경영학석사학위논문

Program Trading and Price Discovery

프로그램매매가 가격발견기능에 미치는 영향

2017년 2월

서울대학교 대학원

경영학과 재무금융전공

강준석

Abstract

Program Trading and Price Discovery

Joonsuk Kang

College of Business Administration

Graduate School of Seoul National University

I examine the role of program trading in price discovery. Program trading is less involved in updating information and adjusting pricing error than non-program trading. For information update, liquidity-demanding program trading is less informed than liquidity-demanding non-program trading, and liquidity-supplying program trading is more adversely selected than liquidity-supplying non-program trading. In terms of pricing error, program trading buys overpriced stocks and sells underpriced stocks. Index arbitrage and non-arbitrage program trading play a similar role in price discovery.

Keywords: Program Trading, Price Discovery, State Space Model

Student Number: 2015-20574

Table of Contents

"Fundamentally, in a system in which the knowledge of the relevant facts is dispersed among many people, prices can act to coordinate the separate actions of different people in the same way as subjective values help the individual to coordinate the parts of his plan."

- Hayek (1945) *The Use of Knowledge in Society*

1. Introduction

The price system functions as the foundation of finance research, and hence of capitalism. Thus, financial markets play a crucial role in price discovery. Price discovery is "the process of determining the price of an asset in the marketplace through the interaction of buyers and sellers."[1] The types and strategies of market participants are complicated and they constantly evolve with changing regulations and the introduction of technological advances.

Recently, there's a growing concern over passive investing. On August 2016, Bernstein & Co. published a report entitled "The Silent Road to Serfdom: Why passive investing is worse than Marxism," which argues that "a supposedly capitalist economy where the only investment is passive is

[1]Nasdaq, Financial Glossary

worse than either a centrally planned economy or an economy with active market led capital management." Active investing involves the identification of undervalued and overvalued stocks, and buying or short selling them. In this process, the price system aggregates information to reflect firms' fundamental values, leading to the efficient allocation of capital and the sustainable growth of an economy. Passive investing, however, aims to replicate the index returns by buying or selling the market simultaneously. When passive investing prevails, market efficiency may be harmed according to the point made by Bernstein & Co.

In response, Vanguard Group founder John C. Bogle claimed that "the issue is at what point is indexing making the market less efficient," and that index investing should be greater to have any noticeable impact on price efficiency or market efficiency. The efficient market hypothesis is strongly supported by the logic that a few arbitragers can exploit all existing mispricing, but in reality, there are some limits of arbitrage that cause them to fail to do so. Thus, there possibly exists a certain type of traders who, if there is a sufficiently large number of them, can facilitate or deteriorate the price discovery process.

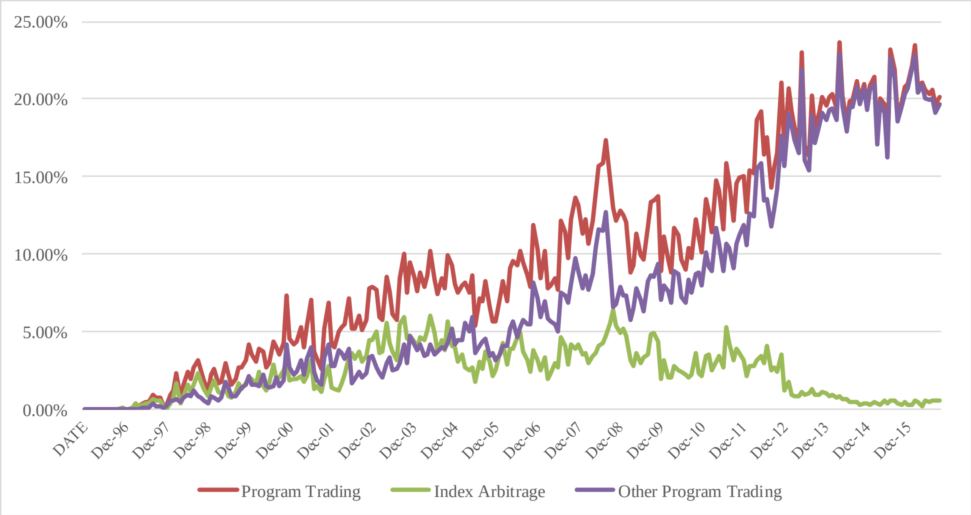

Figure 1
Rise of Program Trading
This figure plots the ratio of program trading in overall trading volume from January 1996 to October 2016. The ratio of program trading is aggregate program trading (buy trading volume plus sell trading volume) divided by two times the total trading volume, calculated monthly.

In fact, program trading as a typical passive investor in stock markets is on the rise at a rapid pace as shown in Figure 1. In KOSPI, program trading accounted for 5% of the overall trading volume in the early 2000s, but its share increased to 20% in 2016. Another noticeable change is the composition of program trading. Historically, the concept of program trading was adopted to regulate index arbitrage or pseudo index arbitrage trades, and index arbitrage outnumbered other types of program trading. As of 2016, almost all program trading was of the non-arbitrage type with only negligible index arbitrage trades. This research is based on the KOSPI data of the Korea Exchange (KRX), but recent publications of the New York Stock Exchange

(NYSE) show that the NYSE shares similar patterns[2]: program trading accounts for almost 20% of the overall trading share with negligible index arbitrage.

According to NYSE Rule 7410 (Definitions), "program trading" is defined as "either (A) index arbitrage or (B) any trading strategy involving the related purchase or sale of a "basket" or group of 15 or more stocks." And "index arbitrage" is defined as "a trading strategy in which pricing is based on discrepancies between a 'basket' or group of stocks and the derivative index product (i.e., a basis trade) involving the purchase or sale of a 'basket' or group of stocks in conjunction with the purchase or sale, or intended purchase or sale, of one or more derivative index products in an attempt to profit by the price difference between the 'basket' or group of stocks and the derivative index products."

The KRX defines "program trading" almost identically. "Program trading" is defined as "all the index profit trading and non-index profit trading through which the same person concurrently makes a trade of more than KOSPI 15 items," and "index profit trading" is defined as a trading strategy "conducted by linking a stock group with futures or options for the purpose of getting

[2] NYSE, Program Trading Reports

profit through price difference between the stock group of KOSPI 200 and KOSPI 200 futures or options."

By definition, program trading in spot stock market trades the *market* or at least *a subset of the market*. Program trading does not have firm-specific information nor does it try to exploit mispricing. It just buys or sells the market as is without consideration of where individual asset prices should be. Program trading acts as a free rider in price discovery.

