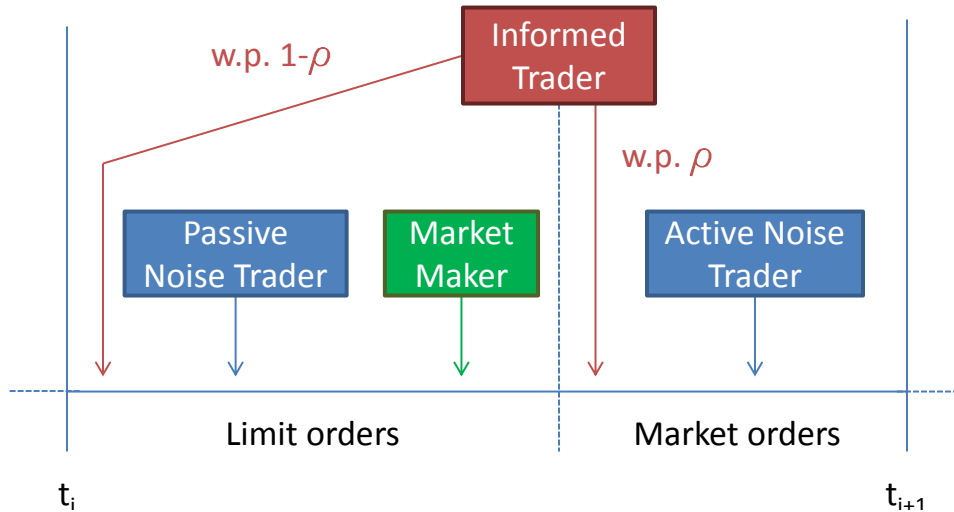
2.4.3 Market Maker

The market maker quotes a bid and ask price at which he is willing to buy and sell, respectively. The market maker is perfectly competitive and therefore chooses prices such that his expected profits are minimal but nonnegative. Depending on traders’ strategies and the Bayesian prior on the asset’s true value, the market maker may be willing to quote bid and ask prices at adjacent ticks $\{x_L, x_l\}$, $\{x_l, x_h\}$, or $\{x_h, x_H\}$. We will focus on the case when the market maker is willing to quote a bid and an ask price at the inner price pair $\{x_l, x_h\}$. This is illustrated in Figure 2.7. When the market maker quotes prices at an adjacent price pair, the bid-ask spread cannot be decreased given the tick constraint. Therefore, a trader never submits a limit buy or sell order

⁴In practice, at any given price, traders may submit any number of orders for any quantity. However, the number of possible strategies that arise with more possibilities for the limit order book makes analysis of equilibrium difficult. The simplified limit order book has the relevant qualities needed for an analysis of equilibrium with order choice: traders can compete with the market maker using limit orders, and an imbalance in limit orders may have some signal value.

⁵Most exchanges operate according to this priority structure; limit orders with earlier time stamps gain execution priority over others at the same price.

Figure 2.6: **Model timing.** A single period is represented by the actions shown from left to right between solid blue lines. In the first sub-period, limit orders are submitted, and in the second, market orders. The informed trader randomized between submitting a limit order or a market order, with probability (w.p.) ρ of submitting a market order. If the informed trader submits a limit order, he does so before the passive noise trader with probability α . If the informed trader submits a market order, he does so before the active noise trader with probability α .



at a price that is not equal to the market maker's buy and sell prices. Indeed, a lower-priced buy order has no chance of execution, as is the case for higher-priced sell orders.

2.4.4 Active Noise Trader

Each period, a new active noise trader arrives at the exchange with a random immediate liquidity need for the risky asset. An active noise trader submits a market buy or market sell order, each with probability $\frac{1}{2}$. A measure β of active noise traders are retail. If trade involves a retail trader, it is identified as such at the end of the trading period. This information will be used by the market maker when updating his prior. The parameter β is the proportion of retail traders within the set of active noise traders. We will also call β the degree of retail segmentation.⁶

2.4.5 Passive Noise Trader

Each period, a new passive noise trader arrives at the exchange with a random liquidity need for the asset. A passive noise trader submits a limit buy or limit sell order, each with probability $\frac{1}{2}$. The passive noise trader chooses the same limit prices as the market maker since a lower bid or higher ask have no chance of being executed, and a higher bid or lower ask are not possible due to the tick constraint.

⁶This is in accordance with how RLPs work in practice where retail market orders are segmented as opposed to limit orders.

2.4.6 Informed Trader

The informed trader is long lived and may trade multiple times over the course of the game. He is endowed with knowledge of the asset's value \tilde{v} . He is risk neutral and therefore maximizes total expected profit. Each period, he chooses either to submit a market order or a limit order. A given order choice may result in higher per-period profit but may have a greater impact on the market maker's prior (we will call this price impact or informational impact, interchangeably), diminishing the potential for future profits. He balances per-period profit with price impact; the rate at which the market maker learns about the value of the risky asset through the pattern of trade.

The informed trader may use a mixed strategy, randomly choosing to either submit a limit order or a market order each period with a strategically chosen probability. Let $\bar{\rho}_i$ and $\underline{\rho}_i$ be the probabilities that the informed trader submits a market order at time t_i , given that the asset's true value is 1 and 0 respectively. When the informed trader submits an order, limit or market, the order is submitted ahead of a noise trader with probability α . We assume α applies to both limit and market orders for simplicity. The informed trader's limit order must be matched with an active noise trader's market order to be executed. Therefore, when the informed trader submits a market order, it will be executed with probability α since there is always a limit order present to execute against; if the informed trader submits a limit order, it is executed with probability $\alpha/2$ where division by two comes from the active noise trader randomly submitting a market buy or sell with probabilities $\frac{1}{2}$. When the informed trader's limit or market order is submitted before a noise trader's, we will call it a "fast" order, and "slow" otherwise.⁷

2.4.7 Market Maker's Strategy

We will show that for certain model parameters, the market maker will be willing to quote the inner price pair $\{x_l, x_h\}$ until a particular level of information is attained about the asset. Denote the market maker's prior on the asset's true value being high $v = 1$ at period i by p_i . Assume the informed trader submits a market order at period i with probability $\underline{\rho}_i$ given the the asset's value is zero, and $\bar{\rho}_i$ given the asset's value is one. Consider the market maker's zero-expected-profit condition if he quotes at the inner price pair $\{x_l, x_h\}$.

$$\begin{aligned}
 & p_i(\alpha(\bar{\rho}_i(\frac{x_h - 1}{2}) + (1 - \bar{\rho}_i)(\frac{x_h - 1}{2}))) + (1 - \alpha)(\frac{x_h - 1}{4} + \frac{1 - x_l}{4}) \\
 & + (1 - p_i)(\alpha(\underline{\rho}_i(\frac{-x_l}{2}) + (1 - \underline{\rho}_i)(\frac{-x_l}{2}))) + (1 - \alpha)(\frac{-x_l}{4} + \frac{x_h}{4}) \\
 & = \frac{1}{4} - \frac{1}{2}x_l - \frac{1}{4}\alpha \geq 0
 \end{aligned} \tag{2.9}$$

where we use the symmetry in prices $x_h = 1 - x_l$ to simplify the formula. Equation 2.9 is an expectation over several random variables: the asset's true value being high p_i ; an informed

⁷In practice, orders are submitted at any time, and limit order submitters compete to have their orders at the "top of the book". That is, when the best bid or ask price changes, high-frequency trading firms race to submit limit orders ahead of others and invest heavily in technology to gain a speed advantage. This could be thought of as an investment in the parameter α , but we treat α as exogenous.

Table 2.1: **Trading outcomes, probabilities and market maker profits.** This table shows combinations of the informed trader’s (Inf.) order type, whether the order is fast or slow, and the noise traders’ (Pass. N. and Act. N) actions, which limit orders are at the “top” of the order book (labelled Best bid and Best ask), each combination’s probability, and the market maker’s (MM) profit given an outcome. Bold entries in the Best bid and Best ask columns indicate the executed limit order. We have left out certain entries when they are irrelevant for computing expected profits. For example, when the informed trader submits a fast limit order, the passive noise trader’s order will never be executed, so the row combines the probabilities of the passive noise trader submitting a buy or a sell. The table is for outcomes given the true value of the asset is high and hence the high probability of a market order $\bar{\rho}$ is used.

Inf. order	Inf. speed	Act. N. order	Pass. N. order	Prob.	Best bid	Best ask	MM profit
limit	fast	buy		$(1 - \bar{\rho}) \cdot \alpha \cdot \frac{1}{2}$	Inf.	MM	$x_h - 1$
limit	fast	sell		$(1 - \bar{\rho}) \cdot \alpha \cdot \frac{1}{2}$	Inf.	MM	0
limit	slow	buy	buy	$(1 - \bar{\rho}) \cdot (1 - \alpha) \cdot \frac{1}{2} \cdot \frac{1}{2}$	N.	MM	0
limit	slow	buy	sell	$(1 - \bar{\rho}) \cdot (1 - \alpha) \cdot \frac{1}{2} \cdot \frac{1}{2}$	N.	MM	0
limit	slow	sell	buy	$(1 - \bar{\rho}) \cdot (1 - \alpha) \cdot \frac{1}{2} \cdot \frac{1}{2}$	MM	N.	0
limit	slow	sell	sell	$(1 - \bar{\rho}) \cdot (1 - \alpha) \cdot \frac{1}{2} \cdot \frac{1}{2}$	MM	N.	0
market	fast		buy	$\bar{\rho} \cdot \alpha \cdot \frac{1}{2}$	N.	MM	$x_h - 1$
market	fast		sell	$\bar{\rho} \cdot \alpha \cdot \frac{1}{2}$	MM	N.	0
market	slow	buy	buy	$\bar{\rho} \cdot (1 - \alpha) \cdot \frac{1}{2} \cdot \frac{1}{2}$	N.	MM	$x_h - 1$
market	slow	sell	buy	$\bar{\rho} \cdot (1 - \alpha) \cdot \frac{1}{2} \cdot \frac{1}{2}$	N.	MM	0
market	slow	buy	sell	$\bar{\rho} \cdot (1 - \alpha) \cdot \frac{1}{2} \cdot \frac{1}{2}$	MM	N.	0
market	slow	sell	sell	$\bar{\rho} \cdot (1 - \alpha) \cdot \frac{1}{2} \cdot \frac{1}{2}$	MM	N.	$1 - x_l$

trader being present α ; the informed trader submitting a market order $\bar{\rho}_i$ and $\underline{\rho}_i$; and probabilities of uninformed traders submitting market buy or sell or limit buy or sell orders which are accounted by in the division by 2 and 4. The details of the computation of Equation 2.9 are shown in Table 2.1 which gives the probabilities for outcomes given the asset’s value is high. Summing the last column of Table 2.1, multiplied by the probabilities in the table gives the first line of Equation 2.9, the market maker’s expected profit when true value is high.

Equation 2.9 shows that the market maker is willing to quote at x_l and x_h if $\alpha \leq 1 - 2x_l$. We will assume that $0 < \alpha \leq 1 - 2x_l$. Note that the market maker’s profits at the inner price pair are independent of his prior p_i about the asset’s value. This is because the market maker will lose the same amount in expectation regardless of whether the informed trader submits a market or limit order due to the symmetry of the prices, $x_p = 1 - x_h$. Even though the expected profit when quoting the inner price pair is independent of the prior, the market maker will learn about the value of the asset by observing the pattern of market orders and limit orders.

It is interesting to note that whenever the informed trader’s order, either limit or market, is accepted the market maker realizes a loss. For market orders, the reasoning is the same as for the model in Section 2.3; the informed trader buys (sells) the asset when the market maker underprices (overprices) the asset. For limit orders, the informed trader submits a limit buy (sell) order when the asset is worth one (zero) for a chance to trade with an active noise trader;

in this case, the only possible trade for the market maker is with an active noise trader who submits a buy (sell) market order resulting in the market maker selling (buying) at a loss.

As described above, the market maker's prior will evolve as he observes trades. Since the market maker is competitive, he quotes the spread that yields the lowest nonnegative profits. The market maker's profits will depend on the prices made available on the tick grid. We would like our model to have the reasonable property that as the market maker's prior moves up (down) sufficiently, he revises his prices upward (downward). The following lemma shows that for certain reasonable values of the prices x_L , x_l , x_H and x_h , this is the case. Figure 2.7 illustrates the market maker's profits at the three price pairs for sample model parameters.

Lemma 2.4.1 *There are fixed prices for which the market maker will quote the inner price pair $\{x_l, x_h\}$. The market maker will prefer to quote upper price pair $\{x_h, x_H\}$ when his prior moves above the threshold $\bar{\theta}$; the market maker will prefer to quote lower price pair $\{x_L, x_l\}$ when his prior moves below the lower threshold $\underline{\theta}$. A proof is found in Appendix A.6.*

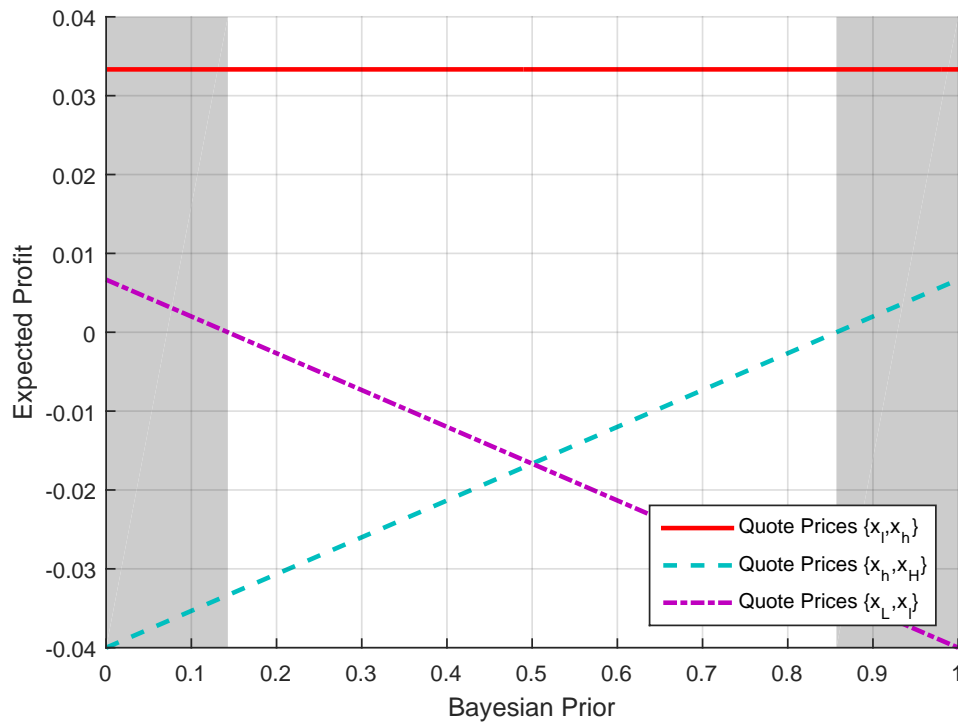
Figure 2.7 shows profits for different possible quote pairs as a function of his Bayesian prior on the asset's true value. Since the red solid line is always greater than zero, the market maker is always willing to quote the inner pair. As the market maker becomes more certain the asset's payoff is high, expected profits at the higher pair $\{x_h, x_H\}$ become greater until they eventually become positive. Profits for the higher pair $\{x_h, x_H\}$ are increasing in the market maker's prior because they are closer to the high asset value of 1 than the low asset value of 0. A similar logic applies to the lower price pair. Since the market maker acts competitively, he will adjust quotes in the natural way: he will switch to quoting the higher prices pair when his prior increases sufficiently, or switch to the lower price pair after his prior decreases sufficiently.

The market maker's Bayesian update will include information about the state of the limit order book observed after a period. Each period there is one observable limit order in addition to the market maker's bid and ask, and one observable market order. If the market order is identified as retail, whether it was a buy or sell conveys no information about the asset's value so we denote distinct cases of retail buy and sell orders with the same variable. Therefore, there are six possible outcomes A_i for trades and quotes each period and six corresponding Bayesian updates:

$$\xi_i = \begin{cases} \frac{\alpha(1+(2\bar{p}-1)\beta)-(\beta-1)}{(1-\beta)(1-\alpha)} & \text{if } A_i = LBMB \\ \frac{1-\alpha(2\bar{p}-1)}{1+\alpha(2\bar{p}-1)} & \text{if } A_i = LBMS \\ \frac{1+\alpha(2\bar{p}-1)}{1-\alpha(2\bar{p}-1)} & \text{if } A_i = LSMB \\ \frac{(1-\beta)(1-\alpha)}{\alpha(1+(2\bar{p}-1)\beta)-(\beta-1)} & \text{if } A_i = LSMS \\ \frac{1-\alpha(2\bar{p}-1)}{1-\alpha} & \text{if } A_i = LBR \\ \frac{1-\alpha}{1-\alpha(2\bar{p}-1)} & \text{if } A_i = LSR \end{cases} \quad (2.10)$$

where *LBMB* is an observation of a limit buy order (*LB*) followed by a market buy order (*MB*), *LBMS* is an observation of a limit buy order followed by a market sell order (*MS*), *LBR* is an observation of a limit buy order followed by a retail market (*R*), etc. The market maker updates his belief by multiplying his likelihood ratio $\lambda_i := \frac{p_i}{1-p_i}$ (as in Section 2.3) by the corresponding factor ξ_i .

Figure 2.7: **Market maker's expected profit.** The figure shows the market maker's expected profit from quoting different bid-ask spreads: the red solid line represents expected profits from the inner pair $\{x_l, x_h\}$, the cyan dashed line the higher pair $\{x_h, x_H\}$, and the purple dash-dot line the lower pair $\{x_L, x_l\}$. The grey regions indicate where the market maker quotes either the upper or lower price pairs. The market maker prefers to quote the upper or lower price pair when they yield positive expected profit because the market maker acts competitively. The figure was created with model parameters: $x_h = \frac{4}{5}$, $x_l = \frac{1}{5}$, $x_L = \frac{1}{3}$, $x_H = \frac{2}{3}$ and $\alpha = \frac{1}{5}$.



The updating factors in Equation 2.10 are ratios of the probability of the outcome given the asset's value is high over the probability of the outcome given the asset's value is low. For example, the probability of *LBMB* given the asset's value is high is:

$$\begin{aligned} & \alpha(\bar{\rho}\frac{1}{2} + (1 - \bar{\rho})((1 - \beta)\frac{1}{2})) \\ & + (1 - \alpha)((1 - \beta)\frac{1}{4}) \\ & = \frac{1}{4}(\alpha(1 + (2\bar{\rho} - 1)\beta) - (\beta - 1)). \end{aligned} \quad (2.11)$$

The first line above corresponds to the informed trader's order being accepted with probability α . The coefficient of α is broken down into the informed trader submitting a maker order or limit order with probabilities $\bar{\rho}$ and $1 - \bar{\rho}$. Each of these multiplies the probabilities of the noise traders taking the action that corresponds to *LBMB*: $\frac{1}{2}$ for a passive noise trader to submit a buy limit order, and $\frac{1-\beta}{2}$ for a nonretail active noise trader to submit a buy market order. The reasoning behind the second summand is similar. The last line above is equal the numerator of ξ_i when $A_i = \text{LBMB}$ in Equation 2.10 up to the factor $\frac{1}{4}$ which cancels with the denominator of ξ_i when computed the same way.

Figure 2.8 shows the results of a simulation of one possible path of the game. The bid and ask prices posted by the market maker remain constant until the market maker learns sufficiently about the true value of the asset and revises his quotes toward the true value of the asset. By trading strategically with market and limit orders, the informed trader would like to maximize profits by preventing the market maker from learning and updating his quotes.

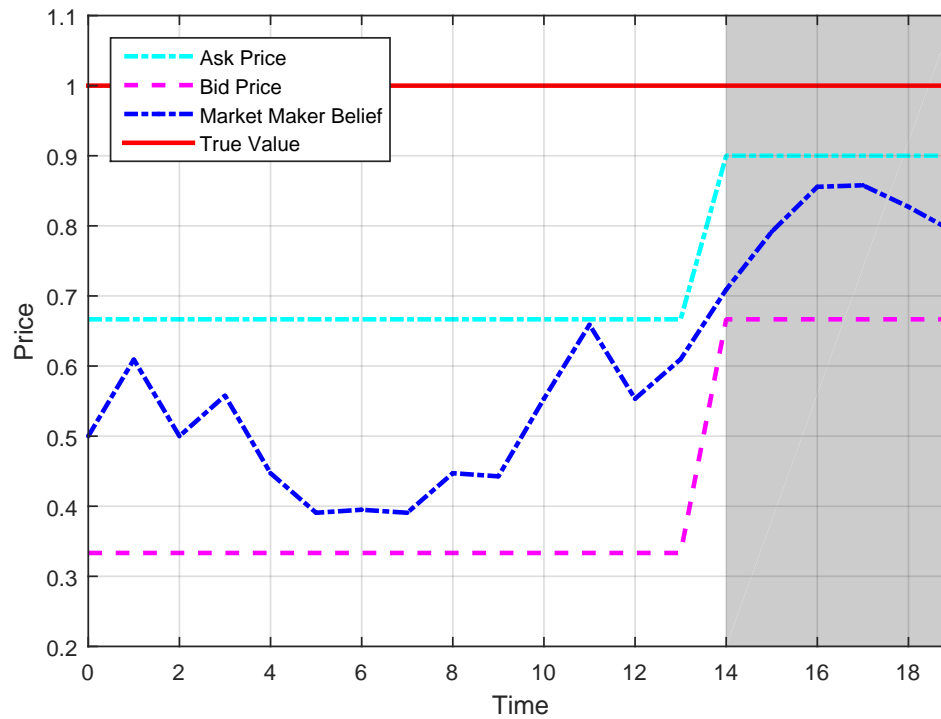
2.4.8 Perfect Bayesian Nash Equilibrium

We look for a Perfect Bayesian Nash Equilibrium (Fudenberg & Tirole, 1991) that corresponds to a strategy used by the informed trader which is restricted to mixing market and limit orders that maximizes expected profits, and a strategy used by the market maker that is restricted to a stopping time when he revises his quotes from the inner price pair $\{x_l, x_h\}$. Let $f(\rho_i)$ be the expected profit for the informed trader in period i as a function of the strategy ρ_i . Formally, an equilibrium is a set of strategies $\bar{\rho}(p_i)$ and $\underline{\rho}(p_i)$ used by the informed trader corresponding to the asset's value being high or low respectively, that satisfy:

$$\begin{aligned} \bar{\rho}_i &= \arg \max_{\{0 \leq \rho_i \leq 1\}} \mathbb{E} \left[\sum_{j=i}^{N(\underline{\theta}, \bar{\theta})} f(\rho_j) \mid v = 1, p_i \right] \\ \underline{\rho}_i &= \arg \max_{\{0 \leq \rho_i \leq 1\}} \mathbb{E} \left[\sum_{j=i}^{N(\underline{\theta}, \bar{\theta})} f(\rho_j) \mid v = 0, p_i \right] \end{aligned} \quad (2.12)$$

for values of the market maker's prior $p_i \in (\underline{\theta}, \bar{\theta})$, for all trading times i . The market maker defines a stopping time $N(\underline{\theta}, \bar{\theta})$ corresponding to p_i exiting the region $(\underline{\theta}, \bar{\theta})$ after updating his

Figure 2.8: **Numerical example of game dynamics.** The figure shows the results of a simulation of one possible path of the game. The blue dashed line represents the market maker's belief about the asset's value which moves randomly according to the pattern of trades and quotes. The purple and cyan dashed lines represent the bid and ask prices quoted by the market maker. The solid red line represents the asset's true value. The grey area marks the end of the game when the market maker revises his quotes.



prior using the formulas in Equation 2.10 defined in Lemma 2.4.1. Details on $\bar{\theta}$ and $\underline{\theta}$ are given in Appendix A.6.

2.4.9 Informed Trader's Problem

Without loss of generality, we will assume the asset's true value is high, $v = 1$, from here forward. The informed trader maximizes profit over all time by choosing an order type each period. He uses a mixed strategy where $\bar{\rho}_i$ is the probability of submitting a market order and $1 - \bar{\rho}_i$ is the probability of submitting a limit order. In general, $\bar{\rho}_i$ will depend on all model parameters, α , β , x_h , x_l , $\bar{\theta}$, $\underline{\theta}$, and the state variable p_i . These parameters are known to the informed trader, as is the state variable since he observes the same information as the market maker and can use it to infer the market maker's prior. In Appendix A.6, we will show that the thresholds $\bar{\theta}$ and $\underline{\theta}$ are independent of $\bar{\rho}_i$ and $\underline{\rho}_i$. Let the informed trader's value function be $V(p_i)$ with the market maker's prior p_i as the state variable. The Bellman equation at time i is given by:

$$V(p_i) = \max_{0 \leq \bar{\rho}_i \leq 1} \mathbb{E}[f(\bar{\rho}_i) + V(p_{i+1})|\bar{\rho}_i] \quad (2.13)$$

$$V(\bar{\theta}) = \bar{V} \quad (2.14)$$

$$V(\underline{\theta}) = \underline{V} \quad (2.15)$$

where $f(\bar{\rho}_i)$ is the per-period profit from trading; $\bar{\theta}$ and $\underline{\theta}$ are the states at which the market maker switches prices; and \bar{V} and \underline{V} are the boundary values corresponding to these states. The "continuation values" \bar{V} and \underline{V} correspond to when prices are revised either upward or downward. These values are arbitrary since we do not model trading after prices are revised. We will assume that $\bar{V} < \underline{V}$. The interpretation of the inequality is that when the prices change, the informed trader is able to make fewer trading profits when prices are revised upward than when they are revised downward.⁸

Explicitly, the expected per-period profits when the asset's true value is high as a function of the strategy $\bar{\rho}_i$ are

$$f(\bar{\rho}_i) = \alpha(\bar{\rho}_i(1 - x_h) + (1 - \bar{\rho}_i)\frac{(1 - x_l)}{2}). \quad (2.16)$$

The Bellman equation describes how the informed trader balances current trading profits while managing the speed at which the market maker is able to learn about the asset's true value. Indeed, a pure strategy may be more informative than a mixed strategy. For example, if the informed trader submits only market orders, the market maker will only update his belief based on observed market orders and will dismiss any imbalances in the order book as coming from passive noise traders. Information will be concentrated in the set of executed market orders and the market maker will learn more efficiently than if information were distributed across the set of limit and market orders. Similar reasoning would apply to the case of a strategy of pure limit orders.

