

전자주문집계장의 가격발견 기능과 유동성 외부효과

이우백*, 최 혁**

<요 약>

본 논문은 2002년 1월부터 2004년 11월까지의 기간동안 유가증권시장에 상장된 460개의 종목을 대상으로 주문집계장에서 최우선호가 이하 단계의 가격발견(price discovery)기능이 시장내에서 횡단면적으로 유동성과 유의적인 관련이 있음을 제시한다. 선행연구인 이우백-최 혁(2006a, 2006b)의 방법론인 벡터오차수정모형을 적용하여 주문집계장의 호가의 가격발견에 대한 공헌도를 추정한 결과, 최우선 호가 이하 단계에서 가격발견의 주도권은 거래량 규모에 따라 체계적으로 증가하는 패턴이 나타났다. 즉, 유동성 소비인 거래량이 높은 종목일수록 주문집계장에서 최우선호가 이하 단계에서 유동성이 충분히 공급된 결과이므로, 이는 가격발견의 효율성을 제고시킨다고 할 수 있다. 유동성이 충분하지 않을 경우 최우선호가 이하에서 투자자들은 높은 거래비용을 부과하게 되며, 이러한 이유로 가격발견과정은 지체되고 비효율적이다. 이러한 실증분석 결과는 호가공개의 범위가 모든 종목에 대해 10단계로 동일하게 적용되지만, 실질적인 가격발견의 주도권은 그 종목의 유동성에 따라 차별화되고 있음을 제시한다. 또한 가격발견의 횡단면의 결정요인을 분석한 결과는 가격발견이 주문집계장의 유동성 변수에 의해 유의적으로 설명됨을 보여 유동성이 높을수록 시장의 질적 수준을 향상시키는 유동성 외부효과(liquidity externality)가 존재함을 제시한다.

핵심단어 : 주문집계장, 가격발견, 유동성 외부효과, 호가공개

* 한국방송통신대학교 경영학과

**서울대학교 경영대학

1. 서론

이 논문은 한국유가증권시장에서 전자주문집계장(electronic limit order book)이 가진 가격발견 기능과 유동성 외부효과(liquidity externality)간의 관계에 대해 실증적으로 분석하였다. 최근에 들어 다수의 증권거래소들은 주문주도형 시장에 기초한 매매방식으로 전환하고 있으며, 주문집계장의 공개범위를 확대하여 거래전 정보의 투명성을 제고시키고자 하는 추세이다.¹ 이러한 주문주도형 시장에서 주문집계장은 유동성 외부효과 또는 네트워크 효과(network effect)를 창출하는 기능을 한다. 특히 주문주도형 시장 설계(market design)에서 유동성 외부효과는 중요한 이슈이다. 유동성 외부효과란 거래자들이 특정한 시간이나 공간에 군집하여 참여함에 따라, 탐색 및 거래비용이 절감됨을 일컫는다. 유동성 외부효과는 유동성이 높은 시장일수록 거래자가 매매에 활발히 참여함으로써 매매성과의 질을 향상시키고, 제고된 시장의 질적 수준으로 인해 거래자들이 다시 매매에 참여하게 한다는 순환적 순기능을 의미한다. 여기에서 매매성과의 질은 암묵적인 거래비용, 매매체결의 효율성, 가격발견의 효율성 등을 포함한다.

본 논문과 관련된 선행 연구들은 비교적 최근에 진행되었으며, 그 특성상 두 가지로 구분된다. 우선적으로 첫번째 연구들은 전자주문집계장과 관련된 핵심적 이슈인 과연 공개단계의 확대가 시장의 질을 긍정적으로 제고시키는 역할을 하는가에 대한 분석이다. 또한 두번째 부류의 논문들은 시장의 질이 유동성 외부 효과의 다각적인 차원에 의해 설명이 될 수 있는가와 관련된 연구이다.

시장에서 유동성 외부효과가 존재하는지에 대해서는 다수의 연구들은 공간적(지역 시장) 및 시간적(거래시간)으로 차원에서 실증적으로 시도한 바 있으며, 또한 특정한 유동성과 관련한 이벤트에 대해서도 연구가 이루어졌다. Hasbrouck(1995)는 NYSE 가 다른 지역 시장(regional market)보다도 가격발견의 유의적으로 높은 주도권을 가지며, 이는 NYSE 의 거래량이 타 시장에 비해 지배적으로 높은 사실과 관련 있음을 제시했다. Huang(2002)은 ECN 과 전통적 시장조성인이 NASDAQ 에 제출하는 호가의 질을 분석한 결과 ECN 의 호가는 정보적(informative)일뿐만 아니라 ECN 과

¹ 세계 42 개국 증권거래소의 매매시스템 유형을 조사한 바에 따르면, 주문주도형 시장의 운영방식인 전자주문집계장을 채택한 거래소가 79%이며, 전자주문집계장에서 거래 전 정보인 잔량을 일부 또는 전부 공개하는 거래소는 95%에 이르는 것으로 나타났다. (World Federation of Exchange Trading Survey 2003, WFE)

같은 주문집계장의 공개정도가 높을수록 호가의 질을 향상시키는데 공헌한다고 주장했다. Hendershott-Jones(2005)은 전자주문집계장인 Island 가 주문공개를 철폐함에 따라 ETF의 거래량과 가격발견이 감소하고, ETF의 거래비용도 증가함으로써 Island의 질적 수준은 저하되었다고 주장한다. Madhavan-Porter-Weaver(2005)는 토론토 주식 시장에서 주문원장의 4 단계 호가와 잔량이 공개된 후 체결비용과 변동성이 증가하고 유동성이 감소하는 등 시장의 질이 저하되어 오히려 투명성 정책의 목적과 상반된 효과를 초래했다고 결론을 내린다. Barclay-Hendershott (2004)은 매매시간과 관련한 유동성 외부효과를 NASDAQ의 정규 매매시간 이후의 시간의 매매의 거래비용을 대상으로 분석했다.

한국주식시장과 관련한 선행연구인 이우백-최 혁(2006a)은 2002 년도에 시행된 주문집계장의 신규공개단계 범위에서 가격발견 기능이 제한적으로 향상되었음을 보고한다. 이러한 결과는 호가공개를 중단한 결과, 거래집중도가 높았던 시장에서 거래량과 가격발견이 저조해지고 거래비용이 증가했음을 제시한 Hendershott-Jones(2005)을 지지한다. 이우백-최 혁(2006b)은 가격발견이 하루중 주문집계장에서 형성되는 유동성의 형태와 높은 상관관계를 가짐을 제시했다.

기존 논문들이 유동성 외부효과와 관련된 시장의 질적 변수를 주로 거래비용에 초점에 두고 있으나, 본 논문은 주문집계장에서 추론할 수 있는 최우선호가 이하 단계의 효율성이 주문집계장의 형태에 내재된 유동성에 의해 횡단면적으로 설명되는가를 분석했다. 최근 저유동성 종목에 대한 관리가 이슈화되고 있는 현실에 비추어, 이 논문이 분석한 유동성 수준과 호가의 가격발견과의 실증적 관계는 정책적 시사점을 제공할 것이다.² 이 논문에서 표본은 2002년 1월부터 2004년 11월까지의 기간에 걸쳐 460 개의 종목으로 구성되었으므로, 일중 자료로는 장기간에 속한다. 따라서 다른 선행 연구에 비해 추정치의 경제적 및 통계적 유의성을 검증할 수 있으며 장기간에 걸친 횡단면-시계열적 변동요인에 대한 분석이 가능하다. 또한 연구에 전자적으로 처리된 특유한 데이터베이스에 기초한 실증적 결과는 추정의 신뢰성을 보장할 수 있다.

² 한국증권선물거래소는 저유동성기업의 유동성 지표 악화현상이 심화됨에 따라 저유동성종목을 중심으로 거래형성일, 거래대금, 가격연속성, 호가스프레드 등의 유동성 수준을 공개하고, 저유동성종목을 별도로 선정해, 이들 종목에 대해 매매방식을 연속경쟁매매에서 단일가경쟁매매로 변경하는 방안도 추진할 것이라고 밝힌바 있다.(한국증권선물거래소 보도자료, 2007.3.14)

본 논문은 주문집계장에서 최우선호가 이하 단계의 가격발견 기능이 시장내에서 횡단면적으로 유동성과 유의적인 관련이 있음을 제시한다. 선행연구인 이우백-최혁(2006a), 이우백-최혁(2006b)의 방법론을 적용하여 추정된 결과, 최우선 호가 이하 단계에서 가격발견의 주도권은 거래량 규모에 따라 체계적으로 증가하는 패턴이 나타났다. 즉, 유동성 소비인 거래량이 높은 종목일수록 주문집계장에서 최우선호가 이하 단계에서 유동성이 충분히 공급된 결과이므로, 이는 가격발견의 효율성을 제고시킨다고 할 수 있다. 유동성이 충분하지 않을 경우 최우선호가 이하에서 투자자들은 높은 거래비용을 부과하게 되며, 이러한 이유로 가격발견과정은 지체되고 비효율적이다. 이러한 실증분석 결과는 호가공개의 범위가 종목별로 10 단계로 동일하게 적용되지만, 실질적인 가격발견의 주도권은 그 종목의 유동성에 따라 차별화되고 있음을 제시한다. 또한 Fama-MacBeth (1973)와 Easley et al. (2002)의 방법론을 따라 가격발견의 횡단면의 결정요인을 분석한 결과도, 주문집계장에서 나타나는 유동성 변수에 의해 유의적으로 설명됨을 보였다.

이하에서 전개될 논문의 순서는 다음과 같이 구성된다. 제 2 장에서는 실증분석에서 사용된 표본자료를 설명하고, 한국유가증권시장의 주문집계장을 소개한다. 제 3 장은 가격발견 측정에 사용된 공헌도인 Hasbrouck (1995)의 방법론을 소개한다. 제 4 장과 제 5 장은 제 3 장의 내용을 기초로 주문집계장에서 최우선 호가 이하 단계에 초점을 두어 가격발견의 기능을 분석하고 유동성과의 관계를 살펴본다. 제 6 장에서는 실증적 분석결과와 의미를 요약하고 이 연구가 가지는 정책적 시사점에 대해 결론을 내리기로 한다.

2. 분석자료와 주문집계장의 특징

2.1. 표본기간 및 종목, 분석자료

본 연구에 사용된 표본종목들은 한국증권선물거래소(Korea Exchange)의 유가증권시장에 상장된 종목들이다. 유가증권시장은 시장조성인(market maker)이 존재하지 않는 순수한 지정가 주문시장으로 주문제출부터 매매체결, 결제에 이루어지는 모든 과정이 전산시스템을 통하여 이루어진다. 따라서 유동성은 투자자들로부터 공급되며, 주문집계장도 모든 투자주체들에게 최우선호가부터 10 단계호가까지 동등하게

공개된다. 주문의 종류에는 지정가주문, 시장가주문, 조건부지정가 주문이 있으며, 2004년부터는 최유리지정가호가와 최우선지정가호가 등이 신설되어 호가종류가 매우 세분화되어 있다. 이외에도 조건을 부여한 주문인 IOC 주문(Immediate-Or-Cancel Order)과 FOK 주문(Fill-Or-Kill Order)이 있다. 유가증권시장에서 호가는 주문을 제출한 당일에만 유효하고, “미공개요청 주문(Hidden Order)”이 없으므로 거래 전 정보의 투명성은 매우 높다.³

한국유가증권시장은 1979년 7월 이후 2003년말까지 10차례의 호가정보의 공개범위를 확대하였으며, 최우선호가 이하의 단계가 실질적으로 외부에 공개된 시기는 1997년 4월 이후이다. 2002년 1월 2일 이후에는 접속매매시간동안 우선 5단계호가 공개에서 우선 10단계호가로 공개범위를 확장하는 대신 총수량 공개를 폐지하고 10단계의 합산수량에 관한 정보만을 공개하고 있다. 집중매매시간에서는 2003년 10월 6일부터 예상체결가격, 예상체결수량 및 매도 매수별 3단계 예상호가의 가격과 그 가격에서의 호가수량을 발표하고 있다.

실증분석의 표본 기간은 접속매매시간의 주문집계장 공개범위가 현행 10단계 공개체계로 변경된 2002년 1월부터 2004년 11월까지 35개월이다. 이우백-최 혁(2006a)에 따르면 새로운 호가공개체계의 도입에 따라 거래자들은 자신의 주문전략을 조정시키고, 이전 5단계 호가공개체계 기간과는 상이한 특성을 보이기 때문에 2002년 1월 이전 자료는 제외했다. 이전 기간이 포함된다면 실증분석 결과에서 혼동효과(confounding effect)가 내포될 것이며, 본 연구의 목적은 선행 연구와 같이 이벤트 전후로 가격발견을 비교하는 것에 초점을 둔 것이 아니기 때문이다.

본 연구의 표본 종목은 표본기간동안 유가증권시장에 연속적으로 상장된 모든 종목들 중에서 저유동성이나 재무적 부실의 이유로 인해 관리종목에 편입된 경험이 없었던 498 종목을 일차적으로 선별했다. 또한 유동성 변화를 초래할 것으로 판단되는 사건인 주식분할이나 주식합병, 유상증자, 무상증자를 표본기간동안 시행한 기업들은 표본종목에서 제외시켰다. 이러한 과정을 거쳐 최종적으로 선정된 종목은 460 개이며, 표본종목의 일평균 전체 시장가치는 유가증권시장의 71.05%를 차지한다.

2.2. 자료

³ 미공개주문(Hidden order)는 주문의 일부수량을 주문집계장에 노출시키지 않을 것을 요구하는 주문으로 Iceberg order 이라고 한다.

본 논문의 실증분석에서 활용된 주요한 자료는 두 가지이다. 일중 주문집계장 및 정보량을 추정하기 위한 자료는 IFB/KSE 데이터베이스이며, 횡단면 회귀분석 수행에 필요한 종목별 회계 및 발행주식수 자료는 FnGudie-Pro에 접속하여 제공받았다. IFB/KSE 데이터베이스는 완전하게 전산화된 주문시스템의 자료로 크게 주문자료, 체결자료, 주문집계장의 3부분으로 구성되어 있다. IFB/KSE 데이터베이스는 1993년부터 2004년까지 유가증권시장에 상장된 모든 종목의 개별 주문과 거래에 대해 1/100초까지 주문집계시각과 매매체결시각이 기록된다. 체결자료에는 체결수량, 체결가격과 각 체결건에 대해 매도 및 매수접수 번호, 매도 및 매수주문수량 등의 정보가 포함된다. 주문자료에는 각 개별 주문건에 대해서 매수 또는 매도 여부를 구분할 수 있으며, 주문수량과 주문가격이 기록된다. 또한 원주문접수번호, 원주문가격, 원주문수량, 원주문잔량 등과 각 주문건에 대해 정정, 취소 수량이 기록되기 때문에 이를 기초로 특정 시점에서의 매순간마다 주문집계장에 담긴 각 단계별 호가와 잔량 데이터를 구축할 수 있다. 본 연구에서 사용한 자료인 주문집계장은 이우백-최혁(2006a, 2006b)처럼 별도로 체결자료와 주문자료를 이용하여 구축하지 않고 직접적으로 주문집계장 자료를 이용했다.⁴ 이러한 주문집계장은 선행연구인 Cao et al. (2005), Kavajecz-Odders-White(2004)가 체결거래와 주문자료를 이용해서 구축한 주문집계장 자료에 비해, 구축 과정에서 발생할 가능성이 있는 오류를 최소화시키고 자료의 정확성을 보장한다는 점에서 장점을 가진다.

2.3. 주문집계장의 형태

거래자의 매수 및 매도 가격과 수량의 스케줄인 주문집계장의 형태를 결정하는 요인은 단계간 거리(step)와 잔량(depth)이다. 거리는 주문집계장에서 특정 호가단계와 인접한 호가단계간의 가격으로 표현된 차이로 측정되며, Kyle(1985)이 제시한 유동성의 개념 중 ‘tightness’와 직접적으로 관련된 개념이다. 유동성이 활발할수록 단계간의 차이인 거리는 축소되고, 특정 단계의 호가는 더욱 경쟁적(competitive)일 것이다. 선행 연구들은 유동성을 측정하는 지표로 주로 주문집계장의 최우선평가의

⁴ IFB/KSE 데이터베이스에서 주문집계장 자료는 공개범위만이 기록되기 때문에 2006년 1월 이전 자료는 5단계 호가와 잔량만이 존재했다. 이우백-최혁(2006a, 2006b)의 연구에서는 10단계 호가 자료가 필요하므로, 주문자료와 체결자료를 이용해서 주문집계장을 구축했지만, 이 논문은 호가공개범위 자료만을 이용하므로 그럴 필요가 없다.

스프레드(spread)에 초점을 두고 있다. 하지만 매매체결의 상당부분도 최우선평가가 이하에서도 성립되므로, 최우선평가가 이하에서 영역을 확대시켜 유동성을 측정하는 지표가 필요하다. 이러한 경우, 최우선평가가 이하에서 측정되는 거리는 거래비용(execution cost)의 의미를 나타낸다.

