

# 한국주요상장사 주가 실현변동성 추정시 시장미시구조 잡음과 최적 추출 빈도수

오로지<sup>1</sup> · 신동완<sup>2</sup>

<sup>1</sup>이화여자대학교 통계학과, <sup>2</sup>이화여자대학교 통계학과

(2011년 12월 24일 접수, 2012년 1월 18일 수정, 2012년 1월 18일 채택)

## 요약

본 논문에서는 KOSPI 시가총액기준 상위 4종목(삼성전자, 현대차, 현대모비스, POSCO)의 고빈도 거래 데이터를 바탕으로 일중 수익률의 실현변동성과 시장미시구조잡음에 대해 연구한다. Volatility signature plot을 통해 실현변동성(Realized Variance; RV)과 편이수정 실현변동성( $RV_{AC_1}$ )의 편의를 확인하고 시장미시구조 잡음의 특징을 실증적으로 파악한다. 또한, 잡음 대 신호비(Noise-to-Signal Ratio; NSR)를 사용하여, 평균제곱오차(Mean Square Error; MSE) 기준의 실현변동성(RV)과 편이수정 실현변동성( $RV_{AC_1}$ )의 최적 추출 빈도수를 추정해본다.

주요용어: 실현변동성(Realized Variance), 변동성, 고빈도 자료, 시장 미시구조 잡음, 편이.

## 1. 서론

변동성(volatility)을 통해 측정되는 위험성은 자산의 포트폴리오 최적화, 각종 투자결정, 자산의 위험 관리와 파생상품의 가격책정을 하는데 핵심이 되는 변수이며 이론적으로나 현실적으로나 금융과 금융공학에서 중요한 위치를 차지하고 있는 분야이다. 이런 추세를 바탕으로, 변동성의 측정에 대한 새로운 접근방식이 Andersen과 Bollerslev (1998), Andersen 등 (1999) 등에 의하여 연구되기 시작하였다. 이들은 Hull과 White (1987)의 적분변동성 개념을 응용하여 고빈도 데이터의 일중 수익률(intraday returns)로부터 일중변동성을 계산하는 방법을 제시하였고, 이를 실현변동성(Realized Variance; RV)이라고 정의하였다.

실현변동성(RV)은 과거 변동성(historical volatility)의 예측을 평가하는 중요한 역할을 한다 (Chen 등, 2010; Pascuala 등, 2006). 또한, 모형을 바탕으로 하는 전통적인 변동성 측정방법과는 달리 변동성에 대한 모형 설정 없이 변동성을 관측변수로서 다룰 수 있게 해주며 계산이 간단하고 정확한 값을 추정하는 이점이 있다. 최근 실현변동성은 많은 학자들의 주목을 받아 다수의 연구결과들이 발표되었다. 중요한 결과들이 Poon과 Granger (2003), Barndorff-Nielsen과 Shephard (2007), McAleer과 Medeiros (2008) 등의 요약논문에 잘 정리되어있다. 또한 Shin과 Park (2011)은 특정주식의 RV를 추정함에 있어서 더 고빈도로 관측되어지는 다른 주식과의 교차상관을 이용하여 효율을 향상시킬 수 있음을 지적하였다.

이 연구는 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (과제번호: 2011-0030811).

<sup>2</sup>교신저자: (120-750) 서울특별시 서대문구 대현동 11-1, 이화여자대학교 통계학과, 교수.

E-mail: shindw@ewha.ac.kr

적분변동성(Integrated Variance; IV)의 추정치로서 실현변동성(RV)의 추정효율은 샘플의 간격 즉 주가 관측간격이 짧을수록 증가하게 되고 그 간격이 0으로 접근함에 따라 일치성을 지니게 된다. 그러나 이것은 수익률이 무자기 상관이라는 전제하에 성립되는 것이다.

지난 10년 동안 NYSE의 TAQ 데이터베이스 등을 통해서 1분, 1초, 틱단위 등 초고빈도 가격데이터의 이용이 가능해졌으나 짧은 간격에 얻어지는 수익률은 비동시적거래, 주문가체결가불일치, 뜬한 거래 등에 의해 야기되는 시장미시구조 잡음에 영향을 받게 되고 이것이 관측 수익률의 잡음에 MA(1) 구조의 자기상관을 유발시키며 나아가 실현변동성이 편의를 갖게 만든다 (Oomen, 2002; Bai 등, 2004).

Andersen 등 (2000)은 실현변동성의 편의는 일중 수익률의 시장미시구조 잡음(market microstructure noise)으로 인한 자기상관으로 발생한 것이며 이는 volatility signature plots로 확인할 수 있음을 보였다. 이와 같은 편의의 문제를 줄이기 위해 Andersen 등 (2001) 등은 5분 단위 데이터 사용을 권장하였지만 이처럼 추출 빈도수를 줄이는 것은 RV의 추정효율을 저하시키는 데이터의 손실이라는 문제를 야기한다.

따라서 Newey와 West (1987)과 Andrews (1991)가 시장미시구조 잡음이 존재하는 고빈도 데이터에서 편의를 야기하는 자기상관을 수정한 추정치를 제안하였고 Zhou (1996)는 kernel 추정치와 부표본 추출(subsampling) 추정치를 제안하였다. 또한, Griffin과 Oomen (2008)는 거래가 일어날 때마다 기록되는 거래시간 단위 추출방법(transaction time sampling)과 거래가격이 변할 때 마다 기록되는 틱 시간 단위 추출방법(tick time sampling)에 대해 논의하였다. Bandi와 Russell (2005)와 Hansen과 Lunde (2006)는 평균제곱오차(MSE)를 최소로 하는 최적 추출 빈도수를 추정하였다.

본 논문에서는 국내 주요 상장회사 주가의 고빈도 자료와 이를 바탕으로 한 일중수익률의 실현변동성(RV)의 편의와 시장미시구조 잡음의 특징을 연구하고 최적 추출 빈도수 추정을 시도한다.

2절에서는 실현변동성과 편의수정 실현변동성의 이론과 자료를 설명하고 3절에서는 실현변동성의 편의를 확인하고 시장미시구조 잡음의 특징을 살펴본다. 또한 잡음의 모수를 추정하고 잡음 대 신호비(Noise-to-Signal Ratio; NSR)를 이용하여, 최적 추출 빈도수를 추정한다.