⁸For details on how \bar{V} and \underline{V} affect the informed trader's strategy, see Appendix A.8, Equation A.33.

To demonstrate the basic intuition behind the effect of segmentation on the informed trader's strategy, consider the extreme case when all uninformed market orders are retail. For example, if the market maker observes a nonretail market order to buy and a limit order to buy, he will know with certainty the informed trader is buying. In this case, the action of the informed trader is perfectly revealing of the asset's true value: the factor ξ_i when $A_i = LBMB$ is infinite when $\beta = 1$. So if the informed trader submits a market order, there is a chance he will forgo all future profits at current prices. If potential future profits in all states other than the perfectly revealing case are sufficiently high, the informed trader will refrain from submitting market orders altogether. We will show that this intuition holds in more general cases.

In general, it is not clear how to solve Equation 2.13, or that it has a tractable solution. We pursue simplifications that allow us to hypothesize that the optimal strategy is a constant, $\rho_i = \bar{\rho}_i = \rho^*$ and independent of the state p_i . We draw on the technique used to prove Proposition 2.3.2 using Wald's lemma to show that the constant solution ρ^* is optimal.

2.4.10 Approximation: Large Switching Threshold

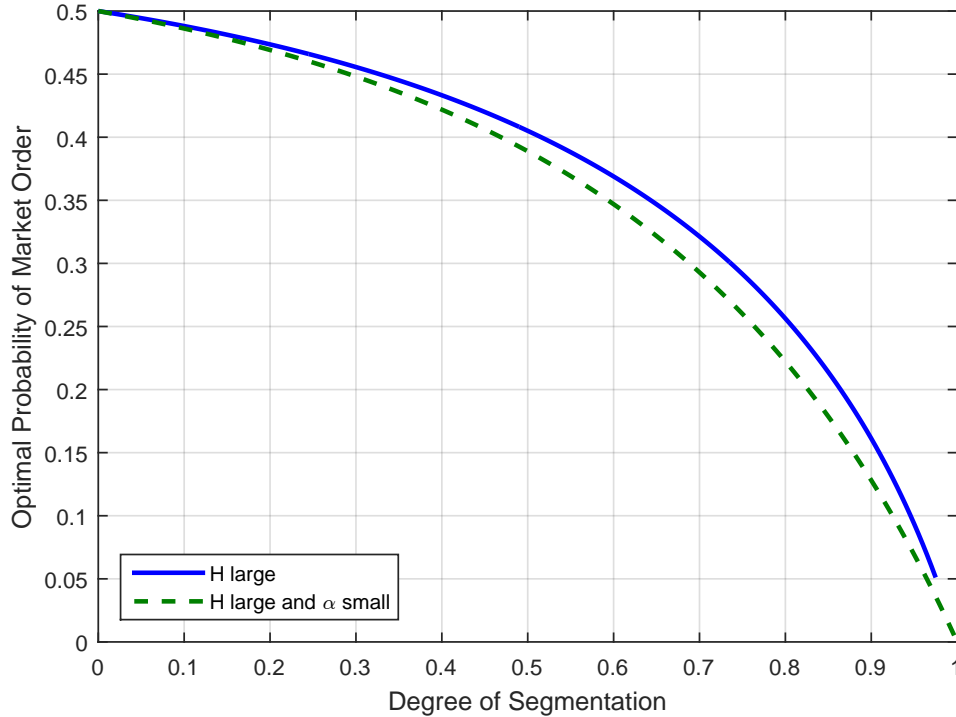
A special case that leads to a useful approximation of the model is when the market maker's thresholds for which he revises his quotes become large in magnitude. We defined the upper and lower thresholds for the market maker's prior as $\bar{\theta}$ and $\underline{\theta}$. Recall that in Section 2.3 the boundaries around the market maker's likelihood ratio were H and $\frac{1}{H}$. Let H and $\frac{1}{H}$ correspond to the values of the likelihood ratio $\frac{p_i}{1-p_i}$ when $p_i = \bar{\theta}$ and $p_i = \underline{\theta}$ respectively. Note that since $\bar{\theta}$ and $\underline{\theta}$ take values in $(0, 1)$, H and $\frac{1}{H}$ take values in $(0, \infty)$. In the limit as H goes to infinity, the informed trader makes infinite profits because there are infinitely many trades before the market maker prefers to revise his quotes. However, it is interesting to derive the informed trader's optimal strategy in the limit as H goes to infinity as an approximation to the case of finite H . The advantage of considering H large is that we are able to show the optimal strategy is a constant ρ^* which is independent of the state variable, the market maker's prior p_i . We will show ρ^* is given by the solution to $D_\rho \left[\frac{\mathbb{E}[f(\rho)]}{\mathbb{E}[\log(\xi_1)|\rho]} \right] = 0$ where $\mathbb{E}[f(\rho)]$ represents expected per-period profits, $\mathbb{E}[\log(\xi_1)|\rho]$ is the expected "first step" taken by the market maker's log-likelihood ratio and D_ρ is the differential operator with respect to ρ .

It is intuitive that the optimal strategy is independent of p_i when H is large. In general, the informed trader balances immediate trading profits with price impact in order to preserve the potential for future profits. When the log-likelihood ratio is close to the market maker's threshold and the potential for future profits is not high relative to immediate profits, the informed trader puts relatively less weight on minimizing price impact. Because the relative contributions of immediate and future profits change with the distance to the boundary, the informed trader will adjust his strategy accordingly. However, when H is large, the log-likelihood ratio is always "far" from the boundary. The balance between immediate and future profits is stable; each trade results in a move toward a boundary that is negligible and therefore no adjustment to the strategy is needed after a trade. We state the result in the following Proposition.

Proposition 2.4.2 *The informed trader's optimal trading strategy, in the sense that it satisfies the first-order condition of the Bellman equation, is given by the solution to $D_\rho \left[\frac{\mathbb{E}[f(\rho)]}{\mathbb{E}[\log(\xi_1)|\rho]} \right] = 0$ in the limit as H goes to infinity. A proof is found in Appendix A.7.*

The denominator of the formula in Proposition 2.4.2 will be nonzero, intuitively, because every trading round will be informative. The informed trader will always submit either a fast limit or market order with positive probability α so the market maker's Bayesian update will be nontrivial.

Figure 2.9: **Optimal strategy.** The figure shows the optimal probability of the informed trader submitting a market on the y-axis and the degree of segmentation on the x-axis. The prices used to generate the figure were $x_l = \frac{1}{3}$ and $x_h = \frac{2}{3}$. The blue line was created using a value of $\frac{1}{10}$ for α .



We can use software to differentiate symbolically and then numerically find the root of the condition in Proposition 2.4.2 that the optimal strategy satisfies. Results for certain parameter values are shown in Figure 2.9. The figure shows that the informed trader submits a higher proportion of limit orders than market orders as the degree of retail segmentation increases. When certain market orders are identified as retail, the relative informativeness of a non-retail market order increases, all else equal. The informed trader mitigates this effect by decreasing the probability he submits a market order as the proportion of retail traders β increases.

Observation 2.4.3 *The informed trader submits a greater proportion limit orders as the measure of retail traders increases; ρ^* is decreasing in β in the limit as H goes to infinity.*

We will prove the result stated in Observation 2.4.3 for a limiting case in next subsection with Proposition 2.4.5.

We are also interested in the effect of segmentation on price efficiency. Recall that in Section 2.3 we defined price efficiency using the expected number of trades before the market maker's prior hit arbitrary upper or lower boundaries. The resulting formula in Proposition

2.3.2 had the interpretation of expected number of trades (time) being equal to expected total change in the market maker's prior before it hits a boundary (distance) over expected "first step" (speed). We will need to change our notion of price efficiency here since the expected total change in the market maker's prior goes to infinity as H goes to infinity. We will focus on the "speed" at which the market maker's prior moves, that is, the expected size of the revision to the market maker's prior after a trade, $\mathbb{E}[\log(\xi_1)|\rho]$. In Section 2.4.11, we will show that this definition of price efficiency is a good proxy the expected number of trades before the maker's prior hits a boundary.

Figure 2.10 shows how price inefficiency, defined as the reciprocal of the expected change in the market maker's prior evolves following a trade, changes with the degree of segmentation. The graph shows that the importance of strategic order submission is low for low degrees of segmentation and high for high degrees of segmentation.

Observation 2.4.4 *Price efficiency, as measured by the rate at which the market maker's prior approaches the asset's true value, 1 or 0, increases with the measure of retail traders β when H is large.*

Observation 2.4.4 shows that despite the ability of the informed trader to mitigate the effect of identifying retail traders there is still an improvement in price efficiency in equilibrium.

The equation solved by the optimal strategy, $D_\rho[\frac{\mathbb{E}[f(\rho)]}{\mathbb{E}[\log(\xi_1)|\rho]}] = 0$, has a natural interpretation. Total profits earned by the informed trader can be loosely described as the expected number of trades he executes n , times the expected profit earned per trade $\mathbb{E}[f(\rho)]$. Wald's lemma tells us that under certain conditions, the expected number of trades before a hitting time is the expected distance travelled by the process divided by the expected move at each step. The expected move at each step is $\mathbb{E}[\log(\xi_1)|\rho]$. Let the expected distance travelled by the market maker's prior be a constant c . The derivative of total profits with respect to the probability of a market order are:

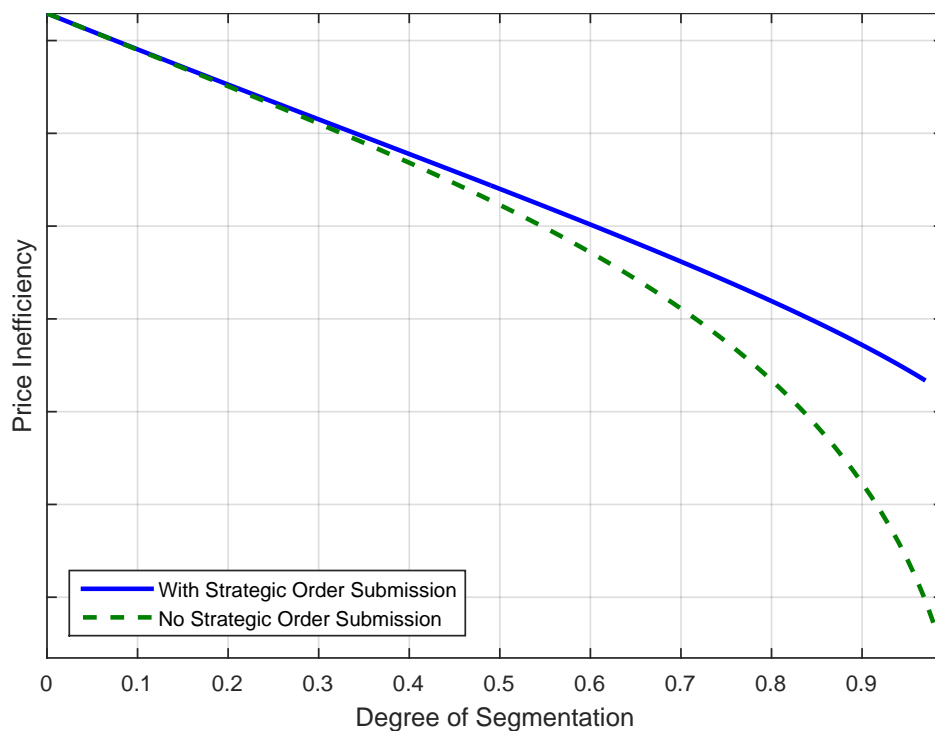
$$\begin{aligned} & D_\rho[n\mathbb{E}[f(\rho)]] \\ &= D_\rho\left[\frac{c\mathbb{E}[f(\rho)]}{\mathbb{E}[\log(\xi_1)|\rho]}\right] \\ &= cD_\rho\left[\frac{\mathbb{E}[f(\rho)]}{\mathbb{E}[\log(\xi_1)|\rho]}\right] \\ &\implies D_\rho\left[\frac{\mathbb{E}[f(\rho)]}{\mathbb{E}[\log(\xi_1)|\rho]}\right] = 0, \end{aligned}$$

when c is finite. The condition $D_\rho[\frac{\mathbb{E}[f(\rho)]}{\mathbb{E}[\log(\xi_1)|\rho]}] = 0$ can be seen as a optimizing a factor that is proportional to total profits. The proof of Proposition 2.4.2 is mostly devoted to showing that c is indeed independent of ρ for large H . Although c will be infinite when H goes to infinity, we will show the first order condition is still meaningful.

Approximation: Large Switching Threshold and Small Probability of Informed Trading

Since the condition in Proposition 2.4.2 is not directly solvable, we pursue a further simplification of the formula; the limit as α goes to zero. When α goes to zero, the result is an analytic

Figure 2.10: **Price efficiency.** The y-axis is price inefficiency, defined as the reciprocal of the expected change in the market maker's prior evolves following a trade. On the x-axis is the degree of segmentation. The solid blue line represents price inefficiency when the informed trader acts strategically, while the dashed green line serves as a benchmark and represents the case when the informed trader submits limit orders and market orders with probabilities $\frac{1}{2}$. The figure shows that despite the informed trader's strategic order choice, price inefficiency decreases increases with the degree of segmentation. The figure was created using values of $\alpha = \frac{1}{10}$, $x_l = \frac{1}{3}$, and $x_h = \frac{2}{3}$



formula for a constant optimal strategy with the same qualitative properties pointed out in Observation 2.4.3. Taking the limit of $D_\rho[\frac{\mathbb{E}[f(\rho)]}{\mathbb{E}[\log(\xi_1)^\rho]}$], the first order condition when H is large, as α goes to zero and solving for ρ^* gives:

$$\rho^* = \frac{(x_l - 1)\beta - 2x_l^2 + 2 - (2(x_l - x_l^2)\beta^2 - (7x_l^2 + 1)\beta + 10x_l^2 - 4x_l + 2)^{\frac{1}{2}}}{(3x_l - 1)(\beta - 2)} \quad (2.17)$$

We use symbolic computing software (Matlab) to take the limit since the condition in Proposition 2.4.2 is difficult to work with by hand when expanded. The resulting strategy in Equation 2.17 is shown in Figure 2.9. It is qualitatively similar to the plot of the more general strategy. We can now state the analogue of Observation 2.4.3 as a provable proposition.

Proposition 2.4.5 *In the limit as both H goes to infinity and α goes to 0, the probability the informed trader submits a limit order strictly decreases as the degree of retail segmentation β increases. A proof is found in Appendix A.8.*

In the next section we will use numerical techniques to show that the limiting results derived above are qualitatively similar to a more general set of model parameters.

2.4.11 Numerical Evidence

We compare the approximations of Subsection 2.4.10 to numerically generated results to show that the results hold more generally; price efficiency increases in the degree of retail segmentation, and informed traders use relatively more limit orders as segmentation increases. To compute numerical results, we use policy function iteration, also called Howard's Improvement Algorithm, combined with Monte Carlo simulation. The algorithm iterates from an initial guess of the informed trader's strategy by simulating the future profits, given the strategy. At each step, a Monte Carlo method is used to estimate future profits, then the strategy is re-optimized given the simulated future profits.

The standard version Howard's Improvement Algorithm (Judd, 1998) is given by the pseudocode in Algorithm 1.

Algorithm 1 Standard Howard's Improvement Algorithm

- 1: Guess a *strategy function*
 - 2: **While the *strategy function* has not converged:**
 - 3: Compute *future profits* given the *strategy function*
 - 4: Replace the *strategy function* with the optimal strategy given *future profits*
-

There are two features of our model that prevent us from using the standard Howard Improvement Algorithm. First, computing the future profits for a given strategy function is intractable because following a trade, the market maker's prior may overshoot the boundaries leading to discontinuities in the optimal strategy. Second, the expected impact of a trade on the market maker's prior involves a Bayesian update (Equation 2.10) which is a function of both the optimal strategy given that the asset's value is high, and the optimal strategy given that the asset's value is low. We will call these the *high strategy* and the *low strategy*. Since we cannot

optimize these strategies simultaneously, we use a separate iterative procedure to perform this optimization.

Altogether, our numerical procedure consists of an outer loop over a Monte Carlo simulation and an inner loop over the high strategy and the low strategy. In the outer loop, the informed trader's strategy, represented by a vector over a discretized state space, is taken as given. Monte Carlo simulation is used to estimate expected future profits that accrue to the informed trader for a given strategy. The Monte Carlo simulation is done by simulating random sequential trading outcomes from the model parameters until the market maker's prior hits one of the boundaries. We note that the simulations are path-dependant because the probability of a market order varies with the market maker's prior.

In the inner loop, we must find a revised strategy given the expected future profits computed in the outer loop. This requires an iterative procedure because the market maker's Bayesian update involves the ratio of the probabilities of market orders given the true value of the asset is high, and given the true value of the asset is low. To know these two probabilities for any given value of the market maker's prior, one must know the informed trader's optimal strategy when given the true value being high, and given the true value being low. As mentioned above, we call these the *high strategy* and the *low strategy*. Since each of the two strategies depend on the market maker's Bayesian update, they depend on each other. We cannot optimize them simultaneously so we iterate by optimizing the *high strategy* for each value of the market maker's prior, given the *low strategy*. We then update *low strategy* exploiting the model's symmetry around the market maker's prior equal to $\frac{1}{2}$. We iterate on optimizing *high strategy* given the previous *low strategy* then updating *low strategy* using symmetry until the strategies converge.

Pseudocode of the above is given in Algorithm 2. As in Section 2.3, without loss of generality, we assume that the true value of the asset is high and focus on the *high strategy* function. For Step 1, we choose initial values of the high and low strategy functions, $\bar{\rho}_0(p_k)$ and $\underline{\rho}_0(p_k)$ respectively, on a grid p_k that take the form of a vector of probabilities of submitting a market order corresponding to grid points over the state space. For p_k , we choose a grid of evenly distributed points between the market maker's switching thresholds 0.1 and 0.9 with length 0.01 between each grid point.

For Step 2, we choose a convergence criteria for our *strategy function* to test whether two consecutive iterations of the outer loop result in sufficiently close strategies. We use the Euclidean norm on the previous two iterations of the strategy function and terminate iteration when $|\bar{\rho}_j(p_k) - \bar{\rho}_{j-1}(p_k)|$ is less than a constant at iteration j .

For Step 3, we must estimate how the candidate *strategy function* performs. To do this, we chose a grid on the state space q_l , possibly different from p_k , and for each grid point, perform Monte Carlo simulation to estimate future profits when the state begins at the grid point. For q_l , we choose the grid [0.1, 0.4, 0.9] because Monte Carlo simulation is computationally expensive. Each Monte Carlo simulation involves generating random trades given the probabilities specified by the model which are partly determined by the iteration of the strategy function used. It may be that when Bayesian updates are applied, the prior p moves off the chosen grid q_l . When this occurs, we interpolate values of the strategy function using piecewise cubic Hermite interpolating polynomials (Matlab's *pchip* function). A constant number of trials are performed for each Monte Carlo simulation; we use 1000 for the figures below. For each grid

point q_l , profits for the informed trader are recorded in a vector of future profits $V_j(q_l)$.

For Steps 4, 5, and 6, we perform an iterative procedure that allows us to find the optimal strategy taking the future profits computed in Step 3 as given. We pursue an equilibrium when the strategy is symmetric around $p = \frac{1}{2}$; the optimal strategy when the asset's value is high as a function of the prior is equal to the optimal strategy when the asset's value is low, mirrored. Since we cannot optimize both the *high strategy* and the *low strategy* simultaneously, we optimize the *high strategy*, taking the *low strategy* as given, then replace the *low strategy* with the mirrored *high strategy*, and repeat until a convergence criteria is met. Again, we use the Euclidean norm and terminate iteration (Steps 5 and 6) when $|\bar{\rho}_j(p_k) - \underline{\rho}_j(p_k)|$ is sufficiently small.

The numerical optimization performed in Step 5 is analogous to the Bellman equation; the optimal strategy balances immediate profits $f(\bar{\rho})$ and future profits $V_j(p_k, \bar{\rho}, \underline{\rho}^{(-)})$, where we use $\underline{\rho}^{(-)}$ to denote the previous iteration of the low strategy. For each point on the grid p_k we numerically solve:

$$\begin{aligned} \max_{\bar{\rho}} \mathbb{E}[f(\bar{\rho}) + V_j(p_k, \bar{\rho}, \underline{\rho}^{(-)}) | \bar{\rho}] & \quad (2.18) \\ V(\bar{\theta}) &= \bar{V} \\ V(\underline{\theta}) &= \underline{V} \end{aligned}$$

Numerical optimization is done with software (Matlab) that uses golden section search and parabolic interpolation. In order to compute the expected future profits from the vector $V_j(p_k)$, the informed trader must consider how the Bayesian updates move the prior from a particular state p_q . If the expectation of the prior is off the grid p_k , we again use piecewise cubic Hermite interpolating polynomials to interpolate.

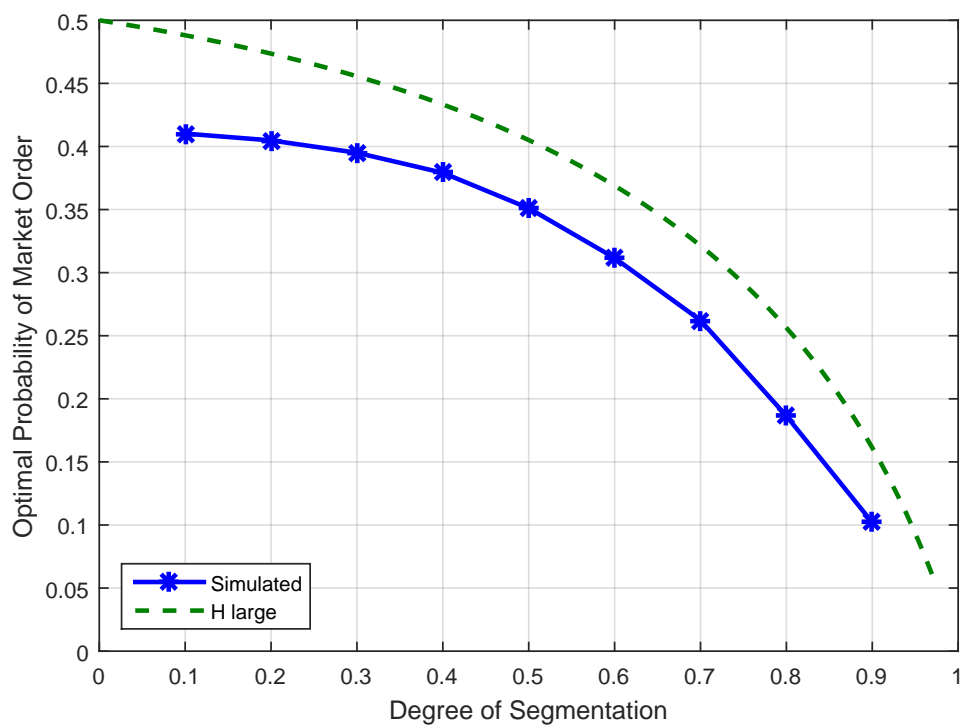
Algorithm 2 Modified Howard's Improvement Algorithm with iteration over high and low strategies and Monte Carlo simulation

- 1: Guess a *high strategy function*
 - 2: **While the *high strategy function* has not converged:**
 - 3: Use Monte Carlo Simulation to estimate *future profits* given the *high strategy function*
 - 4: **While the *high and low strategy functions* have not converged:**
 - 5: Optimize *high strategy function* given *future profits* and *low strategy function*
 - 6: Replace the *low strategy function* with the mirrored *high strategy function*
-

Numerical results are shown in Figures 2.11, 2.12 and 2.13. Figure 2.11 shows that the informed trader's strategy behaves similarly to the approximations derived in subsection 2.4.10.

Figure 2.12 shows the informed trader's strategy as a function of the market maker's prior for various values of β . We have constrained the x -axis to the arbitrary interval $[0.4, 0, 6]$ because outside the interval, the prior is "too close" to the boundaries. When the prior may reach one of the boundaries in only one trade, strategies may be discontinuous and often do not converge in our computations. However, in the interval $[0.4, 0, 6]$, the prior is sufficiently far from the boundaries and the strategies are well-behaved and converge numerically. The probability of a market order is strictly decreasing with β for all values of the market maker's prior, as expected. Strategies are increasing in the market maker's prior for two reasons. First,

Figure 2.11: **Simulated optimal strategy.** The figure shows the numerical optimal strategy (solid blue) and the theoretical optimal strategy when H is large (dashed green) as a function of the degree of retail segmentation. Simulation parameters were $\alpha = \frac{1}{10}$, $x_l = \frac{1}{3}$, $\bar{\theta} = \frac{9}{10}$ and $\beta = \frac{1}{2}$. The probability of a market order is strictly decreasing and convex in the degree of segmentation for both the numerical and theoretical strategies.



price impact is lower in certain cases when the *high strategy* and *low strategy* are opposite. For example, in Equation 2.10, $\xi_i = \frac{1-\alpha(2\bar{\rho}-1)}{1+\alpha(2\rho-1)}$ is equal to 1 when $\bar{\rho} = 1 - \underline{\rho}$ which means the market maker’s prior is unchanged when the outcome *LBMS* is realized. Second, the *high strategy* is more likely to result in the prior moving up, meaning the informed trader will be most concerned with mitigating price impact (submitting more limit orders) for values of p that are less than $\frac{1}{2}$. For larger values of p , the informed trader has fewer trades remaining and mitigating price impact has less weight in his decision. The interaction of these two features of the informed trader’s problem give the high strategy its upward slope.