<표 2>는 표본기간동안 투자자가 실제로 관찰하는 공개 주문집계장의 최우선 단계와 그 이하 특정 단계의 거리를 나타낸다. 가격대에 따라 최소호가단위가 상이하므로, 특정 단계의 거리는 최우선평가와 특정 단계 호가의 차이를 해당 가격대에 대응하는 호가단위로 나누어 표준화시켰다. 즉, 개별 종목에 대해 일별로 특정 단계의 표준화거리(normalized distance)를 측정하여 월평균을 산출하고, 전월 월별 거래량 규모에 따라 매월 5 개의 포트폴리오로 구성했다. 표에 제시된 표준화거리는 횡단면 평균의 시계열 평균과 표준편차이다.

$$ND_s^j = \frac{|P_s^1 - P_s^j|}{tick_p} \quad (1)$$

여기에서 $s \in \{ask, bid\}$

$j = 2, 3, \dots, 10$

<표 1>에서 ND_s^j 는 낮은 거래량 종목일수록 그 값이 확대되므로 매매활동이 부진한 종목일수록 높은 거래비용을 부담함을 의미한다. 거래량이 높은 종목으로 구성된 포트폴리오일수록, ND_s^j 는 축소되는 패턴을 보인다. 결과적으로 유동성이 집중된 종목일수록 거래비용을 적게 부담하므로, 정보거래자는 참여할 유인이 높다 진다고 볼 수 있다. 저유동성 종목에서 나타나는 특징은 상위 단계 보다는 하위 단계일수록 ND_s^j 가 급격히 확대된다는 것이다. 예를 들어, 패널 A의 포트폴리오 1의 ND_{bid}^2 는 2.60으로 포트폴리오 5의 2.36 배이지만, ND_{bid}^{10} 은 3.65 배로 증가한다.

<표 2>에서 5 분 간격별로 측정된 10 단계별 잔량 규모는 거래량과 잔량간에 횡단면적으로 강한 상관관계가 존재함을 제시한다. 전체적으로 상위 단계인 5 단계 이내에서는 매수측 잔량이 매도측 잔량보다 깊게 형성된 반면, 매도측 잔량은 상위 단계보다는 하위 단계인 5 단계 이하에 집중됨을 보여준다. 예를 들어, 패널 A의 매수측 잔량을 볼 때 최우선단계의 잔량은 최저 거래량 포트폴리오 1의 477 주에서

최고 거래량 포트폴리오 5 의 19,405 주까지 증가하는 패턴을 보인다. 대량 주문이 제출될 경우, 최우선호가의 잔량 규모가 대량 주문을 흡수할 정도로 충분히 크다면, 가격충격(price impact)은 낮고, 시장연속성(price continuity)은 높을 것이다. 반면에, 주문집계장의 잔량 규모가 낮은 경우 거래자들은 자신들이 부담할 시장충격비용을 부담하므로 거래에 참가할 유인이 낮아질 것이다. 따라서 횡단면적으로 시장의 심도(depth)를 나타내는 잔량과 거래로 실현된 유동성 지표인 거래량간에는 양의 관계를 가지며, 잔량의 규모는 가격충격을 차별적으로 흡수할 수 있는 능력으로 해석된다고 할 수 있을 것이다. 유동성 소비에 대한 유동성 공급의 관계를 동시에 고려한 유동성 지표는 다음과 같이 표현될 수 있다.⁵

$$LP_s^j / LC = \frac{depth_s^j}{volume} \quad (2)$$

여기에서 $s \in \{ask, bid\}$

$j = 2, 3, \dots, 10$

<표 3>은 유동성 소비(liquidity consumption)에 대한 유동성 공급(liquidity provision)의 비율이 거래량 규모에 따라 큰 차이가 발생함을 나타낸다. 주문집계장의 각 단계의 잔량을 일별 거래량으로 표준화시킨 상대적 잔량(relative depth)인 LP_s^j / LC 은 <표 3>에 제시된 포트폴리오와 관계와는 다른 패턴을 보인다. 주문집계장의 매수측에서 모든 종목의 평균적 LP_B^j / LC 는 최우선호가에서 가장 높으며, 그 이하 단계에 따라 단조적으로 감소하는 형태를 보인다. 반면에 매도측의 LP_A^j / LC 는 5 단계 이내에서 큰 값이 나타나며, <표 2>에서 제시된 결과와는 다소 다른 형태를 나타낸다. 예를 들어, 매수측 최우선호가의 LP_B^1 / LC 는 포트폴리오 1 에서 19.03 이지만 포트폴리오 5 에서는 1.31 까지 감소한다. 또한 포트폴리오 1 과 포트폴리오 2 간의 LP_s^j / LC 의 차이는 다른 포트폴리오들의 LP_s^j / LC 간의 차이보다 상당히 크다.

⁵ 유동성 공급은 지정가 주문흐름(limit order flow)의 개념이지만, 본 논문에서는 지정가 주문흐름의 집합으로부터 시장가 주문흐름(market order flow)이 유동성을 소비하고 주문집계장에 형성된 잔량을 유동성 공급의 대응치로 본다.

따라서 매수측과 매도측의 5 개 포트폴리오를 본다면, 유동성 소비 규모와 LP_s^j / LC 간에는 음의 관계가 존재함을 보여준다. 유동성 공급 규모에 비례해서 유동성을 소비한다면 각 포트폴리오의 LP_s^j / LC 는 일정한 값을 가질 것이다. 하지만 LP_s^j / LC 가 감소하는 패턴은 주문집계장에서 유동성 공급 규모가 높을수록, 보다 큰 규모의 유동성 소비자가 거래에 참여시킬 가능성을 나타내므로 비선형적인 관계를 보여준다. 결과적으로 주문집계장의 최우선단계 이하에서 관찰되는 낮은 거래비용과 높은 유동성 공급 규모는 유동성 소비를 증가시킬 것이며, 이러한 상호작용에 의해 호가의 경쟁과 질은 제고될 것이라는 추측을 할 수 있다.

3. 가격발견 측정의 방법론⁶

본 장에서는 주문집계장에서 각 단계의 호가와 잔량이 가격발견 과정에서 기여하는 정도를 측정하는 방법론인 Hasbrouck(1995)이 제시한 정보량(Information share)을 소개하기로 한다.⁷ 정보량은 지정가들의 상호작용에서 유입된 정보가 균형가격으로 수렴하는 가격발견과정에서 특정 단계의 지정가가 차지하는 상대적 공헌도이며, 다음과 같이 벡터오차수정모형(Vector Error Correction Model : VECM)을 추정하여 산출된다.

$$\Delta p_t = \alpha \beta' p_{t-1} + \sum_{j=1}^k A_j \Delta p_{t-j} + e_t \quad (3-A)$$

$$\Delta p_t = \Psi(L)e_t \quad (3-B)$$

$$p_t = \Psi(1)\left(\sum_{s=1}^t e_s\right) + \Psi^*(L)e_t \quad (3-C)$$

$$p_t = \psi\left(\sum_{s=1}^t e_s\right)t + \Psi^*(L)e_t \quad (3-D)$$

VECM에서 p_t 는 가격벡터이며, 오차조정계수인 α 는 변수 상호간의 feedback의 크기로 균형가격에 수렴하는 속도를 의미한다. $\Psi(1)e_t$ 는 각 호가에서 생성된 이노베

⁶ 본 절의 내용에 대해서는 이우백-최 혁(2006a, 2006b)을 참조할 것.

⁷ Hasbrouck 의 정보량은 벡터오차수정모형을 적용하여 여러 시장에서 거래되는 단일한 주식에 대해 특정 시장의 가격발견을 측정한 것으로, Hasbrouck (1995, 2002), Huang (2002), Cao et al. (2005), Hendershott-Jones(2005)에서 활용되었다.

이션(innovation)의 장기적 충격을 의미하며 정보의 유입에 의하여 가격변동에서 주가에 지속적으로 반영되는 부분으로 해석된다. 따라서 가격변동으로 인한 이노베이션이 전적으로 정보에 의해 발생한다면, $\Psi(1)(\sum_{s=1}^t e_s)$ 에서 매도 및 매수진동(bid-ask bounce)과 같은 일시적인 시장미시구조적인 효과는 제거된다. 벡터오차수정모형에서 Ω 가 이노베이션의 분산-공분산 행렬이라 할 때, 공통확률추세(common stochastic trend)의 총변동성은 $\psi\Omega\psi'$ 이므로 총변동성 중에서 j 번째 변수에 귀속될 수 있는 부분의 비중은 다음과 같이 산출된다.

$$S_j = \frac{\psi_j^2 \Omega_{jj}}{\psi\Omega\psi'} = \frac{\psi_j^2 \sigma_j^2}{\sum_{i=1}^n \psi_i^2 \sigma_i^2} \quad (4)$$

Hasbrouck(1995)이 제시한 정보량(information share)은 바로 (4)와 같이 가격발견에 대한 상대적 기여도(relative contribution to price discovery)로 측정되며, 이 값이 높을수록 가격발견의 주도권(price leadership)을 가진다고 해석할 수 있는 것이다. 실제적인 정보량 계산에 있어서, 이노베이션간의 높은 상관관계가 존재하는 경우 정보량을 측정하기 위해서는 출레스키 요인분해를 이용하여 상관관계를 제거시킨다. (5)에서 F 는 고유한 하삼각행렬(unique lower triangular matrix)이므로 이노베이션은 다음과 같이 직교화된다.

$$\Omega = FF' \quad (5)$$

$$e_t = Fz_t, \quad \text{여기에서 } Ez_t = 0, \quad Var(z_t) = I$$

특정변수의 정보량의 값은 벡터오차수정모형에서 변수들간의 배열순서를 어떻게 결정하느냐에 따라 결정된다. 일반적으로 벡터오차수정모형에서 특정 변수가 순서상 처음 배열될 경우에 정보량의 상한값(upper bound)을 가지며, 가장 나중에 배열될 경우에 하한값(lower bound)를 가진다. 일반적으로 n 개 변수로 구성된 벡터오차수정모형에서 변수의 순서 치환에 따라 각 변수 정보량의 상한값과 하한값은 (6-A), (6-B)로 산출되며, 해당 변수의 정보량은 순서 치환(rotation)에 따라 산출한 정보량

을 평균한 값으로 정의한다.⁸

$$S_1 = \frac{(\psi[F]_1)^2}{\psi\Omega\psi'} = \frac{[\sum_{i=1}^n \psi_i f_{i1}]^2}{[\sum_{i=1}^n \psi_i f_{i1}]^2 + [\sum_{i=2}^n \psi_i f_{i2}]^2 + \dots + (\psi_n f_{nm})^2} \quad (6-A)$$

$$S_n = \frac{(\psi[F]_n)^2}{\psi\Omega\psi'} = \frac{(\psi_n f_{nm})^2}{[\sum_{i=1}^n \psi_i f_{i1}]^2 + [\sum_{i=2}^n \psi_i f_{i2}]^2 + \dots + (\psi_n f_{nm})^2} \quad (6-B)$$

실제적으로 정보량을 추정하기 위한 벡터오차수정모형은 주문집계장의 공개범위 호가정보를 집약하는 변수로 구성되며 Hasbrouck (1995)이 제시한 방법을 적용하여 추정하였다.

Roll(1984)의 모형이 발표된 이후에 시장미시구조를 대상으로 하는 실증연구는 대부분 최우선매도호가와 최우선매수호가의 중간가격(midpoint price)을 주가의 진실한 가격(true price)의 대용치로 간주한다. 본 논문에서는 최우선호가의 중간가격과 Cao et al.(2005), 이우백-최 혁(2006a, 2006b)가 제시한 잔량가중평균가격(depth-weighted average limit price)도 함께 사용한다.⁹ 잔량가중평균가격은 가격변동 뿐만 아니라 취소 및 정정행위와 관련된 잔량의 정보내용도 반영하여 가격추론에 대한 합의를 나타내는 대용치로 사전적인 평균 체결비용(ex-ante cost)을 나타낸다. 주문집계장에서 k단계의 지정가와 잔량이 각각 P^k 와 Q^k 라 할 때, 최우선호가의 중간가격(MP^1)과 잔량가중평균가격(WP^1)은 다음과 같이 정의된다.

$$MP^1 = \frac{P_{ask}^1 + P_{bid}^1}{2} \quad (7-A)$$

$$WP^1 = \frac{P_{ask}^1 Q_{ask}^1 + P_{bid}^1 Q_{bid}^1}{Q_{ask}^1 + Q_{bid}^1} \quad (7-B)$$

이제 최우선호가 이하의 하위단계 호가의 정보를 집약하는 대용치를 보기로 하

⁸ 2 변량 VECM 의 경우에는 변수 순서 치환의 경우가 2 이므로 정보량은 상한치와 하한치의 중간값으로 산출되지만, 3 변량 VECM 은 변수 순서 치환수가 3!이므로 정보량은 6 개의 정보량을 평균한 값이 된다.

⁹ 이러한 잔량가중평균가격은 실무에서 사용되는 거래량가중평균가격(volume-weighted average price)과 같이 주문원장의 수요 및 공급을 반영하는 지시적 가격(indicative price)이라 할 수 있다.

자. 잔량가중평균호가는 하위 단계인 2~10 단계까지의 잔량과 거리의 정보를 총체적으로 반영한 WP^{2-10} 으로 나타낸다. 이 지표는 주문집계장의 형태인 호가의 위치와 잔량에 의해 영향을 받는다. 즉, 다른 조건이 모두 동일하다면 상위 단계에 잔량이 집중되고 호가간 거리가 근접할수록 WP^{2-10} 는 최우선평가의 변동과 유사할 것이며, 하위 단계에 잔량이 집중되고 호가간 거리의 격차가 심하다면 최우선평가 위치에서 이탈할 것이다. 또한 상대적으로 매수측에 잔량이 집중되고 매수측의 호가간 격차가 높다면 $WP^{2-10} < MP^1$ 이지만, 매도측에 잔량이 집중되고 호가간 격차가 크다면 $WP^{2-10} > MP^1$ 의 관계를 가질 것이다.

$$WP^{2-10} = \frac{\sum_{k=2}^{10} P_{ask}^k Q_{ask}^k + \sum_{k=2}^{10} P_{bid}^k Q_{bid}^k}{\sum_{k=2}^{10} Q_{ask}^k + \sum_{k=2}^{10} Q_{bid}^k} \quad (8)$$

WP^{2-10} 은 특정 단계의 과도히 집중된 잔량이나 호가간 거리에 영향을 받을 수 있으므로 이 같은 효과를 완화하기 위해 2~10 단계를 다시 상위 단계와 하위 단계로 구분해보기로 한다. 상위 단계와 하위 단계는 2~5 단계와 6~10 단계로 구분하여 각 영역의 잔량가중평균가격인 WP^{2-5} , WP^{6-10} 을 산출했다.¹⁰

$$WP^{2-5} = \frac{\sum_{k=2}^5 P_{ask}^k Q_{ask}^k + \sum_{k=2}^5 P_{bid}^k Q_{bid}^k}{\sum_{k=2}^5 Q_{ask}^k + \sum_{k=2}^5 Q_{bid}^k} \quad (9-A)$$

$$WP^{6-10} = \frac{\sum_{k=6}^{10} P_{ask}^k Q_{ask}^k + \sum_{k=6}^{10} P_{bid}^k Q_{bid}^k}{\sum_{k=6}^{10} Q_{ask}^k + \sum_{k=6}^{10} Q_{bid}^k} \quad (9-B)$$

(3-A)의 벡터오차수정모형은 오차수정항과 과거 시차변수를 공통적으로 포함하므로 무관회귀분석(seemingly unrelated regression)으로 추정한다. 오차수정모형에서 최

¹⁰ Cao et al. (2005)이 사용한 총잔량가중평균지정가는 개별 단계의 잔량가중평균지정가를 개별 단계의 잔량으로 다시 가중평균한 지정가이다. 잔량가중평균지정가는 평균적인 가격을 나타낸다는 점에서 직관적으로 이해하기 쉽지만, 특정 단계의 잔량규모의 비중에 영향을 많이 받는다는 단점이 있다. 특히 주문의 공격성이 낮고 잔량이 과도히 누적되어 정보가 지체된 단계가 총잔량가중평균지정가에 반영되면 과소추정의 편의가 발생할 가능성이 높다.