## 2. 실현변동성과 자료설명

### 2.1. 실현변동성

실현변동성(RV)은 과거 특정시점 이후 현재까지 실현된 기초자산의 변동성을 의미하며, Protter (2005)과 Barndorff-Nielsen과 Shephard (2002)에 따르면 효율적 일중수익률의 실현변동성(RV)은 시간간격이 0에 접근하게 되면 적분변동성(IV)에 대해 일치 추정량이 된다는 장점을 갖고 있다. 이어서 RV와 IV에 대해 간략히 설명한다.

연속시간에서의 특정자산의 효율적 로그가격과정을  $p^*(t)$ 라고 하고 관측 가능한 로그가격과정을  $p(t)$ 라 하자. 시장 미시구조(market microstructure)의 영향으로 생기는 잡음 과정은  $u(t) = p(t) - p^*(t)$ 와 같이 표현된다. 효율적 로그가격과정( $p^*(t)$ )은 다음의 확률미분방정식  $dp^*(t) = \sigma(t)dw(t)$ 을 만족한다. 여기에서  $w(t)$ 는 표준 브라운운동(standard brownian motion)이고  $\sigma(t)$ 는 시점  $t$ 까지의 정보에 의해 정해지는 함수이다. 일중 수익률의 구간  $[a, b]$ 의 적분변동성(Integrated Variance; IV)를 정의하면 다음과 같다.

$$IV \equiv \int_a^b \sigma^2(t)dt.$$

일중 구간  $[a, b]$ 을  $m$ 개의 구간으로 나누었을 때,  $i$ 번째 구간을  $[t_{i-1,m}, t_{i,m}]$ 라고 하자.  $i$ 번째 구간의 길이는  $\delta_{i,m} = t_{i,m} - t_{i-1,m}$ 와 같이 나타낼 수 있으며 효율적 일중 수익률과 관측가능 일중 수익률을

$$y_{i,m}^* = p^*(t_{i,m}) - p^*(t_{i-1,m}), \quad y_{i,m} = p(t_{i,m}) - p(t_{i-1,m})$$

와 같이 정의 할 수 있다. 또  $e_{i,m}$ 는 일중 수익률의 잡음(noise)을 나타내며 다음의 관계

$$e_{i,m} = u(t_{i,m}) - u(t_{i-1,m})$$

를 만족한다. 잡음( $u(t)$ )이 독립성 가정을 따를 때 이는 MA(1) 과정을 따른다. 이를 바탕으로 하위 구간  $m$ 을 갖는 하루 구간  $[a, b]$ 에서 관측가능 일중수익률의 실현변동성을 정의하면 다음과

$$RV^{(m)} \equiv \sum_{i=1}^m y_{i,m}^2$$

같다. Bandi와 Russell (2005)과 Zhang 등 (2005)에 따르면 관측가능 일중수익률의 실현변동성 ( $RV^{(m)}$ )은 MA(1) 구조의 자기 상관에 의해 야기되는 편의(bias)를 갖고 있으며 적분변동성(IV)에 대해 불일치성(inconsistent)을 갖는다.

## 2.2. 편의 수정 실현변동성( $RV_{AC}$ )

고빈도 데이터 추출 시, 데이터의 정보손실을 막으면서 실현변동성의 편의를 줄이기 위해 Zhou (1996)는 자기상관을 고려한 실현변동성( $RV_{AC}^{(m)}$ )을 제안하였다.

1차 자기공분산(First-order Autocovariance)을 고려한 편의 수정 실현변동성은

$$RV_{AC1}^{(m)} \equiv \sum_{i=1}^m y_{i,m}^2 + \sum_{i=1}^m y_{i,m}y_{i-1,m} + \sum_{i=1}^m y_{i,m}y_{i+1,m}$$

이다. 위의 추정치는 Newey와 West (1987)의 robust 공분산 추정치와 같은 정도로 편의 수정(bias correction)을 만족하며 일치성(consistency) 또한 만족하는 것으로 알려져 있다.

일중 수익률의 자기공분산 시차를  $q$ 차로 확장하여 수정된 실현변동성의 추정 식은

$$RV_{ACq}^{(m)} \equiv \sum_{i=1}^m y_{i,m}^2 + \sum_{h=1}^q \sum_{i=1}^m (y_{i,m}y_{i-h,m} + y_{i,m}y_{i+h,m})$$

와 같이 표현 할 수 있으며 잡음에 자기상관이 존재할 경우 적분변동성(IV)의 불편추정치이다.

다음, Hansen과 Lunde (2006)에 따르면 1tick 간격 데이터의 일반화한 부표본 추정치(subsample-based estimator)는 근사적으로

$$RV_{ACNW_k}^{(1tick)} \equiv \hat{\gamma}_0 + \sum_{j=1}^k (\hat{\gamma}_{-j} + \hat{\gamma}_j) + \sum_{j=1}^k \frac{k-j}{k} (\hat{\gamma}_{-j-k} + \hat{\gamma}_{j+k})$$

와 같이 표현된다. 여기에서,  $\hat{\gamma}_j \equiv \sum_{i=1}^m y_{i,m}y_{i+j,m}$ 는  $j$ 차 자기공분산이며 Bartlett 커널  $(k-j)/k$ 을 사용했다. 위의 추정치는 일중 수익률의 1차자기공분산 편의 수정 실현변동성( $RV_{AC1}^{(1tick)}$ )과  $k$ 번째 차수 까지의 자기공분산 항의 가중합인 Bartlett 유형으로 표현된다.

표 2.1. 거래데이터의 기본 통계량

종목	평균 거래 수 (개)	가격 변화비율 (%)	평균 수익률 (연 기준, %)	평균 변동성 (연 기준, %)
삼성전자	17,000	29.9%	71.09%	42.02%
현대차	17,880	27.0%	87.17%	77.49%
현대모비스	12,855	29.5%	0.81%	43.75%
POSCO	12,188	28.4%	-106.57%	35.27%

### 2.3. 최적 추출 빈도수(Optimal sampling frequency)

잡음과정이  $u(t) \sim iid(0, \omega^2)$ 일 때, 수익률의 잡음( $e_{i,m}$ )은 MA(1)을 따르는 것을 확인할 수 있다. 잡음의 독립성 가정 하에서 MSE를 기준으로 최적 추출 빈도수를 구하기 위한, Hasen과 Lunde (2006)가 제안한 일증변동성의 실현변동성( $RV^{(m)}$ )과 1차 자기공분산 편의 수정 실현변동성( $RV_{AC_1}^{(m)}$ )의 MSE는 다음과 같다.

$$MSE(RV^{(m)}) = IV^2 \left[ 4\lambda^2 m^2 + 12\lambda^2 m + 8\lambda - 4\lambda^2 + 2\frac{1}{m} \right], \quad (2.1)$$

$$MSE(RV_{AC_1}^{(m)}) = IV^2 \left[ 8\lambda^2 m + 8\lambda - 6\lambda^2 + 6\frac{1}{m} - 2\frac{1}{m^2} \right], \quad (2.2)$$

여기서  $\lambda$ 는 거래데이터의 잡음 대 신호비(Noise-to-Signal Ratio; NSR)  $\lambda \equiv \omega^2/IV$ 이며  $\omega^2$ 는 잡음과정의 분산이다( $\omega^2 \equiv E[u(t)^2]$ ). 이에 관한 자세한 추정은 3.4절에서 다루도록 하겠다.