Figure 2.12: **Simulated optimal strategy.** The figure shows the numerical optimal strategy as a function of the market maker’s prior, for various degrees of retail segmentation. Simulation parameters were $\alpha = \frac{1}{10}$, $x_l = \frac{1}{3}$ and $\bar{\theta} = \frac{9}{10}$. The probability of a market order is strictly decreasing in each of the values for segmentation.

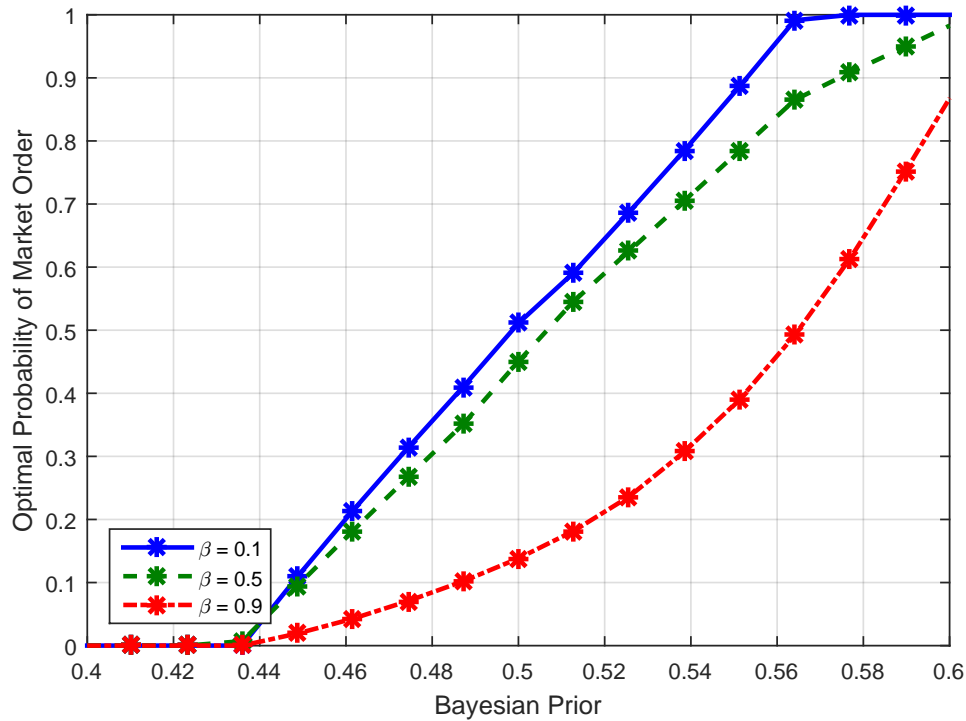
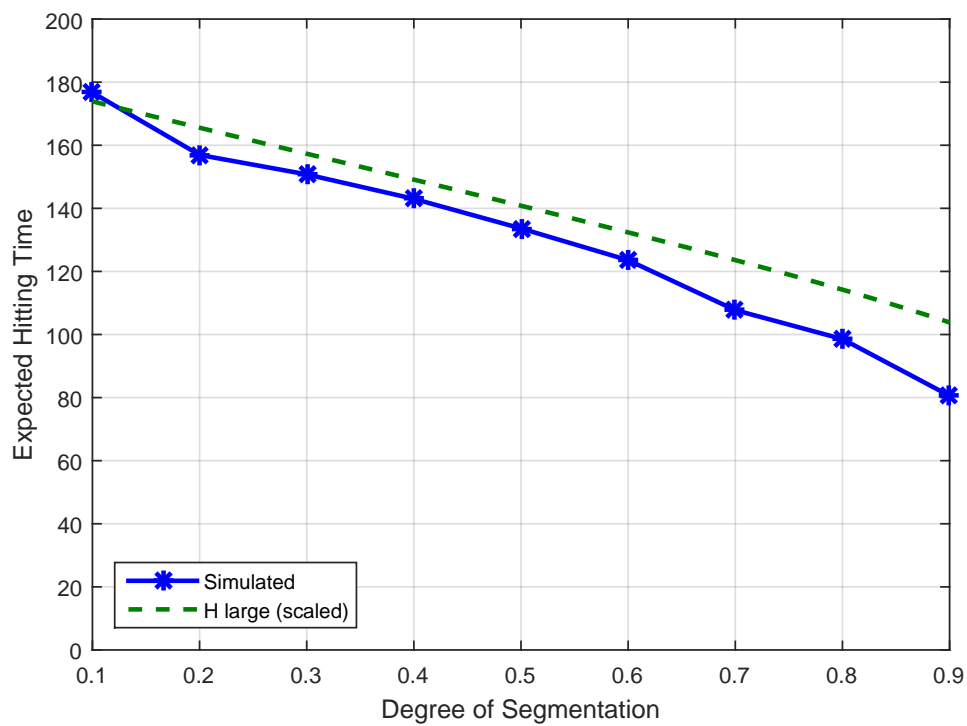


Figure 2.13 shows that price efficiency, defined using the expected number of trades until a boundary is hit, increases with segmentation despite the informed trader’s strategic play. Also shown in the figure is the same line shown in Figure 2.10, scaled for comparison. The figure shows that the expected hitting time decreases roughly at the same rate as the inverse of the size of the market maker’s update, derived for large H .

2.5 Conclusions

We model an exchange-based mechanism of retail order flow segmentation in a model with asymmetric information about the true value of a traded asset and derive results on its potential effects on liquidity, price efficiency, and on trading behaviour. Our work is motivated by the

Figure 2.13: **Simulated hitting time.** The figure compares simulated expected hitting time (solid blue) to the theoretical approximation (dashed green) as the degree of retail segmentation increases. Simulation parameters were $\alpha = \frac{1}{10}$, $x_l = \frac{1}{3}$ and $\bar{\theta} = \frac{9}{10}$. The theoretical line was generated by plotting the reciprocal of the expected update given the parameters. In our approximation when H is large, hitting time is inversely proportional to the magnitude of the expected update. The dashed green line was scaled by a constant to be visually comparable with the solid blue line. Both decrease with the degree of retail segmentation at a comparable rate.



prevalence of facilities on large stock exchanges that allow participants to make offers to trade at prices inside the market wide best bid and offer that are only executable by retail market orders. In Section 2.3, we show that all else equal, when retail traders are given access to better prices, the bid-ask spread for non-retail traders increases. This is because market makers compensate for revenue lost from providing better prices to retail traders by widening the bid-ask spread for non-retail traders. We also study the dynamic effect of segmentation and show that market makers are able to learn more efficiently about the true value of the traded asset when retail traders are segmented. Segmentation allows market makers to distinguish between retail trades, which convey no information about the true value of the asset, and non-retail trades which are informative about the true value of the asset. Segmentation strengthens the signal value of the order flow for market makers who learn about fundamentals from the history of trade. This result is shown empirically in Chapter 1.

In Section 2.4, we study a model with constraints on quotable prices, but where traders can choose to submit either limit orders or market orders. The model environment resembles a situation that is common for highly liquid stocks. For these stocks, bid-ask spreads can be competed down to one cent, which is minimal due to regulatory tick constraints. We study the situation where a market maker is willing to quote a particular, minimal bid-ask spread until sufficiently many trades take place for the market maker to prefer to quote different prices. As in Section 2.3, segmentation allows the market maker to learn more efficiently about the true value of the asset. The informed trader partially mitigates this effect by strategically submitting fewer market orders. Retail market orders, as opposed to retail limit orders, are explicitly segmented, so informed traders avoid market orders and rely more on limit orders to minimize the informational impact of their trades.

The two models presented in Sections 2.3 and 2.4 make empirical predictions that could be analyzed with a sufficiently detailed dataset. All else being equal, a higher degree of retail segmentation in a should result in greater price efficiency and greater use of limit orders by sophisticated traders. However, it could be the case in practice that informed traders scale back their trading activity or their acquisition of new information when segmentation is present since they earn fewer profits from trading. We leave these effects unmodeled but empirical tests could shed light on whether or not our model assumptions are correct. Our model's predictions about liquidity are ambiguous. Bid-ask spreads may widen in a single-period version of our model, but narrow over time and improve on average. The prevailing effect likely depends on the balance of trading activity and the rate of nonpublic changes to fundamentals. In Chapter 1, we find empirical evidence that retail segmentation can cause mild improvements in liquidity.

Chapter 3

The Effect of Hedging on Liquidity

3.1 Introduction

We study the effect of hedging a derivative on liquidity in its underlying security. Existing work on how derivatives affect underlying assets offers different conclusions that depend on model assumptions, market structure, and on the asset class studied. We contribute to this discussion by developing and testing hypotheses about how liquidity is affected by derivatives hedging under certain assumptions that may be informative to market participants and securities regulators. In a market where a component of trading volume is generated by hedging from a derivatives market, a market maker must learn about both the inventories of hedgers and about asset fundamentals through the pattern of trade. We extend the classic model of sequential trade in Kyle (1985) to include a hedger whose trades are determined by a random inventory of derivatives. Assuming the hedger does not have private information about the security, our model predicts that trading volume from hedging results in improved liquidity.

The central tension in the model is between the benefits to a market maker from trading with hedgers, and the need to learn about two quantities simultaneously by executing trades in the underlying security: the true value of the underlying security, and the hedger's inventory. Learning about hedging activity confounds the market maker's ability to learn about the true value of the underlying asset from the history of trades. Despite the additional source of uncertainty, the market maker is willing to provide greater liquidity since a greater proportion of trade is uninformed, providing the market maker with profits that offset losses incurred from trading with informed traders.

The situation we model applies to a market where liquidity provision is highly competitive, when there is information asymmetry between traders about the underlying asset, and when derivatives hedging is largely of an uninformed nature. Equity markets fit this description: there are many highly liquid stocks with multiple liquidity providers who compete on centralized exchanges; information asymmetry has been demonstrated empirically to be a determinant of liquidity (Glosten, 1987); and previous studies have shown that equity options markets tend to lag, rather than lead, equity markets informationally (Muravyev, Pearson, & Broussard, 2013).

We demonstrate our model's predictions empirically. Using a sample of highly liquid US stocks during the 2014 calendar year, we are able to detect large increases in trading volume in the underlying stocks near the monthly expiry of their options. We show that the component of

volume associated with option expiry causes an improvement in standard liquidity measures. The result is consistent with empirical papers which come to similar conclusions about liquidity near option expiries.

Our model includes a market maker, informed traders who have private knowledge about the underlying asset, and a noise trader as in Kyle (1985). We add a fourth trader who acts as a dealer in a derivatives market and hedges by trading in the underlying asset. The dealer's trades can be seen as motivated by a few similar scenarios: option hedging, as in Black and Scholes (1973), exchange-traded fund (ETF) rebalancing trades as in Tuzun (2014), or portfolio insurers as in Gennotte and Leland (1990). The common idea is that each of these strategies may result in large, somewhat predictable trading volumes that contain little or no information about asset fundamentals.

We model sequential trading where traders have discrete opportunities to trade. In each round, traders submit market orders to a market maker who prices and executes the aggregated market orders (which we will call the "order flow" in this chapter). The market maker chooses a price that balances profits made from trading with uninformed traders against losses from trading with informed traders. The price is the market maker's belief about the true value of the asset, given the unexpected volume of the order received. The price depends on volume since it contains information about the value of the asset. It is partly determined the price-impact coefficient (called "Kyle's lambda") which is a proxy for liquidity.¹ The market maker can partially predict hedging trades; the market maker knows the strategy that the hedger uses, but must learn about the size of the hedger's inventory. It is impossible for the market maker to fully distinguish between order flow driven by changes in the underlying asset and changes in the hedger's inventory since the only source of information is the order flow. However, despite uncertainty about the true value of the asset and the hedger's inventory, price impact is less than it would be if the hedging trader was not present.

Our model cannot describe reality over a long time horizon since the variance of beliefs about the asset's true value and the hedging inventories increase with every trade. However, it may apply to periods when hedging or program trading is an important source of trading volume in the underlying asset such as trading days near options or futures expiry, or ETF rebalancing.

The model predicts that price impact, a dimension of liquidity, improves when hedging trades contribute significantly to trading volume in the underlying asset. Our empirical study uses an instrumental variables regression to demonstrate that liquidity improves for a sample of US stocks around option expiry dates. Time-to-expiry is exogenous to liquidity in the underlying asset so we use it as an instrument in a two-stage least squares (2SLS) methodology. In the first stage, we fit a reduced form regression model that relates trading volume to time-to-expiry. In the second stage, we use the component of volume associated with time-to-expiry from the first stage to show that this volume causes an improvement in liquidity, as measured by two price impact measures, an intraday volatility measure, and a measure of the bid-ask spread. We include the bid-ask spread measure and volatility measure for robustness.

The instrumental variables regression methodology is necessary because liquidity, volume in the underlying security and the options market may be coterminous. An improvement in

¹More precisely, the liquidity parameter which is endogenous in the model is the price change per share transacted with the market maker.

liquidity in the underlying asset may result in an improvement in liquidity in the options market, driving up open interest or option trading volumes and increasing volume in the underlying, further improving market quality. The two markets may also be subject to common, unobserved liquidity factors.

3.1.1 Hedging and Program Trading

There is a rich literature on the empirical effects of hedging and program trading on the underlying asset. Studies have found mixed results across markets, and across time periods. Some of the variation in results may come from natural evolution of markets and increasing sophistication of trading venues and participants. Overall, our results are consistent with papers that focus on measures of liquidity such as the bid-ask spread; liquidity is not harmed and may improve when option markets are present.

Stoll and Whaley (1987) and (1991) study the triple witching hour when single stock options, stock index options, and stock index futures simultaneously expire. They find a large increase in volume and that volatility is significantly higher on these days. Hancock (1993) also empirically studies the triple witching hour and finds no evidence of price distortion. The study does find evidence of increased volatility on these days.

Wang (2012) shows empirical evidence that option hedging trades can create price pressure and return reversals. Kumar, Sarin, and Shastri (1998) empirically study the effect of option listing on market quality. They find evidence that option listing improve market quality. Chow, Yung, and Zhang (2003) study the effect of options written on the Hang Seng Index. They find some evidence of negative price pressure and excess return volatility in the underlying stock market. Chamberlain, Cheung, and Kwan (1989) study expiration day effects on the Toronto Stock Exchange. They find evidence of abnormal price behaviour, namely that returns and volatility tend to be higher in the last half-hour of trading on expiry Fridays. Abnormal returns usually reversed the following Mondays.

Chiang (2014) finds evidence that stocks with a large number of deep in-the-money put options may experience price drops on expiration dates. Selling pressure is attributed to option buyers who exercise and immediately sell their shares in the underlying market. Stucki and Wasserfallen (1994) document the effects of the introduction of options to the Swiss stock market. They find evidence of a decrease in volatility and absolute prices. They find no evidence that stock returns are affected by expiration. Jouaber-Snoussi and Tekaya (2012) investigate the effect of option listing on price efficiency. They find no evidence that option listing deteriorates liquidity. Danielsen, Van Ness, and Warr (2007) find evidence that the introduction of options to a stock market improves market quality.

Papers that study other markets include Tuzun (2014) who shows that price moves in major stock indices result in large rebalancing flows in stock markets from leveraged and inverse exchange-traded funds (ETFs). The paper show that rebalancing flows cause price reactions and increase volatility. The paper argues that the destabilizing effect of ETF portfolio rebalancing is similar to portfolio insurers of the 1980s which are believed to have played a major role in the market crash of 1987.

De Jong, Koedijk, and Schnitzlein (2006) run an experiment in a lab setting that examines trading in a market with asymmetric information in the presence of a call option. They find evidence that the introduction of an option improves market quality.

Most of the above papers focus on asset returns and price distortion, while we focus on liquidity. We contribute to this stream of literature methodologically. To the best of our knowledge, the above do not use time to expiry as an instrument for liquidity. Further, we model a previously unstudied mechanism that is consistent with the result of improved liquidity.

Theoretical papers have also shed light on the effect of hedging and program trading on the underlying asset. Papers that study the strategic interaction of traders, as this paper does, include Gennotte and Leland (1990). They develop a rational expectations model in which crashes may arise from small amounts of selling which lowers prices. Their model is consistent with other papers which argue that the crash of 1987 was ignited by selling from portfolio insurers.

In contrast to our approach, in the tradition of Kyle (1985), which treats price impact endogenously, another stream of literature treats price impact as exogenous, and instead focuses on the underlying price process, derivative pricing, and hedging strategies. This literature includes Nayak and Papanicolaou (2008) who study a theoretical model of the effect of portfolio optimizers on underlying asset prices. They find that portfolio optimizers may have a stabilizing effect, decreasing volatility. Wilmott and Schönbucher (2000) develop a theoretical model of dynamic trading strategies with price impact. Their model predicts an increase in liquidity and predicts price jumps that arise as the price of the stock crosses the strike price of an option. Sircar and Papanicolaou (1998) develop a model of dynamic hedging with price feedback. They derive partial differential equations for valuing and pricing options and introduce numerical methods. The model predicts an increase in volatility caused by hedging activity.

Lyukov (2004) develops a pricing model with feedback effects on volatility. The model is able to describe the empirical phenomenon of volatility smiles. Almgren and Li (2016) develop a theoretical model of a hedger who faces price impact. Platen and Schweizer (1998) model feedback effects and numerically show that they can give rise to volatility smiles and skewness effects. Frey and Stremme (1997) analyze feedback effects using a model of hedging demand with feedback and derive closed-form expressions for volatility under feedback. Bank and Baum (2004) develop a continuous-time model of portfolio optimization with price impact and develop trading strategies. We discuss the relationship between our model and this stream of literature in Section 3.2.4.

3.1.2 Transparency and Trading Platforms

Our work is related to transparency in derivatives markets. We model a situation in which there is uncertainty about the number of securities being hedged. This situation might arise if there is a delay in the dissemination of information about trading in derivatives markets that give market participants a way to estimate the number of contracts being hedged. Many derivatives trade over-the-counter (OTC), where trades are bilaterally negotiated and are traded away from organized trading platforms. In this case, many market participants may have no information about the size of markets, the trading prices of derivatives contracts, the nature of the end users since there is no public record of post-trade information.

In 2009, the G20 mandated a set of derivatives market reforms, including that trading information be reported to trade repositories, and standardized contracts be traded on organized

trading platforms which will publically report information about trading.² Trade repository data will contain a record of all derivatives trades, including price, volume, counterparty, jurisdiction, etc. It will only be available to regulators, but it will nonetheless give regulators information with which to judge the potential impact of derivatives markets on other markets.

The other related G20 commitment is that derivatives trades should occur on trading platforms. In the US, swap execution facilities (SEFs) have been launched and carry a mandate that certain large and standardized classes of derivative contracts only be traded on SEFs. These include certain types of interest rate swaps and credit default swaps, each of which may result in hedging in another market.³ Interest rate swaps may generate hedging in the fixed-income market, as may credit default swaps in the corporate bond market. In Canada, the Canadian Securities Administrators (CSA) has released a paper on the Derivatives Trading Facilities (DTFs) for consultation which will also mandate trading of certain classes of derivatives on trading platforms (Canadian Securities Administrators, 2015).

Regulators have not settled on which information should be released publically by derivatives trading platforms. For example, in the US, the Commodities Futures Trading Commission (CFTC) have decided that real-time reporting is appropriate, while in Canada and in other jurisdictions there may be longer delays to protect certain market participants.⁴ This work may help inform this process by showing that when market making is relatively competitive in the underlying security and if derivatives markets are mainly composed of uninformed natural investors, granular and expedient reporting of derivatives trading information may be desirable.

3.1.3 Sunshine Trading vs. Front-Running

Our work is also related to the literature on sunshine trading versus front-running. Harris (2002) defines front-running and sunshine trading and provides empirical examples. The two are opposite sides of the situation in which one or more market participants have knowledge of future trading intentions of a “block trader” who wishes to build a large position in a security. Front-running describes the case where this information is exploited to the detriment of the block trader. If the block trader’s intentions are known to other traders, they may be able to buy before the block trader does, taking advantage of the block trader’s price impact and then selling at the higher price. The frontrunner’s trades may also have price impact and result in worse prices for the block trader.

Sunshine trading describes the case in which a block trader finds it advantageous for its future trading intentions to be known, and possibly because it signals this to other market participants. If the block trader is seen as a desirable counterparty, and can credibly commit to a pattern of trade, then other traders may compete to provide liquidity, resulting in better prices for the block trader than if its trading intentions were unknown. Hedgers may be similar to block traders in this regard in that they may trade in large quantities and without private information about asset values. Option hedgers’ trading intentions may be detectable by other market participants if they have information about their option inventory and their hedging

²The statement is available at www.g20.utoronto.ca/2009/2009communique0925.html.

³Details on SEFs can be found at www.cftc.gov/LawRegulation/DoddFrankAct/Rulemakings/DF_13_SEFRules/index.htm.

⁴Details on reporting of trades on SEFs are available at <http://www.cftc.gov/idc/groups/public/@lrfederalregister/documents/file/2011-33173a.pdf>.

strategy. In the case of equity options, open interest is publically available, and hedging strategies are well-known. We do not model front running and sunshine trading directly, but model a situation where other market participants learning about option inventories is unavoidable for the hedger. Under our assumptions, competitive market making and uninformed hedging, the economics resemble sunshine trading as liquidity improves when the hedger is present in the market.

In some cases hedgers may be concerned about front running when there is knowledge of their trading intentions. For example, in 2014, Southwest Airlines was given permission by the CFTC to delay reporting of commodity derivatives trades that it used to hedge oil prices at long horizons. Many swap contracts are reported to swap data repositories (SDRs) for public dissemination. However, Southwest argued that this data could inform other market participants about its trading intentions which it considered to be unwanted information leakage.⁵ Despite the general possibility for information leakage leading to front running for hedgers, we find our empirical study corroborates our model's prediction that hedging improves liquidity in equity markets around options expiry.

Empirical studies have found evidence of sunshine trading in other markets. Bessembinder, Carrion, Tuttle, and Venkataraman (2016) empirically study predictable ETF roll trades in crude oil futures markets, when the front futures contract expires and the ETF must shift its position to futures of a different expiry. They find evidence that these trades receive improved prices. The result is consistent with sunshine trading, as opposed to front-running.

There are theoretical papers on sunshine trading. Degryse, de Jong, and van Kervel (2014) develop a theoretical Kyle (1985) style model with a block trader who trades over many periods. The block trader can benefit from sunshine trading by signalling his intentions by trading more aggressively early in the game. We take a cue from their model by using short-lived informed traders and steady-state dynamics. Danthine and Moresi (1998) develop a Kyle (1985) style model with predatory trading in which an informed trader has prior knowledge of noise trades. The outcome for noise traders is ambiguous.

3.1.4 Equity Options and Information

For the sunshine trading hypothesis to be valid, hedging traders must be desirable counterparties and attract liquidity providers. Several empirical studies find that in equity markets, option trades contain less information about future prices than trades in the underlying equities. Therefore, option hedging trades are likely to contain little information about future prices and hedgers can be considered desirable counterparties by market makers in underlying equities. Muravyev et al. (2013) use tick-by-tick data on an options market and their corresponding underlying stocks. They find that when the options market and the underlying market disagree about the price of the stock, prices in the options market move to correct the disagreement. They argue that option quotes therefore do not provide economically significant information about the price of the underlying stocks. Cho and Engle (1999) model liquidity in the options market and suggests that market quality in the underlying asset is an important determinant of market quality in options markets. Stucki and Wasserfallen (1994), also mentioned above,

⁵The CFTC's letter allowing delayed reporting is available at www.cftc.gov/idc/groups/public/@l1lettergeneral/documents/letter/14-134.pdf.

find that price discovery in the Swiss equity market leads the options market by roughly ten minutes.

However, other papers have found evidence of informed trading in options markets. Kaul, Nimalendran, and Zhang (2004) empirically test the relationship between bid-ask spreads in the options market with strategic, informed trading. They find evidence of informed trading in options markets by statistically testing the component of the option bid-ask spread that is driven by information asymmetry. Easley, O'Hara, and Srinivas (1998) provide theoretical and empirical evidence that there is informed trade in options markets. These papers highlight that our model's assumption that hedging trades are uninformed may not be realistic, but do not invalidate our model's predictions. Our theoretical result should still hold if hedging trades carry relatively less information and therefore impose lower adverse selection costs on market makers than trades from non-hedgers.

There are other reasons to believe a high degree of options volume comes from natural investors, as opposed to informed speculators who pose risks to market makers. Judd and Leisen (2010) analyze the drivers of demand for options and argue that investors use options to achieve a desired level of skewness. Other sources of natural demand for options may come from covered call writing, and portfolio hedging with put options.

We read the literature on equity options as generally indicating that there is relatively more information about future prices in the underlying stocks than in their options. We note that other papers suggest this may not be true for other asset classes. Charlebois and Sapp (2007) find strategies that use information from foreign exchange options outperform commonly used strategies that use only foreign exchange spot prices. This indicates a high degree of informed trading may occur in the foreign exchange options market.

3.2 Model

We begin with a single period case to illustrate the interaction between the market maker, the hedger, and informed and noise traders. It also eases the exposition of the multiperiod setting in 3.2.2. Our model follows the classic model of Kyle (1985) and is closely related to the model of Degryse et al. (2014).

3.2.1 Single-Period Setting

Trading for a single risky asset takes place once at time t_0 before the asset's payoff is realized at time t_1 . The asset's payoff v is distributed $\mathcal{N}(\mu_v, \sigma_v^2)$. The risk-free rate is normalized to zero. There are four risk-neutral agents: a noise trader, an informed trader, a hedger trader and a market maker.⁶ As in Kyle (1985), agents anonymously submit quantities⁷, positive or

⁶The types of agents in the model have natural analogues: a noise trader needs to increase or decrease exposure to the risky asset but has no private information and resembles a retail trader or a fund manager with random cash inflows and outflows; the informed trader is an arbitrageur such as a hedge fund that searches for mispricings; the hedger resembles a derivatives dealers who hedges by trading directly in the underlying; and the market maker resembled a specialist, or a professional liquidity provider whose business is to profit from transaction costs.