Traditionally, several market qualities were linked to program trading, but the results are mixed. In general, research shows that program trading enhances price discovery by conveying information between the spot market and the futures market (Kawaller, Koch, and Koch 1987), but in the course of index arbitrage, program trading entails a large trade amount, which makes it liable for injecting volatility into the market (Harris, Sofianos, and Shapiro 1994). For price discovery, Hasbrouck investigates the information content of automated orders using techniques suggested by himself earlier (Hasbrouck 1991) and concludes that program trading orders contain information beyond that available from the futures-spot market relationship (Hasbrouck 1996). However, recent program trading is essentially non-arbitrage program trading. The logic and empirics of the effect of program trading on price discovery should be revisited to understand and forecast market evolution.

For program trading in Korean stock market, Choe and Yoon (2007) support the view that non-arbitrage program trading has a positive role in price discovery, but their conclusion is based on relatively small extreme order imbalance samples, and the sample period is from November 1996 to December 2003—when program trading was scarce.

While program trading was not in the interest of researchers, high-frequency trading (HFT) drew a lot of attention recently. The literature on HFT lends valuable and up-to-date tools for analyzing the role of program trading in price discovery. Instead of using aggregate price movement to investigate the impact of program trading on price discovery, I employ the state space model of Hendershott and Menkveld (2014) to decompose aggregate price movement into the information (permanent) part and pricing error (transitory) part. Through price decomposition, I examine the role of program trading in updating information and reducing pricing error. Also, following the methodology of Brogaard, Hendershott, and Riordan (2014), I test the role of program trading in price discovery beyond its order characteristics by splitting program trading trades into liquidity demanding and supplying trades.

The layout of the paper is as follows: Section 2 describes the data and defines program trading variables. Section 3 introduces the state space model for price decomposition. Section 4 examines the role of program trading in information update, and Section 5 examines the role of program trading in reducing pricing error. Section 6 concludes by suggesting that ambiguity aversion plays a role in ever-growing program trading.

2. Data and Descriptive Statistics

The KRX trade and quote (TAQ) sample contains trade and quote data for all trading days in 2010. The trade data includes the trading price and quantity, trading time, and order numbers of the buyer and seller. The quote data includes the quoted price and quantity, quoted time, order number, and program trading type indicator. The trades and quotes are time stamped to the millisecond and are merged using unique order numbers.

According to the quoted time of the bid and ask quotes, the buyer or seller is identified as the liquidity demander or liquidity supplier. When a bid quote arrives prior to a matched ask quote, the buyer of the trade is the liquidity supplier and the seller of the trade is the liquidity demander. Also, my sample identifies every trader as program trading (PT) or non-program trading (nPT). Program trading is again classified as index arbitrage (IA) or other program trading (OP).

Unlike the case of the United States, where a single stock can be traded at multiple trading venues, Korea has no existing alternative trading system (ATS). For each stock listed in the KOSPI, there is a single price discovery process in the unique stock market. Therefore, basket trading is unquestionably captured as program trading in Korea.

One limitation of the data is that classification as index arbitrage or other program trading is reported rather arbitrarily by reporting traders. There is some empirical evidence that the behavior of other program trading (OP) is similar to that of index arbitrage (IA), including the fluctuating pattern with respect to spot-futures basis.

I categorize stocks into three groups by market capitalization: large, medium, and small. All firms in the KOSPI are ranked by market capitalization and partitioned into three groups. It should be noted that many of the firms in the large-size group overlap with firms comprising the KOSPI200 index on which diverse financial products are based.

The KRX TAQ data set is supplemented with DATAGUIDE for daily closing price and market capitalization. I focus on continuous trading during normal trading hours by removing trading before 9:00 and after 15:00, and the period of the closing auction. An observation is made at stock-minute level. For example, Samsung Electronics on March 8th, 2010, 10:45.00-10:45.59 is an observation.

In one-minute intervals, program trading variables are defined as order flow (net trading): buy volume minus sell volume. Note that the order flow of program trading (PT) and the order flow of non-program trading (nPT) always add up to zero by construction. Likewise, when program trading (PT) is divided into index arbitrage (IA) and other program trading (OP), the order flows of program trading (PT), index arbitrage (IA), and other program trading (OP) add up to zero.

Further, to differentiate the effects of liquidity demanding and supplying orders of program trading, I use the order flow (net trading) of liquidity demanding and supplying program trading following Brogaard, Hendershott, and Riordan (2014). PT^D denotes the order flow of liquidity demanding program trading, and it is calculated as the buy volume minus the sell volume of liquidity demanding program trading. PT^S denotes the order flow of liquidity supplying program trading, and it is calculated as the buy volume minus the sell volume of liquidity supplying program trading. $nPT^D, nPT^S, IA^D, IA^S, OP^D, OP^S$ are defined analogously.

To illustrate the definition of trading variables, suppose that for a one-minute interval, ten trades are made as described below.

Trading No.	Price	Quantity	Trading Volume	Buyer		Seller	
				Type	Liquidity	Type	Liquidity
1	1,000	5	5,000	IA	Demand	nPT	Supply
2	950	7	6,650	OP	Demand	nPT	Supply
3	950	10	9,500	IA	Supply	OP	Demand
4	1,000	3	3,000	nPT	Supply	IA	Demand
5	1,100	20	22,000	nPT	Demand	OP	Supply
6	1,050	7	7,350	OP	Supply	IA	Demand
7	1,050	5	5,250	OP	Demand	OP	Supply
8	950	8	7,600	nPT	Demand	nPT	Supply

9	1,000	15	15,000	IA	Supply	nPT	Demand
10	1,000	10	10,000	nPT	Supply	IA	Demand

In this case, $PT^D = (5000 + 6650 + 5250) - (9500 + 3000 + 7350 + 10000) = -12950$; $IA^D = (5000) - (3000 + 7350 + 10000) = -15350$; $OP^D = (6650 + 5250) - (9500) = 2400$; $nPT^D = (22000 + 7600) - (15000) = 14600$. For liquidity supplying trading variables, $PT^S = (9500 + 7350 + 15000) - (22000 + 5250) = 4{,}600$; $IA^S = (9500 + 15000) = 24500$; $OP^S = (7350) - (22000 + 5250) = -19900$; $nPT^S = (3000 + 10000) - (5000 + 6650 + 7600) = -6250$. For overall trading variables, $PT = PT^D + PT^S = -12950 + 4600 = -8350$; $IA = IA^D + IA^S = -15350 + 24500 = 9150$; $OP = OP^D + OP^S = 2400 - 19900 = -17500$; $nPT = nPT^D + nPT^S = 14600 - 6250 = 8350$. Since all trades have buyers and sellers, $PT + nPT = IA + OP + nPT = 0$.