적시차는 Bayesian 정보 기준(Bayesian Information Criteria)에 의해 각 종목별, 월별로 해당하는 시차를 적용했다. 벡터오차수정모형은 접속매매시간 동안 5 분 간격별로 측정된 변수로 구성하였으며, 월별로 추정하였다. <표 4>에서는 비거래(non-trading)의 영향을 파악하고자 측정 간격을 5 분부터 60 분까지 연장할 경우 접속매매시간동안 거래가 체결된 구간의 비율을 나타낸다. 5 분 간격별로 측정할 경우 하루 중 전체 구간 중 80%이상에서 가격이 결정되지만 거래량이 낮은 종목일수록 이 비율은 낮아진다. 측정 구간이 길어질수록 거래 체결 구간 비율은 증가하여, 측정 구간을 60 분으로 설정할 경우에는 모든 포트폴리오에서 체결 비율은 95%이상을 기록한다. 정보의 유입 정도의 관점에서 유동성이 높은 종목과 낮은 종목간의 특성을 구별하기 위해 변수들을 5 분 간격별로 측정하였으며, 측정 구간에 따라 분석결과가 상이한지에 대한 효과는 제 4 장의 강건성 검증에서 재고찰했다.

4. 비최우선호가가는 가격발견에 공헌하는가?

4.1. 호가의 일중 단기적 랜덤워크 특성

주문집계장의 호가들이 주가의 본질가치에 대한 정보내용을 반영하는지를 검증하기 위해, 단위근 검정(unit root test)을 수행했다. 호가에 “효율적 가격(efficient price)”요인이 내포되어 있다면, 호가의 일중 시계열 특성은 근본적으로 불안정성 과정(non-stationary process)을 보이며, 랜덤워크(random walk)를 따르게 된다.¹¹

우선적으로 주문집계장의 호가들이 랜덤워크 과정을 따르는지 확인하기 위해 Engle-Granger(1987)이 제시한 절편이 포함된 Augmented Dickey-Fuller 모형으로 단위근 검정을 수행하였다.¹²

$$\Delta p_t = \alpha + \phi p_{t-1} + \sum_{j=2}^n \gamma_j \Delta p_{t-j} + \eta_t \quad (10)$$

¹¹ 주문집계장의 호가(p_t)는 다음과 같이 표현 가능하다.

$$p_t = m_t + q_t d, \quad m_t = m_{t-1} + u_t$$

여기에서 m_t 는 호가에 내재된 효율적 가격이며, u_t 는 평균이 0이며 i.i.d를 따른다. q_t 는 주문의 방향이며 d 는 효율적 가격과의 거리를 나타낸다.

¹² Dickey-Fuller 검정식에서 오차항의 정규분포 가설이 기각되고 계열 상관성이 존재할 경우에는 Augmented Dickey-Fuller 를 사용하여야 한다.

단위근 검정 모형의 최적시차는 벡터오차수정모형과 마찬가지로 Bayesian 정보 기준을 적용해서 선정되었으며 5분 간격 자료를 이용해서 월별로 추정했다. 단위근에 대한 귀무가설은 $\phi=0$ 이며, Dickey-Fuller(1979)가 제시한 τ -통계치에 대한 임계치로 단위근의 존재여부를 판단한다. 만일 단위근에 대한 귀무가설이 기각되면 호가들은 안정적인 과정을 따르지만, 귀무가설이 기각될 수 없다면 호가들은 랜덤워크를 따른다.

<표 5>는 호가 변수들의 단위근 검정결과를 제시한다. 단위근 검정은 표본기간동안 개별종목별로 매월(month-by-month)수행되었다. 표에 제시된 값들은 단위근 검정을 수행한 결과 전체 종목과 거래량 규모별 포트폴리오의 τ -통계치의 횡단면 평균의 시계열 평균값을 보여준다. %RW는 1%와 5%유의수준에서 귀무가설을 기각할 수 없어서 랜덤워크를 따르는 종목들이 전체 표본종목에서 차지하는 비율이다.

MP^1 의 경우에는 전체 종목×월 표본 중 93.1%가 1%유의수준에서 랜덤워크를 따르는 것으로 나타났으며, 5%유의수준에서는 88.12%가 랜덤워크를 따르는 것으로 보고되었다. 거래량 규모별로 볼 때, 1%유의수준에서는 모든 포트폴리오에서 80%이상의 종목×월 표본이 랜덤워크를 따르지만, 거래량이 높은 기업일수록 랜덤워크의 특성이 강해지는 것이 확인된다. WP^1 에 대한 단위근 검정결과도 MP^1 의 결과와 유사하다. 하지만 나머지 호가 변수인 WP^{2-10} , WP^{6-10} 은 최우선히가에 비해 시계열의 안정성이 높고, 랜덤워크의 특성이 약한 경향을 보였다. WP^{2-10} 가 1% 유의수준에서 랜덤워크를 따르는 표본수는 전체표본 중 90%미만의 비율을 보였으며, τ -통계치의 평균값은 -1.89 이다. WP^{2-10} 를 WP^{2-5} 와 WP^{6-10} 으로 구분한 결과에서는 상위단계의 호가인 WP^{2-5} 는 WP^1 의 랜덤워크 특성과 유사한 통계치를 보이지만, WP^{6-10} 은 랜덤워크 특성이 약해짐을 보인다. 이러한 결과는 공격성이 낮은 하위단계에서는 거래가 지연되거나, 잔량이 오래됨(stale)으로 인해 정보의 효율성이 낮다는 것을 제시한다.

거래량에 의해 분류된 단위근 검정결과의 횡단면적 특성에서는 유동성이 높은 종목일수록 랜덤워크를 따르는 경향이 높다. 최고거래량 포트폴리오(5)에서는, 어떠한 호가 변수에서도 5 개의 포트폴리오 중 귀무가설에 대한 τ -통계치의 평균값이 가장 낮다. 선행 연구들은 고유동성 종목보다는 저유동성 종목에서 정보비대칭이 높은 특성을 보인다고 주장한다. 따라서 단위근 검정결과로 볼 때, 저유동성 종목에

서는 정보의 반영속도가 지연되고, 호가에 내재된 가치는 비효율적으로 형성된다고 볼 수 있다. 하지만 고유동성 종목들의 호가는 대부분 랜덤워크를 따른다는 강한 특성을 나타낸다. 차우선평가가 이하 단계의 어떠한 영역에서도, 단위근이 존재한다는 귀무가설을 기각하지 못한 표본은 포트폴리오 5 에서는 1%유의수준에서 90%이상의 비율을 기록한다. 특히 하위단계의 호가인 WP^{6-10} 가 단위근을 가진 종목 비율은 90.78%이며 τ -통계치는 -1.41 로 나타나 상위단계 호가인 WP^{2-5} 의 시계열 특성과 큰 차이가 없다는 것이 확인된다.

4.2. WP^1 는 균형가격의 대응치인가?

<표 5>에서 제시된 WP^1 과 MP^1 의 단위근 검정결과는 최우선평가의 정보로 표현된 변수인 WP^1 가 횡단면적으로 진정한 주가의 대응치로 역할을 하는가에 대한 의문을 갖게 한다. 전통적으로 시장미시구조의 실증적 연구에서는 주가의 내재가치를 측정하기 위한 대리변수(proxy)는 최우선평가의 중간값으로 사용되었다. 따라서 WP^1 가 MP^1 의 정보내용을 가지는지를 파악하기 위해서는, WP^1 의 가격발견의 정보량과 MP^1 의 가격발견의 정보량의 크기를 비교하면 될 것이다. 이를 검증하기 위해 WP^1 와 MP^1 의 2 변수로 구성된 벡터오차수정모형 (11)로 각 변수의 정보량을 추정했다. (11)에서 오차수정항(z)은 Engel-Granger (1987)의 공적분 관계를 고려해서 최우선단계의 중간가격인 MP^1 에 WP^1 를 회귀분석해서 추정한 잔차항이다.

$$\begin{aligned} \Delta MP_t^1 &= a_{MP}^0 + \alpha_{MP}^1 z_{t-1} + \sum_{j=1}^J \gamma_j^{MP} \Delta MP_{t-j}^1 + \sum_{j=1}^J \lambda_j^{MP} \Delta WP_{t-j}^1 + \varepsilon_t \\ \Delta WP_t^1 &= a_{WP}^0 + \alpha_{WP}^1 z_{t-1} + \sum_{j=1}^J \gamma_j^{WP} \Delta MP_{t-j}^1 + \sum_{j=1}^J \lambda_j^{WP} \Delta WP_{t-j}^1 + \varepsilon_t \end{aligned} \quad (11)$$

<표 6>의 전체 표본종목에서는 WP^1 이 MP^1 보다 가격발견에서 공헌도가 낮다는 것이 확인된다. WP^1 정보량의 상한치와 하한치의 평균값(IS_{WP}^1)은 0.401(40.1%)로 MP^1 정보량의 평균값(IS_{MP}^1)인 0.599(59.9%)보다 열위에 있으므로, 정보 전달에 있어서는 MP^1 에 후행하는 것으로 해석된다. 하지만 보다 유의해서 보아야 할 사항은, 유동성이 낮은 종목들일수록 MP^1 과 WP^1 간의 정보량의 차이가 확대되는 경향이 나

타난다는 것이다. 포트폴리오 1의 WP^1 의 정보량은 27.6%에 그치지만 거래량이 높은 종목일수록 WP^1 의 정보량은 증가한다. 이러한 경향은 저유동성 종목일수록 최우선호가에도 잔량의 정보를 반영할 경우, 잡음적 정보 내용이 내포되어 있음을 의미한다. 하지만 가장 유동성이 높은 종목의 포트폴리오 5에서 WP^1 정보량의 평균 값은 55%로 MP^1 의 정보량보다 높다. 이것은 유동성이 높은 종목에서는 최우선호가의 정보를 구성하는 잔량의 정보가 효율적임을 나타낸다. 따라서 결과를 요약하자면, 주문집계장에서 다른 호가에 비해 정보 내용을 지배적으로 가진 최우선호가라도 유동성이 낮은 종목에서는 정보 내용이 비효율적일 수 있다. 이는 유동성이 높은 종목일수록 최우선호가가 가격발견기능에서 보다 효율적임을 제시한다고 할 수 있다.

4.3. 주문집계장 공개단계의 가격발견의 재고찰

이 절에서는 현행 주문집계장의 공개범위에서 최우선호가를 제외한 나머지 단계 호가가 가격발견에 공헌하는 평균적인 정보량의 횡단면적 특성을 검토해보기로 한다. 진정한 가치의 대용치인 MP^1 와 WP^{2-10} 로 구성된 벡터오차수정모형인 $VECM(MP^1, WP^1)$ 을 이용해서 추정한 결과인 <표 7>에서, WP^{2-10} 의 정보량(IS^{2-10})은 평균적으로 23%이상이다. 월별 정보량의 낮은 표준편차를 고려할 때, 이 결과는 통계적으로도 유의적이다. 이 같은 결과는 이우백·최혁(2006a)가 제시한 호가공개범위 확장 이후 기간의 신규공개단계의 가격발견의 정보내용이 지속적으로 유지된다는 것을 의미한다.

Hasborouck (1995)과 Huang (2002)은 호가의 정보량을 분석한 결과, 가격발견의 주도권은 거래량과 밀접한 연관성을 가진다는 것을 제시한 바 있다. <표 7>은 차우선호가 이하에서 가격발견은 거래가 활발한 종목에서 보다 높은 정보내용을 가진다는 것을 제시한다. <표 7>에서 거래량 포트폴리오로 분류된 정보량의 추정치를 보면 최저거래량 포트폴리오(1)에서 IS^{2-10} 은 14%이지만, 최고거래량 포트폴리오(5)의 IS^{2-10} 은 33.6%까지 증가한다. 즉 최고거래량의 포트폴리오의 IS^{2-10} 은 최저거래량 포트폴리오 IS^{2-10} 의 3배 이상의 높은 값을 보여준다. 비록 고유동성 포트폴리오에서도 가격발견의 주도권은 최우선호가에 있지만, 차우선 단계호가의 변동이 최우선호가를 후행하는 속도는 저유동성 포트폴리오보다 빠르다는 것을 의미한다. 이 같은 결

과는 저거래량 포트폴리오에서는 최우선호가를 제외한 나머지 단계에서는 높은 체결비용과 가격충격을 부담해야 하기 때문에 유동성의 소비가 감소하며, 결과적으로 호가에 내재된 정보의 질이 낮아져서 가격발견 기능도 저하되는 것으로 해석할 수 있다.

Huang (2002)은 비정보거래자(uninformed trader)들은 비용을 감소시킬 목적의 거래를 추구하므로, 유동성이 높은 시장을 선호한다고 주장했다. 이 같은 논리는 차우 선호가단계 이하의 공개범위에서 유동성이 높을수록, 그 호가에 내재된 정보의 질이 더욱더 제고된다는 결과와 연관된다. 비정보거래자가 유동성에 의해 거래의 유인을 가지므로 정보거래자들도 비정보거래자들이 군집한 곳에서 거래할 것이기 때문이다. Admati-Pfleiderer (1998)가 지적했듯이, 정보거래자들은 유동성 거래자들이 시장에 집중되어 있는 시간대에 거래하기를 원할 것이다. 마찬가지로 정보거래자들은 모든 조건이 동일하다면 유동성이 높은 종목들을 선호할 것이다. 유동성이 최우선호가 단계 이하에서도 충분히 공급된다면, 정보거래자들은 주문을 활발하게 제출할 것이며, 이는 가격발견과정을 더욱더 효율적으로 제고시키는 역할을 한다.

IS^l 과 IS^{2-10} 의 월별 횡단면 분포의 시계열 추이는 <그림 1>에서 제시된다. 패널 A 에서 IS^l 는 횡단면적으로 40~90%까지 분포 범위를 형성하는 반면, IS^{2-10} 는 40% 이하에 집중되어 있다. 하지만 두 정보량의 횡단면 분포는 월별로 안정적인 패턴을 나타내며, <표 7>에 제시된 통계치가 유의적임을 제시한다. 패널 B 에서 제시된 거래량 포트폴리오의 정보량도 월별로 안정적인 추세를 가지며, 포트폴리오의 체계적인 순서가 형성되어 있음을 제시한다. 매월마다 최고거래량 포트폴리오(5)의 IS^{2-10} 은 30%~40%에서 정보량이 형성되는 반면 최저유동성 포트폴리오(1)는 월평균 10%~20%에서 변동하는 추세를 보이는데, 이 결과는 <표 7>에서 제시된 값과 일관성 있다는 것을 보여준다.

이상에서 분석한 결과인 IS^{2-10} 는 공개범위인 2~10 단계 호가의 정보량의 평균치이므로, 각 호가가 가진 정보량의 차별적인 가치를 평가할 수는 없다. 따라서 차우 선호가 이하 단계의 공개범위를 최우선호가와의 근접성을 고려해서 2~5 단계, 6~10 단계로 분리해서 추정했다. 즉 공개범위 중에서 상위단계인 2~5 단계는 공개범위에서 상대적으로 공격적인 단계이며, 6~10 단계는 하위단계로 볼 수 있다. 따라서 정보량은 MP^l , WP^{2-5} , WP^{6-10} 의 세변수로 구성된 벡터오차수정모형 $VECM(MP^l, WP^{2-5}, WP^{6-10})$ 으로 추정된다.

<표 8>은 세 호가변수의 정보량인 IS^I (패널 A), IS^{2-5} (패널 B), IS^{6-10} (패널 C)의 460 개 종목의 횡단면 평균에 대한 35 개월 시계열 평균치와 표준편차를 제시한다. 전체 표본종목의 IS^I 은 평균적으로 0.587(58.7%)로 나타나, <표 7>의 0.769(76.9%)보다는 감소했지만 MP^I 가 여전히 가격발견을 주도적으로 수행한다는 증거를 제시한다. IS^{2-5} 는 29.9%, IS^{6-10} 은 11.4%이다. <표 7>에 비해 IS^I 이 감소한 대신, IS^{2-5} 와 IS^{6-10} 를 합한 값이 IS^{2-10} 보다 증가한 이유는 정보량을 구성하는 요인들의 비선형적 결합과, IS^{6-10} 가 IS^{2-5} 보다 낮기 때문에 이를 고려할 경우 IS^{2-10} 은 하향 평균화되는 효과를 나타내기 때문일 것으로 판단된다. 따라서 WP^{2-10} 를 WP^{2-5} , WP^{6-10} 로 분리하는 것은 이와 같은 가중평균효과(weight-average effect)를 제거하는 방식이다. 특히 IS^{2-5} 가 29.9%로 IS^{2-10} 인 11.4%보다 높은 값을 가지는 것은 차우선 호가 단계 이하에서 가격발견의 정보내용은 2~5 단계에 집중되어 있다는 것을 의미한다.

거래량과 가격발견의 정보량간의 관계를 보면, 차우선호가 이하 단계의 정보량은 거래량이 높을수록 체계적으로 증가하는 형태를 보인다. 반면에, 패널 A에서 제시된 IS^I 은 최저거래량 포트폴리오(1)에서 69.6%이지만 최고거래량 포트폴리오(5)에서 43.2%로 감소한다. 패널 B에서 IS^{2-5} 는 22.8%(포트폴리오 1)에서 39.4%(포트폴리오 5)까지 증가하며, 패널 C에서 IS^{6-10} 도 7.6%(포트폴리오 1)에서 17.4%(포트폴리오 5)까지 증가한다. 한편 포트폴리오 내에서 IS^I , IS^{2-5} 와 IS^{6-10} 의 크기를 비교하면, 거래량 규모가 높은 포트폴리오 4와 5에서는 IS^{2-5} 와 IS^{6-10} 의 합산치는 40%이상으로 최우선호가의 정보량과 상당히 근접한 결과를 보인다. 따라서 차우선호가 단계 이하라도 그 종목의 유동성에 따라 가격발견에 공헌하는 정도는 차별적이며, 특히 2~5 단계의 상위단계에서 공헌도가 유의적으로 높다는 것이다.