⁷We can interpret these as market orders in the sense that they are orders to trade a given quantity at the best price offered.

negative for buy and sell respectively, to the market maker who receives an aggregate quantity (the residual after summing over all orders). We will also refer to the aggregate quantity as the “order flow”. The market maker clears the market at the competitive price p , conditional on the information gained from observing the order flow.

An equilibrium in our model is a competitive, Bayesian pricing rule used by the market maker and a quantity that maximizes profit for the informed trader. Only the market maker and the informed trader act strategically since the other agents are concerned with liquidity needs or risk management and not the prices at which they will trade. We will formalize the definition of equilibrium below in 3.2.1.

Noise Trader

At time t_0 , the noise trader realizes a random exogenous liquidity need for u shares with $u \sim \mathcal{N}(0, v_u^2)$ which is independent of v . The noise trader has a perfectly inelastic demand and therefore performs no optimization.⁸ The noise trader submits a market order of size u .

Hedger

The hedger acts as a dealer in a market for a derivative asset and hedges in the market for the underlying asset. We will assume there is no information contained in the demand for the derivative asset.⁹ The hedger manages risk from a random inventory of derivatives of size $H \sim \mathcal{N}(\mu_H, v_H^2)$ which is independent of v and u . He submits a market order for h shares (a constant) of the risky asset for each derivative. His inventory is privately known H .¹⁰ He is only concerned with risk management and not the price at which the hedging position is obtained. The hedging strategy per derivative h is independent of the aggregate inventory of derivatives H . It is assumed that other market participants know the hedging strategy being used and therefore know h but not H . In total, the hedger submits a market order of size hH which to the other agents is distributed $\mathcal{N}(h\mu_H, h^2v_H^2)$.

Informed Trader

At time t_0 , the informed trader receives a perfect signal about the realization of the asset’s future payoff v . He is risk-neutral and submits a market order for \hat{x} shares to maximize expected profits:

$$\hat{x} = \arg \max_x \mathbb{E}[x(v - p(x)) \mid \mathcal{I}_i]. \quad (3.1)$$

⁸In equilibrium, the noise trader’s expected price is equal to the expected value of the asset given only public information μ_v . The competitive nature of the market maker ensures that the noise trader does not provide other traders with infinite profits.

⁹This is a reasonable assumption for equities and equity options, as discussed in Subection 3.1.4. Equity ETFs are another example of where our model is reasonable. Many investors simply want exposure to the market and do not trade for speculative reasons. Since ETFs are passively managed, the underlying assets will be bought and sold periodically for reasons unrelated speculation about fundamentals.

¹⁰We assume there is no credible way to reveal the inventory H . Indeed, the hedger could profit from falsely announcing inventory, or an informed trader could profit from falsely claiming to be hedging.

where $p(x)$ is the price which depends on x through the market maker's pricing rule, \mathcal{I}_i denotes the informed trader's information set $\mathcal{I}_i = \nu$.

Market Maker

The market maker receives market orders from each of the other agents simultaneously and observes the aggregate order flow $y = u + x + hH$ but not the individual order sizes. The market maker learns from the order flow; although orders are submitted anonymously, the market maker can revise his prior about the distribution of the risky asset by using Bayes rule. The market maker acts perfectly competitively¹¹, as if there were other market makers able to undercut prices and therefore clears the market at the price $p(y)$ which results in zero expected profits:

$$p(y) = \mathbb{E}[\nu \mid \mathcal{I}_m]. \quad (3.2)$$

\mathcal{I}_m denotes the market maker's information set. The market maker's information set includes all public information about the risky asset, specifically its mean and variance μ_ν and ν_ν^2 , and aggregate order flow y . We will use the term "belief" to describe the probability distribution of the risky asset's realized value.

Note that in this description of the model, agents submit market orders which are then priced by the market maker conditional on the aggregate order flow. The market maker has no private information before observing the order flow y so other market participants will be able to predict the market maker's pricing rule. This is equivalent to the market maker announcing his price as a function of order flow y before trade occurs, as is the case with a limit order book.

Definition of Equilibrium

We restrict the model to the case of linear strategies.¹² Formally, an equilibrium is a pair of functions

$$\hat{x}(\nu) = \hat{x}(\nu, \hat{p}(y)) \quad (3.3)$$

$$\hat{p}(y) = \hat{p}(y, \hat{x}(\nu)) \quad (3.4)$$

for any (ν, y) in \mathbb{R}^2 , such that $\hat{x}(\nu)$ and $\hat{p}(y)$ satisfy:

$$\hat{x}(\nu) = \arg \max_x \mathbb{E}[x(\nu - \hat{p}(x)) \mid \mathcal{I}_i] \quad (3.5)$$

$$\hat{p}(y) = \mathbb{E}[\nu \mid \mathcal{I}_m]. \quad (3.6)$$

We now derive the informed trader's strategy and the pricing rule explicitly. We will see that if the informed trader's strategy is linear in the true value of the asset ν , then the pricing

¹¹ Assuming perfect competition for the market maker is a standard assumption that greatly increases tractability. It can be seen as a lower bound in the sense that if price impact were less than it is in the model, the market maker would lose money in expectation. Further, market making has become increasingly competitive as barriers to entry have decreased as trading has become more electronic of trading in many asset classes.

¹² Boulatov and Livdan (2005) prove that linear strategies are optimal in the general single-period Kyle (1985) model.

rule is linear in the informed trader's strategy x . Assume a linear pricing rule with the formula:

$$\hat{p}(x) = \mu_v + \lambda(y(x) - \mathbb{E}[y(x) | \mathcal{I}_m]) \quad (3.7)$$

$$= \mu_v + \lambda(y(x) - \tilde{y}). \quad (3.8)$$

as a guess about the form of the market maker's pricing function where we have let \tilde{y} be the market maker's expected aggregate order. We will refer to the parameter λ as the "price impact" coefficient. The formula is intuitive; the market maker's belief about the mean of the asset's true value after trade is the previous belief μ_v , plus a correction term proportional to the unexpected order flow $y(x) - \mathbb{E}[y(x) | \mathcal{I}_m]$.

Informed Trader's Strategy

Substituting the linear pricing rule above into the informed trader's objective:

$$\hat{x}(v) = \arg \max_x \mathbb{E}[x(v - \hat{p}(y)) | \mathcal{I}_i] \quad (3.9)$$

$$= \arg \max_x \mathbb{E}[x(v - (\mu_v + \lambda(y(x) - \tilde{y}))) | \mathcal{I}_i] \quad (3.10)$$

$$= \arg \max_x \mathbb{E}[xv - x\mu_v - x\lambda(x + hH + u - \tilde{y}) | v] \quad (3.11)$$

$$= \arg \max_x xv + x\mu_v - x^2\lambda \quad (3.12)$$

$$= \beta(v - \mu_v) \quad (3.13)$$

where $\beta := \frac{1}{2\lambda}$. The second equality is from the definition of y . The third equality uses the facts that the informed trader's information is the true value of the asset $\mathcal{I}_i = v$, the expectation of u and hH are zero by their independence from v , and that $\mathbb{E}[x + hH + u - \tilde{y} | v] = x + \mathbb{E}[hH | v] - \mathbb{E}[hH | \mathcal{I}_m] = x$ because the informed trader's information about the hedging order hH is no better and no worse than the market maker's information.

Now that we have derived the informed trader's strategy for a pricing rule for a given parameter λ , we can derive the formula for λ explicitly using Bayes' Rule.

Market Maker's Strategy

Note that the aggregate order flow is linear in the fundamental value v : $y = x + u + hH = \beta(v - \mu_v) + u + hH$. This means the market maker's pricing rule is also linear in the fundamental value:

$$\hat{p}(y) = \mu_v + \lambda(y - \mathbb{E}[y | \mathcal{I}_m]) \quad (3.14)$$

$$= \mu_v + \lambda(y - h\mu_H) \quad (3.15)$$

$$= \mu_v + \lambda(\beta(v - \mu_v) + u + hH - h\mu_H) \quad (3.16)$$

$$(3.17)$$

where we have used the fact that the only expected component of the order flow comes from the hedging trade $\mathbb{E}[y | \mathcal{I}_m] = h\mu_H$. This allows us to derive λ using conditional expectation for multivariate normal random variables which in this case is equivalent to the Bayesian update of

the market maker's belief given an observation of a Normal random variable with additive, normally distributed noise. The Bayesian updating coefficient takes the form of a signal-to-noise ratio where the numerator is the variance of the variable being updated and the denominator is the total variance of the observed quantity. The stochastic part of informed trader's order is βv_v , so variance of the "fundamental component" of the order flow is $\beta^2 v_v^2$. The variance of the order flow is $\beta^2 v_v^2 + v_u^2 + h^2 v_H^2$, the sum of the variance of fundamental component, the noise trader's order, and the hedger's order. The signal-to-noise ratio for updating βv_v , is therefore $\frac{\beta^2 v_v^2}{\beta^2 v_v^2 + v_u^2 + h^2 v_H^2}$. Dividing by β gives the Bayesian updating coefficient for the quantity v_v :

$$\lambda = \frac{\beta v_v^2}{\beta^2 v_v^2 + v_u^2 + h^2 v_H^2} \quad (3.18)$$

Using $\beta = \frac{1}{2\lambda}$ from the informed trader's strategy derived above, the Nash equilibrium price impact coefficient is

$$\lambda = \frac{v_v}{2\sqrt{v_u^2 + h^2 v_H^2}}. \quad (3.19)$$

We have derived the same price impact coefficient as in Kyle (1985) but with another term $h^2 v_H^2$ in the denominator showing that the hedger in this setting is essentially additional noise trading. When the variance of the hedging trade $h^2 v_H^2$ increases, liquidity improves in the form of a smaller price impact per share. In the single-period setting above, the market maker only needs to update one quantity after observing the aggregate order flow, the mean of the asset's payoff μ_v . In a multi-period setting, the market maker will need to update four quantities after each trade as he learns about both the fundamental value and the size of the hedger's inventory, the prior means and variances of each quantity.

3.2.2 Multiperiod Setting

Trading takes place at times $n = 1, \dots, N$. The asset's fundamental value follows a random walk with standard deviation v_v per period; $\Delta v_n = v_v \Delta W_n$ where $\Delta W_n \sim \mathcal{N}(0, 1)$. The market clearing mechanism is the same as in the single period case but occurs each period. Agents are otherwise the same as in the single period case, with the differences noted below.

Noise Trader

At the beginning of each period the noise trader realizes a random exogenous liquidity need $v_u U_n$ where $U_n \sim \mathcal{N}(0, 1)$ is a standard normal random walk independent from W_n .

Hedger

The number of contracts in the hedger's portfolio follows a random walk with standard deviation v_H per period; $H_n = v_H V_n$ where V_n is a standard normal random variable. The changes in inventory are due to client-trades in the market for the derivative asset. The hedger trades h per derivative each period. The net inventory H_n is unobservable by other market participants. In total, the hedger submits a market order of size hH_n at period n .

We assume the “hedging strategy” h does not change from period to period, that is, the hedger must trade some constant fraction of their inventory each period. This is not realistic in the sense that most hedging strategies are functions of the underlying price and are quite possibly nonlinear. However, the linear hedging trades hH_n adequately introduce the key friction into the linear Kyle (1985) model setting: that market participants have some knowledge of the level of hedging, but it is imperfect and more information must be inferred from the order flow. It is not clear how to introduce hedging traders that are functions of the underlying price in a tractable way, so we proceed with the assumption that h is constant.

Market Maker

Each period a market maker receives market orders from each of the other agents, observes the aggregate order flow $y_n = u_n + x_n + hH_n$, and clears orders at the price p_n which satisfies the zero expected profit condition

$$p_n = \mathbb{E}[v_n | y_n]. \quad (3.20)$$

Let the market maker’s information set be denoted by \mathcal{I}_n^m .

Informed Trader

A new informed trader arrives at the beginning of each trading period with knowledge of the asset’s value v_n .¹³ Each new informed trader has no private information about the hedger’s inventory and therefore uses the public information set, or equivalently the market maker’s information set \mathcal{I}_n^m . Let the informed trader’s information set be denoted by \mathcal{I}_n^i . They trade x_n in order to maximize profit according to the following objective function:

$$x_n = \arg \max_{\hat{x}_n} \mathbb{E}[\hat{x}_n(v_n - p_{n-1}) | \mathcal{I}_n^i, \mathcal{I}_n^m] \quad (3.21)$$

$$= \arg \max_{\hat{x}_n} \mathbb{E}[\hat{x}_n(v_n - p_{n-1}) | \mathcal{I}_n^i] \quad (3.22)$$

since $\mathcal{I}_n^m \subset \mathcal{I}_n^i$.

As in Subsection 3.2.1, we will assume the game has a linear structure, that is, the market maker’s pricing rule $\hat{p}_t(y_t)$ is linear in the order flow each period, and the informed trader’s strategy $\hat{x}(v_t)$ is linear in the asset’s value.

Definition of Equilibrium

An equilibrium is a sequence of pairs of functions

$$\hat{x}_n(v_n) = \hat{x}_n(v_n, \hat{p}_n(y_n)) \quad (3.23)$$

$$\hat{p}_n(y_n) = \hat{p}_n(y_n, \hat{x}_n(v_n)) \quad (3.24)$$

¹³The assumption that a new informed trader arrives each period is for tractability. It is difficult to solve for the optimal dynamic trading strategy when price impact varies over time, or taking into account how informed traders might learn about the hedger’s inventory. However, in Appendix B.2 it is shown that for a constant price impact the informed trader’s optimal dynamic strategy is similar, the main difference being trades are smaller so as not to move the price too quickly in order to preserve future trading profits. The assumption that a new informed trader arrives each period is unlikely to substantially change nature of the results.

for all $n > 0$ for any (v_n, y_n) in \mathbb{R}^2 , such that $\hat{x}(v)$ and $\hat{p}(y)$ satisfy:

$$\hat{x}_n(v_n) = \arg \max_{x_n} \mathbb{E}[x_n(v_n - \hat{p}_n(x_n)) \mid \mathcal{I}_n^i] \quad (3.25)$$

$$\hat{p}_n(y_n) = \mathbb{E}[v_n \mid \mathcal{I}_n^m]. \quad (3.26)$$

Informed Trader's Strategy

Let the price impact coefficient at period n denoted by λ_n^p . The informed trader's problem each period is similar to single-period case:

$$x_n = \arg \max_{\hat{x}_n} \mathbb{E}[\hat{x}_n(v_n - (p_{n-1} + \lambda_n(y_n - h_n \mu_{Hn}))) \mid \mathcal{I}_n^i, \mathcal{I}_n^m] \quad (3.27)$$

$$= \beta_n(v_n - p_{n-1}) \quad (3.28)$$

where $\beta_n = \frac{1}{2\lambda_n}$, and we have used the fact that the market maker's prior mean about the asset's value v_n is p_n by the equilibrium concept above.

Market Maker's Strategy

As shown above, the amount traded by the informed trader at period n is $\beta_n(v_n - p_{n-1})$, of which $\beta_n v_n$ is unexpected by the market maker. The market maker must maintain beliefs about the distributions of the unexpected components of the order flow that come from the hedger and the informed trader: $\beta_n v_n := A_n$ from the informed trader, and $h_n H_n := B_n$ from the hedger. Denote their "noisy sum" by $C_n := A_n + B_n + u_n$, where u_n is the noise trader's trade at period n . We use C_n for convenience; it is equal to total order flow y_n less a deterministic constant known by the market maker. Let the covariance matrix for A_n , B_n and C_n be

$$\Sigma_n = \begin{bmatrix} (\sigma_n^A)^2 & \sigma_n^{AB} & \sigma_n^{AC} \\ \sigma_n^{AB} & (\sigma_n^B)^2 & \sigma_n^{BC} \\ \sigma_n^{AC} & \sigma_n^{BC} & (\sigma_n^C)^2 \end{bmatrix}$$

and their means be

$$\Upsilon_n = \begin{bmatrix} \mu_n^A \\ \mu_n^B \\ \mu_n^C \end{bmatrix}. \quad (3.29)$$

Together, Σ_n and Υ_n make up the market maker's prior at period n about the components of the order flow.

Define the partitions of Σ and Υ as follows:

$$\Sigma_n = \begin{bmatrix} \Sigma_n^{1,1} & \Sigma_n^{1,2} \\ \Sigma_n^{2,1} & \Sigma_n^{2,2} \end{bmatrix} := \begin{bmatrix} (\sigma_n^A)^2 & \sigma_n^{AB} & \sigma_n^{AC} \\ \sigma_n^{AB} & (\sigma_n^B)^2 & \sigma_n^{BC} \\ \sigma_n^{CA} & \sigma_n^{CB} & (\sigma_n^C)^2 \end{bmatrix}$$

and

$$\Upsilon_n = \begin{bmatrix} \Upsilon_n^1 \\ \Upsilon_n^2 \end{bmatrix} := \begin{bmatrix} \mu_n^A \\ \mu_n^B \\ \mu_n^C \end{bmatrix}.$$

After an observation of C_n the means are updated by the rule

$$\Upsilon_{n+1}^1 = \Upsilon_n^1 + \Sigma_n^{1,2}(\Sigma_n^{2,2})^{-1}(C_n - \Upsilon_n^2) \quad (3.30)$$

which is the multivariate Normal Bayesian update; the means conditional on an observation of C_n .

The factor $\Sigma_n^{1,2}(\Sigma_n^{2,2})^{-1}$ in the Bayesian update above in Equation 3.30 gives a matrix of two ‘‘impact coefficients’’: the impact coefficient for the quantity $\beta_n v_n = A_n$ which is very closely related to the price impact coefficient; and the impact coefficient for the quantity $hH_n = B_n$ which is related to the market maker’s learning about the hedging inventory H_n . Let the impact coefficients for A_n and B_n be $\Sigma_n^{1,2}(\Sigma_n^{2,2})^{-1} := \begin{bmatrix} \lambda_n^A \\ \lambda_n^B \end{bmatrix}$.

We are interested in the price impact coefficient $\lambda_n^p = \frac{\lambda_n^A}{\beta_n}$. In general, λ_n^A evolves over time along a path that is dependent on initial conditions such as the market maker’s initial belief about the asset’s value v_0 . However, we will show below that a steady state exists for λ_n^A . The steady state corresponds to when the information gained by the market maker each period is exactly offset by the information lost through the evolution of the random processes. We proceed with this approach below, and derive results on how the price impact coefficient depends on the level of program trading.

3.2.3 Steady State Beliefs

One can find explicit formulae for the posterior variances and therefore the price impact coefficient if one considers the steady-state distributions as discussed in Appendix B.1. A steady state is a covariance matrix corresponding to a belief of the market maker that results in a price impact coefficient λ_n^p that is constant from one period to the next. This belief is when the reductions in the market maker’s posterior variance resulting from observations are perfectly offset by increases in the market maker’s posterior variance resulting from random innovations in the hedger’s inventory and the fundamental asset value.

As shown in Appendix B.1, the formula for λ_n^X is given by

$$\lambda_A = \frac{v_A^2(v_B^2 + v_A^2 + \sqrt{(v_B^2 + v_A^2)(v_A^2 + v_B^2 + 4v_u^2)})}{(v_A^2 + v_B^2)(\sqrt{(v_B^2 + v_A^2)(v_A^2 + v_B^2 + 4v_u^2)} + (v_A^2 + v_B^2) + 2v_u^2)}$$

Substituting the model parameters and using $\beta_n = \frac{1}{2\lambda_n}$, one can find the game’s equilibrium as

$$\lambda_n^p = \lambda^p = \frac{v_v}{\sqrt{v_u^2 + 2h^2v_H^2 + \sqrt{v_u^4 + 12v_u^2h^2v_H^2 + 4h^4v_H^4}}}. \quad (3.31)$$

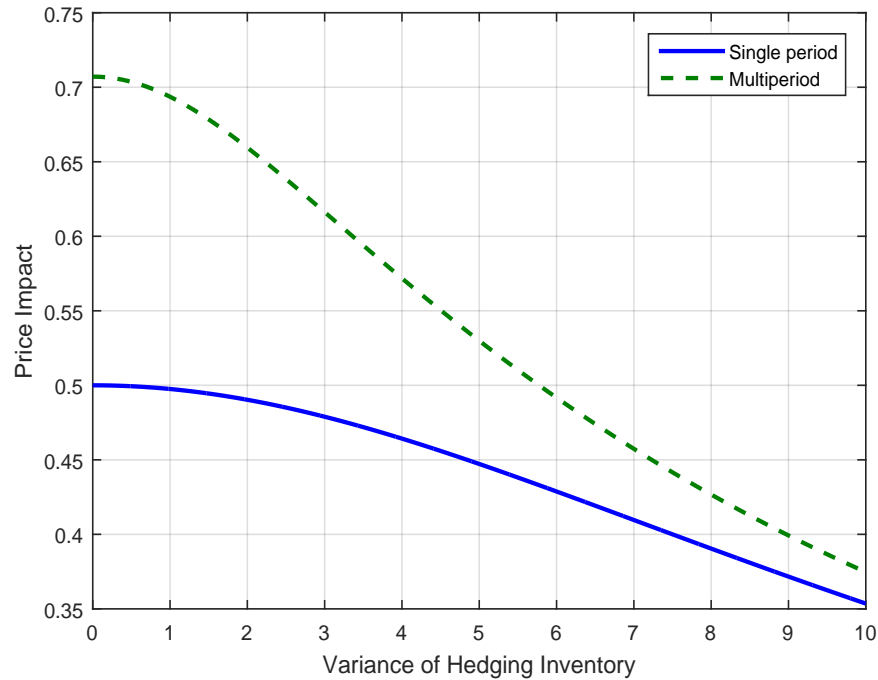
λ^p is decreasing in both variables related to the size of the hedger’s trades h and v_H , meaning that the marginal price impact per share traded is decreasing with the amount of program trading.

Figure 3.1 graphically shows the effect of hedging trades compared to a generic increase in noise trade. Price impact is strictly lower when the market maker does not need to learn about

hedging inventories. When there is uncertainty about hedging activity and the market maker must learn about hedging inventory, price impact is still less than the case without any hedging. This is our main theoretical result, which we emphasize in the following proposition.

Proposition 3.2.1 *Price impact, as given in Equation 3.31, is strictly decreasing in the variance of the hedger's inventory.*

Figure 3.1: **Price impact coefficient.** This figure shows how the price impact coefficient varies with variance the hedging volume. The parameters chosen were $v_u = 0.2$, $v_v = 0.5$ and $h = 0.1$. The solid blue line shows price impact coefficient for the single period model from Equation 3.19 as a function of the variance of hedging trades. The green dashed line shows the price impact coefficient for the multiperiod case from Equation 3.31.



3.2.4 Comparison to Other Models

Models such as those found in Wilmott and Schönbucher (2000), Frey and Stremme (1997), and Sircar and Papanicolaou (1998) consist of a Walrasian market clearing mechanism that determines a price p_t where supply $S(p_t) = S$ is a fixed constant and demand $D(p_t)$ is linear in price:

$$S = D(p_t) = -\theta_t p_t + c_t \quad (3.32)$$

$$\implies p_t = \frac{1}{\theta_t}(c_t - S) \quad (3.33)$$

where θ_t is the slope of the demand function and c_t is an exogenous stochastic process. Typically c_t is taken to be a geometric Brownian motion, supply is normalized to 0 and θ_t to 1 giving

$$p_t = c_t \quad (3.34)$$

$$\implies dp_t = dc_t = \mu p_t dt + \sigma p_t dW_t. \quad (3.35)$$

Adding the demand of program trading strategy $h_t(p_t)$ for and inventory of H contracts to the total demand gives

$$p_t = c_t + Hh_t(p_t) \quad (3.36)$$

$$\implies dp_t = dc_t + Hdh_t = \mu p_t dt + \sigma p_t dW_t + Hdh_t \quad (3.37)$$

$$\implies dp_t = \frac{1}{1 - H \frac{\partial h_t(p_t)}{\partial p_t}} (F dt + G dW_t) \quad (3.38)$$

where the last equation comes from the application of Itô's lemma; $F = \mu p_t + H \left(\frac{\partial h_t}{\partial t} + \frac{\frac{\partial^2 h_t}{\partial p_t^2} p_t^2 \sigma^2}{2(H \frac{\partial h_t}{\partial p_t} - 1)^2} \right)$ and $G = -p_t \sigma$. Supply and demand have multiple intersections when the denominator becomes zero or equivalently $H \frac{\partial h_t}{\partial p_t} = 1$.

Our model endogenizes price impact by modeling how traders act strategically in a similar market clearing mechanism. In contrast to the papers mentioned above where supply is constant and demand functions vary exogenously with time, we study the case where supply and demand functions are endogenous. In the simplest model presented in this paper in Section 3.2.1, a market maker posts a pricing function $p_t = m_t + \lambda_t y_t$ which is linear in the total quantity of market orders y_t submitted by trading agents and has intercept m_t and slope λ_t . This pricing mechanism is similar to the models mentioned above in the sense that price is a linear function of the stochastic process that drives demand.

Frey and Stremme (1997) avoid the case where supply and demand have multiple intersections by imposing that $H \frac{\partial h_t}{\partial p_t} > 1$, while Wilmott and Schönbucher (2000) study the discontinuity in prices that arises when $H \frac{\partial h_t}{\partial p_t} = 1$ with the interpretation that a price jump will occur from one stable Walrasian equilibrium to another when the slope of the demand curve in Equation 3.36 is zero. The authors calculate the prices at which jumps will occur and derive a region in the price-time plane that is unattainable by the process p_t . The intuition behind the unattainable region is that when prices move to a certain level they trigger a feedback loop of hedging trades which pushes prices to the other side of the unattainable region. Our model shows that hedging trades result in lower price impact meaning that prices are less sensitive to changes in demand. This is a reason to believe that predictions of models that do not account for strategic nature of trading may overstate the magnitude of price jumps or unattainable regions.