Note that for each trade, one side is the buyer and the other side is the seller; one side is the liquidity demander and the other side is the liquidity supplier. The buyer and seller can take any type of {IA, OP, nPT}, and the buyer and seller can be of the same trader type (index arbitrage and other program trading are engaged in by multiple traders). Unlike the case of the overall order flow (net trading), PT^D and nPT^D do not add up to zero, and they are also on the liquidity supplying side.

Figure 2 illustrates the order flow of program trading, index arbitrage, other program trading, and non-program trading with the mid-quote price for Samsung Electronics (International Securities Identification Number (ISIN): KR7005930003) on December 3rd, 2010. When price series are stable, the order flows of four types are relatively stable. However, order flows are volatile at times of large price changes.

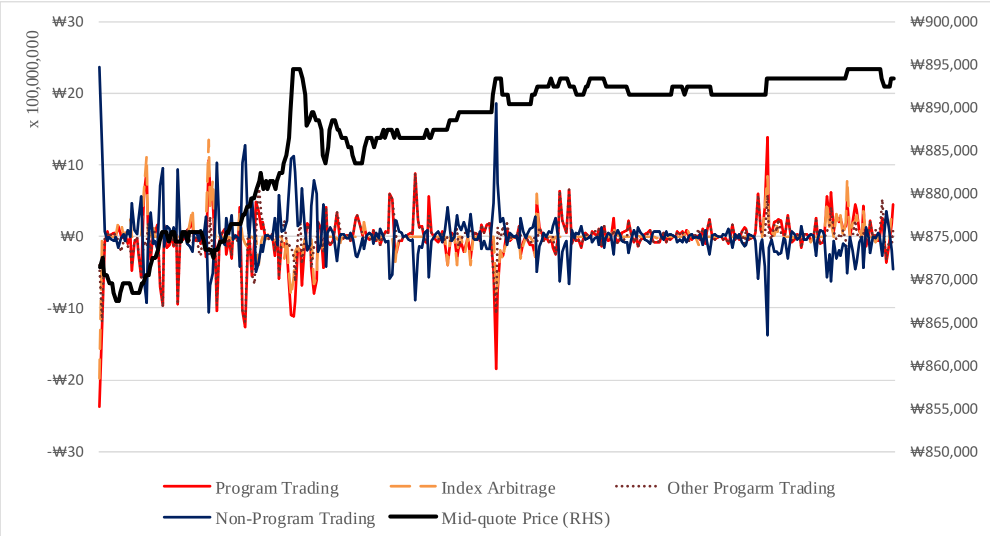

Figure 2
Sample Order Flows and Price Movements
This figure plots the order flow of program trading (PT), index arbitrage (IA), other program trading (OP), non-program trading (nPT), and mid-quote price for Samsung Electronics on December 3rd, 2010. Order flow variables are calculated within a minute in 100 million KRW, and mid-quote price is the end of the interval average of best bid price and best offer price.

Table 1 reports the descriptive statistics by size category. The average market capitalization ranges from 3.75 trillion KRW in large firms to 38 billion KRW in small firms. I report the average daily closing price and mid-quote return volatility. The mid-quote price is calculated as the average of the best bid price and the best offer price at the end of a one-minute interval. Mid-quote volatility is calculated as the realized variance (RV) of ten-second returns, aggregated over the trading day. The bid-ask spread is reported in absolute terms in basis points, and in relative terms in KRW. Relative spread is the largest in small firms and the smallest in large firms, but absolute spread shows the opposite pattern due to minimum tick size regulation. Daily trading volume in the large-size category is 1.86 billion KRW, which is more than 10 times greater than the trading volume in the medium-size category.

Table 1.
Descriptive statistics

Summary Statistics	Units	Large	Medium	Small
Panel A. Key Descriptive Statistics				
Market Capitalization	KRW trillion	3.7583	0.1399	0.0381
Price	KRW	85,876	19,403	7,325
Mid-quote return volatility	bps	1,541	1,691	1,775
Bid-ask spread	KRW	235	162	104
Relative bid-ask spread	%	0.34%	0.75%	1.59%
Trading volume	KRW 100 million	186.30	17.21	7.19
Panel B. Trading Volume Share				
Program Trading (PT) share	%	10.0%	2.7%	1.6%
Index Arbitrage (IA) share	%	2.7%	0.3%	0.0%
Other Program (OP) share	%	7.3%	2.4%	1.5%
liquidity demanding Program Trading (PT) share	%	5.2%	1.2%	0.8%
liquidity demanding Index Arbitrage (IA) share	%	1.4%	0.1%	0.0%
liquidity demanding Other Program (OP) share	%	3.8%	1.1%	0.8%
liquidity supplying Program Trading (PT) share	%	4.8%	1.5%	0.8%
liquidity supplying Index Arbitrage (IA) share	%	1.3%	0.2%	0.0%
liquidity supplying Other Program (OP) share	%	3.5%	1.3%	0.7%
non-Program Trading (nPT) share	%	89.3%	97.2%	98.4%
liquidity demanding non-Program Trading (nPT) share	%	44.6%	48.7%	49.2%
liquidity supplying non-Program Trading (nPT) share	%	44.8%	48.5%	49.2%

This table reports descriptive statistics that are equal-weighted averages across stock-days for all KOSPI stocks in 2010. Each stock belongs to one of three size categories by market capitalization: large, medium, or small. Market capitalization is calculated using daily closing price, and price is daily closing price. Mid-quote return volatility is Realized Variance (RV), calculated in ten-second intervals and aggregated over the trading day. Trading volume is the average daily trading volume in 100 Korean Won. Panel B shows trading volume share of program trading (PT), index arbitrage (IA), other program trading (OP), and non-program trading(nPT).

Program trading makes up 10.0% of the trading volume in large stocks and 1.6% of that in small stocks. By types of program trading, index arbitrage is responsible for 2.7% in large stocks and almost zero percent in small stocks. Other program trading makes up 7.3% in large stocks and 1.5% in small stocks. The trading volume share of non-program trading is simply the residual of program trading share.

Further, trading variables are divided into liquidity demanding and supplying trading variables. Liquidity demanding program trading makes up 5.2% of the trading volume in large stocks and 0.8% in small stocks, while liquidity supplying program trading makes up 4.8% of the trading volume in large stocks and 0.8% in small stocks. Program trading in large stocks is more often executed in liquidity demanding orders.