### 2.4. 자료 설명

본 연구에서는 KOSPI의 시가총액 상위 4위 종목인 삼성전자, 현대차, 현대모비스, POSCO의 거래가격 데이터를 이용하여 분석하였다. 샘플 기간은 2011년 9월 1일부터 9월 30일이며 장외거래를 제외한 오전 9시부터 오후 3시까지 일증 거래 데이터만을 이용하였다.

표 2.1은 거래데이터의 특징을 살펴보기 위한 간단한 통계량이다. 표 2.1의 두 번째 열은 각 종목 별 하루 평균 거래가격의 개수이며 4종목 모두 하루 평균 12,000건이 넘는 거래를 나타내고 있다. 세 번째 열은 거래가격이 이전가격과 다른 경우의 비율을 나타낸 것으로, 4종목 모두 약 30% 정도만 이전가격과 다른 값을 갖고 있다. 실현변동성(RV)을 구하는 경우, 값에 영향을 미치지 않지만, 자기 상관계수(autocorrelation)에 영향을 미치기 때문에 본 연구에서는 관측된 모든 데이터를 사용하였다. 네 번째 열과 다섯 번째 열은 각각 연(year)기준의 평균 수익률과 평균 변동성을 계산한 것이다. 현대차의 평균 수익률은 87.17%(1년기준) 와 평균 변동성은 77.49%(1년기준)로 4종목 중 가장 큰 수준을 보인다.

## 3. 실증 분석

### 3.1. 데이터 추출 방법

데이터는 이어서 설명하는 CTS(Calendar Time Sampling)과 TTS(Transaction Time Sampling)에 의해 수집되었다.  $[a, b]$ 의 하위 간격이  $\delta_{i,m} = (b - a)/m$ 으로 추출 시점  $t_{i,m}, i = 1, \dots, m$ 이 일정한 추출 방법을 달리 시간단위 추출방법(Calendar Time Sampling; CTS)라고 한다. Dacorogna 등 (2001)과 Shin과 Lee (2010)에 따라, 초고빈도 자료를 추출 할 때, 가격이 관측된 시점  $t_0 < \dots < t_N$ 에서, 시점  $\tau \in [t_j, t_{j+1}]$ 에서의 가격 ( $p(\tau)$ )은, 가까운 이전 거래가격으로 대체하는 방법(previous tick method),  $p(\tau) \equiv p_{t_j}$ 을 사용한다. 거래가 일어난 시점 또는 일정 횟수의 거래가 일어난 시점을 추

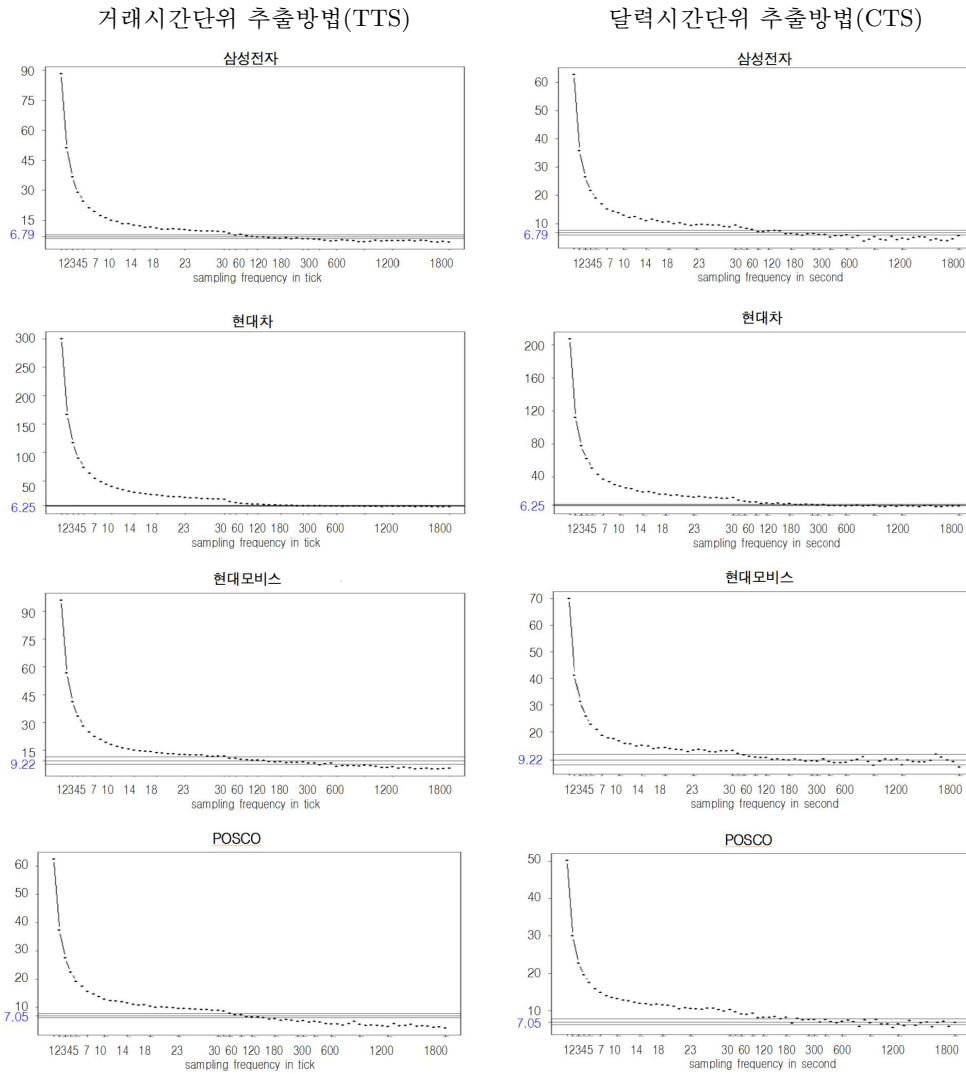


그림 3.1. 실현변동성( $RV^{(m)}$ )의 volatility signature plot

출 시점  $t_{i,m}$ ,  $i = 1, \dots, m$ 으로 사용하는 경우, 거래시간 단위 추출방법(Transaction Time Sampling; TTS)라고 한다. 즉, 거래가 일어난 시점의 거래시간을 그대로 사용하여 데이터를 추출한다.