이제 <그림 2>에서 $VECM(MP^I, WP^{2-5}, WP^{6-10})$ 으로 추정된 정보량의 횡단면 분포와 거래량 포트폴리오의 월별 시계열 추이를 보자. 패널 A에서 IS^I 은 주로 0.4(40%)이상의 범위에서 분포한다. IS^{2-5} 의 추정치는 40%이하에서 주로 밀집되어 있으나, 그 이상의 값에서도 분산되어 있다. IS^{6-10} 도 40%이하에서 대부분 추정치가 밀집되어 있으나, IS^{2-5} 보다 하향해서 분포가 집중되어 있으므로 평균치가 낮다. 패널 B에서 제시된 각 변수 정보량의 월별 시계열 추이는 거래량 포트폴리오에 따라 체계적으로 변동하는 형태를 보이고 있다. 포트폴리오 1과 2의 IS^{2-5} 와 IS^{6-10} 는 특정월에 따라 혼재되는 모습을 보이지만, 포트폴리오 3~5에서는 차이가 뚜렷한 특성을 보인다.

4.4. 일시적 효과와 영구적 효과의 비교

앞서 분석한 정보량 외에도 벡터오차수정모형으로 충격반응함수(impulse response function)를 추정해서 정보 효과에서 일시적 효과와 영구적 효과 중 어느 것이 주로 지배하는지를 분석할 수 있다. 즉, 특정 변수에 충격 1 단위가 가해질 경우 이에 대한 다른 변수의 반응의 크기는 영구적 효과(permanent effect)와 일시적 효과(temporary effect)로 구분되는데, 충격반응이 지속적으로 수렴하는 영구적 효과를 정보 효과(information effect)로 해석할 수 있다. <그림 3>은 $VECM(MP^I, WP^{2-10})$ 으로 추정한 충격반응함수의 그래프를 제시한다. 패널 A의 좌측 그래프에서 MP^I 에 표준편차가 1 단위 가해질 경우, 양 변수의 충격의 크기는 수렴과정을 거쳐서 90 시차 정도에서 0.9 정도로 동일해짐을 볼 수 있다. MP^I 은 즉각적으로 일시적인 효과 없이 반응하지만, WP^{2-10} 는 큰 폭으로 상승과정을 거쳐간다. 즉 이는 WP^{2-10} 가 정보 효과에 반응하기까지 시간이 지연됨을 의미한다. 우측 그래프에서 WP^{2-10} 에 충격이 1 단위 가해질 경우에는 MP^I 는 별다른 변동없이 정체되며, WP^{2-10} 의 반응은 큰 폭으로 하향하며 0에 가까운 값에 수렴함을 확인할 수 있다. 따라서 <그림 3>의 패널 A에서 제시된 결과는 MP^I 는 WP^{2-10} 에 주도적으로 지속적인 영향을 미치지 않지만, 반대로 WP^{2-10} 로부터 MP^I 에 전달되는 정보이전효과는 낮다는 것을 의미한다.

이제 패널 B에서 거래량 포트폴리오로 분류한 충격반응을 보기로 하자. MP^I 에 충격 1 단위가 가해질 경우, WP^{2-10} 이 반응하는 과정은 저거래량 포트폴리오와 고거래량 포트폴리오 모두 유사한 형태를 보이지만, 반응의 강도는 거래량이 큰 포트폴리오일수록 높다. 최고거래량 포트폴리오(5)의 영구적 효과는 1.0으로 수렴하지만, 최저거래량 포트폴리오(1)는 0.8로 상향 조정하여 반응하다. 이 같은 결과는 고유동성 종목일수록, 최우선호가에서 발생하는 예기치 못한 충격에 대해서는 차우선 단계 이하에서 크게 반응함을 의미한다. 반면에 WP^{2-10} 에 충격이 가해질 경우에는 MP^I 는 저유동성 종목에서는 약한 양(+)의 반응을 나타내지만, 포트폴리오 4~5는 음의 반응을 보이는 등 거래량에 따라 상이한 패턴을 나타낸다.

<그림 4>의 패널 A부터 패널 C는 $VECM(MP^I, WP^{2-5}, WP^{6-10})$ 으로 추정하여 120 시차까지 제시한 충격반응함수의 그래프이다. 패널 A에서 MP^I 의 충격에 대해서 WP^{2-5} , WP^{6-10} 는 큰 폭으로 상향 반응하며, 약 0.8로 수렴을 한다. WP^{6-10} 보다는 WP^{2-5}

⁵의 반응의 크기와 수렴속도가 빠른 편이다. 패널 B에서 WP^{2-5} 의 충격에 대해서 MP^I 와 WP^{6-10} 의 반응은 상대적으로 낮은 편이다. 특히 패널 C에서 WP^{6-10} 의 충격이 다른 변수에 미치는 영구적 효과는 발견하기 어렵다. WP^{6-10} 의 반응은 큰 하향 조정을 거치면서 MP 과 WP^{6-10} 에 약간의 영구적 효과를 미칠 뿐이다. <그림 4>와 비교해서 흥미로운 결과는 유동성 규모별로 분류한 결과인 <그림 5>에서 발견된다. 패널 A에서 MP^I 의 충격에 대해서 WP^{2-5} , WP^{6-10} 는 고거래량 종목일수록 영구적 효과가 크다. 이 같은 결과는 $VECM(MP^I, WP^{2-10})$ 에서 추정된 결과에서 보여준 영구적 충격을 보여주는 그래프인 <그림 3>을 확인시킨다. 반면에 패널 B와 패널 C에서는 WP^{2-5} , WP^{6-10} 의 충격에 대해 다른 변수들은 저거래량 종목일수록 반응이 크다는 행태를 보여준다. 패널 B에서 WP^{2-5} 의 충격에 대해서 MP^I 는 완만한 상승 과정이 진행되는 반면, WP^{6-10} 는 20 시차 이내에서 큰 폭의 일시적 효과가 나타나지만, 이후 감소하는 과정을 거쳐 영구적 효과로 수렴한다. 패널 C에서 WP^{6-10} 는 MP^I 와 WP^{2-5} 에 대해 저거래량 종목(포트폴리오 1과 2)에서는 큰 반응을 보이며 수렴하지만, 포트폴리오 3~5와 같은 고거래량 종목들에서는 일시적 효과를 가진 다음 하향 조정되어 낮은 값으로 수렴하고 있다.

4.5. 강건성 검정(Robustness Check)

제 4장의 1절부터 4절까지 밝혀진 가격발견의 정보량의 횡단면적 특성은 추정과정에서 측정 구간 설정에 따른 비거래(non-trading) 효과로 인한 편의일수도 있다. 유동성이 높고 거래가 활발한 종목에서는 5분이라는 호가 측정 간격이 벡터오차수정 모형의 추정에서 심각한 편의를 초래할 가능성이 없지만, 거래량이 저조한 종목에서는 해당 측정 구간에서 호가 조정(quote revision)과 체결이 성립하지 않기 때문에 비거래의 영향이 가격발견의 측정에서 영향을 미칠 가능성도 배제할 수 없다. 따라서 본 절에서는 거래에 따라 호가 조정이 발생하는 구간을 고려하여 변수들의 측정 간격을 10분부터 60분까지 연장하고, 1절부터 4절까지 적용했던 동일한 방법으로 정보량을 추정했다. 또한 이같은 분석의 결과를 바탕으로, 주문집계장 형태에 반영된 정보가 균형가격에 수렴되기까지 소요되는 시간을 간접적으로 추정할 수 있다.

<표 9>의 패널 A에 제시된 $VECM(MP^I, WP^{2-10})$ 과 패널 B의 $VECM(MP^I, WP^{2-5}, WP^{6-10})$ 은 측정 구간이 연장되어도 앞서 <표 7>과 <표 8>에서 제시된 결과와는

큰 차이가 없음을 제시한다. 패널 A와 패널 B에서 전체 표본 종목의 비최우선폭가의 정보량은 측정 구간이 길어질수록 체계적으로 증가하는 패턴을 나타낸다. 이 같은 결과는 시간이 경과에 따라 가격발견 과정에서 최우선폭가의 정보가 그 이하의 포크로 전달된다는 것을 의미한다. 이 같은 결과는 거래량 기준 포트폴리오들에서도 일관성있는 형태로 나타난다. 패널 A에서 최저거래량 포트폴리오(1)부터 최고거래량 포트폴리오(5)까지 IS^{2-10} 은 측정 구간이 연장되면서 증가하지만, 여전히 거래량 규모가 높은 포트폴리오일수록 IS^{2-10} 도 증가한다. 즉 시간의 경과에 따라 정보가 이전된다 하더라도, 유동성이 높은 포트폴리오에서 최우선폭가의 가격발견의 주도권을 흡수하는 효과가 높다. 특히, 30분 간격으로 $VECM(MP^l, WP^{2-10})$ 을 추정한 결과에서 포트폴리오 5를 보면 IS^{2-10} 이 50.6%로 IS^l 를 초과한다.

패널 B에서도 비최우선폭가의 정보량인 IS^{2-5} , IS^{6-10} 은 측정 구간이 연장되면서 포트폴리오가 고거래량일수록 증가한다. 포트폴리오 1~포트폴리오 4에서는 IS^{2-5} , IS^{6-10} 이 동반해서 증가하며, 포트폴리오 5에서는 IS^{2-5} 는 측정구간이 연장되면서 감소하는 추세를 보이지만 IS^{6-10} 는 대폭적으로 증가하는 형태이다. 특히 포트폴리오 5에서 측정구간이 15분 이상이면 IS^{2-5} 는 IS^l 을 초과하고 있음을 보여주는데 이는 거래량이 높은 종목일수록 2~5 단계에서는 최우선폭가 이상으로 가격발견의 주도권을 가질 수 있으며, 특히 시간이 경과할수록 6~10 단계도 후행해서 가격발견에 공헌함을 보여주는 증거이다.

5. 가격발견의 횡단면 결정요인 분석

5.1. 방법론

본 장에서 사용한 가격발견의 정보량의 횡단면 결정요인을 분석하는 회귀분석 방법은 Fama-MacBeth (1973), Fama-French (1992), Easley et al. (2002)의 방법론을 따라 수행되었다. 또한 횡단면적 변동 요인 외에도 시계열 변동 요인이 가격발견에 영향을 미치는지를 분석하기 위해, 횡단면-시계열 통합 회귀분석(time-series and cross-sectional pooling regression)도 보완적인 방법으로 수행했다.

횡단면 회귀분석에서 중점을 두고 있는 사항은 차우선폭가 이하를 구성하는 호크의 정보량의 변동이 제 2장에서 고찰한 주문집계장의 형태를 결정하는 변수인 최

우선호가부터 특정 호가까지 표준화된 거리(ND^{2-10})와 일별 잔량을 일별 거래량으로 나눈 표준화된 거래량(LP^{2-10} / LC)에 의해 유의적으로 포착될 수 있는가의 여부이다. 또한 통제변수(control variable)로 수익률의 횡단면적 변동을 설명하는 변수들로 밝혀진 기업규모($SIZE$)와 장부가치 대 시장가치 비율(BM)을 추가적으로 도입했다. 기업규모($SIZE$)는 전월말 종가에 발행주식수를 곱한 시장가치에 자연로그를 취한 값이며, 장부가치 대 시장가치 비율(BM)은 전월말 대차대조표의 자기자본을 시장가치로 나눈 값에 자연로그를 취했다. 표본기간은 2002년 1월부터 2004년 11월까지이며, 월별로 다음과 같은 횡단면 회귀분석으로 추정했다.

$$IS_{it}^{2-10} = \gamma_{0t} + \gamma_{1t}ND_{it-1}^{2-10} + \gamma_{2t}(LP^{2-10} / LC_{it-1}) + \gamma_{3t}SIZE_{it-1} + \gamma_{4t}BM_{it-1} + \eta_{it} \quad (12)$$

여기에서 종속변수인 IS_{it}^{2-10} 는 t 월의 종목 i 의 2~10단계 정보량이며, $\gamma_j (j=1, \dots, 4)$ 는 추정 회귀계수이다. 또한 η_{it} 는 평균이 0인 잔차항이다. 횡단면 회귀분석으로 추정된 35개의 월별 회귀계수는 Fama-MacBeth(1973)의 방법을 따라 시계열 평균을 구하고 통계적 유의성을 검증했다.

10단계 이내에서도 상위 단계와 하위 단계의 정보량의 횡단면적 변동은 독립변수들에 의해 차별적으로 설명될 것으로 예상되므로, 이를 구분하기 위해 위와 유사한 방법으로 다음과 같은 추가적인 회귀분석 모형을 설정했다. 즉 종속변수가 IS_{it}^{2-5} 인 경우에는 LD^{2-5} 와 LP^{2-5} / LS 가 추가되며, IS_{it}^{6-10} 에는 LD^{6-10} 와 LP^{6-10} / LS 가 도입된다.

$$IS_{it}^{2-5} = \gamma_{0t} + \gamma_{1t}LD_{it-1}^{2-5} + \gamma_{2t}(LP^{2-5} / LS)_{it-1} + \gamma_{3t}SIZE_{it-1} + \gamma_{4t}BM_{it-1} + \eta_{it} \quad (13)$$

$$IS_{it}^{6-10} = \gamma_{0t} + \gamma_{1t}LD_{it-1}^{6-10} + \gamma_{2t}(LP^{6-10} / LS)_{it-1} + \gamma_{3t}SIZE_{it-1} + \gamma_{4t}BM_{it-1} + \eta_{it} \quad (14)$$

5.2. 추정 결과

<표 10>은 횡단면 회귀모형 (12)를 설명변수를 모두 도입한 회귀모형과 개별 설명변수만이 도입된 회귀모형을 월별로 추정한 회귀계수의 통계적 유의성과 부호의 분포를 나타낸다. 패널 A의 회귀모형에서 LD^{2-10} 의 횡단면 회귀계수의 시계열 평균치는 -0.066이며 이에 대한 t-통계치는 -20.212이다. 모든 월에 대해서 회귀계수

의 부호가 5% 유의수준에서 음수로 추정된 결과는 최우선후가 10단계까지의 거리의 격차가 클수록, 가격발견의 효율성은 낮다는 것을 의미한다. (LP^{2-10} / LS)의 회귀계수의 평균치가 음수로 추정된 것은 잔량에 비해 거래량이 높은 종목일수록 공개범위의 비최우선후가에서는 가격발견이 효율적으로 나타나는 것으로 해석된다. IS^{2-10} 은 기업규모($SIZE$)과 유의적인 양의 상관관계가 존재하며, BM 과는 평균적으로 음의 상관관계가 나타나지만, 5%수준에서 통계적으로 유의적인 음의 부호를 가진 월수는 절반 정도였다.

기업규모 변수인 $SIZE$ 에서 양의 부호가 추정된 것은 대기업일수록 투자자간 정보비대칭 정도가 낮기 때문에 주문집계장에서 가격발견을 개선시키는 것으로 해석할 수 있다. 반면에 BM 이 높다는 것은 상대적으로 시장가치가 저평가(undervalued)되었거나, 재무적 곤경(financial distress)에 처한 기업이라는 것을 의미한다. 이러한 종목에 대해서는 가격발견과정이 비효율적인 것으로 해석된다. 설명변수를 모두 도입해서 추정한 다중회귀분석의 결과는 특정 설명변수만을 도입해서 회귀분석으로 추정한 단일변량 회귀분석 결과와 질적으로 큰 차이가 없다.

패널 B에서는 시계열-횡단면 통합 회귀분석의 결과를 제시한다. 이 추정 결과 역시 회귀계수의 부호와 통계적 유의성에서 패널 A의 결과를 지지한다. 기업규모는 IS^{2-10} 와 유의적인 양의 상관관계를 갖지만, LD^{2-10} , (LP^{2-10} / LS), BM 은 가격발견과 유의적인 음의 관계를 가지는 것으로 분석되었다.