Program trading, especially index arbitrage, is concentrated in large stocks and less so in small, less liquid stocks. One possible reason for this is that underlying assets for financial products such as stock futures, mutual funds, and ETFs are concentrated in large stocks. It should be noted that the sample period is 2010 when program trading makes up 10% of the trading volume, but in 2016, the program trading’s ratio more than doubled.

3. State Space Model of Program Trading and Prices

To understand the role of program trading in price discovery, I estimate the state space model of Hendershott and Menkveld (2014) to decompose aggregate price movement into the permanent part (information part) and the transitory part (pricing error). The price of a stock can be represented as:

$$p_{i,t} = m_{i,t} + s_{i,t},$$

where $\mathrm{p_{i,t}}$ is the log mid-quote price at time *t* for stock *i* and is composed of a permanent component $m_{i,t}$ and a transitory component $s_{i,t}$. The permanent part is modeled as a martingale:

$$m_{i,t} = m_{i,t-1} + w_{i,t}.$$

where $\mathrm{w_{i,t}}$, the change in the permanent (informational) part, represents information update. Information update in the aggregate model is modeled as:

$$\mathrm{w_{i,t}} = \kappa_i \widetilde{PT_{\iota,t}} + \mu_{i,t},$$

where $\widetilde{PT_{\iota,t}}$ is the residual of an autoregressive model in the order flow of program trading to remove autocorrelation. A lag length of ten (ten minutes) is used following Brogaard, Hendershott, and Riordan (2014). For the disaggregate model, $\mathrm{w_{i,t}}$ is modeled as:

$$\mathrm{w_{i,t}} = \kappa_{i,PT}^{D}\widetilde{PT_{\iota,t}^{D}} + \kappa_{i,nPT}^{D}\widetilde{nPT_{\iota,t}^{D}} + \mu_{i,t},$$

where $\widetilde{PT_{\iota,t}^{D}}$ and $\widetilde{nPT_{\iota,t}^{D}}$ are the residuals of autoregressive models in liquidity demanding program trading and liquidity demanding non-program trading. For liquidity supplying program trading and non-program trading, the analogous disaggregate model is estimated:

$$\mathrm{w_{i,t}} = \kappa_{i,PT}^{S}\widetilde{PT_{\iota,t}^{S}} + \kappa_{i,nPT}^{S}\widetilde{nPT_{\iota,t}^{S}} + \mu_{i,t},$$

These models capture the role of program trading in information update. $\mu_{i,t}$ captures information update unrelated to trading. Further, to distinguish the role of index arbitrage and other program trading, models with these two trading variables are estimated:

$$\mathrm{w_{i,t}} = \kappa_{i,IA}\widetilde{IA_{\iota,t}} + \kappa_{i,OP}\widetilde{OP_{\iota,t}} + \mu_{i,t},$$

$$\mathrm{w_{i,t}} = \kappa_{i,IA}^{D}\widetilde{IA_{\iota,t}^{D}} + \kappa_{i,OP}^{D}\widetilde{OP_{\iota,t}^{D}} + \kappa_{i,nPT}^{D}\widetilde{nPT_{\iota,t}^{D}} + \mu_{i,t},$$

$$\mathrm{w_{i,t}} = \kappa_{i,IA}^{S}\widetilde{IA_{\iota,t}^{S}} + \kappa_{i,OP}^{S}\widetilde{OP_{\iota,t}^{S}} + \kappa_{i,nPT}^{S}\widetilde{nPT_{\iota,t}^{S}} + \mu_{i,t}$$

In the state space model, the transitory part is assumed to be stationary. Following Hendershott and Menkveld (2014), I model $\mathrm{s_{i,t}}$ with an autoregressive component and trading variables. For the aggregate model, $\mathrm{s_{i,t}}$ is modeled as:

$$\mathrm{s_{i,t}} = \phi s_{i,t-1} + \psi_i PT_{i,t} + \nu_{i,t}$$

and for the disaggregate model as:

$$\mathrm{s_{i,t}} = \phi s_{i,t-1} + \psi_{i,PT}^{D} PT_{i,t}^{D} + \psi_{i,nPT}^{D} nPT_{i,t}^{D} + \nu_{i,t}.$$

These models measure the aggregate role of program trading in pricing error movements. Also, the disaggregate model for liquidity supplying trading variables and models including index arbitrage and other program trading variables are estimated analogously.

It should be noted that the inclusion of non-program trading variables in the disaggregate model equations makes possible the comparison between the impact of program trading and the impact of non-program trading in updating information and reducing pricing error. For identification, I assume that the innovations in the permanent part and the transitory part are uncorrelated. I estimate the state space model for 350 one-minute intervals on a stock-day-by-stock-day basis using maximum likelihood, using the KRX TAQ mid-quote price and the order flow of trading variables.

The KRX TAQ sample contains roughly 700 stocks for 251 trading days in 2010. The estimation uses the sum of the order flow (buy volume minus sell volume) within a minute. For a stock-day sample to be valid, I require at least ten one-minute intervals where liquidity demanding program trading, liquidity supplying program trading, and price change are all nonzero. To eliminate erroneously optimized parameters due to singular Hessian matrix in optimization, I deleted stock days that fail to satisfy either one of two criteria or both based on the estimated transitory series.

a. Out of 350 one-minute intervals, no less than 100 intervals should have an estimated transitory part with an absolute value larger than 0.0001.[3]

b. The autoregressive coefficient of the estimated transitory series should be greater than zero.[4]

Qualifying samples are 681 stock days for large stocks, 982 for medium stocks, and 673 for small stocks. I use these stock days for the analysis in the remainder of the paper.

---

[3] It should be noted that price series used for the state space model is the log value of the mid-quote price. Thus, 0.0001 implies 0. 01%p deviation from the efficient (information) series.

[4] Some ill-optimized transitory series have negative autoregressive coefficients. Autoregressive specification is introduced to capture the stationary process, which naturally restricts the AR (1) coefficient to be positive.

4. Information in Prices and the Order Flow of Program Trading

Table 2 reports the result of the permanent (information) price component for each specification by each size category. Panel A of Table 2 shows that program trading is negatively related to efficient price changes. Program trading seems to be poor at predicting permanent price changes, implying that program trading is less involved in updating information in asset prices.