### 3.2. 실현변동성의 편의(Bias)

3.2.1. 실현변동성( $RV^{(m)}$ )의 편의 Fang (1996)에 의해 소개된 volatility signature plot은 일평균실현변동

$$\overline{RV}^{(m)} \equiv \frac{1}{n} \sum_{t=1}^n RV_t^{(m)}$$

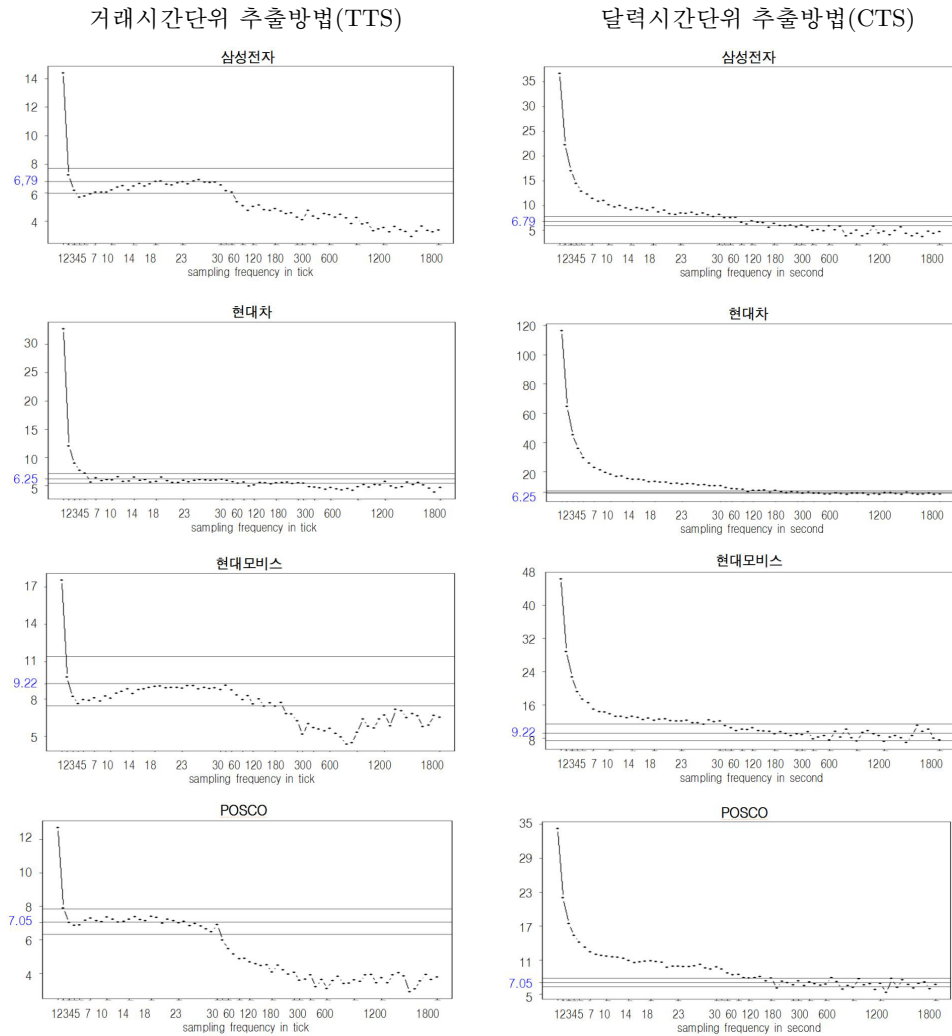


그림 3.2. 1차 편이수정 실현변동성( $RV_{AC1}^{(m)}$ )의 volatility signature plot

을 샘플링 빈도에 따라 표시한 것으로 실현변동성(RV)에 내재된 편이(bias)문제를 시각적으로 보여준다. 본 논문에서 사용한 자료의 경우 9월 한 달 간 자료를 사용하기 때문에  $n = 20$ 이다.

그림 3.1은 매일 실현변동성( $RV_t^{(m)}$ )의 평균한 값( $\overline{RV}^{(m)}$ )을 추출의 빈도수를 1부터 1800까지 다르게 하여 나타낸 것이다. 처음 열은 TTS로 추출된 표본을 바탕으로 한 것이며 두 번째 열은 CTS로 추출된 표본으로 구한 것이다. 수평선은 평균 적분변동성(IV)의 추정치로써 2.2절에 정의된  $\sigma^2 \equiv \overline{RV}_{ACNW_{30}}^{(1tick)}$  과 평균 변동성에 대한 95% 신뢰구간이다.  $\overline{RV}_{ACNW_{30}}^{(1tick)}$  는 일중수익률의 30차 자기상관을 고려한 실현변동성으로 편이가 거의 없을 것으로 간주한다. 또  $\overline{RV}^{(m)}$  가 신뢰구간 밖에 있는 경우 편이가 존재한다고 본다. 4종목 모두 시간 간격이 넓어질수록  $\overline{RV}^{(m)}$  가  $\overline{RV}_{ACNW_{30}}^{(1tick)}$  의 95% 신뢰구간으로 접근하였다. 삼성전자의 경우, TTS의 경우 60tick 간격 이상, CTS의 경우 1분보다 긴 시간 간격으로 추출할 때 신뢰구간에 근접하며 안정적인 값을 보이고 있다. 하지만 그 이상의 고빈도로 추출한 데이터의 경