종속변수인 IS^{2-10} 를 IS^{2-5} 와 IS^{6-10} 으로 구분하고 이에 대응하는 설명변수를 도입해서 추정한 결과는 <표 11>에 제시된다. 패널 A는 Fama-French(1992)의 방법론을 적용해서 추정한 결과를 제시한다. 유동성 변수인 LD 와 (LP / LS)의 회귀계수의 시계열 평균치는 모두 유의적인 수준에서 음의 값을 가진다. 즉 최우선후단계와 거리가 짧고, 유동성 공급보다 유동성 소비가 더 활성화되는 종목일수록 가격발견이 제고된다는 사실이 확인된다. 모든 설명변수를 도입한 다중회귀모형에서 <표 10>과 차이가 있는 부분은 기업규모($SIZE$)의 회귀계수가 유의적인 음의 부호로 추정된다는 것이다. 종속변수가 IS^{2-5} 인 경우, 각각의 개별 설명변수만을 도입해서 추정한 단순회귀분석의 결과를 보면 기업규모는 평균적으로 양의 부호를 가진다. 하지만 t통계치의 평균은 0.914로 낮은 편이며, 5%수준에서 유의적으로 추정된 월은 약 35%에 그친다. 따라서 이는 상대적으로 가격발견의 변동에 대해 보다 높은 설명력을

가진 다른 변수의 도입으로 인해 효과가 흡수되어 나타나는 다중공선성(multicollinearity)로 해석할 수 있을 것이다.

전반적으로 수익률의 횡단면 특성을 설명하는 변수인 기업규모와 장부가-시가비율을 통제변수로 도입하더라도, 주문집계장의 가격발견은 유동성을 나타내는 변수에 의해 높은 설명력을 보이고 있음을 제시한다.

6. 결 론

한국증권선물거래소와 같은 지정가주문시장에서 시장조성 기능을 수행하는 전자주문집계장은 시장의 투명성을 강화하려는 추세에 따라 그 역할에 대한 중요성이 강조되고 있다. 주문집계장의 공개 및 확대를 옹호하는 핵심적인 경제적 논리는 바로 공적정보인 주문집계장의 공개로 인해 거래자의 군집성(congregation)이 높아짐으로써 시장의 유동성이 향상되고 시장의 전체 질적 수준이 향상되는 네트워크 효과라고 할 수 있다. 본 논문은 이러한 주문집계장이 유동성 외부효과를 창출하는지를 차우선호가 이하 단계 호가정보의 가격발견에 대한 공헌도를 중심으로 호가공개가 5단계에서 10단계로 확대된 시점 이후인 2002년 1월 이후 35개월의 표본기간을 대상으로 분석하였다.

벡터오차수정모형으로 추정된 본 논문의 실증분석 결과는 공개 주문집계장이 가진 가격발견 기능이 동일한 시장내에서도 횡단면적으로 유동성과 유의적인 관계가 있음을 제시한다. 선행연구인 이우백-최 혁(2006a, 2006b)의 방법론을 적용한 결과, 차우선호가 이하 단계에서 나타나는 가격발견의 정보량은 거래량 규모가 증가할수록 유의적으로 높아지는 경향을 보였다. 즉, 주문집계장의 차우선호가 이하 단계에서 유동성이 풍부히 공급되고 호가간 거리가 짧을수록 유동성을 소비할 유인이 높아지며, 이는 차우선호가 이하에서 가격발견의 효율성을 제고시킨다고 설명할 수 있다. 이러한 실증분석 결과는 현행 호가공개범위는 모든 종목마다 동일하게 적용된다 하더라도, 실질적인 가격발견 기능은 유동성 공급 수준에 따라 차별화된다는 것을 제시한다. 특히 유동성 소비규모가 높은 종목에서 2~5 단계 호가의 가격발견에 대한 공헌도는 최우선호가와 유사한 수준까지 추정되었으며, 비거래 효과를 완화시킨 결과에서는 6~10 단계도 유의적인 공헌을 하는 것으로 결론을 내릴 수가 있다.

특히 유동성이 낮은 종목일수록 최우선호가에 형성된 잔량의 변동은 가격발견에서 유동성이 높은 종목에 비해 정보내용에서 열등한 것으로 나타났다.

Fama-MacBeth(1973)의 방법론을 이용해서 최우선호가 이하 단계의 주문집계장의 형태가 가격발견의 정보량에 미치는 영향을 분석한 결과에서도 호가간 거리와 정보량은 유의적인 음의 관계를 보였다. 또한 유동성 공급 수준에 비해 유동성 소비가 높은 종목일수록, 정보량은 유의적으로 높았다.

이상과 같이 본 연구의 실증적 증거는 호가공개가 가격발견에서 정보 내용을 가지며, 특히 횡단면적으로 유동성과 높은 상관관계가 존재함을 제시한다. 이 결과는 공개범위가 10단계라 하더라도, 유동성 공급과 소비가 저조한 종목들은 그 정보 내용이 낮기 때문에, 호가공개가 확대된다 하더라도 그 효익은 제한될 것이다. 반면에 유동성이 높은 종목은 추가적인 호가공개시 유동성 외부효과가 높기 때문에 정보 내용을 가질 가능성이 높을 것이다.

본 연구 결과는 동일 시장 내에서 유동성 외부효과를 검정한 것이지만, 동일 종목이 다수의 시장에 상장된 경우 시장의 유동성 수준과 가격발견의 질적 수준을 비교하는 것이 보다 직접적인 시장의 네트워크 효과를 비교하는 타당한 방법일 것이다. 추후 연구과제에서는 이러한 측면에서 유동성 외부 효과를 검증하는 것이 바람직하다고 본다.

참 고 문 헌

- 박종호, 엄경식, “한국주식시장에서 투명성과 질적 수준과의 관계 : 호가공개범위 확대를 중심으로”, 재무연구, 제18호 1권 (2005), pp. 157~198.
- 이우백, 최 혁, “거래 전 정보공개와 가격발견”, 증권학회지, 제35권 4호 (2006), 143-190
- 이우백, 최 혁, “하루중 유동성 형성 과정과 가격발견”, 제19권 제2호 (2006), 1-38
- 정재만, “한국증권거래소 종가의 정보효율성”, 증권학회지, 제33권 2호 (2004), pp. 107-153.
- 최 혁, 이우백, “주문 공격성의 비용”, 경영논집, 제36권 4호 (2002), 서울대학교 경영대학 경영연구소, pp. 61-114.
- Admati, A.R., Pfleiderer, P., 1988, A theory of intraday patterns : Volume and price variability, *Review of Financial Studies* 1, 3~40
- Ahn, Hee-Jun, Kee-Hong Bae, and Kalok Chan, 2001, Limit Orders, Depth and Volatility : Evidence from the Stock Exchange of Hong Kong, *Journal of Finance* 56, pp. 767-788.
- Baruch, Shmuel, 2005, Who Benefits from an Open Limit-order Book?, *Journal of Business* 78, pp. 1267-1306.
- Barclay, Michael J., and Jerold B. Warner, 1993, Stealth Trading and Volatility : Which Trades Move Prices?, *Journal of Financial Economics* 34, pp. 281-305.
- Barclay, Michael J. and Terrence Hendershott , 2004, Liquidity Externalities and Adverse Selection: Evidence from Trading after Hours, *Journal of Finance* 59, pp. 681~710.
- Boehmer, Ekkehart, Gideon Saar, and Lei yu, 2005, Lifting the Veil : An Analysis of Pre-trade Transparency at the NYSE, *Journal of Finance* 60, pp. 783-815.
- Biais, Bruno, Pierre Hillion, and Chester Spatt, 1995, An Empirical Analysis of the Limit-order Book and the Order Flow in the Paris Bourse, *Journal of Finance* 50, pp. 1665-1689.
- Bloomfield, Robert, Maureen O’Hara, and Gideon Saar, 2005, The “Make or Take” Decision in an Electronic Market : Evidence on the Evolution of Liquidity, *Journal of Financial Economics* 75, pp. 165-199.
- Cao, Charles, Oliver Hansch, and Xiaoxin Wang, 2005, The Informational Content of an Open Limit Order Book, *working paper*, Pennsylvania State University
- Copeland, Thomas E., and Dan Galai, 1983, Informational Effects and the Bid-ask Spread, *Journal of Finance* 38, pp. 1457-1469.
- Dickey, David A., and Wayne A Fuller, 1979, Distribution of the estimators for autoregressive

- time-series with a unit root, *Journal of American Statistical Association* 74, 427~431
- Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, Is Information Risk a Determinant of Asset Returns?, *Journal of Finance* 57, pp. 2185-2221.
- Engel, R.F., and Clive W.J. Granger, 1987, Co-integration and Error Correction : Representation, Estimation, and Testing, *Econometrica* 55, pp. 251-276.
- Fama, Eugene F., and James D. MacBeth, 1973, Risk, Return and Equilibrium : Empirical Test, *Journal of Political Economy* 81, pp. 607-636.
- Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, pp. 427-465.
- Harris, Lawrence E., and Venkatesh Panchapagesan, 2005, The Information Content of the Limit Order Book : Evidence from NYSE Specialist Trading Decisions, *Journal of Financial Markets* 8, pp.25-67.
- Hasbrouck, Joel, 1995, One Security, Many Markets : Determining the Contributions to Price Discovery, *Journal of Finance* 50, pp. 1175-1199.
- Hasbrouck, Joel, and Gideon Saar, 2002, Limit Orders and Volatility in a Hybrid Market : The Island ECN, *Working Paper*
- Hasbrouck, Joel, 2002, Intraday Price Formation in US Equity Index Markets, *Journal of Finance* 58, pp. 2375-2400.
- Hasbrouck, Joel, 2002, Stalking the "Efficient Price" in Market Microstructure Specifications : an Overview, *Journal of Financial markets* 5, pp. 329-339.
- Hendershott, Terrence and Charles M. Jones, 2005, Island Goes Dark: Transparency, Fragmentation, and Regulation, *Review of Financial Studies* 18, pp. 743-793
- Huang, Roger D., 2002, The Quality of ECN and Nasdaq Market Maker Quotes, *Journal of Finance* 57, pp. 1285-1319.
- Kavajecz, Kenneth A., and Elizabeth R. Odders-White, 2004, Technical Analysis and Liquidity Provision, *Review of Financial Studies* 17, pp.1043-1071.
- Madhavan, Ananth, 1996, Security Prices and Market Transparency, *Journal of financial intermediation* 5, pp. 255-283.
- Madhavan, Ananth, David Porter and Daniel Weaver, 2005, Should Securities Markets Be Transparent?, *Journal of Financial Markets* 8, pp. 265-287.
- Pagano, Marco and Ailsa Roell, 1996, Transparency and Liquidity : A comparison of Auction and Dealer Markets with Informed Trading, *Journal of finance* 51, pp. 579-611.
- Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* 39, pp. 1127-1140

Table I
Tick size-normalized distance between best step and remaining step
in limit order book

This table presents time-series average of the monthly distance between best step and remaining step normalized by tick size in the limit order book on the basis of monthly volume. Tick size-normalized distance is measured as difference between best quote and other steps in limit order book, divided by tick size corresponding to price level. For a stock, daily tick size-normalized distance is computed at each 5 minute-interval during continuous auction from 9:00 to 14:50 in a day, and then averaged over 70 intervals. Monthly tick size-normalized distance is averaged across days for each month during sample period from January 2002 to November 2004. The cross-sectional means of all 460 sample stocks and 5 portfolios are obtained for each month. Portfolios are formed on the basis of share volume of prior month (1[low]~5[high]). The time-series average of cross-sectional mean with standard deviation in parenthesis is reported for all sample stocks and 5 portfolios on the basis of volume.

step	All	1(low)	2	3	4	5(high)
<i>Panel A. Bid side</i>						
2	1.52 (0.27)	2.60 (0.39)	2.00 (0.23)	1.66 (0.14)	1.31 (0.07)	1.10 (0.03)
3	2.92 (0.45)	5.12 (0.73)	3.89 (0.45)	3.20 (0.28)	2.56 (0.13)	2.18 (0.05)
4	4.32 (0.65)	7.90 (1.20)	5.87 (0.70)	4.79 (0.41)	3.83 (0.18)	3.26 (0.06)
5	5.79 (0.85)	11.04 (1.77)	8.02 (0.98)	6.48 (0.59)	5.13 (0.25)	4.36 (0.08)
6	7.36 (1.07)	14.62 (2.49)	10.40 (1.33)	8.28 (0.78)	6.49 (0.32)	5.47 (0.10)
7	9.07 (1.33)	18.97 (3.49)	13.15 (1.82)	10.28 (1.02)	7.92 (0.40)	6.61 (0.11)
8	11.00 (1.65)	24.06 (4.61)	16.40 (2.49)	12.54 (1.34)	9.44 (0.50)	7.77 (0.14)
9	13.14 (2.11)	30.25 (6.00)	20.31 (3.39)	15.12 (1.75)	11.07 (0.63)	8.98 (0.16)
10	15.64 (2.79)	37.36 (7.41)	25.06 (4.61)	18.12 (2.27)	12.87 (0.78)	10.22 (0.19)
<i>Panel B. Ask side</i>						
2	1.85 (0.36)	2.74 (0.30)	2.21 (0.23)	1.82 (0.16)	1.44(0.09)	1.15 (0.04)
3	3.52 (0.68)	5.50 (0.70)	4.27 (0.47)	3.48 (0.30)	2.77(0.17)	2.25 (0.06)
4	5.17 (1.02)	8.48 (1.19)	6.38 (0.74)	5.16 (0.46)	4.08(0.24)	3.34 (0.09)
5	6.84 (1.38)	11.68 (1.71)	8.61 (1.06)	6.85 (0.63)	5.40(0.31)	4.43 (0.12)
6	8.53 (1.72)	15.26 (2.40)	10.99 (1.44)	8.61 (0.82)	6.72(0.39)	5.52 (0.14)
7	10.29 (2.13)	19.27 (3.19)	13.57 (1.91)	10.45 (1.05)	8.07(0.47)	6.61 (0.17)
8	12.12 (2.50)	23.80 (4.15)	16.36 (2.46)	12.39 (1.31)	9.44(0.57)	7.70 (0.20)
9	14.01 (2.90)	28.62 (5.17)	19.38 (3.12)	14.43 (1.62)	10.84(0.67)	8.80 (0.23)
10	15.96 (3.36)	33.63 (6.26)	22.60 (3.83)	16.54 (1.94)	12.28(0.78)	9.90 (0.26)

Table II
Depth in visible steps in limit order book

This table presents time-series average of the monthly estimation of depth within 10 steps in the limit order book on the basis of monthly volume. For a stock, daily depth is observed at each 5 minute-interval during continuous auction from 9:00 to 14:50 in a day, and then averaged over 70 intervals. Monthly depth is averaged across days for each month during sample period from January 2002 to November 2004. The cross-sectional means of all 460 sample stocks and 5 portfolios are obtained for each month. Portfolios are formed on the basis of share volume of prior month (1[low]~5[high]). The time-series average of cross-sectional mean with standard deviation in parenthesis is reported for all sample stocks and 5 portfolios on the basis of volume.

	All	1(low)	2	3	4	5(high)
<i>Panel A. Bid side</i>						
1	631 (288)	477 (530)	939 (811)	1,610 (692)	4,480 (2,052)	19,405 (9,272)
2	1,014 (587)	340 (131)	716 (112)	1,478 (250)	4,692 (1,447)	21,490 (5,797)
3	1,092 (612)	345 (121)	755 (129)	1,587 (254)	4,559 (1,076)	19,688 (3,915)
4	1,105 (568)	319 (110)	736 (133)	1,549 (255)	4,145 (811)	17,214 (2,807)
5	1,044 (504)	313 (102)	735 (148)	1,501 (256)	3,786 (668)	14,917 (2,139)
6	961 (485)	300 (96)	699 (136)	1,421 (250)	3,396 (574)	13,022 (1,971)
7	889 (430)	291 (92)	674 (138)	1,365 (240)	3,138 (532)	11,556 (1,881)
8	830 (374)	285 (88)	647 (131)	1,304 (237)	2,902 (479)	10,527 (1,755)
9	783 (345)	276 (87)	624 (125)	1,240 (231)	2,703 (454)	9,575 (1,780)
10	733 (292)	262 (83)	586 (122)	1,154 (223)	2,467 (443)	8,678 (1,780)
<i>Panel B. Ask side</i>						
1	461 (226)	236 (70)	450 (88)	838 (212)	2,332 (714)	11,514 (6,498)
2	622 (458)	267 (87)	525 (85)	1,016 (171)	3,043 (875)	14,915 (4,185)
3	676 (600)	277 (97)	549 (96)	1,105 (193)	3,252 (836)	16,052 (3,993)
4	721 (679)	261 (105)	545 (103)	1,112 (201)	3,241 (783)	16,242 (3,760)
5	760 (700)	265 (106)	552 (112)	1,148 (228)	3,294 (752)	16,379 (3,387)
6	775 (717)	265 (106)	546 (114)	1,121 (222)	3,200 (703)	16,059 (3,219)
7	769 (678)	259 (103)	543 (118)	1,120 (231)	3,222 (723)	16,082 (3,089)
8	773 (716)	253 (99)	530 (117)	1,103 (218)	3,209 (763)	15,343 (2,756)
9	779 (773)	249 (99)	522 (117)	1,113 (233)	3,233 (747)	15,011 (2,561)
10	776 (763)	243 (95)	510 (116)	1,091 (233)	3,056 (670)	14,088 (2,603)

Table III
Ratio of liquidity provision to liquidity consumption

This table presents time-series average of the monthly estimation of ratio of liquidity provision to liquidity consumption within 10 steps in the limit order book on the basis of monthly volume. Ratio of liquidity provision to consumption is computed as the depth divided by daily share volume, and then multiplied by 10,000. For a stock, daily ratio of liquidity provision to consumption is computed at each 5 minute-interval during continuous auction from 9:00 to 14:50 in a day, and then averaged over 70 intervals. Monthly Ratio of liquidity provision to consumption is averaged across days for each month during sample period from January 2002 to November 2004. The cross-sectional means of all 460 sample stocks and 5 portfolios are obtained for each month. Portfolios are formed on the basis of share volume of prior month (1[low]~5[high]). The time-series average of cross-sectional mean with standard deviation in parenthesis is reported for all sample stocks and 5 portfolios on the basis of volume.