3.2.5 Empirical Study

In the Sections 3.2.1 and 3.2.2, we showed that liquidity increases with the amount of uninformed hedging activity. To test the effect empirically, we study option hedging in equity markets. In our model, the hedger submitted orders of hH where h is the hedging strategy per derivative and H is the inventory to be hedged. In this section we focus on factors that

change hedging per derivative h rather than the inventory H . The reason for doing so is that there may be unobserved variables that affect both the demand for derivatives and liquidity affecting inventory H and price impact λ simultaneously. Although we do not model demand for derivatives, one could imagine a change in volatility affecting demand for options and affecting liquidity. Indeed, the price impact coefficient in Equations 3.19 and 3.31 depend on the fundamental variance of the asset v_v .

Option time-to-expiry serves as a variable affects hedging strategies; as option expiry dates approach, option hedgers must trade in the underlying equities more frequently or in greater volume to maintain their hedge. We know of no economic reason that time to expiry should directly affect liquidity in the underlying stocks. We show that the increase in trading volume in the underlying stocks associated with option expiry results in improved liquidity measures, including two proxies for price impact, as predicted by our model.

Institutional Details

Options on US stocks mainly trade on the Chicago Board Options Exchange (CBOE). There are both European- and American-style index options while most single stock options are American-style.¹⁴ Index options are cash-settled during the last trading day before the third Saturdays of each month, while single stock options are physically-settled on the same days.

We will need to account for the significant trading volume generated from hedging stock index futures. These contracts trade on the Chicago Mercantile Exchange (CME) and expire quarterly on the last trading day before the third Saturday of the months of March, June, September, and December. These contracts are cash-settled.¹⁵

There are intermediaries in each of these markets. Electronic market makers provide liquidity on options exchanges by continually maintaining bid and offer prices. When an options market maker's order is filled, a new option enters their inventory. A common strategy is to immediately hedge the new option by buying the appropriate delta hedge in the underlying stock or index. The CBOE's floor traders play a similar role to market makers.

Each of these derivative markets generates increased trading volume in the underlying equities as they approach expiry. For cash-settled options, hedgers will deliver the cash to their counterparty if the option expires in the money. However, if they are delta-hedged at the time of expiry, they will need to liquidate this position which generates trading volume in the underlying stock. For physically-settled options, the counterparty in receipt of the stock may hold the shares or immediately sell them. If the shares are sold, this generates an increase in trading volume in the underlying stock.

Hedging index options or futures may be done by either trading directly in the underlying stocks that make up the index, trading an ETF, or trading futures. Futures floor traders may also hedge their exposures using the underlying stocks or an ETF. When a futures contract expires, the trader must unwind their hedging position, generating trading volume in the underlying stocks.

¹⁴Details on CBOE options can be found at www.cboe.com.

¹⁵Details on futures can be found at www.cmegroup.com.

Data

We use a dataset from Bloomberg that spans the calendar year of 2014. The sample contains trading data that is aggregated across trading venues and includes: daily opening, closing, high, and low prices; total daily trading volume; total daily put option open interest; total daily call option open interest; total daily put option trading volume; total daily call option trading volume; for all stocks that were in the S&P 100 Index during the sample, as well as the aggregated data for the S&P 100 Index.¹⁶ We choose US stocks because they are highly liquid and are therefore less likely to be subject to unobservable factors that might affect their liquidity. One stock was dropped from the sample since it had a stock split during the sample period which resulted in incorrect data for its corresponding options. We obtain data on options and futures expiries from the CBOE website. There are twelve option expiry events in our sample.

Liquidity Measures

We compute standard liquidity measures: the Amihud liquidity measure (Amihud, 2002) which is a proxy for price impact, and the Roll liquidity measure (Roll, 1984) which is a measure of the bid-ask spread. We also compute “Relative High-Low” makes use of the daily high and low price in the dataset as a proxy for daily volatility by taking the difference of the two relative to the daily opening price. We also define “High-Low Amihud” which is similar to the high-low volatility measure defined above but is normalized by the number of shares traded. It can be interpreted as another price impact measure which we include for robustness. For each measure, a lower value indicates greater liquidity. We use the following formulae:

$$Amihud_t = \frac{|Return_t|}{Volume_t} \quad (3.39)$$

$$AmihudHighLow_t = \frac{HighPrice_t - LowPrice_t}{Volume_t} \quad (3.40)$$

$$HighLow_t = \frac{HighPrice_t - LowPrice_t}{OpeningPrice_t} \quad (3.41)$$

$$RollMeasure_t = 2 \sqrt{-AutoCov(Return_t, Return_{t-1}, Return_{t-2})}, \quad (3.42)$$

where $Return_t$ is the return computed using the closing price of days t and $t - 1$. The $AutoCov$ operator define above is the autocovariance of the vector of three consecutive returns. The Amihud and High-Low Amihud measures are displayed in units of percent per million shares traded; Relative High-Low is displayed as percent; and the Roll measure is displayed as percent.

Summary Statistics

Table 3.3 shows summary statistics for all stocks in the sample, the stocks that were S&P 100, during the sample period, the calendar year of 2014. Mean Put Volume was roughly 11030 contracts per day while mean Call Volume was roughly double, roughly 18954 contracts per day. Open interest shows a less significant difference between put and call options. Mean Put Open Interest was roughly 273589 while mean Call Open Interest was roughly 342579. Mean

¹⁶A historical list of the S&P 100 constituents can be found at siblisresearch.com/data/historical-components-sp-100.

Volume was 4850768. The mean Relative High-Low price change was roughly 1.5%. Mean volatility was roughly 19%. On average the Amihud liquidity measure was roughly 0.7 percent per million shares, while the Amihud High-Low liquidity measure was roughly 1.3 percent per million shares. The Roll spread measure was roughly 0.6%.

Table 3.2 shows summary statistics as in Table 3.3, but only for trading days which were the last Friday before an option expiry. Table 3 shows complementary summary statistics for the sample excluding the last Friday before an option expiry. We do not report volatility since we only compute it using the entire sample period. Tables 2 and 3 show that expiry days are different than non-expiry days. On expiry days, mean put and call volume are greater than for non-expiry days: roughly 13923 to 10883 contracts per day for puts and 24728 to 18660 contracts per day for calls. Little difference between expiry and non-expiry days is observed for put and call open interest: roughly 272770 to 289683 contracts per day for puts and 342327 to 347530 contracts per day for calls. Volume is significantly higher on expiry days: roughly 9206417 compared to 4629294. Some of this increased volume is due to futures expiries that occur contemporaneously. We control for futures expiry in the analysis below.

The only market quality measure that was higher on expiry days was the Relative High-Low measure: it was roughly 0.1 percentage point points higher on expiry days, 1.6% per millions shares compared to 1.5% per millions shares on non-expiry days. However, we will show that after controlling for other factors in our regressions hedging volume is associated with a lower Relative High-Low measure. All other market quality measures were lower on expiry days. The Amihud price impact measure was significantly lower on expiry days: roughly 0.4% compared to 0.7% per millions shares. The Amihud High-Low liquidity measure was also significantly lower on expiry days: roughly 0.8% compared to 1.3% per millions shares. The Roll measure was lower on expiration days: 0.5% compared to 0.6%.

Figure 3.2 shows average trading volume for stocks in the S&P 100 Index over the sample period. Red vertical lines mark options expiry dates. Volume is higher than average near the 12 option expiries in the sample and significantly higher around the four futures expiries in March, June, September, and December. Figure 3.3 shows average open interest for both calls and puts for stocks in the S&P 100 Index over the sample period. Open interest for calls and puts is also greater around expiry dates. However, we cannot say that all changes in open interest are an exogenous determinant of trading volume so we control for open interest and change in open interest in our regressions.

Table 3.1: Summary Statistics. The rows of this table display summary statistics for the stocks in the sample period, stocks that were part of the S&P 100 for the 2014 calendar year.

Variable	Obs	Mean	Std. Dev.	Min	Max	P25	P50	P75
Put Volume	24549	11030.24	21156.76	4	435505	1728	4662	11498
Call Volume	24552	18953.89	44494.57	22	1336920	2650	6798.5	17499
Put Open Interest	24551	273588.5	407371.8	199	4094143	74069	140430	305116
Call Open Interest	24551	342578.8	565508.1	236	5572699	81693	160192	378141
Volume	24552	4850768	2.59e+07	104478	7.85e+08	803098.5	1376536	2738804
Volatility	24552	19.293	5.214	11.168	35.923	15.222	18.263	21.409
Relative High-Low	24552	1.532	.869	.243	14.189	.965	1.323	1.848
Amihud	24552	.714	.904	0	12.874	.152	.41	.93
Amihud High-Low	24552	1.291	1.198	0	11.967	.496	.966	1.681
Roll Spread Measure	24552	.56	.598	0	7.574	0	.432	.823

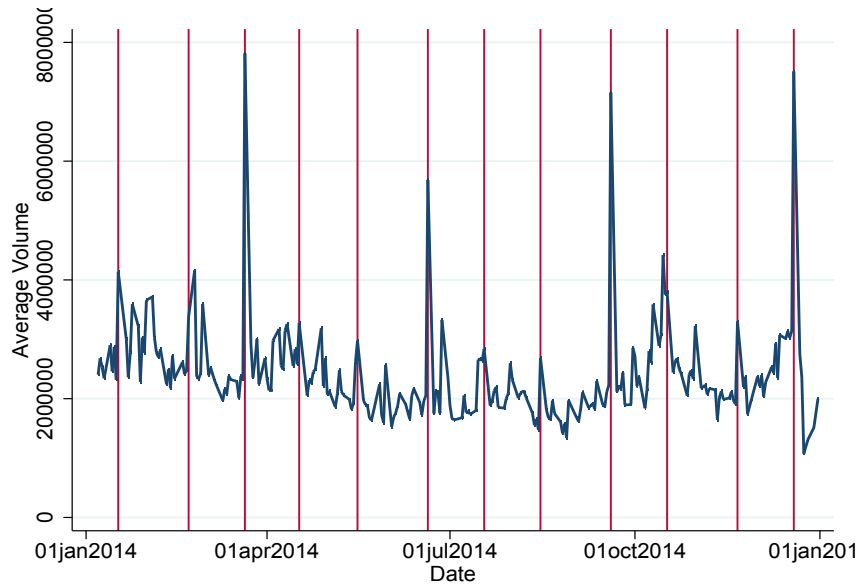
Table 3.2: **Summary Statistics Expiration Days.** The rows of the table shows summary statistics as in Table 1, only for trading days that were the last before an option expiry date. *Volatility* has been excluded since it is computed only using all observations in the sample period.

Variable	Obs	Mean	Std. Dev.	Min	Max	P25	P50	P75
Put Volume	1188	13922.98	23596.63	41	263480	2273.5	6354.5	14598
Call Volume	1188	24727.8	46490.39	316	569825	4786.5	10742.5	25122.5
Put Open Interest	1188	289682.5	424967	13848	4081913	79337	149361	326581.5
Call Open Interest	1188	347530.2	571795.8	9309	5501615	82379	162936.5	380733
Volume	1188	9206417	5.04e+07	291295	7.85e+08	1453632	2658246	5310385
Relative High-Low	1188	1.583	.889	.254	13.154	1.069	1.392	1.853
Amihud	1188	.419	.578	0	5.306	.074	.203	.526
Amihud High-Low	1188	.757	.757	0	6.527	.263	.534	.992
Roll Spread Measure	1188	.527	.575	0	5.948	0	.425	.794

Table 3.3: **Summary Statistics Non-Expiration Days.** This rows of this table shows summary statistics as in Table 1, excluding trading days that were the last before an option expiry date. *Volatility* has been excluded since it is computed only using all observations in the sample period.

Variable	Obs	Mean	Std. Dev.	Min	Max	P25	P50	P75
Put Volume	23361	10883.13	21015.04	4	435505	1700	4595	11317
Call Volume	23364	18660.3	44371.65	22	1336920	2564.5	6610.5	17129
Put Open Interest	23363	272770.1	406449.1	199	4094143	73893	140068	303856
Call Open Interest	23363	342327	565197.7	236	5572699	81675	160094	378015
Volume	23364	4629294	2.39e+07	104478	4.62e+08	787406	1337523	2633981
Relative High-Low	23364	1.53	.867	.243	14.189	.961	1.317	1.848
Amihud	23364	.729	.915	0	12.874	.16	.422	.948
Amihud High-Low	23364	1.318	1.21	.001	11.967	.517	.991	1.718
Roll Spread Measure	23364	.561	.599	0	7.574	0	.433	.824

Figure 3.2: **Average trading volume.** This figure shows average daily trading volume for stocks the S&P 100 Index over the sample period. Red vertical lines mark options expiry dates.



3.2.6 Methodology

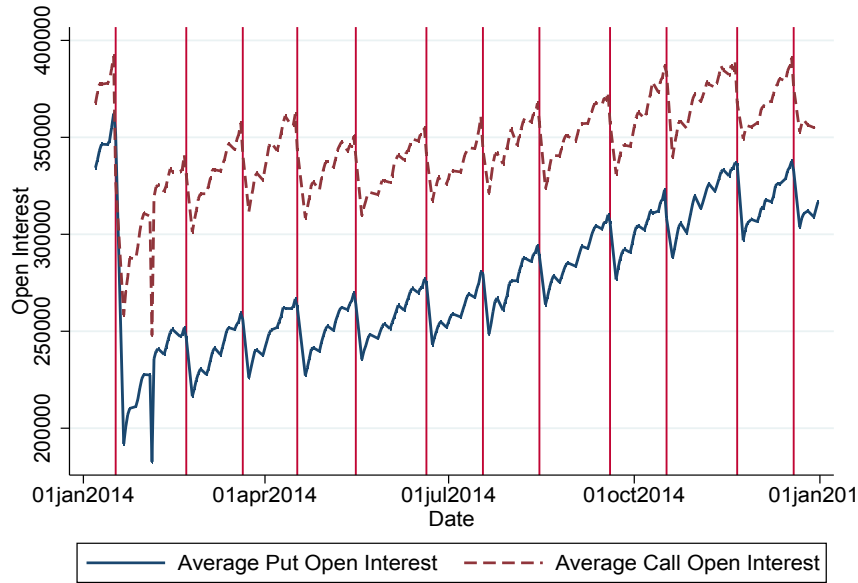
To measure the effect of option hedging trades on liquidity, we use time to expiry as an instrument for volume. Market quality may be co-determined with option hedging volume. Indeed, an improvement in market quality in the underlying security may trigger more trading in the options market, which in turn generates hedging volume in the underlying. Further, there may be unobserved factors that effect both the options market and the underlying market. Figure 3.4 shows a causation diagram.

The component of trading volume in underlying equities attributable to its options' time-to-expiry is exogenous, that is, trading activity in the underlying stock is only affected by the approach of an option expiry date indirectly through option hedgers. As expiry approaches, gamma for both put and call options near the money tends to increase resulting in larger or more frequent delta hedging trades in the underlying stocks. We capture the hedging trades that arise from an increase in gamma by finding the excess trading volume in the underlying stocks during the last five trading days before an option expiry after accounting for volume associated with a number of control variates including option open interest. This does not use all the variation from our volume variable, but only the component attributable to a change associated with time to expiry.

We use a two-stage least squares (2SLS) approach (Cameron & Trivedi, 2005). The method will generally attenuate any corresponding result where liquidity is regressed directly on volume. In the first stage, volume is regressed on dummy variables for each day of options expiry weeks and a set of control variates to control for endogenous variation in volume:

$$Volume_{i,t} = \beta_1 ExpiryMonday_t + \dots + \beta_5 ExpiryFriday_t + \Gamma' X_{i,t} + FE_{i,t} + \varepsilon_{i,t} \quad (3.43)$$

Figure 3.3: **Open interest.** This figure shows average put open interest (the solid blue line) and average call open interest (the dashed red line) for stocks the S&P 100 Index over the sample period. Red vertical lines mark options expiry dates.



The β_j are the regression coefficients of interest for the first stage. All variables are aggregated to the stock-date level. Γ' is a set of regression coefficients for the matrix of covariates $X_{i,t}$ which are: dummy variables for each day of the week; a dummy variable for futures expiry Fridays; daily open interest for put options; daily open interest for call options; daily open interest for S&P 100 Index put options; daily open interest for S&P 100 Index call options; put daily option trading volume; daily call option trading volume; daily trading volume for S&P 100 Index put options; daily trading volume for S&P 100 Index call options; daily absolute change in open interest for S&P 100 Index put options; daily absolute change in open interest for S&P 100 Index call options; daily absolute change in open interest for put options; daily absolute change in open interest for call options; and a linear time trend. We use open interest and option trading volumes to explain hedging volume that may be generated by activity in the options market because these may be driven by unobserved factors that also drive liquidity in the underlying equities. We also control for changes in open interest since a change in open inventory necessitates a hedging adjustment and may cause trading volume in the underlying stocks. We compute the absolute change in open interest as below:

$$|ChangeOpenInterest_t| = |OpenInterest_t - OpenInterest_{t-1}| \quad (3.44)$$

We include fixed effects by stock and option period, and cluster our standard errors at the stock level.

We use the results of the first stage to compute the exogenous component of volume \widehat{Volume}_t by using the β_j regression coefficients and their corresponding dummy variables.

$$\widehat{Volume}_t = \beta_1 ExpiryMonday_t + \dots + \beta_5 ExpiryFriday_t \quad (3.45)$$

Figure 3.4: **Causation diagram.** This causation diagram shows the channels through which liquidity in the underlying stock and hedging volume are affected. Liquidity in the underlying stock and hedging volume may directly affect each other; if liquidity improves options hedging become less costly. This may lower the cost of options and lead to greater open interest which results in further hedging volume in the stock. This volume may again improve liquidity. Unobserved variables may affect both the activity in the options market and the liquidity of the underlying stock. Time to expiry qualifies as an instrument because it does not directly affect liquidity in the underlying stock, but only indirectly through the hedging channel.

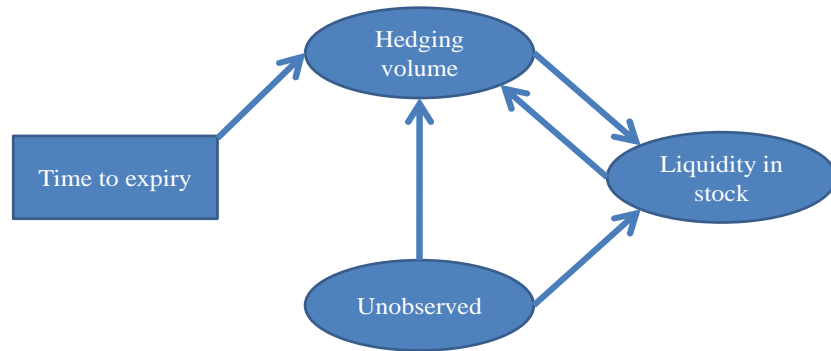


Figure 3.5 shows a plot of \widehat{Volume}_t versus days to expiry. There is a large increase beginning two days from expiry while the values are small elsewhere. The nonlinear shape is expected since options' gammas tend to increase nonlinearly leading up to expiry. The peak of the graph is 600000 shares at one day to expiry meaning we estimate option hedging generates around 600000 shares of trading volume on option expiry Fridays.

In the second stage, the four liquidity measures are regressed on fitted values for volume, \widehat{Volume}_t .

$$Liquidity_{i,t} = \beta \widehat{Volume}_t + \Gamma' X_{i,t} + FE_{i,t} + \varepsilon_{i,t} \tag{3.46}$$

β is the coefficient of interest and the control variables and fixed effects are identical to the first stage. β is may be interpreted as the average marginal effect of hedging volume on liquidity.

3.2.7 Results

Table 3.4 shows the results of the first stage of the 2SLS regressions. The R^2 for the first stage regression was 0.75 and the F -statistic was 14.73, greater than the critical value of 10, indicating that we do not have a weak instrument (Stock & Yogo, 2002).

Table 3.5 shows the results of the second stage of the 2SLS regressions. Each column shows the regression coefficients for the four market quality measures on the variable of interest, fitted values for volume, and the set of covariates. Each market quality measure had a negative coefficient with statistical significance for the fitted values for volume: coefficients were roughly -0.00000032, -0.00000053, -0.00000074, and -0.00000012 respectively for Relative High-Low, Amihud, and Amihud High-Low and thr Roll spread measure. This allows

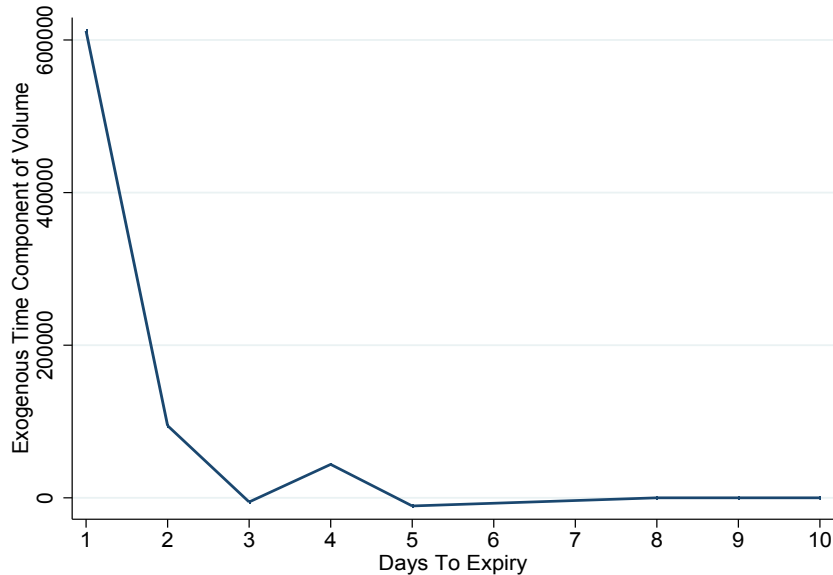
Table 3.4: **Instrumental Variables Regression - First Stage.** The columns of this table show the regression coefficients for the first stage of a two-stage instrumental variables methodology where volume is regressed on a set of covariates.

	(1) Volume
Monday Expiry Week	-10642.8 (-0.22)
Tuesday Expiry Week	43814.4 (1.33)
Wednesday Expiry Week	-5428.9 (-0.13)
Thursday Expiry Week	94544.2* (2.10)
Friday Expiry Week	611947.0*** (8.11)
Change Open Put Interest	0.929 (0.27)
Change Open Call Interest	1.846 (0.60)
Change Index Open Put Interest	-6.155 (-1.04)
Change Index Open Call Interest	14.56 (1.53)
Futures Expiry	3949165.2*** (8.09)
Call Open Interest	-0.163 (-0.14)
Put Open Interest	-0.0179 (-0.02)
Put Volume	50.98*** (8.76)
Call Volume	9.671** (2.64)
Index Put Volume	60.05*** (5.67)
Index Call Volume	18.26** (2.74)
Index Put Open Interest	-2.736 (-0.62)
Index Call Open Interest	28.03*** (3.77)
Observations	24201

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 3.5: **The exogenous component of volume.** The figure shows the exogenous component of volume $Volume_t$, versus time to expiry.



us to make the inference that an increase in hedging volume caused by time to expiry results in an improvement in liquidity. This result is consistent with our hypothesis that price impact decreases near option expiry as market markets compete for uninformed order flow. In the first stage of the 2SLS regressions we estimated that hedging volume was an average of around 600000 shares on option expiry Fridays. Multiplying this average volume increase by our estimates for the changes in our four liquidity measures we estimate the average marginal effect of option hedging on liquidity to be roughly 0.2 percentage points for the Relative High-Low measure (unconditional average of 1.5), 0.3 and 0.4 percentage points per million shares for the Amihud and Amihud High-Low measure (unconditional averages of 0.7 and 1.3 respectively), and 0.1 percentage points for the Roll measure (unconditional average of 0.6).

3.3 Conclusions

We study the effect of hedging derivatives on liquidity in the underlying assets. We develop a theoretical model where traders' interactions are strategic and liquidity is endogenous, built on the classic model of market microstructure of Kyle (1985). We assume that hedging trades are not driven by private information, and that a market maker must learn about the hedging activity through trading. Price impact is due to the presence of informed trades who pose adverse selection risks to the market maker. Our model predicts that despite the need to learn about the asset's true value and hedging inventories through one anonymous order flow, liquidity improves when there is more hedging. The improvement in liquidity arises from the uninformed component of the order flow becoming larger, lessening the need for the market maker to update prices following a trade.

Table 3.5: **Instrumental Variables Regression.** The columns of this table show the regression coefficients for the final stage of a two-stage instrumental variables methodology where four market quality measures are regressed on fitted values of trading volume and a set of covariates. \widehat{Volume} is the exogenous component of volume constructed from the first stage regression.