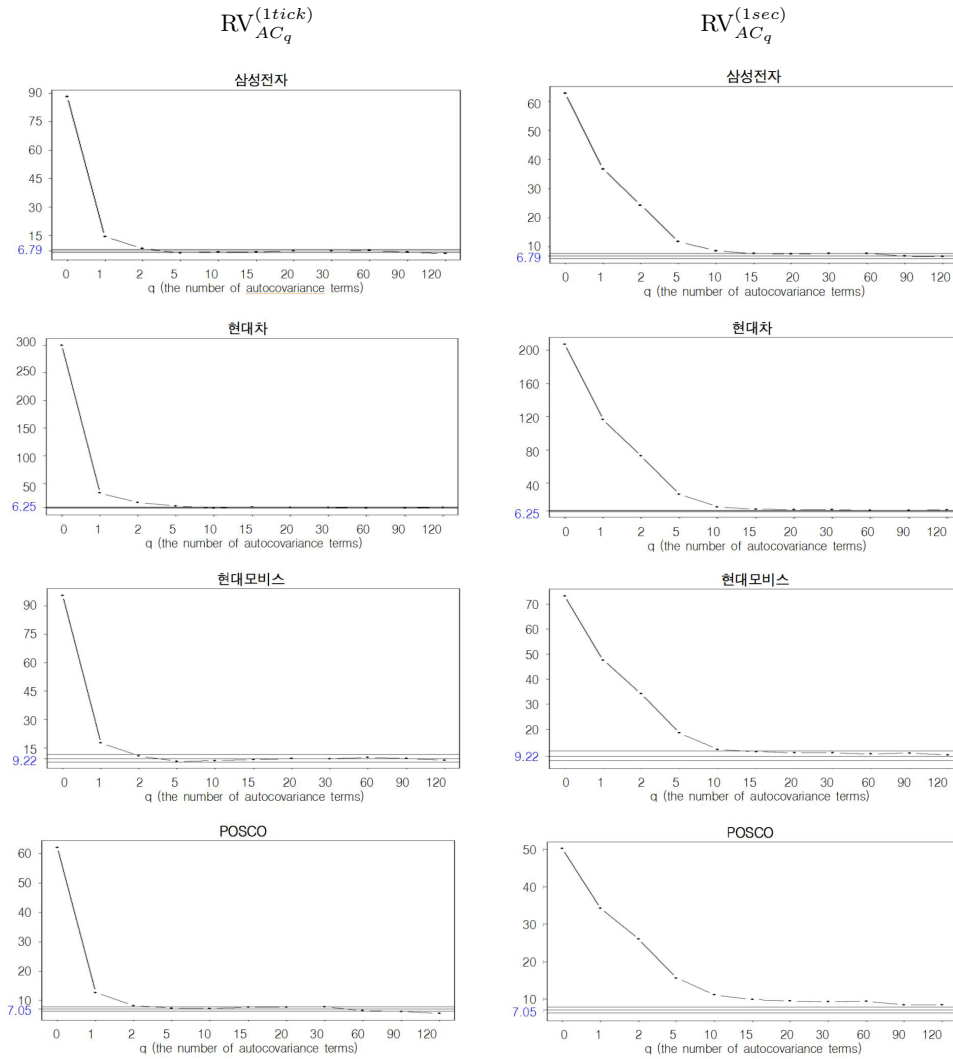


그림 3.3. q차 편의수정 실현변동성( $RV_{AC_q}$ )의 volatility signature plot

우 실현변동성은 믿을만한 정보를 갖고 있지 않으며 이는 시장 미시구조(market microstructure)의 영향으로 인한 편의임을 확인 할 수 있다.

**3.2.2. 편의 수정 실현변동성( $RV_{AC_1}^{(m)}$ )의 편의 실증분석**에서는  $h(> 0)$ 차 공분산을 구하는 경우 Zhou (1996)과 Hansen과 Lunde (2006)에서 제안한 커널(kernel) 추정치를 사용하여  $\tilde{\gamma}_h \equiv m/(m-h) \sum_{i=1}^{m-h} y_{i,m} y_{i+h,m}$ 을 이용하였으며  $h$ 가 음수인 경우에도  $\tilde{\gamma}_h$  식을 이용해서 구하였다.

그림 3.2는 매일 1차 자기상관을 고려한 실현변동성을 평균한 값( $\overline{RV_{AC_1}^{(m)}}$ )을 추출의 빈도수를 1부터 1800까지 다르게 하여 나타낸 것이다. 처음 열은 TTS로 추출된 표본으로 구한 실현변동성이며 두 번째 열은 CTS로 추출된 표본으로 구한 것으로 이 둘 모두 실현변동성의 한 달 평균값이다.

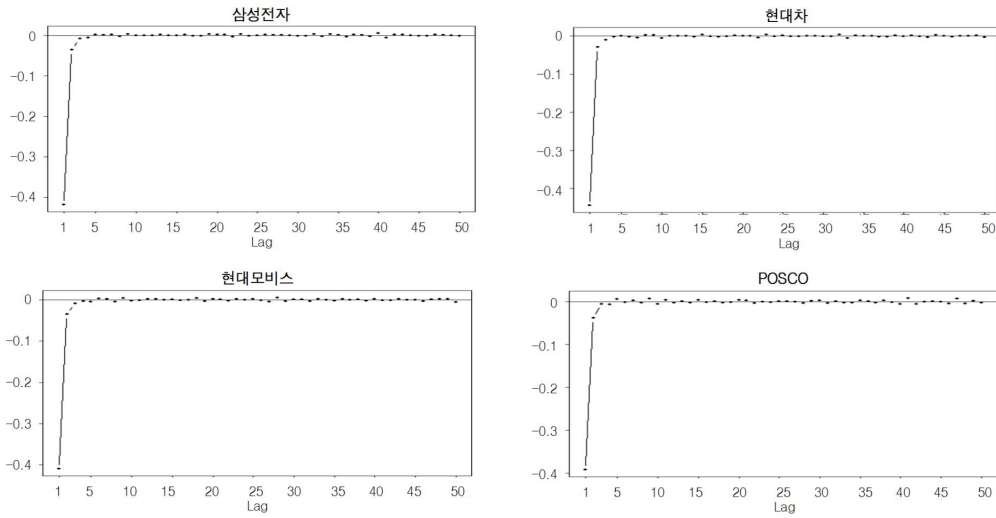


그림 3.4. 주가 4종목의 일중수익률( $y_{i,m}$ )의 평균 ACF

TTS의 경우, 삼성전자, 현대모비스, POSCO는 고빈도 추출을 할 때 편이가 거의 없었지만 오히려 60tick 이상으로 추출할 때 실현변동성이 적분변동성(IV)추정함에 있어서 편이가 존재함을 나타냈다. 반면에, CTS의 경우 4종목 모두 추출의 간격이 1분보다 더 짧게 고빈도 추출을 할 때 편이수정 실현변동성( $RV_{AC_1}^{(m)}$ )에 편이가 발생함을 확인 할 수 있다. CTS에서 추출 빈도수( $m$ )가 증가 할수록 여러 시점에서 같은 가격을 갖게 되기 때문에 자기상관을 야기하게 된다. 따라서 CTS 방법으로 수익률을 구할 때, 샘플수가 증가 할수록 더 높은 차수의 자기상관 조정이 필요한 것으로 보인다.