	All	1(low)	2	3	4	5(how)
<i>Panel A. Bid side</i>						
1	6.75 (4.52)	19.03 (17.36)	7.81 (9.92)	3.37 (1.82)	2.06 (1.03)	1.31 (0.63)
2	5.50 (2.92)	14.77 (9.83)	5.71 (2.72)	3.11 (1.24)	2.20 (0.94)	1.59 (0.77)
3	5.54 (2.79)	14.74 (9.21)	5.94 (2.75)	3.34 (1.31)	2.15 (0.80)	1.43 (0.54)
4	4.89 (1.99)	12.24 (5.79)	5.66 (2.45)	3.25 (1.25)	1.95 (0.63)	1.22 (0.37)
5	4.75 (1.84)	11.96 (5.16)	5.69 (2.70)	3.14 (1.18)	1.79 (0.54)	1.05 (0.27)
6	4.51 (1.68)	11.60 (4.75)	5.34 (2.30)	2.96 (1.12)	1.62 (0.47)	0.92 (0.21)
7	4.37 (1.63)	11.39 (4.66)	5.17 (2.33)	2.85 (1.08)	1.50 (0.43)	0.82 (0.18)
8	4.23 (1.62)	11.24 (4.77)	4.94 (2.20)	2.72 (1.02)	1.40 (0.41)	0.75 (0.15)
9	4.03 (1.47)	10.77 (4.24)	4.72 (1.99)	2.58 (0.95)	1.30 (0.38)	0.68 (0.14)
10	3.76 (1.37)	10.12 (4.00)	4.37 (1.77)	2.40 (0.87)	1.19 (0.34)	0.62 (0.11)
<i>Panel B. Ask side</i>						
1	3.93 (2.12)	12.28 (7.62)	3.63 (1.87)	1.77 (0.74)	1.11 (0.50)	0.77 (0.38)
2	4.48 (2.51)	13.37 (8.87)	4.23 (2.15)	2.19 (0.93)	1.44 (0.60)	1.07 (0.51)
3	4.59 (2.43)	13.47 (8.40)	4.36 (2.13)	2.37 (1.02)	1.52 (0.59)	1.13 (0.47)
4	4.02 (1.72)	10.75 (5.05)	4.24 (1.97)	2.37 (0.98)	1.51 (0.55)	1.13 (0.43)
5	4.03 (1.66)	10.75 (4.79)	4.24 (1.93)	2.43 (0.98)	1.53 (0.53)	1.13 (0.41)
6	3.99 (1.62)	10.70 (4.70)	4.19 (1.85)	2.36 (0.94)	1.49 (0.50)	1.12 (0.41)
7	3.93 (1.61)	10.42 (4.60)	4.15 (1.84)	2.35 (0.93)	1.50 (0.50)	1.13 (0.42)
8	3.82 (1.61)	10.08 (4.66)	4.04 (1.81)	2.31 (0.93)	1.50 (0.52)	1.07 (0.39)
9	3.75 (1.52)	9.83 (4.29)	3.96 (1.76)	2.33 (0.97)	1.51 (0.52)	1.04 (0.32)
10	3.60 (1.49)	9.38 (4.28)	3.86 (1.72)	2.28 (0.95)	1.43 (0.48)	0.96 (0.26)

Table IV
Relative execution frequency on basis of time interval

Relative execution frequency(%) is computed as the ratio of intervals when trade is executed to total number of intervals during continuous auction from 9:00 to 14:50. Relative execution is observed at 5 minutes, 10 minutes, 15 minutes, 30 minutes and 60 minutes interval, respectively. Mean of daily relative execution frequency is calculated for a stock during sample period from January 2002 to November 2004 and then cross-sectional mean of relative execution frequency is computed across 460 sample stocks, with standard deviation in parenthesis. Trading volume is computed as mean of monthly trading volume for a stock, and all sample stocks are categorized into 5 portfolios on the ranking of trading volume (1[low]~5[high]).

	5 minute	10 minute	15 minute	30 minute	60 minute
All	80.92 (19.22)	89.01 (13.53)	92.54 (10.43)	96.83 (5.61)	98.92 (2.49)
1(low)	51.73 (11.79)	67.66 (11.52)	76.04 (10.58)	88.37 (7.40)	95.66 (4.10)
2	75.55 (13.34)	87.13 (8.73)	92.00 (6.20)	97.31 (2.73)	99.36 (0.88)
3	85.32 (9.85)	93.44 (5.69)	96.33 (3.68)	98.96 (1.32)	99.77 (0.32)
4	93.74 (7.44)	97.42 (3.80)	98.61 (2.31)	99.61 (0.73)	99.89 (0.15)
5(high)	98.29 (3.56)	99.42 (1.51)	99.70 (0.82)	99.89 (0.19)	99.92 (0.04)

Table V
Unit root test on midpoint quote and depth-weighted quotes in limit order book

This table presents the unit root test on the quote price of steps in the limit order book. Using Augmented Dickey-Fuller model with drift term, unit root test is conducted month-by-month during sample period across 460 sample stocks and 5 portfolios. Portfolios are formed on the basis of share volume of prior month (1[low]~5[high]). In implementation of test, at each 5 minute interval quote price is observed during continuous auction from 9:00 to 14:50. Lag is chosen on the basis of Bayesian Information Criteria (BIC). Mean of t-stat indicates monthly time-series average of cross-sectional mean of t-statistic obtained from ADF model for 460 sample stocks. %RW indicates that the percentages of number of sample stocks×sample months without rejection of null hypothesis on unit root at 1% and 5%, respectively.

		MP^l	WP^l	WP^{2-10}	WP^{2-5}	WP^{2-10}
All	Mean of t-statistic	-1.45	-1.65	-1.89	-1.74	-2.34
	%RW (signif at 1%)	93.41	91.97	87.17	90.60	77.36
	%RW (signif at 5%)	88.12	84.56	78.37	82.78	67.50
1(low)	Mean of t-statistic	-1.75	-2.28	-2.69	-2.47	-3.43
	%RW (signif at 1%)	82.73	76.00	67.19	71.94	49.03
	%RW (signif at 5%)	75.32	63.51	53.06	59.78	34.72
2	Mean of t-statistic	-1.51	-1.85	-2.29	-1.98	-2.96
	%RW (signif at 1%)	87.10	84.22	74.97	82.08	59.64
	%RW (signif at 5%)	81.96	75.73	64.51	72.88	47.94
3	Mean of t-statistic	-1.45	-1.55	-1.82	-1.61	-2.28
	%RW (signif at 1%)	89.34	89.28	85.37	88.60	76.88
	%RW (signif at 5%)	84.84	83.05	77.50	82.11	67.16
4	Mean of t-statistic	-1.30	-1.31	-1.39	-1.35	-1.63
	%RW (signif at 1%)	91.25	92.30	91.39	92.36	88.63
	%RW (signif at 5%)	87.25	88.31	86.66	87.54	82.46
5(high)	Mean of t-statistic	-1.27	-1.25	-1.28	-1.28	-1.41
	%RW (signif at 1%)	90.25	92.07	92.27	92.42	90.78
	%RW (signif at 5%)	86.31	88.28	87.96	88.19	86.13

Table VI
Information share of best quote sorted by volume

This table presents time-series average of the monthly estimated information share of best quote price and depth-weighted average price from step 2 to 10 in limit order book. Monthly information share is estimated by using Vector Error Correction Model specification with inclusion of midpoint best quote price(MP^l) and depth-weighted average best quote price (WP^l) observed at 5 minute-intervals during continuous auction from 9:00 to 14:50. VECM is estimated with lag selection on the basis of Bayesian Information Criteria(BIC) month-by-month during sample period from January 2002 to November 2004 across sample stocks. The cross-sectional means of lower bound, upper bound and mean of estimated information share of all 460 sample stocks and 5 portfolios are obtained for each month. Portfolios are formed on the basis of share volume of prior month (1[low]~5[high]). The time-series average of cross-sectional mean with standard deviation in parenthesis is reported for all sample stocks and 5 portfolios on the basis of volume. Figures in bracket indicate the percentage of number of sample stocks and months with $\max[IS_{WP}^1, IS_{MP}^1]$.

	All		1(low)		2		3		4		5(high)	
	IS_{WP}^1	IS_{MP}^1										
Lower	0.179 (0.041) [15.68]	0.377 (0.063) [74.32]	0.134 (0.064) [12.20]	0.582 (0.076) [87.80]	0.143 (0.051) [12.07]	0.509 (0.094) [87.93]	0.174 (0.048) [14.97]	0.387 (0.077) [85.03]	0.191 (0.059) [16.31]	0.255 (0.057) [83.69]	0.250 (0.065) [22.93]	0.148 (0.041) [77.07]
Upper	0.623 (0.063) [15.68]	0.821 (0.041) [74.32]	0.418 (0.076) [12.20]	0.866 (0.064) [87.80]	0.491 (0.094) [12.07]	0.857 (0.051) [87.93]	0.613 (0.077) [14.97]	0.826 (0.048) [85.03]	0.745 (0.057) [16.31]	0.809 (0.059) [83.69]	0.852 (0.041) [22.93]	0.750 (0.065) [77.07]
Mean	0.401 (0.050) [15.68]	0.599 (0.050) [74.32]	0.276 (0.068) [12.20]	0.724 (0.068) [87.80]	0.317 (0.069) [12.07]	0.683 (0.069) [87.93]	0.394 (0.059) [14.97]	0.606 (0.059) [85.03]	0.468 (0.053) [16.31]	0.532 (0.053) [83.69]	0.551 (0.049) [22.93]	0.449 (0.049) [77.07]

Table VII
Information share of best step and remaining step 2 to step 10 sorted by volume

This table presents time-series average of the monthly estimated information share of best quote price and depth-weighted average price from step 2 to 10 in limit order book. Monthly information share is estimated by using Vector Error Correction Model specification with inclusion of best quote price(MP^l) and depth-weighted average price from step 2 to 10(WP^{2-10}) observed at 5 minute-intervals during continuous auction from 9:00 to 14:50. Midpoint price is selected as proxy for best quote price. VECM is estimated with lag selection on the basis of Bayesian Information Criteria(BIC) month-by-month during sample period from January 2002 to November 2004 across 460 sample stocks. The cross-sectional means of lower bound, upper bound and mean of estimated information share of all 460 sample stocks and 5 portfolios are obtained for each month. Portfolios are formed on the basis of share volume of prior month(1[low]~5[high]). The time-series average of cross-sectional mean with standard deviation in parenthesis is reported for all sample stocks and 5 portfolios on the basis of volume. Figures in bracket indicate the percentage of number of sample stocks and months with $\max[IS^l, IS^{2-10}]$.

	All		1(low)		2		3		4		5(high)	
	IS^l	IS^{2-10}										
Lower	0.567	0.028	0.759	0.039	0.676	0.029	0.579	0.023	0.461	0.021	0.355	0.028
	(0.051)	(0.005)	(0.059)	(0.011)	(0.066)	(0.007)	(0.061)	(0.006)	(0.055)	(0.005)	(0.042)	(0.007)
	[99.52]	[0.48]	[99.84]	[0.16]	[99.94]	[0.06]	[99.81]	[0.19]	[99.91]	[0.09]	[98.16]	[1.84]
Upper	0.972	0.433	0.961	0.241	0.971	0.324	0.977	0.421	0.979	0.539	0.972	0.645
	(0.005)	(0.051)	(0.011)	(0.059)	(0.007)	(0.066)	(0.006)	(0.061)	(0.005)	(0.055)	(0.007)	(0.042)
	[99.52]	[0.48]	[99.84]	[0.16]	[99.94]	[0.06]	[99.81]	[0.19]	[99.91]	[0.09]	[98.16]	[1.84]
Mean	0.769	0.231	0.860	0.140	0.823	0.177	0.778	0.222	0.720	0.280	0.664	0.336
	(0.024)	(0.024)	(0.028)	(0.028)	(0.031)	(0.031)	(0.029)	(0.029)	(0.026)	(0.026)	(0.019)	(0.019)
	[99.52]	[0.48]	[99.84]	[0.16]	[99.94]	[0.06]	[99.81]	[0.19]	[99.91]	[0.09]	[98.16]	[1.84]

Table VIII

Information share of best quote, the upper quotes and the lower quotes sorted by volume

This table presents the result for Hasbrouck information share of best quote (MP^I) and depth-weighted price of the upper part and of the lower part. Hasbrouck information share is estimated by using VECM specification with inclusion of midpoint price, depth-weighted price of the upper part and of the lower part. Depth-weighted price of the upper part is measured based on step 2 from step 5 (WP^{2-5}) and depth-weighted price of the lower part is measured based on step 6 from step 10 (WP^{6-10}). VECM is composed of midpoint price, depth-weighted price of the upper part and of the lower part observed at 5 minute-intervals during continuous auction from 9:00 to 14:50. VECM is estimated with lag selection on the basis of Bayesian Information Criteria (BIC) month-by-month during sample period from January 2002 to November 2004 across 460 sample stocks. The cross-sectional means of lower bound, upper bound and mean of estimated information share of all 460 sample stocks and 5 portfolios are obtained for each month. Portfolios are formed on the basis of share volume of prior month (1[low]~5[high]). The time-series average of cross-sectional mean with standard deviation in parenthesis is reported for all sample stocks and 5 portfolios on the basis of volume. Figures in bracket indicate the percentage of number of sample stocks and months with $\max[IS^1, IS^{2-5}, IS^{6-10}]$.

	All	1(low)	2	3	4	5(high)
<i>Panel A. Information share of $MP^I : IS^I$</i>						
Lower	0.383 (0.050) [85.42]	0.551 0.067 [89.61]	0.488 0.070 [89.07]	0.395 0.060 [86.39]	0.293 0.045 [84.93]	0.185 0.035 [77.18]
Upper	0.832 (0.041) [85.37]	0.850 0.052 [89.58]	0.863 0.045 [88.85]	0.848 0.050 [86.48]	0.835 0.059 [84.77]	0.766 0.069 [77.24]
Mean	0.587 (0.046) [85.44]	0.696 (0.059) [89.60]	0.666 (0.058) [88.94]	0.604 (0.056) [86.42]	0.534 (0.051) [84.90]	0.432 (0.049) [77.43]
<i>Panel B. Information share of $WP^{2-5} : IS^{2-5}$</i>						
Lower	0.091 (0.020) [14.14]	0.094 0.035 [9.89]	0.082 0.027 [10.81]	0.082 0.025 [13.46]	0.084 0.028 [14.82]	0.110 0.032 [21.85]
Upper	0.555 (0.051) [14.47]	0.378 0.064 [10.11]	0.440 0.069 [11.12]	0.536 0.062 [13.49]	0.650 0.047 [15.20]	0.772 0.040 [22.57]
Mean	0.299 (0.031) [14.38]	0.228 (0.046) [9.99]	0.249 (0.042) [11.02]	0.290 (0.037) [13.55]	0.334 (0.032) [15.07]	0.394 (0.032) [22.38]
<i>Panel C. Information share of $WP^{6-10} : IS^{6-10}$</i>						
Lower	0.019 (0.002) [0.40]	0.028 0.006 [0.50]	0.022 0.005 [0.12]	0.020 0.004 [0.16]	0.014 0.003 [0.25]	0.013 0.004 [0.97]
Upper	0.254 (0.038) [0.12]	0.138 0.029 [0.31]	0.170 0.041 [0.03]	0.229 0.047 [0.03]	0.310 0.049 [0.03]	0.425 0.051 [0.19]
Mean	0.114 (0.016) [0.14]	0.076 (0.014) [0.40]	0.085 (0.017) [0.03]	0.106 (0.019) [0.03]	0.131 (0.021) [0.03]	0.174 (0.022) [0.19]

Table IX
Robustness Check

This table presents the result for Hasbrouck information share estimated from VECM specification with different time interval. In Panel A, VECM is composed of midpoint quote price and depth-weighted average price from step 2 to 10 (WP^{2-10}) observed at 10, 15, 30 and 60 minute interval during continuous auction from 9:00 to 14:50. In Panel B, VECM is composed of midpoint quote price (MP^l), depth-weighted average price from step 2 to 5 (WP^{2-5}) and depth-weighted average price from step 6 to 10 (WP^{6-10}) observed at 10, 15, 30 and 60 minute interval during continuous auction from 9:00 to 14:50. VECM is estimated with lag selection on the basis of Bayesian Information Criteria (BIC) month-by-month during sample period from January 2002 to November 2004 across 460 sample stocks. The cross-sectional means of lower bound, upper bound and mean of estimated information share of all 460 sample stocks and 5 portfolios are obtained for each month. Portfolios are formed on the basis of share volume of prior month (1[low]~5[high]). The time-series average of cross-sectional mean with standard deviation in parenthesis is reported for all sample stocks and 5 portfolios on the basis of volume.

