Chapter 19
미시 구조적 특징

1

목차

- 동기
- 문헌 고찰
- 1세대 : 가격 시퀀스
  - 틱 규칙
  - 롤 모델[1984]
  - 고-저변동성 추정기[1983]
  - 코윈과 슐츠[2012]
- 2세대 : 전략적 거래 모델
  - 카일의 람다[1985]
  - 아미후드의 람다[2002]
  - 하스부룩의 람다[2009]
- 3세대 : 순차적 거래 모델
  - 정보-기반 거래의 확률[1996]
  - 정보-기반 거래의 거래량-동기화 확률[2011]
- 미시 구조적 데이터 세트로부터의 추가 특징
  - 주문 크기의 분포
  - 취소율, 지정가 주문, 시장가 주문
  - 시간-가중 평균 가격 실행 알고리즘
  - 옵션 시장
  - 부호가 있는 주문 흐름의 계열 상관관계
- 미시구조적 정보란 무엇인가?

1 동기

- **목적**
  - 명시적 거래 규칙 아래 자산 거래의 프로세스와 결과 연구
- **데이터 세트**
  - 주문 취소, 이중 경매 예약, 대기열, 부분 집행, 조정, 주문 대체

2 문헌 고찰

- **흐름 요약**
  - 가용 데이터양과 종류 증가
- **1세대**
  - 가격 정보만을 사용
  - 결과
    - 거래 분류 모델, 롤 모델
- **2세대**
  - 가격 정보 + 거래량 데이터 세트
  - 거래량이 가격에 미치는 영향 연구
  - 결과
    - 카일, 아미후드
- **3세대**
  - PIN 이론
    - 매매가 차이를 유동성 공급자와 전략적 투자가 사이의 run된 전략 결정의 결과로 설명
    - 전문 투자가들은 정보를 가진 거래자들이 역선택을 통해 옵션을 사도록 함
      - 매매가의 차이는 그 옵션 가격

1세대 2세대

	Date and Time	Price	Volume
0	2020-01-28 09:00:15	59400	1130369
1	2020-01-28 09:00:15	59400	100
2	2020-01-28 09:00:15	59400	10
3	2020-01-28 09:00:15	59400	8333
4	2020-01-28 09:00:15	59400	100
...	...	...	...
17484738	2020-10-23 15:19:58	60100	100
17484739	2020-10-23 15:19:58	60100	3
17484740	2020-10-23 15:19:59	60100	10
17484741	2020-10-23 15:19:59	60000	1
17484742	2020-10-23 15:30:21	60200	846611

17484743 rows × 3 columns

3

1세대 가격 시퀀스

- **목적**
  - 매매가 차이와 변동성을 계산하여 유동성 파악
  - 거래 프로세스에 전략이나 순차적 구조 없이 제한된 데이터를 이용해 계산
- **Input**
  - Price

1세대 가격 시퀀스

- **틱 규칙**
  - 판매가는 항상 사려는 가격을 앞섬
    - 거래는 **구매자가 판매가를** 맞추거나 **매도자가 구매가**를 맞출 경우에만 일어나기 때문
      - 그렇지 않으면 이미 거래가 되어서 반영되었기 때문
  - **틱 규칙** : 구매로 체결된 거래는 1, 판매로 체결된 거래는 -1로 라벨링
    - **수식**
      - $b_t = \begin{cases} 1 & if\ \Delta \mathrm{p_t} > 0 \\ -1 & if\ \Delta \mathrm{p_t} < 0 \\ b_{t-1} & if\ \Delta \mathrm{p_t} = 0 \end{cases}$
    - **해석**
      - 정보성 특징 추출 가능
        - 1) 미래 기댓값$E_t[b_{t+1}]$에 대한 칼만 필터
        - 2) 이런 추정(17장)에 대한 구조적 변화
        - 3) $\{b_t\}$ 시퀀스의 엔트로피
        - 4) $\{b_t\}$의 런에 대한 월드-포보비츠 검정으로부터 t값
        - 5)누적 $\{b_t\}$ 계열 $\sum b_i$ 의 분수 미분 등

```python
def tick_rule(tick_prices):
    price_change = tick_prices.diff()
    aggressor = pd.Series(index=tick_prices.index, data=np.nan)

    aggressor.iloc[0] = 1.
    aggressor[price_change < 0] = -1.
    aggressor[price_change > 0] = 1.
    aggressor = aggressor.fillna(method='ffill')
    return aggressor
aggressor = tick_rule(df['Price'])
```

```python
pd.concat([df["Price"], df["Price"].diff(), aggressor], axis = 1)
```

	Price	Price	0
0	59400	NaN	1.0
1	59400	0.0	1.0
2	59400	0.0	1.0
3	59400	0.0	1.0
4	59400	0.0	1.0
...	...	...	...
17484738	60100	0.0	1.0
17484739	60100	0.0	1.0
17484740	60100	0.0	1.0
17484741	60000	-100.0	-1.0
17484742	60200	200.0	1.0