### 3.3. 시장미시구조 잡음(Market microstructure noise)의 특성 및 모수 추정

**3.3.1. 시간 의존성(Time dependence)** 그림 3.3은 2.2절에서 다룬  $q$ 차 편이수정 실현변동성( $RV_{AC_q}^{(m)}$ )의 그래프이며  $x$ 축은 자기공분산의 차수( $q$ )이다. 왼쪽 옆은 자기공분산의 차수( $q$ )가 변화함에 따른 1tick 단위 일중수익률의 실현변동성( $RV_{AC_q}^{(1tick)}$ )의 변화이다. 오른쪽 옆은 차수( $q$ )가 변화에 따른 1초 단위로 추출한 데이터의 실현변동성( $RV_{AC_q}^{(1sec)}$ )의 변화 그래프이다.  $RV_{AC_q}^{(1tick)}$  또는  $RV_{AC_q}^{(1sec)}$ 의 수준이 일정하지 않으면 잡음의 시간 의존성을 나타낸다. 1tick 단위의 데이터의 경우, 1차 자기공분산 수정이 실현변동성 값에 큰 영향을 주었고 2차 자기공분산 수정부터는 안정적인 값을 보인다. 1초단위의 데이터의 경우에는 10차 자기공분산을 수정한 실현변동성이전에 시간 의존성을 보인다.

그림 3.4는 일중 거래데이터의 수익률의 자기상관 함수(AutoCorrelation Function; ACF)의 평균값이며 이 그래프를 통해서 또한 잡음과정의 시간 의존성을 확인 할 수 있다. 그림 3.4는 평균값을 나타낸 것이기 때문에 몇몇 일(day)은 더 긴 자기상관 차수동안 시간의존성을 내재하고 있을 수 있지만, 평균적으로 4종목 모두 1차수 이후에 0에 수렴하는 값을 가지고 있어 수익률의 잡음이 MA(1) 과정을 따르는 것으로 판단되며, 즉, 잡음의 독립성 가정을 만족하는 것으로 보인다. 하지만 그림 3.3에서  $RV_{AC_1}^{(1tick)}$ 과  $RV_{AC_{30}}^{(1tick)}$ 의 차이가 존재하기 때문에 높은 차수의 자기공분산을 무시할 수 없다.

**3.3.2. 잡음의 모수 추정(Estimator of noise)** Bandi와 Russell (2005)와 Zhang 등 (2005)에 따



표 3.1. 잡음분산 모수  $\omega^2$ 의 추정 결과

종목	$\tilde{\omega}^2$	$\hat{\omega}^2$	$\hat{\omega}^2$
삼성전자	0.0027	0.0025	0.0022
현대차	0.0086	0.0085	0.0077
현대모비스	0.0038	0.0035	0.0031
POSCO	0.0028	0.0025	0.0022

표 3.2. 잡음 대 신호비( $\lambda$ )의 추정 결과

종목	$\hat{\omega}^2 \cdot 100$	$\overline{IV}$	$\hat{\lambda} \cdot 100$
삼성전자	0.2228	14.4030	0.0155
현대차	0.7679	32.7426	0.0235
현대모비스	0.3081	17.5626	0.0175
POSCO	0.2161	12.7198	0.0170

르면 잡음(noise)의 독립성 가정 하에서,  $E(RV^{(m)}) = IV + 2m\omega^2$ 로부터

$$\tilde{\omega}^2 \equiv \frac{RV^{(m)}}{2m}$$

와 같이 잡음의 모수( $\omega^2$ )를 추정 할 수 있다. 또한, Hansen과 Lunde (2006)은 다음과 같은

$$\hat{\omega}^2 \equiv \frac{RV^{(m)} - RV^{(13)}}{2(m-13)}, \quad \hat{\omega}^2 \equiv \frac{RV^{(m)} - \widehat{IV}}{2m}.$$

추정 방법을 이용하였다.  $RV^{(13)}$ 은 30분 간격으로 추출한 데이터의 실현변동성이며,  $\widehat{IV}$ 는  $RV_{AC_1}^{(1tick)}$ 로 적분변동성(IV)의 불편추정치이다. 다음은 각 4종목의 거래데이터의 잡음의 모수( $\omega^2$ )를 세 가지 방법으로 추정한 값이다.

표 3.1을 보면,  $\tilde{\omega}^2$ ,  $\hat{\omega}^2$ ,  $\hat{\omega}^2$  세 가지 방법의 잡음의 모수 추정 값 모두 0.01보다 작은 값을 가지고 있으며  $\tilde{\omega}^2$ ,  $\hat{\omega}^2$ ,  $\hat{\omega}^2$ 의 차이가 크지 않음을 확인 할 수 있다.

### 3.4. 최적 추출 빈도수(Optimal sampling frequency)

본 논문에서는 잡음이 과대 추정 되었을 경우 최적 추출수가 낮아지는 것을 막기 위해  $\hat{\omega}^2$ 를 사용하였다. 다음으로 잡음의 독립성 가정 하에 잡음 대 신호비(NSR)의 추정 식은 다음과 같다.

$$\hat{\lambda} \equiv \frac{\hat{\omega}^2}{\overline{IV}}, \quad \hat{\omega}^2 = n^{-1} \sum_{t=1}^n \hat{\omega}_t^2, \quad \overline{IV} = n^{-1} \sum_{t=1}^n RV_{AC_1,t}^{(1tick)}.$$

다음 표 3.2는 잡음 대 신호비(NSR,  $\lambda$ ) 추정결과이며 4종목 모두 약 0.02% 값을 가진다.

평균제곱오차 제곱근(Root Mean Square Error; RMSE)를 최소로 하는 최적추출 빈도수를 구하고자 한다. 2.3절 식 (2.1)과 식 (2.2)의 RMSE에는 IV가 포함되어 있기 때문에  $RV^{(m)}$ 와  $RV_{AC_1}^{(m)}$ 의 RMSE를 IV로 나눈 부분식

$$r_0(\hat{\lambda}, m) \equiv \sqrt{4\hat{\lambda}^2 m^2 + 12\hat{\lambda}^2 m + 8\hat{\lambda} - 4\hat{\lambda}^2 + \frac{2}{m}}, \quad r_1(\hat{\lambda}, m) \equiv \sqrt{8\hat{\lambda}^2 m + 8\hat{\lambda} - 6\hat{\lambda}^2 + \frac{6}{m} - \frac{2}{m^2}}$$

을 이용하여 이를 최소로 하는  $RV^{(m)}$ 의 최적 추출 빈도수( $m_0^*$ )와  $RV_{AC_1}^{(m)}$ 의 최적 추출 빈도수( $m_1^*$ )를 계산하였다. 다음 표 3.3과 그림 3.5는 최적 추출 빈도수 추정의 결과이다.