	All	1(low)	2	3	4	5(high)
<i>Panel A. VECM (MP^l, WP^{2-10})</i>						
<i>10 minute interval</i>						
IS^l	0.663 (0.053)	0.775 (0.064)	0.738 (0.066)	0.675 (0.062)	0.608 (0.062)	0.516 (0.053)
IS^{2-5}	0.337 (0.053)	0.225 (0.064)	0.262 (0.066)	0.325 (0.062)	0.392 (0.062)	0.484 (0.053)
<i>15 minute interval</i>						
IS^l	0.654 (0.053)	0.774 (0.066)	0.734 (0.069)	0.660 (0.062)	0.591 (0.057)	0.509 (0.053)
IS^{2-5}	0.346 (0.053)	0.226 (0.066)	0.266 (0.069)	0.340 (0.062)	0.409 (0.057)	0.491 (0.053)
<i>30 minute interval</i>						
IS^l	0.639 (0.055)	0.764 (0.067)	0.719 (0.072)	0.642 (0.063)	0.572 (0.058)	0.494 (0.052)
IS^{2-5}	0.361 (0.055)	0.236 (0.067)	0.281 (0.072)	0.358 (0.063)	0.428 (0.058)	0.506 (0.052)
<i>60 minute interval</i>						
IS^l	0.614 (0.054)	0.744 (0.073)	0.686 (0.074)	0.611 (0.059)	0.545 (0.053)	0.478 (0.050)
IS^{2-5}	0.386 (0.054)	0.256 (0.073)	0.314 (0.074)	0.389 (0.059)	0.455 (0.053)	0.522 (0.050)
<i>Panel B. VECM ($MP^l, WP^{2-5}, WP^{6-10}$)</i>						
<i>10 minute interval</i>						
IS^l	0.562 (0.048)	0.680 (0.056)	0.645 (0.062)	0.579 (0.059)	0.502 (0.054)	0.404 (0.044)
IS^{2-5}	0.306 (0.030)	0.233 (0.042)	0.258 (0.043)	0.300 (0.037)	0.344 (0.031)	0.393 (0.029)
IS^{6-10}	0.132 (0.019)	0.087 (0.015)	0.097 (0.020)	0.121 (0.023)	0.154 (0.026)	0.203 (0.022)
<i>15 minute interval</i>						
IS^l	0.545 (0.049)	0.667 (0.055)	0.629 (0.064)	0.559 (0.060)	0.482 (0.051)	0.389 (0.040)
IS^{2-5}	0.308 (0.029)	0.236 (0.041)	0.263 (0.043)	0.305 (0.035)	0.346 (0.027)	0.388 (0.025)
IS^{6-10}	0.147 (0.021)	0.096 (0.017)	0.108 (0.023)	0.137 (0.027)	0.172 (0.027)	0.223 (0.022)
<i>30 minute interval</i>						
IS^l	0.522 (0.050)	0.647 (0.055)	0.607 (0.066)	0.534 (0.064)	0.455 (0.051)	0.369 (0.039)
IS^{2-5}	0.314 (0.027)	0.244 (0.041)	0.272 (0.042)	0.313 (0.034)	0.352 (0.024)	0.388 (0.023)
IS^{6-10}	0.164 (0.023)	0.109 (0.017)	0.122 (0.026)	0.153 (0.031)	0.193 (0.029)	0.243 (0.024)
<i>60 minute interval</i>						
IS^l	0.485 (0.048)	0.609 (0.057)	0.560 (0.068)	0.487 (0.060)	0.416 (0.045)	0.348 (0.036)
IS^{2-5}	0.322 (0.024)	0.259 (0.039)	0.290 (0.039)	0.327 (0.029)	0.355 (0.019)	0.382 (0.022)
IS^{6-10}	0.193 (0.025)	0.132 (0.022)	0.150 (0.031)	0.186 (0.033)	0.229 (0.029)	0.270 (0.024)

Table X
Month-by-month cross-sectional determinants of information share of non-best depth-weighted average quotes in limit order book

This table reports the result for determinants of Hasbrouck information share of non-best quotes using standard Fama-MacBeth (1973) regression and time-series and cross-sectional pooling regression. Dependant variable is Hasbrouck information share of depth-weighted average price from step 2 to 10 (IS^{2-10}) estimated month-by-month by using VECM specification during sample period from January 2002 to November 2004 across 460 sample stocks. Explanatory variables are tick size-normalized distance (ND^{2-10}), liquidity provision-to-liquidity consumption ratio (LP^{2-10}/LS), firm size, and book-to-market ratio. Tick size-normalized distance is measured as difference between best quote and other steps in limit order book, divided by tick size corresponding to price level. Ratio of liquidity provision to consumption is computed as the depth divided by monthly share volume. Firm size is monthly log market value, multiplying closing price by total outstanding shares at the end of month. Panel A contains time-series means and t-value of the coefficients and time-series mean of t-value for the coefficients in cross-sectional regression. + (%) indicates the percentage of number of months with positive sign of coefficient. % signif (+) indicates the the percentage of number of months with positive sign of coefficient significant at 5% level. % signif (-) indicates the the percentage of number of months with negative sign of coefficient significant at 5% level. Panel B contains coefficients and t-value in parenthesis from time-series and cross-sectional pooling regression.

	<i>intercept</i>	ND^{2-10}	(LP^{2-10} / LS)	<i>SIZE</i>	<i>BM</i>	$adjR^2$
<i>Panel A. Fama-MacBeth regression</i>						
Mean	0.334	-0.066	-0.027	0.010	-0.013	0.314
t-value	18.286	-20.212	-7.415	8.729	-9.585	
Mean of t-value	5.579	-5.586	-3.936	2.792	-1.706	
%(+)	100.000	0.000	0.000	94.118	2.941	
%signif (+)	100.000	0.000	0.000	73.529	0.000	
%signif (-)	0.000	100.000	94.118	0.000	50.000	
mean	0.552	-0.115				0.250
t(mean)	69.968	-45.616				
mean of t-value	20.686	-12.304				
%(+)	100.000	0.000				
%signif (+)	100.000	0.000				
%signif (-)	0.000	100.000				
Mean	0.269		-0.057			0.184
t-value	68.560		-9.875			
Mean of t-value	43.203		-10.095			
%(+)	100.000		0.000			
%signif (+)	100.000		0.000			
%signif (-)	0.000		100.000			
Mean	-0.082			0.028		0.167
t-value	-7.038			31.211		
Mean of t-value	-2.483			9.484		
%(+)	14.706			100.000		
%signif (+)	5.882			100.000		
%signif (-)	70.588			0.000		
Mean	0.270				-0.053	0.113
t-value	106.690				-30.070	
Mean of t-value	37.479				-7.601	
%(+)	100.000				0.000	
%signif (+)	100.000				0.000	
%signif (-)	0.000				100.000	
<i>Panel B. time-series and cross-sectional regression</i>						
coefficient	0.341	-0.070	-0.015	0.009	-0.014	0.313
t-value	(33.39)	(-34.83)	(-20.61)	(16.32)	(-11.79)	
coefficient	0.555	-0.116				0.259
t-value	(123.14)	(-73.32)				
coefficient	0.261		-0.038			0.165
t-value	(250.44)		(-55.13)			
coefficient	-0.088			0.028		0.165
t-value	(-15.17)			(55.21)		
coefficient	0.271				-0.055	0.124
t-value	(215.31)				(-46.69)	

Table XI

Month-by-month cross-sectional determinants of information share of best quote, upper quotes and lower quotes in limit order book

This table reports the result for determinants of Hasbrouck information share of non-best quotes using standard Fama-MacBeth (1973) regression and time-series and cross-sectional pooling regression. Dependant variables are Hasbrouck information share of depth-weighted average price from step 2 to 5 (IS^{2-5}) and depth-weighted average price from step 2 to 10 (IS^{2-10}) estimated month-by-month by using VECM specification during sample period from January 2002 to November 2004 across 460 sample stocks. Explanatory variables are tick size-normalized distance (ND), liquidity provision-to-liquidity consumption ratio (LP/LS), firm size, and book-to-market ratio. Tick size normalized-distance is measured as difference between best quote and other steps in limit order book, divided by tick size corresponding to price level. Ratio of liquidity provision to consumption is computed as the depth divided by monthly share volume. Firm size is monthly log market value, multiplying closing price by total outstanding shares at the end of month. Panel A contains time-series means and t-value of the coefficients and time-series mean of t-value for the coefficients in cross-sectional regression. + (%) indicates the percentage of number of months with positive sign of coefficient. % signif (+) indicates the the percentage of number of months with positive sign of coefficient significant at 5% level. % signif (-) indicates the the percentage of number of months with negative sign of coefficient significant at 5% level. Panel B contains coefficients and t-value in parenthesis from time-series and cross-sectional pooling regression.

	<i>dependant variable : IS²⁻⁵</i>						<i>dependant variable : IS⁶⁻¹⁰</i>					
	<i>intercept</i>	<i>ND²⁻⁵</i>	<i>(LP²⁻⁵ / LS)</i>	<i>SIZE</i>	<i>BM</i>	<i>adjR²</i>	<i>intercept</i>	<i>ND⁶⁻¹⁰</i>	<i>(LP⁶⁻¹⁰ / LS)</i>	<i>SIZE</i>	<i>BM</i>	<i>adjR²</i>
<i>Panel A. Fama-MacBeth regression</i>												
Mean	0.659	-0.074	-0.038	-0.017	-0.034	0.079	0.204	-0.026	-0.036	-0.001	-0.011	0.125
t-value	22.590	-11.113	-5.270	-9.076	-12.221		13.673	-10.127	-6.181	-1.103	-10.135	
Mean of t-value	6.727	-3.083	-2.105	-2.660	-2.645		4.539	-3.042	-2.813	-0.291	-1.738	
%(+)	100.000	2.941	8.824	2.941	0.000		100.000	5.882	2.941	41.176	2.941	
%signif (+)	100.000	0.000	0.000	0.000	0.000		94.118	0.000	0.000	23.529	0.000	
%signif (-)	0.000	82.353	70.588	70.588	76.471		0.000	76.471	79.412	38.235	52.941	
mean	0.448	-0.083				0.042	0.215	-0.045				0.093
t(mean)	40.584	-12.718					42.777	-24.354				
mean of t-value	11.681	-4.190					13.548	-6.701				
%(+)	100.000	2.941					100.000	0.000				
%signif (+)	100.000	0.000					100.000	0.000				
%signif (-)	0.000	88.235					0.000	100.000				
Mean	0.318		-0.061			0.034	0.132		-0.059			0.081
t-value	65.364		-9.559				43.364		-8.796			
Mean of t-value	30.955		-3.809				27.067		-6.286			
%(+)	100.000		0.000				100.000		0.000			
%signif (+)	100.000		0.000				100.000		0.000			
%signif (-)	0.000		88.235				0.000		100.000			
Mean	0.250			0.004		0.006	0.010			0.009		0.041
t-value	12.093			2.616			1.057			11.616		
Mean of t-value	4.372			0.914			0.268			4.066		
%(+)	100.000			70.588			52.941			97.059		
%signif (+)	97.059			35.294			29.412			91.176		
%signif (-)	0.000			11.765			17.647			0.000		
Mean	0.323				-0.037	0.027	0.131				-0.024	0.045
t-value	70.911				-13.783		59.545				-23.156	
Mean of t-value	27.765				-3.399		23.773				-4.598	
%(+)	100.000				2.941		100.000				0.000	
%signif (+)	100.000				0.000		100.000				0.000	
%signif (-)	0.000				88.235		0.000				97.059	

Panel B. time-series and cross-sectional regression

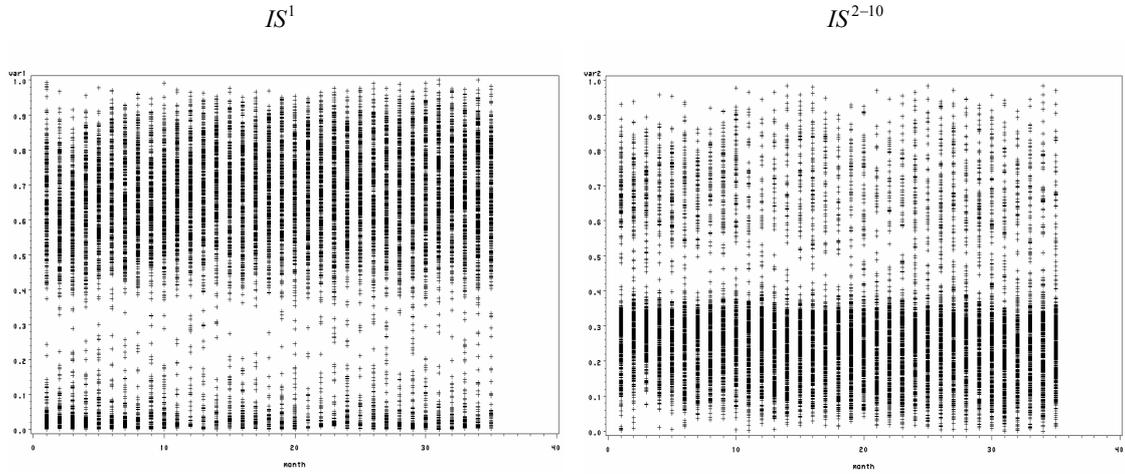
coefficient	0.661	-0.074	-0.027	-0.016	-0.038	0.072	0.211	-0.030	-0.019	-0.001	-0.012	0.117
t-value	(39.62)	(-17.66)	(-12.51)	(-15.92)	(-17.16)		(27.93)	(-20.28)	(-13.86)	(-2.22)	(-11.61)	
coefficient	0.454	-0.086				0.038	0.217	-0.045				0.096
t-value	(68.60)	(-24.68)					(81.40)	(-40.49)				
coefficient	0.314		-0.045			0.031	0.128		-0.040			0.070
t-value	(187.00)		(-22.20)				(158.83)		(-34.22)			
coefficient	0.241			0.004		0.001	0.006			0.009		0.035
t-value	(24.85)			(5.52)			(1.32)			(23.64)		
coefficient	0.325				-0.040	0.029	0.132				-0.025	0.050
t-value	(160.81)				(-21.50)		(138.28)				(-28.74)	

Figure I

Empirical distribution and monthly time-series trend of information share of best quote and non-best quotes in limit order book

This graph depicts monthly empirical distribution and time-series plot for Hasbrouck information share of best quote (MP^I) and depth-weighted average price from step 2 to 10 (WP^{2-10}) on the basis of monthly volume. All sample stocks are categorized into 5 groups on the ranking of volume (low[1]~high[5]). Hasbrouck information share is estimated by using VECM specification with inclusion of best quote and depth-weighted average price from step 2 to 10 observed at 5 minute-intervals during continuous auction from 9:00 to 14:50. VECM is estimated with lag selection on the basis of Bayesian Information Criteria (BIC) month-by-month during sample period from January 2002 to November 2004 across 460 sample stocks. Panel A graphs the monthly empirical distribution for Hasbrouck information share of midpoint price (IS^I) and depth-weighted average price from step 2 to 10 (IS^{2-10}) from 460 sample stocks. Panel B depicts the monthly trend of volume-based portfolio.