	(1) Relative High-Low	(2) Amihud	(3) Amihud High-Low	(4) Roll Spread Measure
\widehat{Volume}	-0.00000322*** (-4.86)	-0.00000525*** (-5.24)	-0.00000736*** (-5.28)	-0.00000123* (-2.34)
Change Open Put Interest	0.00000262 (0.29)	-2.38e-09 (-0.00)	0.00000692 (0.28)	-0.00000297 (-0.76)
Change Open Call Interest	0.0000146 (1.54)	0.0000115 (0.73)	0.0000162 (0.70)	0.00000991 (1.55)
Change Index Open Put Interest	-0.0000323*** (-8.71)	-0.0000775 (-1.93)	-0.000258*** (-4.42)	-0.0000274 (-1.11)
Change Index Open Call Interest	0.0000378*** (8.43)	0.0000142* (2.46)	0.000273*** (3.44)	0.0000114** (3.26)
Futures Expiry	1.611*** (6.26)	1.758*** (5.89)	2.540*** (6.15)	0.550* (2.42)
Call Open Interest	0.00000452 (1.14)	-7.27e-08 (-0.13)	-0.00000190 (-0.23)	0.00000260 (1.59)
Put Open Interest	-0.00000700 (-1.81)	-0.00000221 (-0.36)	-0.00000171 (-0.20)	-0.00000306 (-1.60)
Put Volume	0.0000303*** (6.21)	0.0000283*** (5.13)	0.0000374*** (4.96)	0.0000101*** (3.36)
Call Volume	0.00000464** (2.62)	0.00000540* (2.38)	0.00000706* (2.26)	0.00000142 (1.83)
Index Put Volume	0.000101*** (15.10)	0.0000377*** (5.13)	0.0000614*** (5.73)	0.00000819* (2.00)
Index Call Volume	0.0000309*** (5.12)	0.0000643*** (8.34)	0.0000232*** (3.61)	0.0000319*** (8.80)
Index Put Open Interest	-0.00000864* (-2.53)	0.00000272 (0.70)	-0.0000147*** (-3.51)	0.0000115*** (3.79)
Index Call Open Interest	0.0000426*** (8.83)	0.00000614 (1.37)	0.0000323*** (4.84)	0.0000167*** (5.49)
Observations	24201	24201	24201	24201

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We test the model's predictions empirically and verify the effect for a set of stocks which are used to hedge options. Using a 2SLS regression, we find the component of trading volume that is driven by options' time to expiry. We assume that as option expiries approach, an increase in options' gamma results in an increase in hedging volumes, and that this time-to-expiry effect is exogenous to liquidity in the underlying stocks. We find increased trading volumes in the last two days before option expiries, and that the increased volume results in improved measures of liquidity, including a measure of price impact.

Our model of liquidity and option hedging is at the intersection of two streams of literature: market microstructure, which tends to derive liquidity parameters endogenously while simplifying other assumptions, and the options pricing literature, which tends to assume liquidity is exogenous. Our model provides evidence that in certain situations, hedgers should expect improved market quality through lower price impact and lower bid-ask spreads. Some studies with exogenous liquidity derive that option hedging may result in an increase in volatility and price jumps when asset prices and option strike prices approach certain critical values. Our model provides evidence of a natural mechanism that should partially correct for these predicted effects. When hedgers are present, prices may be less sensitive to trading volume if it is known by market makers to be generated by hedging rather than informed speculation.

There are potentially ways to generalize our model framework that would be interesting in future work. We assume that informed traders are short-lived, but one could relax the assumption and try to include dynamic strategies for informed traders. One challenge that arises is that the informed traders will have better information about hedging inventories than the market maker. Since informed traders know the size of their own trades, they will be able to infer hedging traders by observing aggregate trading volumes using a less noisy signal than the market maker. This may result in a further improvement in liquidity if informed traders adjust their strategies to offset rather than amplify hedging trades. Another generalization could be for hedgers to trade strategically. If improved liquidity causes hedgers to trade more, liquidity may improve further. Last, it would be interesting to derive results with a risk-averse market maker. Since the variance of the market maker's belief about fundamentals increases with each trade, it may be that a risk-averse market maker offers worse liquidity, or that results on liquidity are ambiguous.

References

- Alfonsi, A., & Schied, A. (2010). Optimal trade execution and absence of price manipulations in limit order book models. *SIAM Journal on Financial Mathematics*, 1(1), 490–522.
- Almgren, R. (2003). Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10(1), 1–18.
- Almgren, R., & Li, T. (2016). Option hedging with smooth market impact. *Market Microstructure and Liquidity*, 2(01), 1650002.
- Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31–56.
- Bank, P., & Baum, D. (2004). Hedging and portfolio optimization in financial markets with a large trader. *Mathematical Finance*, 14(1), 1–18.
- Battalio, R. (1997). Third market broker-dealers: Cost competitors or cream skimmers? *The Journal of Finance*, 52(1), 341–352.
- Battalio, R., & Holden, C. (2001). A simple model of payment for order flow, internalization, and total trading cost. *Journal of Financial Markets*, 4(1), 33–71.
- Bessembinder, H., Carrion, A., Tuttle, L., & Venkataraman, K. (2016). Liquidity, resiliency and market quality around predictable trades: Theory and evidence. *Journal of Financial Economics*, 121(1), 142–166.
- Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy*, 637–654.
- Boehmer, E., & Kelley, E. (2009). Institutional investors and the informational efficiency of prices. *Review of Financial Studies*, 22(9), 3563–3594.
- Boni, L., Brown, D., & Leach, C. (2013). Dark pool exclusivity matters. Available at SSRN 2055808.
- Boulatov, A., & George, T. (2013). Hidden and displayed liquidity in securities markets with informed liquidity providers. *Review of Financial Studies*, 26(8), 2096–2137.
- Boulatov, A., & Livdan, D. (2005). Strategic trading with market closures. Available at SSRN 870608.
- Brogaard, J., Hendershott, T., & Riordan, R. (2015). Price discovery without trading: Evidence from limit orders. Available at SSRN 2655927.
- Brolley, M. (2016). Mandatory price improvement in dark pools. Available at SSRN 2661722.
- Buti, S., Rindi, B., & Werner, I. (2010). Diving into dark pools. *Charles A. Dice Center Working Paper*(2010-10).
- Buti, S., Rindi, B., & Werner, I. (2015). Dark pool trading strategies, market quality and welfare. *Charles A. Dice Center Working Paper*(2010-6).
- Cameron, C., & Trivedi, P. (2005). *Microeconometrics: Methods and applications*. Cambridge

- University Press.
- Canadian Securities Administrators. (2015). Consultation paper 92 - 401 - derivatives trading facilities. Available at www.osc.gov.on.ca/documents/en/Securities-Category9/csa_20150129_92-401_derivative-trading.pdf.
- Cartea, Á., & Jaimungal, S. (2015). Optimal execution with limit and market orders. *Quantitative Finance*(ahead-of-print), 1–13.
- Chamberlain, W., Cheung, S., & Kwan, C. (1989). Expiration-day effects of index futures and options: Some Canadian evidence. *Financial Analysts Journal*, 45(5), 67–71.
- Charlebois, M., & Sapp, S. (2007). Temporal patterns in foreign exchange returns and options. *Journal of Money, Credit and Banking*, 39(2-3), 443–470.
- Chau, M., & Vayanos, D. (2008). Strong-form efficiency with monopolistic insiders. *Review of Financial Studies*, 21(5), 2275–2306.
- Chiang, C. (2014). Stock returns on option expiration dates: Price impact of liquidity trading. *Journal of Empirical Finance*, 28, 273–290.
- Cho, Y., & Engle, R. (1999). Modeling the impacts of market activity on bid-ask spreads in the option market. Available at www.nber.org/papers/w7331.pdf.
- Chow, Y., Yung, H., & Zhang, H. (2003). Expiration day effects: The case of Hong Kong. *Journal of Futures Markets*, 23(1), 67–86.
- Comerton-Forde, C., Malinova, K., & Park, A. (2016). Regulating dark trading: Order flow segmentation and market quality. Available at SSRN 2755392.
- Comerton-Forde, C., & Putniņš, T. (2014). Dark trading and price discovery. Available at SSRN 2183392.
- Danielsen, B., Van Ness, B., & Warr, R. (2007). Reassessing the impact of option introductions on market quality: A less restrictive test for event-date effects. *Journal of Financial and Quantitative Analysis*, 42(04), 1041–1062.
- Danthine, J., & Moresi, S. (1998). Front-running by mutual fund managers: a mixed bag. *European Finance Review*, 2(1), 29–56.
- De Jong, C., Koedijk, K., & Schnitzlein, C. (2006). Stock market quality in the presence of a traded option*. *The Journal of Business*, 79(4), 2243–2274.
- Degryse, H., de Jong, F., & van Kervel, V. (2014). Does order splitting signal uninformed order flow? Available at http://www.rsm.nl/fileadmin/home/Department_of_Finance_VG5_/LQ2013/Vincent_van_Kervel.pdf.
- Degryse, H., Tombeur, G., Van Achter, M., & Wuyts, G. (2013). *Dark trading*. Wiley Online Library.
- Degryse, H., Van Achter, M., & Wuyts, G. (2009). Dynamic order submission strategies with competition between a dealer market and a crossing network. *Journal of Financial Economics*, 91(3), 319–338.
- Diamond, D., & Verrecchia, R. (1987). Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics*, 18(2), 277–311.
- Easley, D., Kiefer, N., & O'Hara, M. (1996). Cream-skimming or profit-sharing? The curious role of purchased order flow. *The Journal of Finance*, 51(3), 811–833.
- Easley, D., O'Hara, M., & Srinivas, P. (1998). Option volume and stock prices: Evidence on where informed traders trade. *The Journal of Finance*, 53(2), 431–465.
- Fleming, M., & Nguyen, G. (2013). Order flow segmentation and the role of dark trading in

- the price discovery of US treasury securities. *Federal Reserve Board of New York Staff Report*(624).
- Foley, S., & Putniņš, T. (2016). Should we be afraid of the dark? Dark trading and market quality. *Journal of Financial Economics*.
- Forsyth, P., Kennedy, S., Tse, S., & Windcliff, H. (2012). Optimal trade execution: a mean quadratic variation approach. *Journal of Economic Dynamics and Control*, 36(12), 1971–1991.
- Frey, R., & Stremme, A. (1997). Market volatility and feedback effects from dynamic hedging. *Mathematical Finance*, 7(4), 351–374.
- Fudenberg, D., & Tirole, J. (1991). *Game theory*. MIT Press.
- Garriott, C., & Walton, A. (2016). Retail order flow segmentation. *Bank of Canada Staff Working Paper*, 20.
- Gennotte, G., & Leland, H. (1990). Market liquidity, hedging, and crashes. *The American Economic Review*, 999–1021.
- Glosten, L. (1987). Components of the bid-ask spread and the statistical properties of transaction prices. *The Journal of Finance*, 42(5), 1293–1307.
- Glosten, L., & Milgrom, P. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1), 71–100.
- Goettler, R., Parlour, C., & Rajan, U. (2005). Equilibrium in a dynamic limit order market. *The Journal of Finance*, 60(5), 2149–2192.
- Hancock, G. (1993). Whatever happened to the triple witching hour? *Financial Analysts Journal*, 49(3), 66–72.
- Harris, L. (2002). *Trading and exchanges: Market microstructure for practitioners*. Oxford University Press.
- Hasbrouck, J. (1991). Measuring the information content of stock trades. *The Journal of Finance*, 46(1), 179–207.
- Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The Journal of Finance*, 50(4), 1175–1199.
- International Organization of Securities Commissions. (2010). Issues raised by dark liquidity. Available at www.iosco.org/library/pubdocs/pdf/IOSCO336.pdf.
- Jouaber-Snoussi, K., & Tekaya, R. (2012). How does option listing affect underlying stock efficiency? Evidence from a duration model. Available at SSRN 2181391.
- Jovanovic, B., & Menkveld, A. (2011). Middlemen in limit-order markets. *Western Finance Association*.
- Judd, K. (1998). *Numerical methods in economics*. MIT Press.
- Judd, K., & Leisen, D. (2010). Equilibrium open interest. *Journal of Economic Dynamics and Control*, 34(12), 2578–2600.
- Kaul, G., Nimalendran, M., & Zhang, D. (2004). Informed trading and option spreads. Available at SSRN 547462.
- Kumar, R., Sarin, A., & Shastri, K. (1998). The impact of options trading on the market quality of the underlying security: An empirical analysis. *The Journal of Finance*, 53(2), 717–732.
- Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.

- Lee, C., & Ready, M. (1991). Inferring trade direction from intraday data. *The Journal of Finance*, 46(2), 733–746.
- Liu, H., & Kaniel, R. (2004). So what orders do informed traders use? *Available at SSRN 602581*.
- Lo, I., & Sapp, S. (2003). Order submission: The choice between limit and market orders. In *EFMA 2004 Basel Meetings*.
- Lyukov, A. (2004). Option pricing with feedback effects. *International Journal of Theoretical and Applied Finance*, 7(06), 757–768.
- Menkveld, A., Yueshen, B., & Zhu, H. (2015). Shades of darkness: A pecking order of trading venues. *Available at SSRN 2510707*.
- Muravyev, D., Pearson, N., & Broussard, J. (2013). Is there price discovery in equity options? *Journal of Financial Economics*, 107(2), 259–283.
- Nayak, S., & Papanicolaou, G. (2008). Market influence of portfolio optimizers. *Applied Mathematical Finance*, 15(1), 21–40.
- Nimalendran, M., & Ray, S. (2014). Informational linkages between dark and lit trading venues. *Journal of Financial Markets*, 17, 230–261.
- Obizhaeva, A., & Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1), 1–32.
- O'Donoghue, S. (2015). The effect of maker-taker fees on investor order choice and execution quality in US stock markets. *Kelley School of Business Research Paper*(15-44).
- Ontario Securities Commission. (2010). Position paper 23-405 dark liquidity in the Canadian market. *Available at www.osc.gov.on.ca/documents/en/Securities-Category2/csa_20101119_23-405_dark-liquidity.pdf*.
- Parlour, C., & Rajan, U. (2003). Payment for order flow. *Journal of Financial Economics*, 68(3), 379–411.
- Platen, E., & Schweizer, M. (1998). On feedback effects from hedging derivatives. *Mathematical Finance*, 8(1), 67–84.
- Ready, M. (2014). Determinants of volume in dark pool crossing networks. In *AFA 2010 Atlanta Meetings*.
- Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39(4), 1127–1139.
- Rosu, I. (2015). Liquidity and information in order driven markets. *Available at SSRN 1286193*.
- Sircar, R., & Papanicolaou, G. (1998). General Black-Scholes models accounting for increased market volatility from hedging strategies. *Applied Mathematical Finance*, 5(1), 45–82.
- Stock, J., & Yogo, M. (2002). *Testing for weak instruments in linear IV regression*. National Bureau of Economic Research Cambridge, Mass., USA.
- Stokey, N., Lucas, R., & Prescott, E. (1989). *Recursive methods in economic dynamics*. Harvard University Press.
- Stoll, H., & Whaley, R. (1987). Program trading and expiration-day effects. *Financial Analysts Journal*, 43(2), 16–28.
- Stoll, H., & Whaley, R. (1991). Expiration-day effects: what has changed? *Financial Analysts Journal*, 47(1), 58–72.
- Stucki, T., & Wasserfallen, W. (1994). Stock and option markets: The Swiss evidence. *Journal of Banking & Finance*, 18(5), 881–893.

- Tuzun, T. (2014). Are leveraged and inverse ETFs the new portfolio insurers? In *Paris December 2014 Finance Meeting EUROFIDAI-AFFI*.
- Wang, T. (2012). The price impacts of index option hedging and the anomalous weekly reversals. Available at SSRN 2413058.
- Wilmott, P., & Schönbucher, P. (2000). The feedback effect of hedging in illiquid markets. *SIAM Journal on Applied Mathematics*, 61(1), 232–272.
- Zhu, H. (2014). Do dark pools harm price discovery? *Review of Financial Studies*, 27(3), 747–789.

Appendix A

Chapter 2 Proofs

A.1 Bid and Ask Prices

We can substitute $\mathbb{E}_m^i[v] = p_i$ since v takes the value of either 1 or 0, where \mathbb{E}_m^i is the expectation with respect to the market maker's information set. After distributing the expectation operator, Equations 2.1 and 2.2 may be written as

$$\left(\left(\frac{\beta}{2} - \frac{1}{2} \right) (a_i - p_i) + \frac{\beta (p_i - a_i + c (a_i - b_i))}{2} \right) (\alpha - 1) + \alpha p_i (a_i - 1) \quad (\text{A.1})$$

$$\alpha b_i (p_i - 1) - \left(\left(\frac{\beta}{2} - \frac{1}{2} \right) (b_i - p_i) - \frac{\beta (b_i - p_i + c (a_i - b_i))}{2} \right) (\alpha - 1) \quad (\text{A.2})$$

and solving yields the formulas in 2.3.

A.2 Proof of Proposition 2.3.1

Differentiating $s_i(\beta)$ gives

$$\frac{8 \alpha c p_i (1 - \alpha)(1 - p_i)}{(4 \alpha^2 p_i^2 - 4 \alpha^2 p_i + \alpha^2 - 2 \beta c \alpha + 2 \beta c - 1)^2} \quad (\text{A.3})$$

which is greater than zero; the denominator is greater than zero as it squared and the numerator because all variables are between zero and one.

A.3 Proof of Proposition 2.3.2

This proof uses the technique of Glosten and Milgrom (1985) and Diamond and Verrecchia (1987). Let $\kappa_{\tilde{v}}^{A_i}$ be the conditional probability of the action A_i given $\tilde{v} = \bar{v}$ or $\tilde{v} = \underline{v}$ and let $\phi_i := \log\left(\frac{\kappa_{\bar{v}}^{A_i}}{\kappa_{\underline{v}}^{A_i}}\right) = \log(\xi_i)$. The values of $\kappa_{\tilde{v}}^{A_i}$ are given in Table A.1. Given the relation

$$\log(\lambda_i) = \sum_{i=n} \phi_i, \quad (\text{A.4})$$

Table A.1: **Conditional Probabilities.** This table shows the probabilities of each of the probabilities $\kappa_{\tilde{v}}$, which are conditional on the asset's true value.

Action A_i	$\kappa_{\tilde{v}}(A_i)$	$\kappa_{\underline{v}}(A_i)$
Institutional buy I_{buy}	$\frac{2\alpha+(1-\alpha)(1-\beta)}{2}$	$\frac{(1-\alpha)(1-\beta)}{2}$
Institutional sell I_{sell}	$\frac{(1-\alpha)(1-\beta)}{2}$	$\frac{2\alpha+(1-\alpha)(1-\beta)}{2}$
Retail buy R_{buy}	$\frac{\beta(1-\alpha)}{2}$	$\frac{\beta(1-\alpha)}{2}$
Retail sell R_{sell}	$\frac{\beta(1-\alpha)}{2}$	$\frac{\beta(1-\alpha)}{2}$

the strategy is to apply Wald's lemma to the two cases of $\tilde{v} = \underline{v}$ and $\tilde{v} = \bar{v}$ separately.

Without loss of generality consider the hypothesis $\tilde{v} = \bar{v}$. In this case the market maker is able to confirm this hypothesis with some certainty if λ_i reaches the upper boundary H and reject the hypothesis with some certainty if λ_i reaches the lower boundary L . Let N be the stopping time corresponding to λ_i reaching either H or L and applying Wald's lemma to Equation A.4¹,

$$\mathbb{E}_{\tilde{v}}[\log(\lambda_i)] = \mathbb{E}_{\tilde{v}}[N]\mathbb{E}_{\tilde{v}}[\phi_1] \quad (\text{A.5})$$

where $\mathbb{E}_{\tilde{v}}$ denotes the expectation conditional on $\tilde{v} = \bar{v}$.

The expected hitting time is finite because the probability a path that stays within the two boundaries is zero. The probability of a path that stays in state k forever is the probability of failing to ever leave the state $\prod_{n=1}^{\infty}(1-c) = 0$ for $c > 0$. An inductive argument can be used to show that this is true for any set of neighbouring states defined by two boundaries.

Consider $\mathbb{E}_{\tilde{v}}[\log(\lambda_i)]$. Given $N = \tilde{N}$, $\log(\lambda_{\tilde{N}}) = \log(H)$ or $\log(\lambda_{\tilde{N}}) = \log(L)$. The probability of either value can be calculated path-wise over the process λ_i . Let the set of paths of λ_i that reach H first given $N = \tilde{N}$ be $\{U_j\}$ and the set of paths that reach L first given $N = \tilde{N}$ be $\{L_j\}$. Using the above notation,

$$\mathbb{E}_{\tilde{v}}[\log(\lambda_i) | N = \tilde{N}] = \sum_j \Pr(U_j | \tilde{v} = \bar{v}) \log(H) + \sum_j \Pr(L_j | \tilde{v} = \bar{v}) \log(L). \quad (\text{A.6})$$

The probability of any particular path U_j given $\tilde{v} = \bar{v}$ is the product of probabilities of the elements in the sequence of corresponding actions $\{A_i^j\}$ given $\tilde{v} = \bar{v}$;

$$\Pr_{\tilde{v}}(U_j | \tilde{v} = \bar{v}) = \prod_i \Pr(A_i^j | \tilde{v} = \bar{v}) = \prod_i \kappa_{\tilde{v}}(A_i^j) = \prod_i \frac{\kappa_{\tilde{v}}(A_i^j)}{\kappa_{\underline{v}}(A_i^j)} \kappa_{\underline{v}}(A_i^j) \quad (\text{A.7})$$

$$= \prod_i \lambda_i \kappa_{\underline{v}}(A_i^j) = H \prod_i \kappa_{\underline{v}}(A_i^j) = H \Pr(U_j | \tilde{v} = \underline{v}) \quad (\text{A.8})$$

Summing over all paths U_j ,

$$\Pr(\lambda_{\tilde{N}} = H | \tilde{v} = \bar{v}) = H \sum_j \prod_i \kappa_{\underline{v}}(A_i^j) = H \Pr(\lambda_{\tilde{N}} = H | \tilde{v} = \underline{v}). \quad (\text{A.9})$$

¹A statement of Wald's lemma can be found here: https://en.wikipedia.org/wiki/Wald%27s_equation.

A similar argument gives the relation $\Pr(\lambda_{\bar{N}} = L \mid \tilde{\nu} = \bar{\nu}) = L \Pr(\lambda_{\bar{N}} = L \mid \tilde{\nu} = \underline{\nu})$ which can be used to find $\Pr(\lambda_{\bar{N}} = H \mid \tilde{\nu} = \bar{\nu})$;

$$\Pr(\lambda_{\bar{N}} = H \mid \tilde{\nu} = \bar{\nu}) = H \Pr(\lambda_{\bar{N}} = H \mid \tilde{\nu} = \underline{\nu}) \quad (\text{A.10})$$

$$= H(1 - \Pr(\lambda_{\bar{N}} = L \mid \tilde{\nu} = \underline{\nu}))$$

$$= H\left(1 - \frac{1}{L} \Pr(\lambda_{\bar{N}} = L \mid \tilde{\nu} = \bar{\nu})\right) \quad (\text{A.11})$$

$$= H\left(1 - \frac{1}{L}(1 - \Pr(\lambda_{\bar{N}} = H \mid \tilde{\nu} = \bar{\nu}))\right)$$

$$\Rightarrow \Pr(\lambda_{\bar{N}} = H \mid \tilde{\nu} = \bar{\nu}) = \frac{H(1-L)}{H-L}.$$

The remaining probabilities can be computed in a similar way as

$$\Pr(\lambda_{\bar{N}} = L \mid \tilde{\nu} = \bar{\nu}) = \frac{L(H-1)}{H-L}, \quad (\text{A.12})$$

$$\Pr(\lambda_{\bar{N}} = H \mid \tilde{\nu} = \underline{\nu}) = \frac{1-L}{H-L},$$

$$\Pr(\lambda_{\bar{N}} = L \mid \tilde{\nu} = \underline{\nu}) = \frac{H-1}{H-L}.$$

The above relations allow $\mathbb{E}_{\tilde{\nu}}[\log(\lambda_i)]$ to be explicitly written as

$$\frac{H(1-L)}{H-L} \log(H) + \frac{L(H-1)}{H-L} \log(L). \quad (\text{A.13})$$

Next, consider $\mathbb{E}_{\tilde{\nu}}[\phi_1]$. This can be calculated using Table A.1 and Equation 2.6;

$$\begin{aligned} \mathbb{E}_{\tilde{\nu}}[\phi_1] &= \sum_{A \in \mathcal{A}} \Pr(A \mid \tilde{\nu} = \bar{\nu}) \log\left(\frac{\kappa_{\bar{\nu}}(A)}{\kappa_{\underline{\nu}}(A)}\right) = \sum_{A \in \mathcal{A}} \kappa_{\bar{\nu}}(A) \log\left(\frac{\kappa_{\bar{\nu}}(A)}{\kappa_{\underline{\nu}}(A)}\right) \\ &= \alpha \log\left(\frac{2\alpha + (1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)}\right). \end{aligned} \quad (\text{A.14})$$

Substituting Equations A.13 and A.14 into Equation A.5 and rearranging yields

$$\mathbb{E}_{\bar{\nu}}[N] = \frac{\frac{H(1-L)}{H-L} \log(H) + \frac{L(H-1)}{H-L} \log(L)}{\alpha \log\left(\frac{2\alpha + (1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)}\right)}. \quad (\text{A.15})$$

Similarly,

$$\mathbb{E}_{\underline{\nu}}[N] = \frac{\frac{(1-L)}{H-L} \log(H) + \frac{(H-1)}{H-L} \log(L)}{\alpha \log\left(\frac{2\alpha + (1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)}\right)}. \quad (\text{A.16})$$

Combining the above into one ratio test to accept $\nu = \bar{\nu}$ and reject $\nu = \underline{\nu}$ simultaneously or accept $\nu = \underline{\nu}$ and reject $\nu = \bar{\nu}$ simultaneously by equating $\mathbb{E}_{\underline{\nu}}[N]$ and $\mathbb{E}_{\bar{\nu}}[N]$ yields $L = \frac{1}{H}$ and the desired result.

A.4 Alternate Proof of Proposition 2.3.2

We can also prove the Proposition 2.3.1 using a difference equation. Let E_k be the expected number of rounds of trading until the log-likelihood $\log(\lambda_i)$ hits either the upper boundary $\log(H)$ or the lower boundary $\log(1/H)$ beginning from state k . Here, k belongs to the natural numbers corresponding to an enumeration of the set of possible states with $\log(\lambda_i) = 0$ corresponding to $k = 0$.