표 3.3. 최적 추출 빈도수( $m_0^*$ ,  $m_1^*$ )의 추정 결과

종목	$m_0^*$	$\delta_{i,m_0^*}$ (초)	$m_1^*$	$\delta_{i,m_1^*}$ (초)	$\Delta RMSE$
삼성전자	218	99	5598	4	52.6%
현대차	165	130	3692	6	49.5%
현대모비스	201	107	4937	4	51.7%
POSCO	205	105	5096	4	51.9%

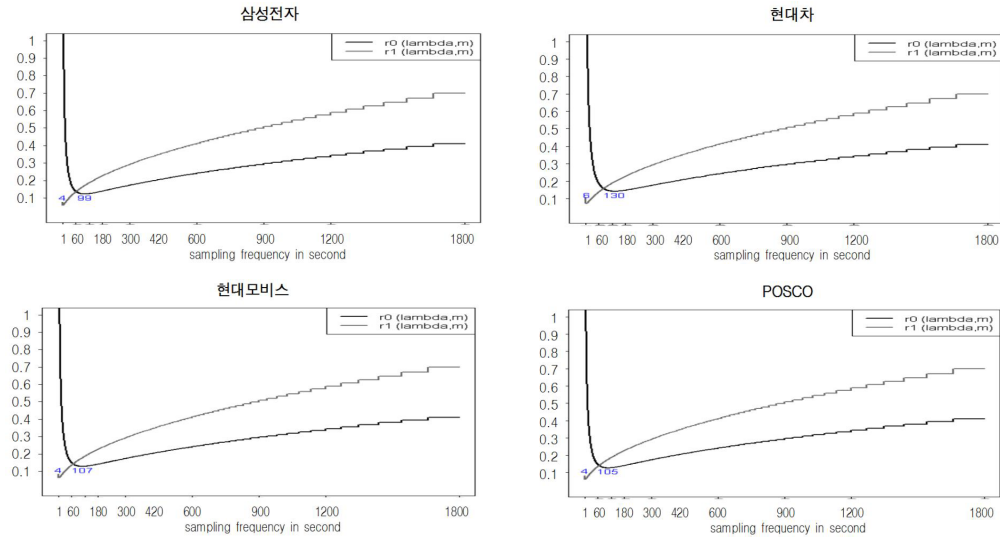
그림 3.5. 실현변동성(RV)과 1차 편이수정 실현변동성( $RV_{AC_1}$ )의 RMSE

표 3.3의 첫 번째와 세 번째 열( $m_0^*$ ,  $m_1^*$ )은 이를 이용하여 구한 최적추출 빈도수이며 두 번째와 네 번째 열( $\delta_{i,m_0^*}$ ,  $\delta_{i,m_1^*}$ )은  $m_0^*$ ,  $m_1^*$ 에 대응하는 시간간격을 나타낸다.  $m_0^*$ 는 대체로 하루에 200개이며  $m_1^*$ 는 대체로 하루에 5000개로  $m_1^*$ 가 더 큰 값을 갖는다. 이는 일중 수익률의 1차 자기공분산을 수정하여 고빈도 데이터 추출 시 발생하는 편의를 어느 정도 제거하였기 때문이다. 표 3.3의 마지막 열은 각각 최적 추출 빈도수로 추출한 일중 수익률의 실현변동성( $RV^{(m_0^*)}$ )에서 편이 수정 실현변동성( $RV_{AC_1}^{(m_1^*)}$ )으로의 이론적 RMSE의 감소 비율이다. 그 결과로 4종목 모두 50% 정도의 감소함을 확인하였다.

그림 3.5는  $r_0(\hat{\lambda}, m)$ 와  $r_1(\hat{\lambda}, m)$ 의 그래프이며  $x$ 축의 단위는 초단위로, 최적 추출 빈도수에 대응하는 시간 간격을 시각적으로 확인 할 수 있다. 이를 통해 실현변동성( $RV^{(m)}$ )의 경우 최소값은 2분 간격에 대응하며 편이 수정 실현변동성( $RV_{AC_1}^{(m)}$ )의 경우 최소값은 약 6초 간격에 대응한다. 또한, 그림 3.5를 통해 편이 수정 실현변동성( $RV_{AC_1}$ )의 경우 추출 간격이 길수록 1분 단위 이내의 고빈도 추출을 할 때보다 MSE가 급격히 증가하여 오히려 추출 간격이 길수록 문제를 야기하는 것을 확인 할 수 있었다. 반면에 실현변동성(RV)의 경우 추출간격이 상대적으로 짧은 1분이내의 고빈도 추출을 할 때 MSE가 높은 수준을 보여 편이 수정 실현변동성( $RV_{AC_1}$ )과 상반되는 결과를 나타냈다.

#### 4. 결론 및 논의

본 연구에서는 KOSPI의 시가총액 상위 4종목(삼성전자, 현대차, 현대모비스, POSCO)의 거래가격 데이터를 이용하여 실현변동성의 편이와 최적 추출 빈도수에 대해 알아보았다. 먼저, volatility signature