Table 2
Permanent price component of prices and program trading variables (PT)

Panel A: Aggregate program trading and prices

	Units	Large	Medium	Small
κ	bps/KRW 100m	-50.47	-166.77	-452.65
(t-statistics)		-5.28	-8.39	-10.25

Panel B: Liquidity demanding program trading, non-program trading, and prices

	Units	Large	Medium	Small
κ_{PT}^{D}	bps/KRW 100m	10.54	-29.10	-146.62
(t-statistics)		0.66	-0.53	-2.89
κ_{nPT}^{D}	bps/KRW 100m	35.45	90.49	158.01
(t-statistics)		6.71	13.69	21.77
p-value of difference		0.14	0.03	0.00

Panel C: Liquidity supplying program trading, non-program trading, and prices

	Units	Large	Medium	Small
κ_{PT}^{S}	bps/KRW 100m	-76.32	-307.87	-49802.79
(t-statistics)		-3.31	-6.04	-1.03
κ_{nPT}^{S}	bps/KRW 100m	-35.98	-89.32	-154.49
(t-statistics)		-8.97	-10.32	-22.42
p-value of difference		0.09	0.00	0.30

The model is estimated by stock-day using order flow variables to decompose the observed mid-quote price $p_{i,t}$ for stock *i* at time *t* into two components: the information part $m_{i,t}$ and the pricing error part $\mathrm{s}_{\mathrm{i,t}}$:

$$p_{i,t} = m_{i,t} + s_{i,t},$$

$$m_{i,t} = m_{i,t-1} + w_{i,t}$$

Specification for Panel A. Aggregate model

$$w_{i,t} = \kappa_i \widetilde{PT_{i,t}} + \mu_{i,t},$$

$$s_{i,t} = \phi_i s_{i,t-1} + \psi_i PT_{i,t} + \nu_{i,t}$$

Specification for Panel B. liquidity demand model

$$w_{i,t} = \kappa_{i,PT}^{D} \widetilde{PT_{i,t}^{D}} + \kappa_{i,nPT}^{D} \widetilde{nPT_{i,t}^{D}} + \mu_{i,t},$$

$$s_{i,t} = \phi_i s_{i,t-1} + \psi_{i,PT}^{D} PT_{i,t}^{D} + \psi_{i,nPT}^{D} nPT_{i,t}^{D} + \nu_{i,t}$$

Specification for Panel C. liquidity supply model

$$w_{i,t} = \kappa_{i,PT}^{S} \widetilde{PT_{i,t}^{S}} + \kappa_{i,nPT}^{S} \widetilde{nPT_{i,t}^{S}} + \mu_{i,t},$$

$$s_{i,t} = \phi_i s_{i,t-1} + \psi_{i,PT}^{S} PT_{i,t}^{S} + \psi_{i,nPT}^{S} nPT_{i,t}^{S} + \nu_{i,t}$$

$PT_{i,t}$ is program trading's overall order flow; $\widetilde{PT_{i,t}}$ is the surprise component of the order flow. $PT_{i,t}^{D}$ and $nPT_{i,t}^{D}$ are program trading's and non-program trading's liquidity demanding order flow; $\widetilde{PT_{i,t}^{D}}$ and $\widetilde{nPT_{i,t}^{D}}$ are the surprise components of those order flows. $PT_{i,t}^{S}$ and $nPT_{i,t}^{S}$ are program trading's and non-program trading's liquidity supplying order flow; $\widetilde{PT_{i,t}^{S}}$ and $\widetilde{nPT_{i,t}^{S}}$ are the surprise components of those order flows. Each stock is in one of three market capitalization categories: large, medium, or small. Columns 3-5 report coefficients for the entire sample at one-minute frequencies using KRX TAQ.

Panel B of Table 2 reports the result of the disaggregate model of program trading and non-program trading's liquidity demanding trades. The specification includes both PT^D and nPT^D variables to identify the different roles in price discovery. For large firms, surprise innovation of non-program trading is significantly positively correlated with efficient price changes, while surprise innovation of program trading is less positively but insignificantly correlated. A positive κ implies informed trading because order flow is correlated with positive information update. The less positive (and statistically insignificant) κ on PT^D suggests that on a per KRW basis, program trading is less informed or uninformed when it trades.

Panel C of Table 2 reports the result of the disaggregate model of program trading and non-program trading's liquidity supplying trades. A negative coefficient κ implies adverse selection because order flow is correlated with negative information update. The coefficients are all negative and significant with the exception of program trading in the small size category.

For large- and medium-size categories, program trading is more adversely selected than non-program trading. It seems that program trading is bad at placing and managing liquidity supplying orders. Brogaard, Hendershott, and Riordan (2016) argue that limit orders, especially those of HFTs, play an important role in price discovery. Considering their remark, the active and efficient management of limit orders in liquidity supplying orders also plays an important role in price discovery. Therefore, program trading plays a small role in updating information in prices, not only through liquidity demanding orders but also through liquidity supplying orders.

Table 3 reports the result of the permanent (information) price component for each specification by each size category using index arbitrage (IA) and the other program trading (OP) instead of the program trading (PT) variable. The effects of index arbitrage and other program trading in price discovery seem to be similar, and the difference between the two is insignificant over all specifications and all size categories.

Table 3
Permanent price component of prices and program trading variables (IA, OP)

Panel A: Aggregate program trading and prices

	Units	Large	Medium	Small
κ_{IA}	bps/KRW 100m	-5.37	-11831.61	-2902.50
(t-statistics)		-0.57	-1.03	-1.25
κ_{OP}	bps/KRW 100m	2.86	-43.20	-6.18
(t-statistics)		1.19	-2.65	-0.03
p-value of difference		0.40	0.30	0.22

Panel B: Liquidity demanding program trading, non-program trading, and prices

	Units	Large	Medium	Small
κ_{IA}^{D}	bps/KRW 100m	14883.38	-312121.65	662.31
(t-statistics)		1.02	-1.06	0.15
κ_{OP}^{D}	bps/KRW 100m	2.66	-10.48	1378.96
(t-statistics)		1.26	-0.51	1.10
κ_{nPT}^{D}	bps/KRW 100m	0.40	12.86	47.55
(t-statistics)		0.47	6.14	4.60
p-value of difference (IA-nPT)		0.21	0.19	0.87
p-value of difference (OP-nPT)		1.00	1.00	0.73
p-value of difference (IA-OP)		0.21	0.19	0.85

Panel C: Liquidity supplying program trading, non-program trading, and prices

	Units	Large	Medium	Small
κ_{IA}^{S}	bps/KRW 100m	14.35	-8025.10	-1649.71
(t-statistics)		1.20	-0.67	-0.91
κ_{OP}^{S}	bps/KRW 100m	7.81	-182.68	-812.90
(t-statistics)		2.01	-2.15	-2.38
κ_{nPT}^{S}	bps/KRW 100m	-8.72	-36.21	-88.05
(t-statistics)		-8.27	-7.70	-11.51
p-value of difference (IA-nPT)		0.02	0.41	0.30
p-value of difference (OP-nPT)		0.11	0.99	0.63
p-value of difference (IA-OP)		0.52	0.42	0.58