17484743 rows × 3 columns

```python
aggressor.mean()
```

```
0.17736703364756348
```

3

1세대 가격 시퀀스

- **롤 모델**
  - 가격 변동 없이 랜덤 워크를 따르는 중간가의 계열 m_t
    - $m_t = m_{t=1} + u_t$
      - $\Delta \mathrm{m_t} \sim N[0, \sigma_u^2]$
      - 가격 변동을 갖고, 이분산성이며, 계열 상관성을 갖고, 수익률은 정규분포가 아님
    - $p_t = m_t + b_t c$
      - 관측 가격 p_t는 매매가 차이에 대한 순차적 거래의 결과
      - c는 매매가 차이의 절반, $b_t \in \{-1,1\}$ 는 틱 규칙에 의해 생성(1 : 매수, 0 : 매도)
  - **가정**
    - 구매와 판매가 동일하게 발생 가능
      - $P[b_t = 1] = P[b_t = -1] = \frac{1}{2}$
      - 계열 독립($E[b_t, b_{t-1}] = 0$)
      - 잡음으로부터 독립($E[b_t u_t] = 0$)

3

1세대 가격 시퀀스

- **롤 모델**
  - $m_t = m_{t=1} + u_t$
    - $\Delta \mathrm{m_t} \sim N[0, \sigma_u^2]$
  - $p_t = m_t + b_t c$
    - $P[b_t = 1] = P[b_t = -1] = \frac{1}{2}$
      - 구매와 판매가 동일하게 발생 가능하다고 가정
    - $E[b_t, b_{t-1}] = 0,\ E[b_t u_t] = 0$
      - 계열 독립, 잡음으로부터 독립
  - $\sigma^2[\Delta \mathrm{p_t}] = \mathrm{E}[(\Delta \mathrm{p_t})^2] - (\mathrm{E}[(\Delta \mathrm{p_t})])^2 = 2\mathrm{c}^2 + \sigma_u^2$
    - $\sigma[\Delta p_t, \Delta \mathrm{p}_{t-1}] = -c^2$
      - $c = \sqrt{max\{0, -\sigma[\Delta \mathrm{p_t}, \Delta \mathrm{p}_{t-1}]\}}$
      - $\sigma_u^2 = \sigma^2[\Delta \mathrm{p_t}] + 2\sigma[\Delta \mathrm{p_t}, \Delta p_{t-1}]$
  - **결론**
    - 매매가의 차이는 가격 변화의 계열 공분산 함수
    - (관측되지 않음)참 가격의 잡음에서 미시 구조적 잡음을 제외한 것으로 관측된 잡음과 가격 변화의 계열 공분산 함수

```python
def roll_model(prices):
    # delta_price : price_change
    price_change = prices.diff()
    # \sigma[\delta p_t, \delta p_{t-1}]
    autocorr = price_change.autocorr(lag=1)
    # max{0, -\sigma[delta p_t, \delta p_{t-1}]}
    spread_squared = np.max([-autocorr, 0])
    # sqrt{max{0, -\sigma[delta p_t, \delta p_{t-1}]}}

    # c : spread : \sqrt{-\sigma[\delta p_t, \detla p_{t-1}]}
    spread = np.sqrt(spread_squared)

    # var(\delta p_t) -2 * c^2
    noise = price_change.var() - 2 * (spread ** 2)
    return spread, noise

spread, noise = roll_model(df['Price'])
spread, noise
```

```
(0.6927467926939844, 2874.193458816462)
```

3

1세대 가격 시퀀스

- 고-저변동성 추정기
  - 등장 배경
    - 고-저 가격에 근거한 변동성 추정기는 종가에 기초한 변동성 표준 추정기보다 더 정확함
    - $E\left[\frac{1}{T}\sum\left(log\left[\frac{H_t}{L_t}\right]\right)^2\right] = k_1\sigma_{HL}^2$
    - $E\left[\frac{1}{T}\sum(log[\frac{H_t}{L_T}])\right] = k_2\sigma_{HL}$
      - $k_1 = 4log[2], k_2 = \sqrt{\frac{8}{\pi}}$
      - H_t : 바 t에 대한 고가
      - L_t : 바 t에 대한 저가

```python
def value_bar(df, value_size):
    df = df.reset_index()

    df["value_num"] = df["거래대금"].cumsum() // value_size
    groupby = df.groupby("value_num")

    bars = groupby["현재가"].ohlc()
    bars[["volume", "value"]] = groupby[["체결수량", "거래대금"]].sum()
    bars["시각"] = groupby["시각"].first()
    bars.set_index("시각", inplace = True)
    return bars
```

```python
val_2_500 = value_bar(data, 2_500_000_000)
val_2_500.head(5)
```

	open	high	low	close	volume	value
시각						
2020-01-28 09:00:15	59400	59400	59400	59400	1130479	67150452600
2020-01-28 09:00:15	59400	59400	59200	59300	48004	2847474800
2020-01-28 09:00:16	59300	59300	59200	59300	37965	2250102000
2020-01-28 09:00:22	59300	59400	59200	59400	43078	2554359900
2020-01-28 09:00:25	59400	59400	59300	59300	25305	1501407400

```python
def high_low_estimator(high, low, window):
    log_high_low = np.log(high / low)
    volatility = log_high_low.rolling(window=window).mean() / np.sqrt(8. / np.pi)
    return volatility
```