plot을 살펴본 결과로 실현변동성( $RV^{(m)}$ )과 편이 수정 실현변동성( $RV_{AC_1}^{(m)}$ )은 대체로 2분보다 짧은 시간간격으로 고빈도 데이터를 추출할 때 편의를 확인 할 수 있었다. 또한, CTS로 추출 간격을 늘여가며 실현변동성( $RV^{(m)}$ )과 편이 수정 실현변동성( $RV_{AC_1}^{(m)}$ )을 구하였을 때, 편이가 더 오래 지속됨을 보였다. 더불어, TTS를 사용하는 경우보다, CTS를 사용하는 경우 잡음의 시간의존성이 더 큰 특징을 갖는다. 따라서 TTS가 편의기준에서 더 나은 추출 방법임을 나타냈다. 다음, 잡음 대 신호비(Noise-to-Signal Ratio; NSR)를 이용하여 평균제곱오차(MSE)기준의 실현변동성( $RV^{(m)}$ )과 편이 수정 실현변동성( $RV_{AC_1}^{(m)}$ )의 최적 추출 빈도수를 추정하였다. 이에 대한 결과로, 실현변동성( $RV^{(m)}$ )의 최적 추출 빈도수는 하루당 200개정도이며 약 2분 간격 추출에 대응하며 1차 자기공분산 편이수정 실현변동성( $RV_{AC_1}^{(m)}$ )은 대체로 하루당 5000개이며 이는 약 6초 간격에 대응함을 확인하였다. 또한, 편이 수정 실현변동성( $RV_{AC_1}^{(m)}$ )의 경우 추출 간격이 넓을수록 MSE가 급격히 증가하여 오히려 넓은 간격으로 데이터를 추출 할 때 문제를 야기하는 것을 확인 할 수 있었다. 반면에 실현변동성( $RV^{(m)}$ )의 경우 추출 간격이 1분(또는 60tick)이내의 고빈도 추출을 할 때 MSE가 높은 수준을 보여 편이 수정 실현변동성( $RV_{AC_1}^{(m)}$ )과 상반되는 결과를 나타냈다.

## 참고문헌

- Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review*, **39**, 885–905.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001). The distribution of realized stock return volatility, *Journal of Financial Economics*, **61**, 43–76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (1999). (Understanding, optimizing, using and forecasting) Realized volatility and correlation, unpublished manuscript, Northwestern University, Duke University and University of Pennsylvania.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2000). Great realizations, *Risk*, 105–108.
- Andrews, D. W. K. (1991). Heteroskedasticity- and autocorrelation-consistent covariance matrix estimation, *Econometrica*, **59**, 817–858.
- Bai, X., Russell, J. R. and Tiao, G. C. (2004). Effect of non-normality and dependence on the precision of variance estimates using high-frequency financial data, working paper, University of Chicago, Graduate school of business.
- Bandi, F. M. and Russell, J. R. (2005). Microstructure noise, realized volatility, and optimal sampling, working paper, University of Chicago, Graduate School of Business.
- Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of Royal Statistical Society, Series B*, **64**, 253–280.
- Barndorff-Nielsen, O. E. and Shephard, N. (2007). Variation, jumps and high frequency data in financial econometrics, In: Blundell, R., Torsten, P. Newey, W. K., eds. *Advances in Economics and Econometrics: Theory and Applications*, Ninth World Congress. Econometric Society Monograph, Cambridge University Press.
- Chen, B., Gel, Y. R., Balakrishna, N. and Abraham, B. (2010). Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes, *Journal of Forecasting*, **30**, 51–71.
- Dacorogna, M. M., Gencay, R., Muller, U., Olsen, R. B. and Pictet, O. V. (2001). *An Introduction to High-Frequency Finance*, Academic Press, London.
- Fang, Y. (1996). Volatility modeling and estimation of high-frequency data with gaussian noise, unpublished doctoral thesis, MIT, Sloan School of Management.
- Griffin, J. E. and Oomen, R. C. A. (2008). Sampling returns for realized variance calculations: tick time or transaction time?, *Econometric Reviews*, **27**, 230–253.
- Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise, *Journal of Business & Economic Statistics*, **24**, 127–161.
- Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities, *Journal of Finance*, **42**, 381–400.

- McAleer, M. and Medeiros, M. C. (2008). Realized volatility: A review, *Econometric Reviews*, **27**, 10–45.
- Newey, W. and West, K. (1987). A simple positive semi-definite, heteroskedasticity- and autocorrelation-consistent covariance matrix, *Econometrica*, **55**, 703–708.
- Oomen, R. A. C. (2002). Modeling realized variance when returns are serially correlated, unpublished manuscript, University of Warwick, Warwick Business School.
- Pascuala, L., Romo, J. and Ruiz, E. (2006). Bootstrap prediction for returns and volatilities in GARCH models, *Computational Statistics & Data Analysis*, **50**, 2293–2312.
- Poon, S. and Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review, *Journal of Economic Literature*, **41**, 478–539.
- Protter, P. (2005). *Stochastic Integration and Differential Equations*, Springer-Verlag, New York.
- Shin, D. W. and Lee, O. (2010). Nonstationary time series with missing data, *The Korean Journal of Applied Statistics*, **23**, 73–79.
- Shin, D. W. and Park, S. (2011). Efficient realized variance, regression coefficient, and correlation coefficient under different sampling frequencies, *Economics Letters*, Accepted.
- Zhang, L., Mykland, P. A. and Ait-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of American Statistical Association*, **100**, 1394–1411.
- Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates, *Journal of Business & Economic Statistics*, **14**, 45–52.

# Market Microstructure Noise and Optimal Sampling Frequencies for the Realized Variances of Stock Prices of Four Leading Korean Companies

Rosy Oh<sup>1</sup> · Dong Wan Shin<sup>2</sup>

<sup>1</sup>Department of Statistics, Ewha Womans University

<sup>2</sup>Department of Statistics, Ewha Womans University

(Received December 24, 2011; Revised January 18, 2012; Accepted January 18, 2012)

---

## Abstract

We have studied the realized variance(RV) of intra-day returns and market microstructure noise based on high-frequency stock transaction data for the four largest companies in terms of market capitalization in the KOSPI. First, non-negligible biases are observed for the RV and for the bias-corrected realized variance( $RV_{AC_1}$ ) which is constructed by adjusting RV for the first order autocorrelation in intra-day returns. Bias is more obvious for the RV and the  $RV_{AC_1}$  when intra-day returns are sampled more frequently than every 2 minutes. Transaction Time Sampling(TTS) is shown to be better than Calendar Time Sampling(CTS) in terms of biases of the RV and the  $RV_{AC_1}$  for the 4 companies. The analysis reveals that market microstructure noise is temporally dependent. Second, by using the Noise-to-Signal Ratio(NSR), we estimate sampling frequencies that are optimal in terms of the Mean Square Errors(MSE) of the RV and the  $RV_{AC_1}$ . The optimal sampling frequencies are around 200 for RV and is around 5000 for the  $RV_{AC_1}$  for all the four stock prices. For the 6 hour transaction period of the Korean stock trading, these correspond to about 2 minutes and 6 seconds.

Keywords: Realized Variance, Volatility, High-frequency data, Market microstructure noise, Bias.

---

---

This research is supported by the National Research foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2011-0030811).

<sup>2</sup>Corresponding author: Professor, Department of Statistics, Ewha Woman's University, Seoul 120-750, Korea. E-mail: shindw@ewha.ac.kr