The model is estimated by stock-day using order flow variables to decompose the observed mid-quote price $p_{i,t}$ for stock *i* at time *t* into two components: the information part $m_{i,t}$ and the pricing error part $s_{i,t}$:

$$p_{i,t} = m_{i,t} + s_{i,t},$$

$$m_{i,t} = m_{i,t-1} + w_{i,t}$$

Specification for Panel A. Aggregate model

$$w_{i,t} = \kappa_{i,IA}\widetilde{IA_{i,t}} + \kappa_{i,OP}\widetilde{OP_{i,t}} + \mu_{i,t},$$

$$s_{i,t} = \phi_i s_{i,t-1} + \psi_{i,IA}IA_{i,t} + \psi_{i,OP}OP_{i,t} + \nu_{i,t}$$

Specification for Panel B. liquidity demand model

$$w_{i,t} = \kappa_{i,IA}^{D}\widetilde{IA_{i,t}^{D}} + \kappa_{i,OP}^{D}\widetilde{OP_{i,t}^{D}} + \kappa_{i,nPT}^{D}\widetilde{nPT_{i,t}^{D}} + \mu_{i,t},$$

$$s_{i,t} = \phi_i s_{i,t-1} + \psi_{i,IA}^{D} IA_{i,t}^{D} + \psi_{i,OP}^{D} OP_{i,t}^{D} + \psi_{i,nPT}^{D} nPT_{i,t}^{D} + \nu_{i,t}$$

Specification for Panel C. liquidity supply model

$$w_{i,t} = \kappa_{i,IA}^{S}\widetilde{IA_{i,t}^{S}} + \kappa_{i,OP}^{S}\widetilde{OP_{i,t}^{S}} + \kappa_{i,nPT}^{S}\widetilde{nPT_{i,t}^{S}} + \mu_{i,t},$$

$$s_{i,t} = \phi_i s_{i,t-1} + \psi_{i,IA}^{S} IA_{i,t}^{S} + \psi_{i,OP}^{S} OP_{i,t}^{S} + \psi_{i,nPT}^{S} nPT_{i,t}^{S} + \nu_{i,t}$$

$IA_{i,t}$ is index arbitrage's overall order flow, $OP_{i,t}$ is other program trading's overall order flow; $\widetilde{IA_{i,t}}$ and $\widetilde{OP_{i,t}}$ are the surprise components of those order flows. $IA_{i,t}^{D}$, $OP_{i,t}^{D}$ and $nPT_{i,t}^{D}$ are liquidity demanding order flows of index arbitrage, other program trading, and non-program trading; $\widetilde{IA_{i,t}^{D}}, \widetilde{OP_{i,t}^{D}}$ and $\widetilde{nPT_{i,t}^{D}}$ are the surprise components of those order flows. $IA_{i,t}^{S}$, $OP_{i,t}^{S}$ and $nPT_{i,t}^{S}$ are liquidity supplying order flows of index arbitrage, other program trading, and non-program trading; $\widetilde{IA_{i,t}^{S}}, \widetilde{OP_{i,t}^{S}}$ and $\widetilde{nPT_{i,t}^{S}}$ are the surprise components of those order flows. Each stock is in one of three market capitalization categories: large, medium, or small. Columns 3-5 report coefficients for the entire sample at one-minute frequencies using KRX TAQ. p-value of pairwise differences in parameters are shown at the bottom of each panel.

5. Noise in Prices and the Order Flow of Program Trading

Table 4 reports the result of the transitory (noise) price component for each specification by each size category. Panel A of Table 2 reports that program trading is negatively related to transitory price changes. The pricing errors are persistent with an AR(1) coefficient between 0.23 and 0.48. The ψ coefficients are in basis points per 100 million KRW traded. The -35.77 coefficient in large stocks implies that per 100 million of positive program trading order flow (buy volume minus sell volume) is related to a 35.77 basis points decrease in pricing error.

Negative coefficients over three size categories imply that program trading is related to a decrease in pricing error. When program trading has positive (or negative) order flow, pricing error decreases (or increases). It seems that program trading is also less informed about the pricing error level.

Panel B of Table 4 reports the result of the disaggregate model of liquidity demanding trades and finds that program trading is negatively correlated with transitory price movements, while non-program trading is positively correlated with transitory price movements. Though the pricing error process reverts to zero, the process is still persistent. Positive coefficients for non-program trading imply that non-program trading exploits temporary pricing error, while negative coefficients for program trading imply that program trading fails to do so.

Table 4

Transitory price component of prices and program trading variables (PT)

Panel A: Aggregate program trading and prices

	Units	Large	Medium	Small
ϕ		0.23	0.37	0.48
(t-statistics)		21.63	38.26	39.40
ψ	bps/KRW 100m	-35.77	-88.01	-458.69
(t-statistics)		-1.33	-2.57	-6.76

Panel B: Liquidity demanding program trading, non-program trading, and prices

	Units	Large	Medium	Small
ϕ		0.26	0.41	0.52
(t-statistics)		23.35	42.41	43.79
ψ_{PT}^{D}	bps/KRW 100m	-28.11	-34.80	-286.35
(t-statistics)		-0.78	-0.38	-2.98
ψ_{nPT}^{D}	bps/KRW 100m	24.42	77.23	166.34
(t-statistics)		4.65	4.36	8.00
p-value of difference		0.15	0.23	0.00

Panel C: Liquidity supplying program trading, non-program trading, and prices

	Units	Large	Medium	Small
ϕ		0.26	0.42	0.53
(t-statistics)		23.70	42.48	43.99
ψ_{PT}^{S}	bps/KRW 100m	-57.17	-138.51	-624.03
(t-statistics)		-1.51	-2.44	-1.48
ψ_{nPT}^{S}	bps/KRW 100m	-25.35	-93.45	-164.00
(t-statistics)		-5.54	-3.95	-7.95
p-value of difference		0.40	0.46	0.28