We have the difference equation

$$\begin{aligned} E_k &= 1 + uE_{k+1} + dE_{k-1} + (1 - u - d)E_k \\ \Rightarrow E_k &= \frac{1 + uE_{k+1} + dE_{k-1}}{u + d} \end{aligned} \quad (\text{A.17})$$

where k is the initial state, u is the probability of an upward move on the lattice, d is the probability of a downward move on the lattice, and $1 - u - d$ is therefore the probability of remaining in a given state which happens in the event of a retail trade. The general solution of the homogeneous equation is of the form $w^k + b$. Substituting into the homogeneous difference equation and solving for b yields 1 and $\frac{d}{u}$ and therefore the general solution $A + B(\frac{d}{u})^k$. To obtain a particular solution, we try the ansatz Ck . Substituting into the equation above gives $C = \frac{1}{d-u}$. So the solution of the difference equation has the form $A + B(\frac{d}{u})^k + \frac{k}{d-u}$.

Let the upper barrier in units of k be h and the lower barrier in units of k be l . We can use the boundary conditions $E_h = E_l = 0$ to find A and B . Solving for A and B yields

$$\begin{aligned} A &= \frac{l(\frac{d}{u})^h - h(\frac{d}{u})^l}{(d-u)((\frac{d}{u})^h - (\frac{d}{u})^l)} \\ B &= \frac{h-l}{(u-d)((\frac{d}{u})^h - (\frac{d}{u})^l)}. \end{aligned}$$

We are interested in E_0 , for which the solution is

$$E_0 = \frac{h(\frac{d}{u})^l - l(\frac{d}{u})^h + l - h}{(d-u)((\frac{d}{u})^h + (\frac{d}{u})^l)}.$$

Choosing the lower boundary to be $-h$, as explained above, simplifies the expectation to be

$$\begin{aligned} E_0 &= \frac{h(\frac{d}{u})^{-h} + h(\frac{d}{u})^h - 2h}{(d-u)((\frac{d}{u})^h + (\frac{d}{u})^{-h})} \\ &= \frac{h((\frac{d}{u})^h - 1)}{(d-u)((\frac{d}{u})^h + 1)}. \end{aligned} \quad (\text{A.18})$$

To make this comparable to the formula in Proof of Proposition 2.3.2, we must translate the probabilities u and d into probabilities of informed, uninformed and institutional trades, and translate the boundary h to the step sizes of the log-likelihood ratio rather than unit steps that were used to set up the difference equation. We can consider the market maker's Bayesian updates for institutional and retail trades given in Equation 2.6. Taking logs, we see that

the log-likelihood ratio is either unchanged in the event of a retail trade, or increases by $\pm \log\left(\frac{2\alpha+(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)}\right)$. Therefore, we have the relation $h \log\left(\frac{2\alpha+(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)}\right) = \log(H)$ to convert the number of steps h into units of the lattice of the log-likelihood ratio.

The difference equation in Equation A.17 gives the probability of the log-likelihood ratio remaining constant as $1 - u - d$. This corresponds to a retail trade so we have the relation $\beta = 1 - u - d$. Institutional buy and sell orders correspond to u and d respectively. The probability of an institutional buy order given $\nu = \bar{\nu}$ is given by $u = \frac{2\alpha+(1-\alpha)(1-\beta)}{2}$, and an institutional sell order given $\nu = \bar{\nu}$ by $d = \frac{(1-\alpha)(1-\beta)}{2}$. Therefore, the ratio $\frac{u}{d}$ that appears in Equation A.18 is equal to $\frac{2\alpha+(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)}$. The expression $(\frac{u}{d})^h$ in Equation A.18 becomes $(\frac{2\alpha+(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)})^{\frac{\log(H)}{\log(\frac{2\alpha+(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)})}}$. Using the change of base formula for logarithms, the expression becomes simply $\frac{1}{H}$.

Using the above simplifications, Equation A.18 becomes

$$\frac{\log(H)\left(\frac{1-H}{1+H}\right)}{\alpha \log\left(1 + \frac{2\alpha}{(1-\alpha)(1-\beta)}\right)}$$

which matches Equation 2.8.

A.5 Proof of Proposition 2.3.3

Let G_k be the probability of $\log(\lambda_i)$ hitting the upper boundary $\log(H)$ before the lower boundary $\log(\frac{1}{H})$ given the initial state is k . The difference equation below for G_k described how these probabilities evolve as the log-likelihood ratio moves from state to state:

$$G_k = uG_{k+1} + dG_{k-1} + (1 - u - d)G_k \quad (\text{A.19})$$

$$= \frac{uG_{k+1} + dG_{k-1}}{u + d}, \quad (\text{A.20})$$

The equation is homogeneous with general solution $A + B\left(\frac{u}{d}\right)^k$. The boundary conditions $G_h = 1$ and $G_l = 0$ can be used to find A and B .

$$A = -\frac{1}{\left(\frac{d}{u}\right)^{2h} - 1} \quad (\text{A.21})$$

$$B = \frac{\left(\frac{d}{u}\right)^h}{\left(\frac{d}{u}\right)^{2h} - 1}. \quad (\text{A.22})$$

We can use A and B above to find G_0

$$G_0 = \frac{1}{\left(\frac{d}{u}\right)^h + 1}.$$

Using $\left(\frac{d}{u}\right)^h = \frac{1}{H}$ we have

$$G_0 = \frac{H}{H + 1}.$$

Therefore, the probability of an error in accepting the hypothesis is $1 - \frac{H}{H+1} = \frac{1}{H+1}$.

A.6 Proof of Lemma 2.4.1

As was shown at in Equation 2.9 profits at inner price pair $\{x_l, x_h\}$ are independent of the market maker's prior p_i and are positive as long as $\alpha < 1 - 2x_l$. The market maker is competitive, so if there is a bid-ask pair that yields positive expected profits that are lower than they are for the pair $\{x_l, x_h\}$, he will adjust prices. Since expected profits at the pair $\{x_l, x_h\}$ are constant in the prior p_i , then if expected profits at another pair are increasing or decreasing in p_i there will be a point at which the market maker will prefer to quote the other pair. Consider $\{x_h, x_H\}$; the market maker's expected profit is:

$$\frac{(x_H + x_h - 1)\alpha p_i - (x_H + x_h)\frac{\alpha}{2} + \frac{x_H - x_h}{2}}{2}$$

which is linear in p_i . The expected profits from the higher price pair are computed using the same logic as for Equation 2.9 where the expectation is over the probability of an informed order being fast α , the probabilities of the informed trader submitting a market order $\bar{\rho}$ and $\underline{\rho}$, the asset's value being high p_i , and the probabilities of the noise trader's actions. Note that the expected profit is independent of $\bar{\rho}$ and $\underline{\rho}$. To see why, consider the expected profit from quoting $\{x_h, x_H\}$ given that the asset's value is high:

$$\begin{aligned} & \alpha(\bar{\rho}\frac{(x_H - 1)}{2} + (1 - \bar{\rho})\frac{(x_H - 1)}{2}) + (1 - \alpha)(\frac{x_H - 1}{4} + \frac{1 - x_h}{4}) \\ &= \frac{\frac{x_H - x_h}{2} + \alpha(\frac{x_H + x_h}{2} - 1)}{2}. \end{aligned}$$

The strategy $\bar{\rho}$ cancels because the market maker's losses are the same regardless of whether the informed trader submits a market order or a limit order; he is selling an asset worth a for the price $h_H < 1$. The same is true for the expected value given the asset's true value is low; the market maker's expected profits are independent of ρ . Similar logic holds for the lower price pair. The fact that the market maker's expected profits are independent of the informed trader's strategy is important because it means the thresholds $\bar{\theta}$ and $\underline{\theta}$ do not depend on $\bar{\rho}$ and ρ . Otherwise, we would have a free boundary problem that would likely lead to intractability.

A numerical example is shown in Figure 2.7; the market maker will prefer to quote the higher prices when his prior rises above a certain threshold $\bar{\theta}$ because expected profits for the higher price pair are increasing in his prior. The figure shows a region where expected profits from the higher price pair are greater than zero and less than for the inner price pair.

A.7 Proof of Proposition 2.4.2

The general approach of the proof is as follows: we begin by using a result in Stokey, Lucas, and Prescott (1989) to show that a Bellman equation can be applied to solve the informed trader's problem of maximizing the sequence of per-period profits. We then conjecture that the optimal strategy is independent of the state variable in the limit as H goes to infinity. We will conjecture a constant strategy function $\rho_i^*(p_i) = \rho^*$ as the solution to a particular maximization problem based on Wald's lemma, and derive the first-order condition that ρ^* must satisfy. We will then show that the solution to the Bellman equation must satisfy the same

first-order condition in the limit as H becomes infinitely large. We will conclude that when H is large, we can approximate the optimal strategy with our conjecture ρ_i^* . The more general optimization when H is finite is intractable because the state may move with different step sizes which complicates outcomes near the boundaries, and makes techniques like backward induction difficult. In taking the limit as H becomes large, we are able avoid the intractable nature of the boundaries because the state is always sufficiently far from them.

We must define the elements of the optimization problem more formally so we can follow Stokey et al. (1989) Chapter 4 to prove the equivalence of optimizing the sequence of per period profits and solving the corresponding Bellman equation. Let $f(\rho)$ be per period profits as defined in Equation 2.16 when market maker's log-likelihood ratio $\log(\frac{p_i}{1-p_i})$ is strictly within the boundaries $\log(H)$ and $-\log(H)$, and zero otherwise. Let the state variable be the market maker's prior p_i for $p_i \in [\frac{H}{H+1}, \frac{1}{H+1}]$ (the interval corresponds to the boundaries at which the market maker moves the prices) with absorbing boundaries corresponding to the log-likelihood ratio $\log(\frac{p_i}{1-p_i})$ at $\log(H)$ and $-\log(H)$; $p = \frac{H}{H+1}$ when $\log(\frac{p_i}{1-p_i})$ hits or passes $\log(H)$ and $p = \frac{1}{H+1}$ when $\log(\frac{p_i}{1-p_i})$ hits or passes $-\log(H)$.

Let the continuation values for the informed trader corresponding to state variable hitting either the upper or lower a boundary respectively be \bar{V} and \underline{V} . Let the time of hitting be N , so that $V(p_N) = \underline{V}$ or $V(p_N) = \bar{V}$ and $V(p_i) = 0$ for all $i > N$. The assumption is that the informed trader receives a one-time payment of the continuation value at the hitting time and nothing thereafter. The continuation values can be thought of as the value the beginning a new game at revised prices, either higher or lower than current prices.

The informed trader's problem is to maximize the expected sequence of per-period profits:

$$\sup_{\{0 \leq \rho_i \leq 1\}} \mathbb{E}[\sum_i f(\rho_i)] \quad (\text{A.23})$$

subject to the market maker updating the prior p_i according to the Bayesian updating rules defined in Equation 2.10, given some initial prior p_0 , and where i indexes the times for which the log-likelihood ratio is within the boundary. The associated Bellman equation is

$$V(p_i) = \sup_{0 \leq \rho_i \leq 1} \mathbb{E}[f(\rho_i) + V(p_{i+1}) | \rho_i] \quad (\text{A.24})$$

again, subject to the market maker updating the prior p_i according to the Bayesian updating rules defined in Equation 2.10.

From Stokey et al. (1989) Chapter 4, to show the equivalence of the above equations, the following must hold: for any value of the state variable, the following state as determined by the "law of motion" (in this case the Bayesian updates in Equation 2.10) is in the state space; and $\lim_{n \rightarrow \infty} \mathbb{E}[\sum_{i=1}^n f(\rho_i)]$ exists for all possible strategies and all initial values p_0 . The former is true in our case since we have absorbing boundaries at $\frac{H}{H+1}$ and $\frac{1}{H+1}$ which are part of the state space; the latter is less obvious but still straightforward. The limit exists because all per-period profits are finite and positive and, by similar logic to Proposition 2.3.2, the expected number of trades before either an upper or lower boundary is hit is finite. From Stokey et al. (1989) Chapter 4, we also need the following boundedness condition on $V(p_i)$: $\lim_{n \rightarrow \infty} V(\rho_n) = 0$ for all possible sequences $\{p_i\}$. The assumption holds because the expectation of the hitting time N is finite and by definition $V(p_i)$ for $i > N$ is 0. Given the above, we can say that a solution to

the Bellman equation, that is, a function $\rho_i(p_i)$, is unique and maximizes the expected sequence of per-period profits. The reverse also holds, that a function that maximizes the expected per-period profits also satisfies the Bellman equation.

We conjecture that the optimal strategy is the constant $\bar{\rho} = \underline{\rho} = \rho^*$. If this is true, the informed trader's profits are given by $\mathbb{E}[\sum_{n=1}^N f(\rho^*) + V(p_N)|\rho^*]$, the expected sum of per-period profits plus the continuation value, where N is the stopping time corresponding to the upper or lower thresholds at which the market maker prefers to adjust his quotes, and $V(p_N)$ is the continuation value at N , either at the high or low boundary. We can use Wald's lemma to find a guess for the form of the sum of the informed trader's per-period trading profits

$$\begin{aligned} \mathbb{E}\left[\sum_{n=1}^N f(\rho^*)\right] &= \mathbb{E}[f(\rho^*)]\mathbb{E}[N|\rho^*] \\ &= \mathbb{E}[f(\rho^*)] \frac{\mathbb{E}[h(p_i)|\rho^*]}{\mathbb{E}[g(\rho^*)|\rho^*]} \\ &:= f^*(\rho^*) \frac{h^*(p_i)}{g^*(\rho^*)} \end{aligned} \tag{A.25}$$

where $h(p_i) = \log(\frac{p_N}{1-p_N})$, the distance the log-likelihood ratio moves from a given state p_i to the upper or lower boundaries; $g(\rho^*)$ is the additive update to the market maker's log-likelihood ratio following a trade; and $f(\rho^*)$ is per-period profit. Let starred functions f^* , g^* , and h^* denote the corresponding functions under expectation given ρ^* . The first equality in Equation A.25 is possible because the assumption of a constant solution ρ^* results in independence of per-period profit the expected value of N . The second equality is due to the application of Wald's lemma, as used in the proof of Proposition 2.3.2. Putting the above together with the expected boundary condition we guess the form of the value function which we denote $V^*(p_i)$:

$$V^*(p_i) := f^*(\rho^*) \frac{h^*(p_i)}{g^*(\rho^*)} + k^*(p_i) \tag{A.26}$$

where $k^*(p_i)$ denotes the expected value to the informed trader when the market maker moves his quote, the expected continuation value. We will assume ρ^* maximizes the function in Equation A.26 and use the functional form to prove that ρ^* is an optimal strategy under certain conditions.

Note that only h^* and k^* are functions of the state p_i . This is because expected per-period profits f^* and the expected first step g^* are independent of the market maker's prior, while the expected value of the log-likelihood ratio h^* depends on the distance of the prior p_i to the boundary H , as does the expected continuation value k^* . Also note that h^* and k^* are independent of ρ^* . The reasoning has been shown for an analogous case in the proof of Proposition 2.3.2 where the numerator is only a function of H . We will give details on the functional forms further below.

To prove the optimal strategy is the constant ρ^* , it is sufficient to show that if ρ^* maximizes

the value function V^* conjectured above, it simultaneously solves the Bellman equation,

$$\begin{aligned}\hat{\rho} &= \arg \max_{0 \leq \rho \leq 1} \mathbb{E}[f(\rho) + V^*(p_{i+1})|\rho] \\ &= \arg \max_{0 \leq \rho \leq 1} \mathbb{E}[f(\rho) + f^*(\rho^*) \frac{h^*(p_{i+1}(\rho))}{g^*(\rho^*)} + k^*(p_{i+1}(\rho))|\rho]\end{aligned}\tag{A.27}$$

We have written h^* and k^* as functions of ρ to show the dependence which is due to the fact that the state p_{i+1} is not taken as given, but depends on the choice of ρ . Towards showing the equivalence of Equation A.26 and Equation A.27, it is useful to write the first-order condition satisfied by ρ^* in Equation A.26. Differentiating with respect to ρ^* and using the product rule,

$$\begin{aligned}D_{\rho^*}[f^*(\rho^*) \frac{h^*(p_i)}{g^*(\rho^*)} + k^*(p_i)] \\ &= D_{\rho^*}[f^*(\rho^*)] \frac{h^*(p_i)}{g^*(\rho^*)} - f^*(\rho^*) \frac{h^*(p_i)}{g^{*2}(\rho^*)} D_{\rho^*}[g^*(\rho^*)] = 0 \\ &\Rightarrow D_{\rho^*}[f^*(\rho^*)] - \frac{f^*(\rho^*)}{g^*(\rho^*)} D_{\rho^*}[g^*(\rho^*)] = 0.\end{aligned}\tag{A.28}$$

where D is the differential operator. We have used the fact that k^* and h^* are independent of ρ^* . We will show that the first-order condition necessary to solve Equation A.27 is the same as the first-order condition above in Equation A.28 when H is large. Consider the first-order condition for Equation A.27. After distributing the expectation operator we have

$$\begin{aligned}\mathbb{E}[V(p_i)] &= \mathbb{E}[f(\rho)] + \mathbb{E}[f^*(\rho^*) \frac{h^*(p_{i+1}(\rho))}{g^*(\rho^*)} + k^*(p_{i+1}(\rho))|\rho] \\ &= f^*(\rho) + \frac{f^*(\rho^*)}{g^*(\rho^*)} \mathbb{E}[h^*(p_{i+1}(\rho)) + g^*(\rho^*)k^*(p_{i+1}(\rho))|\rho].\end{aligned}\tag{A.29}$$

In the equalities above we have used the definition of f^* and the fact that $f^*(\rho^*)$ and $g^*(\rho^*)$ are independent of p_{i+1} and therefore ρ . Differentiating with respect to ρ yields

$$D_{\rho}[f^*(\rho)] + \frac{f^*(\rho^*)}{g^*(\rho^*)} D_{\rho}[\mathbb{E}[h^*(p_{i+1}(\rho)) + g^*(\rho^*)k^*(p_{i+1}(\rho))|\rho]].\tag{A.30}$$

Comparing Equation A.30 with Equation A.28, we see that the terms in the first terms are equivalent, as are the coefficients for second. We will show $D_{\rho}[\mathbb{E}[h^*(p_{i+1}(\rho)) + k^*(p_{i+1}(\rho))g^*(\rho^*)|\rho]]$ and $-D_{\rho^*}[g^*(\rho^*)]$ are equal when H is large to complete the proof.

First, consider $-D_{\rho^*}[g^*(\rho^*)] = -D_{\rho^*}[\mathbb{E}[g(\rho^*)|\rho^*]]$. g^* is the expected ‘‘first step’’ taken by the log-likelihood ratio. Since each random variable is discrete, the expectation takes the form of a sum of probabilities times corresponding update factors to the log-likelihood ratio $\xi_j(\rho^*)$ over possible outcomes $LBMB$ etc. Let the probabilities be $\eta_j(\rho^*)$. Expanding,

$$\begin{aligned}-D_{\rho^*}[g^*(\rho^*)] &= -D_{\rho^*}[\mathbb{E}[g(\rho^*)]] \\ &= -D_{\rho^*}[\sum_j \eta_j(\rho^*) \log(\xi_j(\rho^*))] \\ &= -\sum_j D_{\rho^*}[\eta_j(\rho^*) \log(\xi_j(\rho^*))] \\ &= -\sum_j (D_{\rho^*}[\eta_j(\rho^*)] \log(\xi_j(\rho^*)) + \frac{\eta_j(\rho^*)}{\xi_j(\rho^*)} D_{\rho^*}[\xi_j(\rho^*)]).\end{aligned}\tag{A.31}$$

Second, consider $D_\rho[\mathbb{E}[h^*(p_{i+1}(\rho)) + g^*(\rho^*)k^*(p_{i+1}(\rho))|\rho]]$. We will need details on the form of $h^*(p_{i+1}(\rho))$ and $k^*(p_{i+1}(\rho))$. h^* is the expected value of the log-likelihood ratio and k^* is the expected continuation value when the market maker revises his prices, either up or down. We can rewrite these expectations using the probability that the log-likelihood ratio crosses the upper boundary before the lower. Since our guess of the value function was based on the constant strategy ρ^* , the probability of revising prices up rather than down is analogous to the formula proved in Proposition 2.3.3 and is a function of the state variable alone (the market maker's prior) and not the strategy ρ^* . Recall from the proof of Proposition 2.3.2, the probability of the market maker's likelihood ratio λ_i hitting the upper boundary H first, beginning from the state of $\lambda_i = 1$, is $\frac{H(1-L)}{H-L}$ (using the notation of that proof). Therefore, after the market maker revises his likelihood ratio λ_i by multiplying by a factor $\xi_i(\rho)$, the probability becomes $\frac{\frac{H}{\xi_i(\rho)}(1-\xi_i(\rho)L)}{\frac{H}{\xi_i(\rho)} - \xi_i(\rho)L}$ which simplifies to $\frac{H}{H+\xi_i(\rho)}$ using $H = \frac{1}{L}$. Note that each ξ_i depend on ρ .

We will use the probabilities of quoted prices being revised up, $\frac{H}{H+\xi_i(\rho)}$, and down, $\frac{\xi_i(\rho)}{H+\xi_i(\rho)}$, to rewrite $h^*(p_{i+1}(\rho)) + g^*(\rho^*)k^*(p_{i+1}(\rho))$ as a binary outcome. The expected values of the log-likelihood ratio $\mathbb{E}[\log(\frac{p_N}{1-p_N})|\frac{p_N}{1-p_N} \geq H]$ when a boundary is crossed are not straightforward to calculate. In the proof of Proposition 2.3.2 we were able to use that fact that $\mathbb{E}[\log(\frac{p_N}{1-p_N})|\frac{p_N}{1-p_N} \geq H] = \log(H)$ because in the model developed in Section 2.3, the log-likelihood ratio moved on a lattice; since $\log(H)$ was chosen on the lattice, the value of the log-likelihood ratio was exactly the boundary at its hitting time. However, in model in Section 2.4 the log-likelihood ratio is not confined to a simple lattice. The log-likelihood ratio can take any value of

$$\begin{aligned} w_0 \log(LBMB) + w_1 \log(QBMS) + w_2 \log(QSMB) \\ + w_3 \log(QSMS) + w_4 \log(QSR) + w_5 \log(QBR) \end{aligned} \quad (\text{A.32})$$

for any value of $w_0, \dots, w_5 \in \mathbb{Q}$. Since each of the Bayesian update factors $LBMB$ etc. are real functions of ρ , it is not necessarily possible to constrain $\log(H)$ to be both the boundary and the value of the log-likelihood ratio. The process will almost surely overshoot the boundary.

However, when H is large the problem of overshooting becomes negligible. The expected value of the log-likelihood ratio at its hitting time is at most greater than a boundary by the maximum of the summands written in A.32. Let $\mathbb{E}[\log(\frac{p_N}{1-p_N})|\frac{p_N}{1-p_N} \geq H] = \log(H) + d_0$ and $\mathbb{E}[\log(\frac{p_N}{1-p_N})|\frac{p_N}{1-p_N} \leq \frac{1}{H}] = -\log(H) - d_1$ where d_0 and d_1 are finite adjustments in the expectation needed to account for the variety of values taken on by the log-likelihood ratio at its hitting time. We will see that d_0 and d_1 are negligible when H is large.

The numerator derived from Wald's lemma $h^*(p_j)$ represents the expected distance the log-likelihood ratio moves from some initial state p_j which we had previously taken to be $\frac{1}{2}$ in the proof of Proposition 2.3.2. We will need adjust the formula to incorporate the possible future states p_{j+1} after a trade to use the formula appropriately. After an update to the log-likelihood ratio by adding a factor $\log(\xi_j)$, the distance from the upper boundary has decreased by that amount and so the relevant term that should enter into the value function at p_{i+1} is $\log(H) - \log(\xi_j(\rho))$. The values \bar{V} and \underline{V} in the expression k^* are the high and low continuation

values. We can now write $h^*(p_{i+1}(\rho)) + g^*(\rho^*)k^*(p_{i+1}(\rho))$ in terms of H and $\xi_j(\rho)$:

$$\begin{aligned}
& h^*(p_{i+1}(\rho)) + g^*(\rho^*)k^*(p_{i+1}(\rho)) \\
&= \frac{H}{H + \xi_j(\rho)} \left(\log(H) - \log(\xi_j(\rho)) + d_0 + g^*(\rho^*)\bar{V} \right) \\
&\quad + \frac{\xi_j(\rho)}{H + \xi_j(\rho)} \left(-\log(H) + \log(\xi_j(\rho)) - d_1 + g^*(\rho^*)\underline{V} \right) \\
&= \frac{Hd_0 - d_1\xi_j(\rho) + g^*(\rho^*)(H\bar{V} + \xi_j(\rho)\underline{V}) + (H - \xi_j(\rho)) \log\left(\frac{H}{\xi_j(\rho)}\right)}{H + \xi_j(\rho)}. \tag{A.33}
\end{aligned}$$

Note that the continuation values \bar{V} and \underline{V} appear grouped together in Equation A.33 as $\frac{H\bar{V} + \xi_j(\rho)\underline{V}}{H + \xi_j(\rho)}$, which is proportional to the value function. The variables \bar{V} and \underline{V} have an intuitive impact on the value function: when \underline{V} is large relative to \bar{V} , the value function is more sensitive to the informational impact a given trade because of the Bayesian update factor $\xi_j(\rho)$; future profits diminish quickly when the state variable moves closer to the smaller continuation value \bar{V} . Therefore, the informed trader puts more weight on the informational impact of trades rather than current profit the larger $\underline{V} - \bar{V}$ is. Indeed, the term $\underline{V} - \bar{V}$ appears in the derivative of the value function in Equation A.34.