Panel A. monthly empirical distribution



Panel B. monthly trend of volume-based portfolio

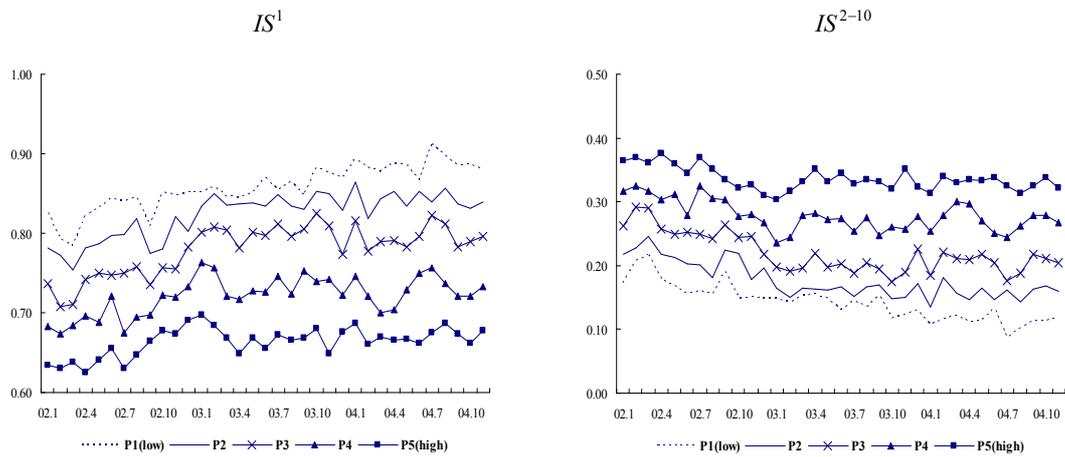
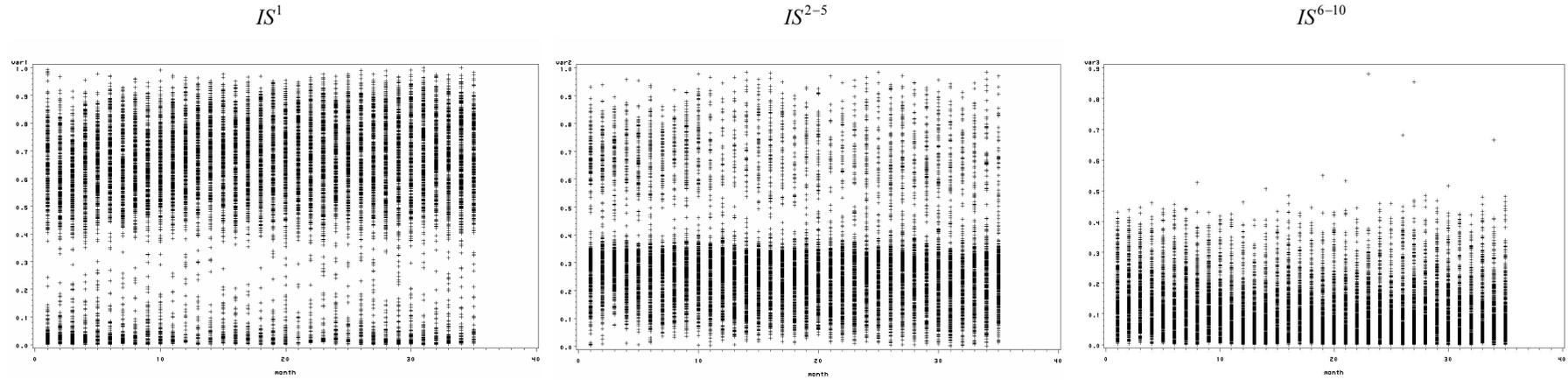


Figure II

Empirical distribution of information share of best quote, upper quote, and lower quote in limit order book

This graph depicts monthly empirical distribution and time-series plot for Hasbrouck information share of best quote (MP^I), depth-weighted average price from step 2 to 5 (WP^{2-5}) and depth-weighted average price from step 6 to 10 (WP^{6-10}) on the basis of monthly volume. All sample stocks are categorized into 5 groups on the ranking of volume (low[1]-high[5]). Hasbrouck information share is estimated by using VECM specification with inclusion of MP^I , WP^{2-5} and WP^{6-10} observed at 5 minute-intervals during continuous auction from 9:00 to 14:50. VECM is estimated with lag selection on the basis of Bayesian Information Criteria (BIC) month-by-month during sample period from January 2002 to November 2004 across 460 sample stocks. Panel A graphs the monthly empirical distribution for Hasbrouck information share of midpoint price (IS^I), depth-weighted average price from step 2 to 5 (IS^{2-5}) and depth-weighted average price from step 2 to 10 (IS^{2-10}) from 460 sample stocks. Panel B depicts the monthly trend of volume-based portfolio.

Panel A. monthly empirical distribution



Panel B. monthly trend of volume-based portfolio

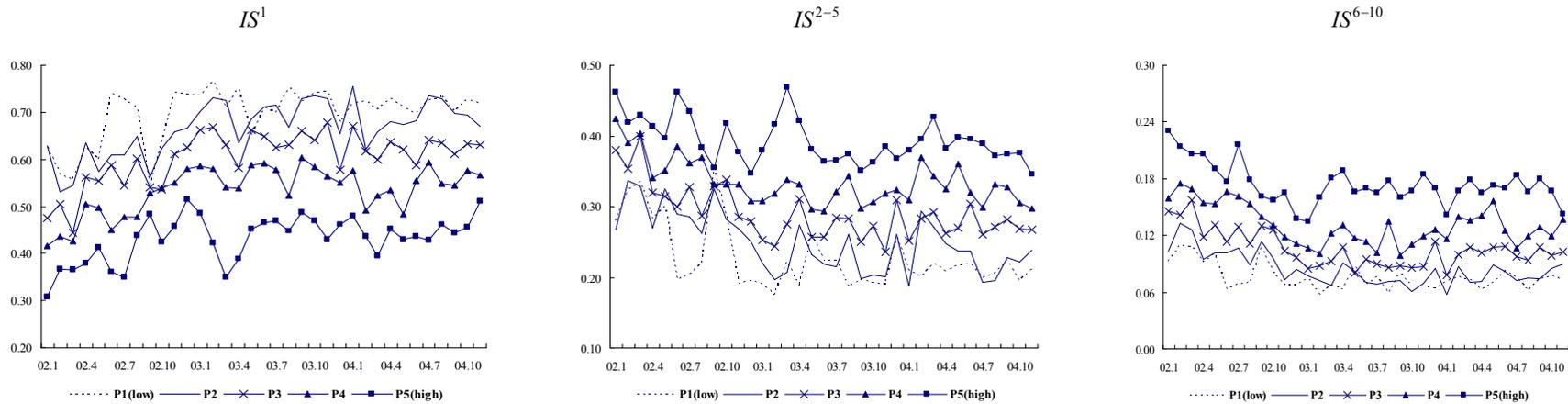
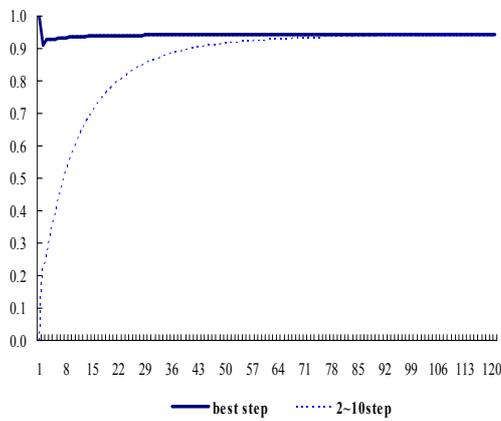


Figure III

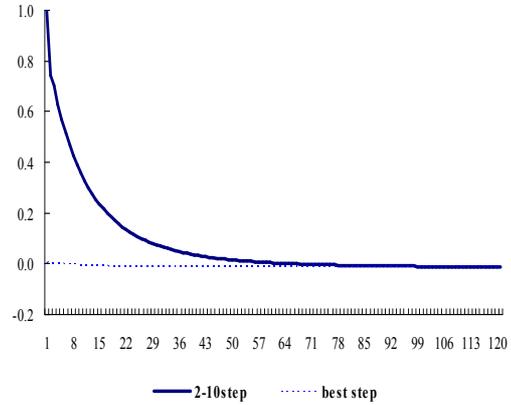
Impulse response function analysis of best quote and non-best quote in limit order book

This graph depicts the result for impulse response function analysis of best quote(MP^l) and depth-weighted average price from step 2 to 10(WP^{2-10}). Cumulative impulse response function is estimated by using VECM specification with inclusion of best quote and depth-weighted average price from step 2 to 10 observed at 5 minute-intervals during continuous auction from 9:00 to 14:50. Midpoint quote price proxies for best quote. VECM is estimated with lag selection on the basis of Bayesian Information Criteria (BIC) month-by-month during sample period from January 2002 to November 2004 across 460 sample stocks. The cross-sectional means of impulse response function are calculated from 460 sample stocks and then averaged across 35 months. In panel A, cumulative impulse response is depicted when initial shock is given to MP^l and WP^{2-10} . In panel B, for 5 portfolios on the basis of volume, cumulative impulse response of WP^{2-10} is depicted when initial shock is given to MP^l (left) and cumulative impulse response of MP^l is depicted when initial shock is given to WP^{2-10} (right).

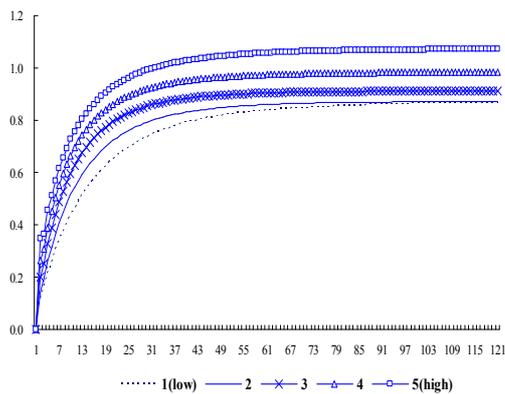
Panel A. Response of MP^l and WP^{2-10} to initial shock to MP^l



Response of MP^l and WP^{2-10} to initial shock to WP^{2-10}



Panel B. Response of WP^{2-10} to initial shock to MP^l sorted by volume



Response of MP^l to initial shock to WP^{2-10} sorted by volume

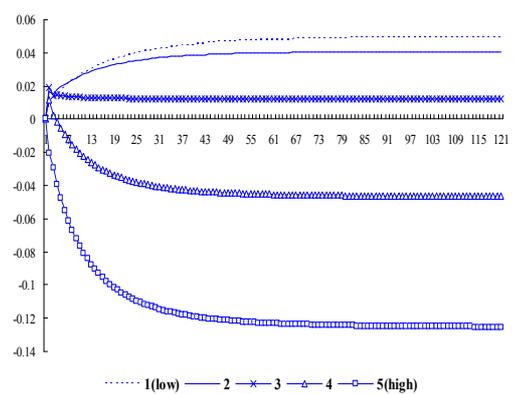


Figure IV

Impulse response function analysis of best quote, upper quote, and lower quote in limit order book

This graph depicts the result for impulse response function analysis of best quote(MP^1) and depth-weighted average price from step 2 to 5(WP^{2-5}), and depth-weighted average price from step 6 to 10(WP^{6-10}). Cumulative impulse response function is estimated by using VECM specification with inclusion of MP^1 , WP^{2-5} and WP^{6-10} observed at 5 minute-intervals during continuous auction from 9:00 to 14:50. Midpoint quote price proxies for best quote. VECM is estimated with lag selection on the basis of Bayesian Information Criteria (BIC) month-by-month during sample period from January 2002 to November 2004 across 460 sample stocks. The cross-sectional means of impulse response function are calculated from 460 samples stocks and then averaged across 35 months.

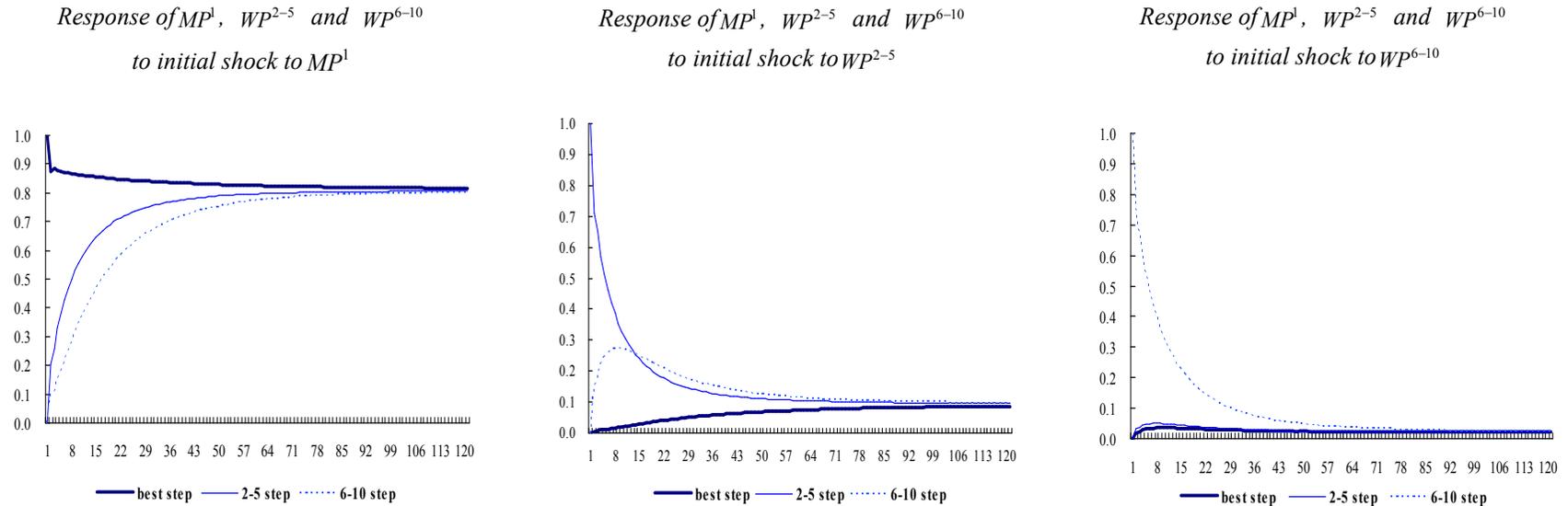
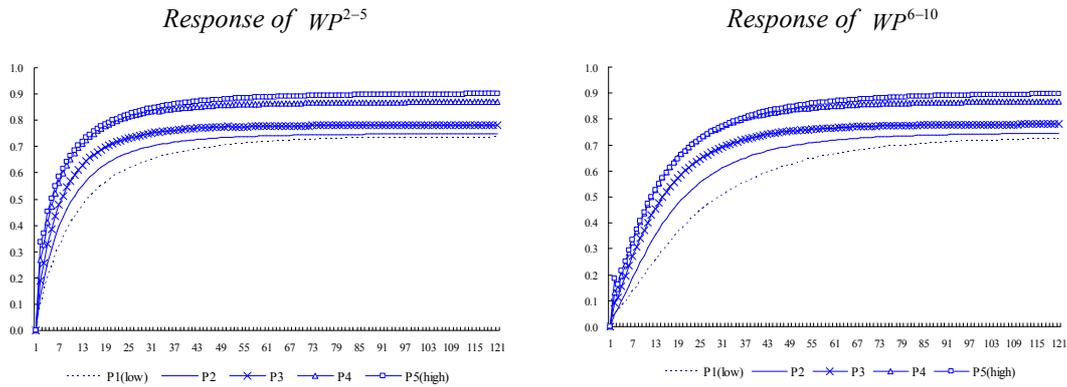


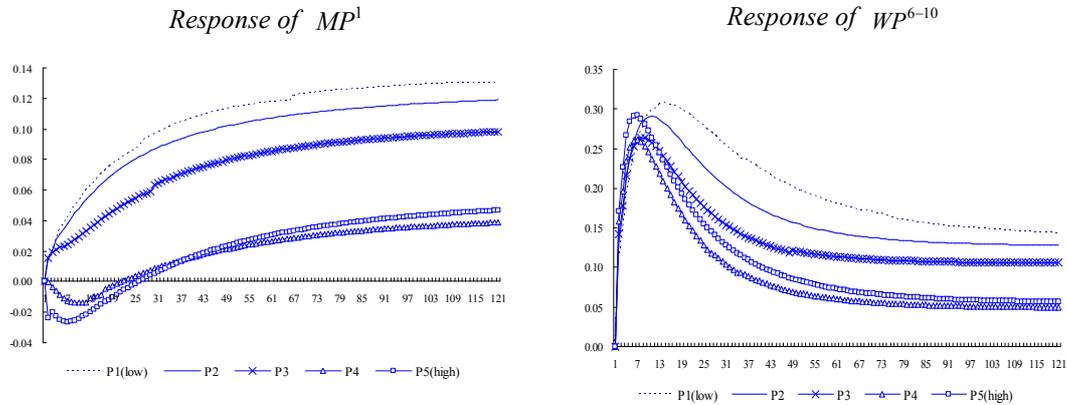
Figure V
Impulse response function analysis of best quote, upper quote, and lower quote
in limit order book on the basis of volume

This graph depicts the result for impulse response function analysis of best quote(MP^1) and depth-weighted average price from step 2 to 5(WP^{2-5}), and depth-weighted average price from step 6 to 10(WP^{6-10}) on the basis of volume. Cumulative impulse response function is estimated by using VECM specification with inclusion of MP^1 , WP^{2-5} and WP^{6-10} observed at 5 minute-intervals during continuous auction from 9:00 to 14:50. Midpoint quote price proxies for best quote. VECM is estimated with lag selection on the basis of Bayesian Information Criteria (BIC) month-by-month during sample period from January 2002 to November 2004 across 460 sample stocks. The cross-sectional means of impulse response function are calculated for 5 portfolios and then averaged across 35 months.

Panel A. initial shock to MP^1



Panel B. initial shock to WP^{2-5}



Panel C. initial shock to WP^{6-10}