The model is estimated by stock-day using order flow variables to decompose the observed mid-quote price $p_{i,t}$ for stock *i* at time *t* into two components: the information part $m_{i,t}$ and the pricing error part $s_{i,t}$:

$$p_{i,t} = m_{i,t} + s_{i,t},$$

$$m_{i,t} = m_{i,t-1} + w_{i,t}$$

Specification for Panel A. Aggregate model

$$w_{i,t} = \kappa_i \widetilde{PT_{i,t}} + \mu_{i,t},$$

$$s_{i,t} = \phi_i s_{i,t-1} + \psi_i PT_{i,t} + \nu_{i,t}$$

Specification for Panel B. liquidity demand model

$$w_{i,t} = \kappa_{i,PT}^{D} \widetilde{PT_{i,t}^{D}} + \kappa_{i,nPT}^{D} \widetilde{nPT_{i,t}^{D}} + \mu_{i,t},$$

$$s_{i,t} = \phi_i s_{i,t-1} + \psi_{i,PT}^{D} PT_{i,t}^{D} + \psi_{i,nPT}^{D} nPT_{i,t}^{D} + \nu_{i,t}$$

Specification for Panel C. liquidity supply model

$$w_{i,t} = \kappa_{i,PT}^{S} \widetilde{PT_{i,t}^{S}} + \kappa_{i,nPT}^{S} \widetilde{nPT_{i,t}^{S}} + \mu_{i,t},$$

$$s_{i,t} = \phi_i s_{i,t-1} + \psi_{i,PT}^{S} PT_{i,t}^{S} + \psi_{i,nPT}^{S} nPT_{i,t}^{S} + \nu_{i,t}$$

$PT_{i,t}$ is program trading's overall order flow; $\widetilde{PT_{i,t}}$ is the surprise component of the order flow. $PT_{i,t}^{D}$ and $nPT_{i,t}^{D}$ are program trading's and non-program trading's liquidity demanding order flow; $\widetilde{PT_{i,t}^{D}}$ and $\widetilde{nPT_{i,t}^{D}}$ are the surprise components of those order flows. $PT_{i,t}^{S}$ and $nPT_{i,t}^{S}$ are program trading's and non-program trading's liquidity supplying order flow; $\widetilde{PT_{i,t}^{S}}$ and $\widetilde{nPT_{i,t}^{S}}$ are the surprise components of those order flows. Each stock is in one of three market capitalization categories: large, medium, or small. Columns 3-5 report coefficients for the entire sample at one-minute frequencies using KRX TAQ.

Similarly, Panel C of Table 4 reports that liquidity supplying trades of both program trading and non-program trading are negatively correlated with transitory price movements. Though the coefficients are all negative, the coefficients for program trading are greater in magnitude. For Panels B and C, though not statistically significant, the coefficients for program trading are more negative than the coefficients for their non-program trading counterparts.

Table 5 reports the result of the temporary price component for each specification by each size category using index arbitrage (IA) and the other program trading (OP) instead of the program trading (PT) variable.

The negative coefficients of program trading seem to be driven by the effects of other program trading (OP) rather than index arbitrage (IA).

Table 5

Transitory price component of prices and program trading variables (PT)

Panel A: Aggregate program trading and prices

	Units	Large	Medium	Small
ϕ		0.23	0.37	0.48
(t-statistics)		21.77	38.30	39.41
ψ_{IA}	bps/KRW 100m	-2.58	-376.76	-315.19
(t-statistics)		-0.27	-1.55	-1.20
ψ_{OP}	bps/KRW 100m	-61.24	-73.71	-446.17
(t-statistics)		-1.91	-2.19	-6.54
p-value of difference		0.08	0.22	0.63

Panel B: Liquidity demanding program trading, non-program trading, and prices

	Units	Large	Medium	Small
ϕ		0.26	0.41	0.52
(t-statistics)		23.37	42.45	43.78
ψ_{IA}^{D}	bps/KRW 100m	32.49	-272.77	156.47
(t-statistics)		2.90	-1.72	1.08
ψ_{OP}^{D}	bps/KRW 100m	-34.21	-29.83	-305.25
(t-statistics)		-0.94	-0.32	-3.13
ψ_{nPT}^{D}	bps/KRW 100m	24.45	77.33	166.03
(t-statistics)		4.57	4.37	7.98
p-value of difference (IA-nPT)		0.80	0.02	0.95
p-value of difference (OP-nPT)		0.06	0.48	0.00
p-value of difference (IA-OP)		0.03	0.11	0.00

Panel C: Liquidity supplying program trading, non-program trading, and prices

	Units	Large	Medium	Small
ϕ		0.26	0.42	0.53
(t-statistics)		23.75	42.51	43.98
ψ_{IA}^{S}	bps/KRW 100m	-16.42	-159.00	-216.66
(t-statistics)		-0.90	-0.49	-1.06
ψ_{OP}^{S}	bps/KRW 100m	-72.66	-307.65	-627.26
(t-statistics)		-1.89	-1.92	-1.49
ψ_{nPT}^{S}	bps/KRW 100m	-25.19	-93.41	-163.95
(t-statistics)		-5.46	-3.95	-7.94
p-value of difference (IA-nPT)		0.80	0.82	0.89
p-value of difference (OP-nPT)		0.17	0.47	0.23
p-value of difference (IA-OP)		0.11	0.61	0.28

The model is estimated by stock-day using order flow variables to decompose the observed mid-quote price $p_{i,t}$ for stock *i* at time *t* into two components: the information part $m_{i,t}$ and the pricing error part $\mathrm{s}_{\mathrm{i,t}}$:

$$p_{i,t} = m_{i,t} + s_{i,t},$$
$$m_{i,t} = m_{i,t-1} + w_{i,t}$$

Specification for Panel A. Aggregate model

$$w_{i,t} = \kappa_{i,IA}\widetilde{IA_{i,t}} + \kappa_{i,OP}\widetilde{OP_{i,t}} + \mu_{i,t},$$
$$s_{i,t} = \phi_i s_{i,t-1} + \psi_{i,IA}IA_{i,t} + \psi_{i,OP}OP_{i,t} + \nu_{i,t}$$

Specification for Panel B. liquidity demand model

$$w_{i,t} = \kappa_{i,IA}^{D}\widetilde{IA_{i,t}^{D}} + \kappa_{i,OP}^{D}\widetilde{OP_{i,t}^{D}} + \kappa_{i,nPT}^{D}\widetilde{nPT_{i,t}^{D}} + \mu_{i,t},$$
$$s_{i,t} = \phi_i s_{i,t-1} + \psi_{i,IA}^{D}IA_{i,t}^{D} + \psi_{i,OP}^{D}OP_{i,t}^{D} + \psi_{i,nPT}^{D}nPT_{i,t}^{D} + \nu_{i,t}$$