Rewriting $D_\rho[\mathbb{E}[h^*(p_{i+1}(\rho)) + g^*(\rho^*)k^*(p_{i+1}(\rho))|\rho]]$ using Equation A.33 we have:

$$\begin{aligned}
& D_\rho[\mathbb{E}[h^*(p_{i+1}(\rho)) + g^*(\rho^*)k^*(p_{i+1}(\rho))|\rho]] \\
&= D_\rho \left[\sum_j \eta_j(\rho) \left(\frac{Hd_0 - d_1\xi_j(\rho) + g^*(\rho^*)(H\bar{V} + \xi_j(\rho)\underline{V}) + (H - \xi_j(\rho)) \log\left(\frac{H}{\xi_j(\rho)}\right)}{H + \xi_j(\rho)} \right) \right] \\
&= \sum_j D_\rho[\eta_j(\rho)] \left(\frac{Hd_0 - d_1\xi_j(\rho) + g^*(\rho^*)(H\bar{V} + \xi_j(\rho)\underline{V}) + (H - \xi_j(\rho)) \log\left(\frac{H}{\xi_j(\rho)}\right)}{H + \xi_j(\rho)} \right) \\
&\quad + \eta_j(\rho) \left(\frac{-H\xi_j(\rho)(d_0 + d_1 - g^*(\rho^*)(\underline{V} - \bar{V})) - H^2 + 2H\xi_j(\rho) \log\left(\frac{H}{\xi_j(\rho)}\right) + \xi_j(\rho)^2}{\xi_j(\rho)(H + \xi_j(\rho))^2} \right) D_\rho[\xi_j(\rho)] \tag{A.34}
\end{aligned}$$

where we have used the product rule over $\eta_j(\rho)$ and the expression in large parentheses, then the chain rule on the expression in large parentheses.

We will now take the limit as H goes to infinity. Consider the derivative of each of the $\eta_j(\rho)$. The expectation over the $\eta_j(\rho)$ has the form

$$\alpha(\rho\left(\frac{A}{2} + \frac{B}{2}\right) + (1 - \rho)((1 - \beta)\left(\frac{C}{2} + \frac{D}{2}\right) + \beta(E))) \tag{A.35}$$

which comes from the probability of the informed trader's order being accepted α ; the probability of the informed trader submitting a market order ρ ; the probability of a retail market order β ; and the probabilities of uninformed buy and sell limit and market orders, each of which are $\frac{1}{2}$. The variables A through E represent possible outcomes in the expectation corresponding to *LBMB*, *QBMS* etc. The derivative with respect to ρ is $\alpha\left(\frac{A}{2} + \frac{B}{2} - (1 - \beta)\left(\frac{C}{2} + \frac{D}{2}\right) - \beta(E)\right)$. Notice that is if $A = B = C = D = E$ then the derivative is zero. Therefore any common additive

component over j of $\frac{Hd_0 - d_1\xi_j(\rho) + g^*(\rho^*)(H\bar{V} + \xi_j(\rho)\underline{V}) + (H - \xi_j(\rho))\log(\frac{H}{\xi_j(\rho)})}{H + \xi_j(\rho)}$ over j will cancel. Reorganizing the expression:

$$\begin{aligned} & \frac{Hd_0 - d_1\xi_j(\rho) + g^*(\rho^*)(H\bar{V} + \xi_j(\rho)\underline{V}) + (H - \xi_j(\rho))\log(\frac{H}{\xi_j(\rho)})}{H + \xi_j(\rho)} \\ &= \frac{Hd_0 + g^*(\rho^*)H\bar{V}}{H + \xi_j(\rho)} + \frac{-d_1\xi_j(\rho) + g^*(\rho^*)\xi_j(\rho)\underline{V} - \xi_j(\rho)\log(\frac{H}{\xi_j(\rho)})}{H + \xi_j(\rho)} \\ & \quad - \frac{H\log(\xi_j(\rho))}{H + \xi_j(\rho)} + \frac{H\log(H)}{H + \xi_j(\rho)}. \end{aligned} \tag{A.36}$$

Consider each summand above separately: $\frac{Hd_0 + g^*(\rho^*)H\bar{V}}{H + \xi_j(\rho)}$ becomes $d_0 + g^*(\rho^*)\bar{V}$ as H goes to infinity which is independent of j and therefore cancels as a common additive component over j ; $\frac{-d_1\xi_j(\rho) + g^*(\rho^*)\xi_j(\rho)\underline{V} - \xi_j(\rho)\log(\frac{H}{\xi_j(\rho)})}{H + \xi_j(\rho)}$ goes to zero as H goes to infinity; and $\frac{H\log(\xi_j(\rho))}{H + \xi_j(\rho)}$ goes to $\log(\xi_j(\rho))$ as H goes to infinity. Finally, we will show that $\frac{H\log(H)}{H + \xi_j(\rho)}$ cancels as a common additive component over j as H goes to infinity. As an illustration, consider $\frac{H\log(H)}{H + \xi_1(\rho)} - \frac{H\log(H)}{H + \xi_2(\rho)}$:

$$\begin{aligned} & \frac{H\log(H)}{H + \xi_1(\rho)} - \frac{H\log(H)}{H + \xi_2(\rho)} \\ &= \frac{\log(H)(1 + \frac{\xi_2(\rho)}{H}) - \log(H)(1 + \frac{\xi_1(\rho)}{H})}{(1 + \frac{\xi_1(\rho)}{H})(1 + \frac{\xi_2(\rho)}{H})} \\ &= \frac{\log(H)\frac{\xi_2(\rho)}{H} - \log(H)\frac{\xi_1(\rho)}{H}}{(1 + \frac{\xi_1(\rho)}{H})(1 + \frac{\xi_2(\rho)}{H})}, \end{aligned} \tag{A.37}$$

which goes to zero as H goes to infinity. The mechanics of Equation A.37 are the similar in the sum $\alpha(\frac{A}{2} + \frac{B}{2} - (1 - \beta)(\frac{C}{2} + \frac{D}{2}) - \beta(E))$ if A through E are replaced by $\frac{H\log(H)}{H + \xi_j(\rho)}$ for j equal to 1 through 5; the $\log(H)$ terms cancel. Putting the above together, we have in the limit as H goes to infinity,

$$\begin{aligned} & \lim_{H \rightarrow +\infty} \sum_j D_\rho[\eta_j(\rho)] \left(\frac{Hd_0 - d_1\xi_j(\rho) + g^*(\rho^*)(H\bar{V} + \xi_j(\rho)\underline{V}) + (H - \xi_j(\rho))\log(\frac{H}{\xi_j(\rho)})}{H + \xi_j(\rho)} \right) \\ &= - \sum_j D_\rho[\eta_j(\rho)] \log(\xi_j(\rho)). \end{aligned} \tag{A.38}$$

Taking the limit of $\frac{-H\xi_j(\rho)(d_0 + d_1 + g^*(\rho^*)(\bar{V} - \underline{V})) - H^2 + 2H\xi_j(\rho)\log(\frac{H}{\xi_j(\rho)}) + \xi_j(\rho)^2}{\xi_j(\rho)(H + \xi_j(\rho))^2}$ as H goes to infinity gives

$\frac{1}{\xi_j(\rho)}$. Substituting into and Equation A.38 and into Equation A.34, we have

$$\begin{aligned} & \lim_{H \rightarrow +\infty} \sum_j D_\rho[\eta_j(\rho)] \left(\frac{Hd_0 - d_1\xi_j(\rho) + g^*(\rho^*)(H\bar{V} + \xi_j(\rho)V) + (H - \xi_j(\rho)) \log\left(\frac{H}{\xi_j(\rho)}\right)}{H + \xi_j(\rho)} \right) \\ & + \eta_j(\rho) \left(\frac{-H\xi_j(\rho)(d_0 + d_1 + g^*(\rho^*)(\bar{V} - V)) - H^2 + 2H\xi_j(\rho) \log\left(\frac{H}{\xi_j(\rho)}\right) + \xi_j(\rho)^2}{\xi_j(\rho)(H + \xi_j(\rho))^2} \right) D_\rho[\xi_j(\rho)] \\ & = - \sum_j \left(D_\rho[\eta_j(\rho)] \log(\xi_j(\rho)) + \frac{\eta_j(\rho)}{\xi_j(\rho)} D_\rho[\xi_j(\rho)] \right) \end{aligned} \quad (\text{A.39})$$

which is equivalent to Equation A.31.

Summarizing the proof to this point, we conjectured the value function is $V^*(p_i) := f^*(\rho^*) \frac{h^*(p_i)}{g^*(\rho^*)} + k^*(p_i)$ in Equation A.26 and showed that the first-order condition for optimality is

$$D_{\rho^*}[f^*(\rho^*)] + \frac{f^*(\rho^*)}{g^*(\rho^*)} \sum_j \left(D_{\rho^*}[\eta_j(\rho^*)] \log(\xi_j(\rho^*)) + \frac{\eta_j(\rho^*)}{\xi_j(\rho^*)} D_{\rho^*}[\xi_j(\rho^*)] \right) = 0 \quad (\text{A.40})$$

using Equations A.28 and A.31. Substituting $V^*(p_i)$ into the Bellman equation in Equation A.27, we showed that as H goes to infinity, the first-order condition for optimality is

$$D_\rho[f^*(\rho)] + \frac{f^*(\rho^*)}{g^*(\rho^*)} \sum_j \left(D_\rho[\eta_j(\rho)] \log(\xi_j(\rho)) + \frac{\eta_j(\rho)}{\xi_j(\rho)} D_\rho[\xi_j(\rho)] \right) = 0 \quad (\text{A.41})$$

using Equations A.30 and the limits derived above. By definition, ρ^* is a root of Equation A.40 and therefore is also a root of Equation A.41. We have shown that when H goes to infinity, the conjectured strategy ρ^* satisfies the first-order condition for the Bellman equation. We can conclude that when H is large, the behaviour of the value function is approximated by $V^*(p_i)$ and the optimal strategy is approximated by ρ^* .

A.8 Proof of Proposition 2.4.5

The derivative of the informed trader's strategy, shown Equation 2.17, with respect to β is:

$$\frac{\beta + 2x_l - 5\beta x_l - 2}{2(\beta - 2)^2(-2\beta^2 x_l^2 + 2\beta^2 x_l - 7\beta x_l^2 - \beta + 10x_l^2 - 4x_l + 2)^{1/2}}.$$

Since the denominator is strictly positive, it suffices to show the numerator is negative:

$$\begin{aligned} & \beta + 2x_l - 5\beta x_l - 2 \\ & < \beta + 2x_l - \beta x_l - 2 \\ & = (\beta - 2)(1 - x_l) \\ & < 0 \end{aligned} \quad (\text{A.42})$$

where the final inequality is due to the fact that $\beta < 2$ and $x_l < 1$.

Appendix B

Chapter 3 Proofs

B.1 Steady States

Let A_n and B_n be unobserved random walks; $A_n = A_{n-1} + \eta_n$ and $B_n = B_{n-1} + \zeta_n$ where η_n and ζ_n are independent $\mathcal{N}(0, v_A^2)$ and $\mathcal{N}(0, v_B^2)$ respectively. Observations of the noisy sum $C_n = A_n + B_n + U_n$ are made for each n where U_n is a $\mathcal{N}(0, v_U^2)$ independent normal white noise. Let Σ_n be the covariance matrix of the joint Bayesian prior probability distribution (we sometimes refer to this distribution as a belief) of A_n , B_n , and C_n :¹

$$\Sigma_n = \begin{bmatrix} (\sigma_n^A)^2 & \sigma_n^{AB} & \sigma_n^{AC} \\ \sigma_n^{AB} & (\sigma_n^B)^2 & \sigma_n^{BC} \\ \sigma_n^{CA} & \sigma_n^{CB} & (\sigma_n^C)^2 \end{bmatrix}.$$

Let Υ_n be the means of the beliefs about A_n , B_n , and C_n ;

$$\Upsilon_n = \begin{bmatrix} \mu_n^A \\ \mu_n^B \\ \mu_n^C \end{bmatrix} \quad (\text{B.1})$$

We now specify the timing of the iterative process defined by the model for which we write the corresponding step in parentheses. For each $n > 0$ the following occurs:

1. C_n is observed (the market maker observes the aggregate order flow);
2. The means and variances of the beliefs are conditioned on the observation of C_n (the market maker updates belief). Call the intermediate mean and variance Σ_n^+ and Υ_n^+ .
3. A_n , B_n and C_n evolve according to the random-walk innovations η_n , ζ_n and white noise U_n (random innovations in the price and inventory processes and noise trading are realized before the next order flow);
4. and as a result the variances of the beliefs increase by v_A^2 , v_B^2 , and v_U^2 respectively (the market maker accounts for innovations in the price and inventory processes and noise trading before observing the next order flow).

¹Note that even though the two random walks are independent, beliefs of A_n and B_n about them are not necessarily independent.

Since only the noisy sum of A_n and B_n is observed each period, the variances of the beliefs about A_n and B_n increase each period as the random walks evolve. Indeed, fewer variables are observed than evolve randomly each period causing a steady decay in the information about A_n and B_n . The variances of the beliefs about A_n and B_n will increase linearly in n and their covariance will decrease linearly with n . However, the variance of the belief about the *sum* of A_n and B_n will approach a constant which will be referred to as the steady state. We proceed below by showing a steady state exists by demonstrating one.

In step 2) means and variances are updated by conditioning the multivariate normal distribution on an observation of C_n . Define the partitions of Σ and Υ as follows:

$$\Sigma_n = \left[\begin{array}{c|c} \Sigma_n^{1,1} & \Sigma_n^{1,2} \\ \hline \Sigma_n^{2,1} & \Sigma_n^{2,2} \end{array} \right] := \left[\begin{array}{cc|c} (\sigma_n^A)^2 & \sigma_n^{AB} & \sigma_n^{AC} \\ \sigma_n^{AB} & (\sigma_n^B)^2 & \sigma_n^{BC} \\ \hline \sigma_n^{CA} & \sigma_n^{CB} & (\sigma_n^C)^2 \end{array} \right]$$

and

$$\Upsilon_n = \left[\begin{array}{c} \Upsilon_n^1 \\ \hline \Upsilon_n^2 \end{array} \right] := \left[\begin{array}{c} \mu_n^A \\ \hline \mu_n^B \\ \hline \mu_n^C \end{array} \right].$$

Upon observing C_n , the bivariate normal distribution of A_n and B_n conditional on C_n is distributed $\mathcal{N}(\Upsilon_n^{+1}, \Sigma_n^{+1,1})$:

$$\Sigma_n^{+1,1} = \Sigma_n^{1,1} - \Sigma_n^{1,2}(\Sigma_n^{2,2})^{-1}\Sigma_n^{2,1} \quad (\text{B.2})$$

$$\Upsilon_n^{+1} = \Upsilon_n^1 + \Sigma_n^{1,2}(\Sigma_n^{2,2})^{-1}(C_n - \Upsilon_n^2) \quad (\text{B.3})$$

The two elements of $\Sigma_n^{1,2}(\Sigma_n^{2,2})^{-1}$ define two linear ‘‘impact coefficients’’ λ_A and λ_B . The mean of the beliefs about A and B are updated by adding the unexpected component of the observation of C times λ_A and λ_B . We will show that λ_A and λ_B , will also approach a steady state. These are closely related to the object of interest in the model, the price impact coefficient.

In order to update the covariance matrix Σ_n^+ to Σ_{n+1} the covariances of the beliefs about A_{n+1} and B_{n+1} with C_{n+1} must be calculated. Without loss of generality consider the covariance of A_{n+1} with C_{n+1} . C_{n+1} is the sum of three components so the covariances may be broken down component-wise: the covariance of A_{n+1} and A_{n+1} is the variance of A_{n+1} , $(\sigma_n^A)^2$; the covariance of A_{n+1} and B_{n+1} is σ_n^{AB} ; and the covariance of A_{n+1} and U_{n+1} is 0 since they are independent. The covariance of A_{n+1} and C_{n+1} is therefore $(\sigma_n^A)^2 + \sigma_n^{AB}$ with a similar rule holding for the covariance of B_{n+1} and C_{n+1} .

Let the function that updates Σ_n to Σ_{n+1} be denoted by Φ , defined by

$$\Phi(\Sigma_n) = \left[\begin{array}{c|c} \Sigma_n^{1,1} - \Sigma_n^{1,2}(\Sigma_n^{2,2})^{-1}\Sigma_n^{2,1} + \begin{bmatrix} v_A^2 & 0 \\ 0 & v_B^2 \end{bmatrix} & \begin{bmatrix} (\sigma_{n+1}^A)^2 + \sigma_{n+1}^{AB} \\ \sigma_{n+1}^{AB} + (\sigma_{n+1}^B)^2 \end{bmatrix} \\ \hline \begin{bmatrix} (\sigma_{n+1}^A)^2 + \sigma_{n+1}^{AB} & \sigma_{n+1}^{AB} + (\sigma_{n+1}^B)^2 \end{bmatrix} & (\sigma_{n+1}^A)^2 + 2\sigma_{n+1}^{AB} + (\sigma_{n+1}^B)^2 + v_U^2 \end{array} \right] \quad (\text{B.4})$$

Definition: A *steady state* is a fixed point of the function Φ in the sense that each entry of Σ_n excluding $\Sigma_n^{1,1}$ are invariant under Φ ; $\Phi(\Sigma_n^{1,2}) = \Sigma_n^{1,2}$, $\Phi(\Sigma_n^{2,1}) = \Sigma_n^{2,1}$, and $\Phi(\Sigma_n^{2,2}) = \Sigma_n^{2,2}$.

We will now find a steady-state solution explicitly. Conjecture that the solution has the form:

$$\Sigma_n = \begin{bmatrix} n\alpha + c_A & -n\alpha & c_A \\ -n\alpha & n\alpha + c_B & c_B \\ c_A & c_B & c_A + c_B + v_U^2 \end{bmatrix}.$$

The result of applying the function Φ on $\Sigma_n^{1,1}$ is then

$$\Sigma_{n+1}^{1,1} = \begin{bmatrix} \frac{c_A\alpha + c_B\alpha + \alpha v_U^2 + c_A c_B + c_A v_U^2 + v_A^2 c_A + v_A^2 c_B + v_A^2 v_U^2}{c_A + c_B + v_U^2} & -\frac{c_A\alpha + c_B\alpha + \alpha v_U^2 + c_A c_B}{c_A + c_B + v_U^2} \\ -\frac{c_A\alpha + c_B\alpha + \alpha v_U^2 + c_A c_B}{c_A + c_B + v_U^2} & \frac{c_A\alpha + c_B\alpha + \alpha v_U^2 + c_A c_B + c_B v_U^2 + v_B^2 c_A + v_B^2 c_B + v_B^2 v_U^2}{c_A + c_B + v_U^2} \end{bmatrix}.$$

Since

$$\Sigma_{n+1}^{1,1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} c_A \\ c_B \end{bmatrix}$$

by conjecture, one can solve for c_A and c_B

$$c_A = \frac{v_A^2 v_B^2 + v_A^2 + \sqrt{(v_B^2 + v_A^2)(v_A^2 + v_B^2 + 4v_U^2)}}{2(v_B^2 + v_A^2)} \quad (\text{B.5})$$

$$c_B = \frac{v_B^2 v_A^2 + v_B^2 + \sqrt{(v_B^2 + v_A^2)(v_A^2 + v_B^2 + 4v_U^2)}}{2(v_B^2 + v_A^2)}. \quad (\text{B.6})$$

Using the above to solve for α gives

$$\alpha = \frac{v_B^2 v_A^2}{v_B^2 + v_A^2}.$$

An important consequence is that when in the steady state, one updates the means of A and B using the linear impact coefficient λ_A and λ_B :

$$\lambda_A = \frac{v_A^2(v_B^2 + v_A^2 + \sqrt{(v_B^2 + v_A^2)(v_A^2 + v_B^2 + 4v_U^2)})}{(v_A^2 + v_B^2)(\sqrt{(v_B^2 + v_A^2)(v_A^2 + v_B^2 + 4v_U^2)} + v_A^2 + v_B^2 + 2v_U^2)} \quad (\text{B.7})$$

$$\lambda_B = \frac{v_B^2(v_B^2 + v_A^2 + \sqrt{(v_B^2 + v_A^2)(v_A^2 + v_B^2 + 4v_U^2)})}{(v_A^2 + v_B^2)(\sqrt{(v_B^2 + v_A^2)(v_A^2 + v_B^2 + 4v_U^2)} + v_A^2 + v_B^2 + 2v_U^2)}. \quad (\text{B.8})$$

It is instructive to write the formulas for in terms of the signal to noise ratio. Without loss of generality consider let $r_A := \frac{v_A^2}{v_B^2 + v_A^2}$.

$$\lambda_A = r_A \frac{(v_B^2 + v_A^2 + \sqrt{(v_B^2 + v_A^2)(v_A^2 + v_B^2 + 4v_U^2)})}{(\sqrt{(v_B^2 + v_A^2)(v_A^2 + v_B^2 + 4v_U^2)} + v_A^2 + v_B^2 + 2v_U^2)} \quad (\text{B.9})$$

Assume v_U is zero and therefore $\lambda_A = r_A$. This formula is intuitive; when updating the mean of A one must account for the variation in the noise that interferes with direct observation of A . Here, the noise is B . Notice that if v_B were zero, then updating the mean is trivial because the impact coefficient $\lambda_A = r_A = 1$ which is equivalent to directly observing A . When adding the noise of B to the observation of A , one must account for the fact that there is more variation present and scale down the impact coefficient by r_A , the ratio of the variance of A to the total variance of $A + B$.

When the third source of variation v_U is nonzero, the impact coefficient must be scaled down further due to the addition to total variation. One can see this from the equation above; since the r_A is multiplied by a factor less than one, we have taken the impact coefficient from the standard learning problem and scaled it down according to the noise from v .

We simulate the Bayesian updating process with different initial conditions to provide evidence that the steady state is unique. An initial condition is defined by prior variances for A and B and their covariance which populate the covariance matrix $\Sigma_0^{1,1}$ defined above. We take 1000 random draws of each of entries in $\Sigma_0^{1,1}$, each drawn from a normal distribution with a mean of 0 and variance of 25. We ensure variances of A and B are positive by taking the absolute value of the random draws. We then iterate the Bayesian updating process 50 steps corresponding to 50 trades in the model setting. Table B.1 shows the mean and standard deviation of the price impact coefficient λ_A initially and after 50 Bayesian updates.

Table B.1: **Numerical simulation of convergence to the steady state.** This table shows the mean and standard deviation of the impact coefficient for random initial conditions and after 50 Bayesian updates, as well as the model's predicted value for the steady state.

	Initial	After 50 Updates	Model
Mean	0.0866	0.2030	0.2030
Standard Deviation	2.0884	7.9692e-07	0

B.2 Example of Dynamic Optimization

We will show that the informed trader's strategy takes on a similar form under dynamic optimization compared to the one-period optimization in the model. Consider a simplified version of the model with an asset that follows a normal random walk with innovations distributed $\mathcal{N}(0, \sigma_v^2 h)$, time increment h , and noise traders with $\sigma_u^2 h$. We leave out program trading for tractability. If the informed trader was long-lived in the presence of program trading, he would learn about the program trader's inventory using an information set superior to the market maker's as he would know the size of his own trader and could infer the remainder. This situation is complicated to model and is simplified in the main body of the paper by allowing a new informed trader to arrive each period and trade only once.

The informed trader chooses x_t in order to optimize the sum of current expected profits and future profits discounted with factor $e^{-\gamma h}$. Let the value of all maximized expected future and

current profits be $V(v_t)$. The Bellman equation is then,

$$V(v_t) = \max_{\hat{x}_t} \mathbb{E}[\hat{x}_t(v_t - p_t) + e^{-\gamma h} V(v_{t+h}) | \mathcal{I}_t]. \quad (\text{B.10})$$

The trader must balance profit taken now with forgone profit later due to price impact. Towards applying the method of undermined coefficients, guess that the value function is $V(v_t) = A(v_t - p_{t-h})^2 + B$ as in Chau and Vayanos (2008). Crucial to this value function is the assumption of a steady-state belief meaning the variance of the market maker's belief about v_t is constant over time. This gives constant price impact $\lambda_t = \lambda$.

Substituting the pricing rule in for p_t and the conjectured value function gives

$$\begin{aligned} V(v_t) &= \max_{\hat{x}_t} \mathbb{E}[\hat{x}_t(v_t - \mathbb{E}[p_t | \mathcal{I}_t]) + e^{-\gamma h} (A(v_{t+1} - p_t)^2 + B) | \mathcal{I}_t] \\ &= \max_{\hat{x}_t} \mathbb{E}[\hat{x}_t(v_t - (p_{t-h} + \lambda(\hat{x}_t + u_t))) + e^{-\gamma h} (A(v_{t+1} - (p_{t-h} + \lambda(\hat{x}_t + u_t)))^2 + B) | \mathcal{I}_t] \end{aligned} \quad (\text{B.11})$$

Evaluating the expectation and differentiating gives

$$x_t = \frac{1}{2} \frac{(2A\lambda_t - e^{\gamma h})(v_t - p_{t-h})}{\lambda(A\lambda - e^{\gamma h})}. \quad (\text{B.12})$$

Substituting back into the value function and matching coefficients results in the optimal trading strategy,

$$x_t = \frac{1}{\lambda} \frac{v_t - p_{t-h}}{\frac{1}{\sqrt{1-e^{-\gamma h}}} + 1}. \quad (\text{B.13})$$

Note that in the case of an infinite discount factor, $\gamma \rightarrow \infty$, the strategy is identical to the single-period strategy. The dynamic strategy and the single-period strategy are similar in that they both involve trades that are proportional to the difference between the market maker's prior and the true value of the asset times a constant. The constant is inversely proportional to the price impact coefficient in both cases.

Curriculum Vitae

Name: Adrian Walton

Post-Secondary Education and Degrees: B.Sc. Honours Mathematics, B.Mus.
McGill University
2003 - 2009

M.Sc. Applied Mathematics
The University of Western Ontario
2009 - 2010

Ph.D. Applied Mathematics
The University of Western Ontario
2010 - 2016

Honours and Awards: Ontario Graduate Scholarship
2010-2011

NSERC PGS D
2011-2013

Related Work Experience: Bank of Canada
2013 - present

Publications:

Garriott, C. & Walton, A. (2016). Retail Order Flow Segmentation. *Bank of Canada Staff Working Paper*, 20.