Specification for Panel C. liquidity supply model

$$w_{i,t} = \kappa_{i,IA}^{S}\widetilde{IA_{i,t}^{S}} + \kappa_{i,OP}^{S}\widetilde{OP_{i,t}^{S}} + \kappa_{i,nPT}^{S}\widetilde{nPT_{i,t}^{S}} + \mu_{i,t},$$
$$s_{i,t} = \phi_i s_{i,t-1} + \psi_{i,IA}^{S}IA_{i,t}^{S} + \psi_{i,OP}^{S}OP_{i,t}^{S} + \psi_{i,nPT}^{S}nPT_{i,t}^{S} + \nu_{i,t}$$

$IA_{i,t}$ is index arbitrage's overall order flow, $OP_{i,t}$ is other program trading's overall order flow; $\widetilde{IA_{i,t}}$ and $\widetilde{OP_{i,t}}$ are the surprise components of those order flows. $IA_{i,t}^{D}$, $OP_{i,t}^{D}$ and $nPT_{i,t}^{D}$ are liquidity demanding order flows of index arbitrage, other program trading, and non-program trading; $\widetilde{IA_{i,t}^{D}}, \widetilde{OP_{i,t}^{D}}$ and $\widetilde{nPT_{i,t}^{D}}$ are the surprise components of those order flows. $IA_{i,t}^{S}$, $OP_{i,t}^{S}$ and $nPT_{i,t}^{S}$ are liquidity supplying order flows of index arbitrage, other program trading, and non-program trading; $\widetilde{IA_{i,t}^{S}}, \widetilde{OP_{i,t}^{S}}$ and $\widetilde{nPT_{i,t}^{S}}$ are the surprise components of those order flows. Each stock is in one of three market capitalization categories: large, medium, or small. Columns 3-5 report coefficients for the entire sample at one-minute frequencies using KRX TAQ. p-value of pairwise differences in parameters are shown at the bottom of each panel.

Considering the dominant role of other program trading in the recent surge of program trading, the effect of program trading in price discovery with respect to adjusting pricing error may be getting worse.

6. Conclusion: Who are behind the scenes?

The results of the state space model imply that overall program trading plays a minimal role in updating information in stock prices. When submitting liquidity demanding orders, program trading is less informed or not informed. When submitting liquidity supplying orders, program trading is more adversely selected, showing that program trading does not care about or is bad at placing liquidity supplying orders. Whatever the reason is, program trading's low involvement in information aggregation is still valid.

Contemporaneous pricing error is negatively related to the order flow of program trading for all specifications and all size categories. Program trading is poor at exploiting temporary pricing error; therefore, it also plays a minimal role in price discovery with respect to adjusting pricing error.

Everything put together, program trading is less informed about information update and pricing error level, and it suffers more from adverse selection due to poor book management. Then, why do traders make trades through program trading? In the KRX sample, program trading is made almost entirely by institutional investors (40.4%) and foreign investors (57.8%). And they seem to trade to manage their risk exposure after selling related products. The growth of ETF is regarded as the driver of the rise of program trading.

Then, why do individual investors buy market-wide investment instruments? Caskey (2009) presents a model where ambiguity-averse investors prefer to trade based on an aggregate signal that reduces ambiguity at the cost of a loss in information. This leads to equilibrium prices failing to impound publicly available information. While this creates profit opportunities for ambiguity-neutral investors, ambiguity-averse investors perceive that the benefit of ambiguity reduction outweighs the cost of trading against investors who have superior information. In this context, ambiguity-averse investors trade on aggregate, and less informative at firm-level, signal. This leads to program trading which is less informed about the firm-level fundamentals and also less informed about firm-level pricing errors.

In a laboratory portfolio choice experiment, Ahn, Choi, Gale, and Kariv (2014) estimate ambiguity-averse subjects as 10.4% of the sample. Using a representative US household survey, Dimmock, Kouwenberg, Mitchell, and Peijnenburg (2016) estimate that 52% of the respondents are ambiguity-averse, which is large enough to make pricing impacts in the real world.

References

Ahn, D., Choi, S., Gale, D., & Kariv, S. (2014). Estimating ambiguity aversion in a portfolio choice experiment. *Quantitative Economics*, 5(2), 195-223.

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-frequency trading and price discovery. *Review of Financial Studies*, 27(8), 2267-2306.

Brogaard, J., Hendershott, T., & Riordan, R. (2016). Price discovery without trading: Evidence from limit orders. Available at SSRN 2655927.

Caskey, J. A. (2009). Information in equity markets with ambiguity-averse investors. *Review of Financial Studies*, 22(9), 3595-3627.

Dimmock, S. G., Kouwenberg, R., Mitchell, O. S., & Peijnenburg, K. (2016). Ambiguity aversion and household portfolio choice puzzles: Empirical evidence. *Journal of Financial Economics*, 119(3), 559-577.

Harris, L., Sofianos, G., & Shapiro, J. E. (1994). Program trading and intraday volatility. *Review of Financial Studies*, 7(4), 653-685.

Hasbrouck, J. (1991). Measuring the information content of stock trades. *The Journal of Finance*, 46(1), 179-207.

Hasbrouck, J. (2002). Stalking the “efficient price” in market microstructure specifications: an overview. *Journal of Financial Markets*, 5(3), 329-339.

Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review*, 35(4), 519-530.

Hendershott, T., & Menkveld, A. J. (2014). Price pressures. *Journal of Financial Economics*, 114(3), 405-423.

Kawaller, I. G., Koch, P. D., & Koch, T. W. (1987). The temporal price relationship between S&P 500 futures and the S&P 500 index. *The Journal of Finance*, 42(5), 1309-1329.

국문 초록

프로그램매매가 가격발견기능에 미치는 영향

서울대학교 대학원

경영학과 재무금융전공

강준석

본 연구는 프로그램매매가 가격발견기능에 미치는 영향을 살펴본다. 프로그램매매는 비프로그램매매에 비해 정보가 가격에 반영되는 과정과 가격 오차가 수정되는 과정에서 소극적인 역할을 한다. 정보 반영 측면에서 유동성 수요 프로그램매매는 유동성 수요 비프로그램매매보다 정보에 덜 기반하며, 유동성 공급 프로그램매매는 유동성 공급 비프로그램매매에 비해 역선택에 취약하다. 가격 오차 수정 과정에서 프로그램매매는 고평가된 주식을 매수하며, 저평가된 주식을 매도한다. 프로그램매매를 구성하는 지수차익거래와 비차익거래는 유사한 매매 특성을 지닌다.

주요어: 프로그램매매, 가격발견기능, 상태공간모형

학번: 2015-20574