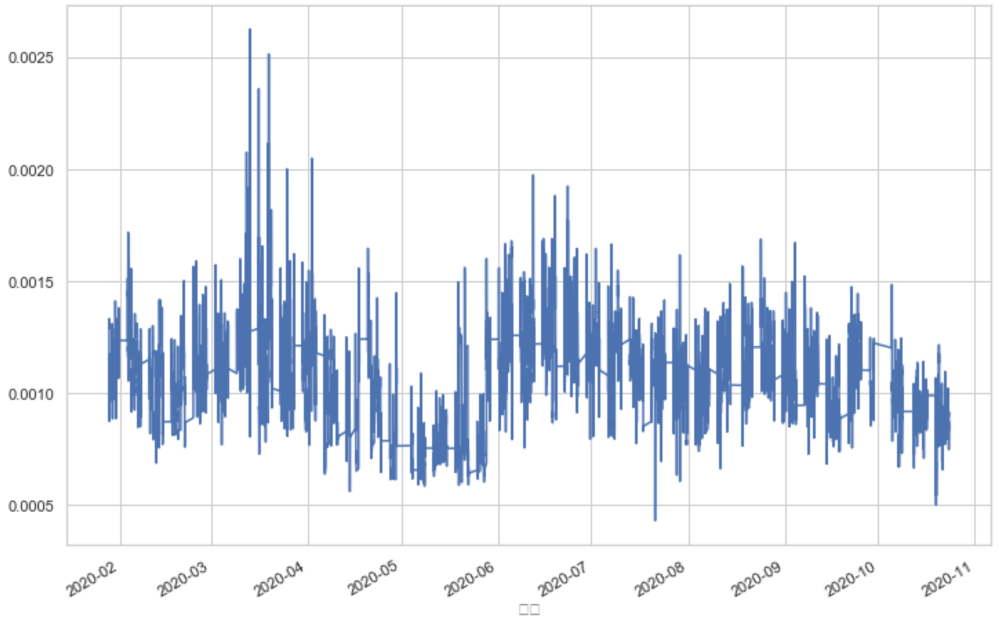
```python
vol = high_low_estimator(val_2_500.high, val_2_500.low, window=50)
```

3

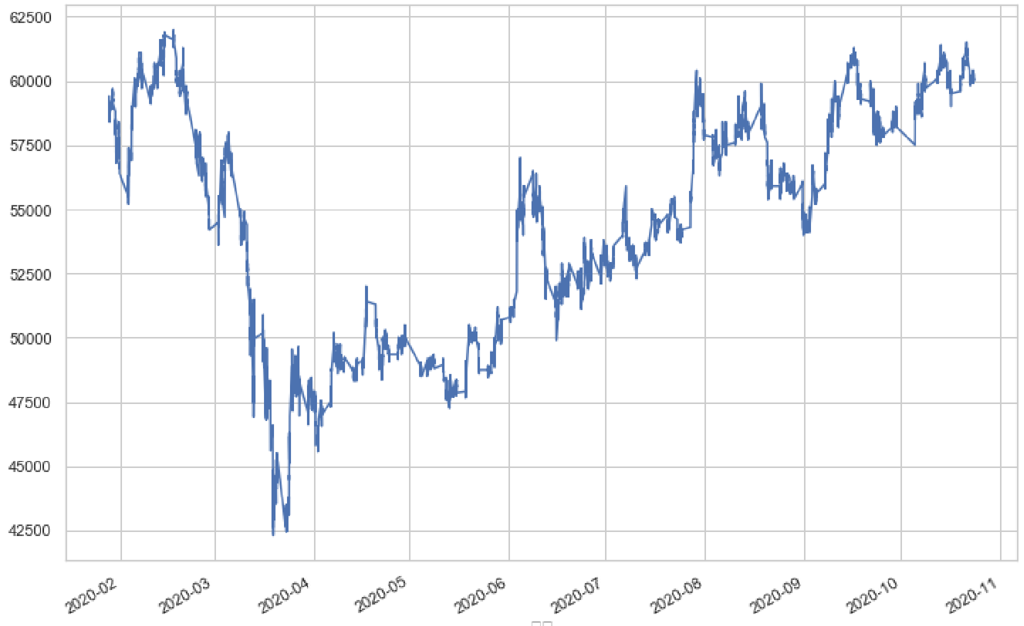
1세대 가격 시퀀스

- 고-저변동성 추정기

변동성



종가



3

1세대 가격 시퀀스

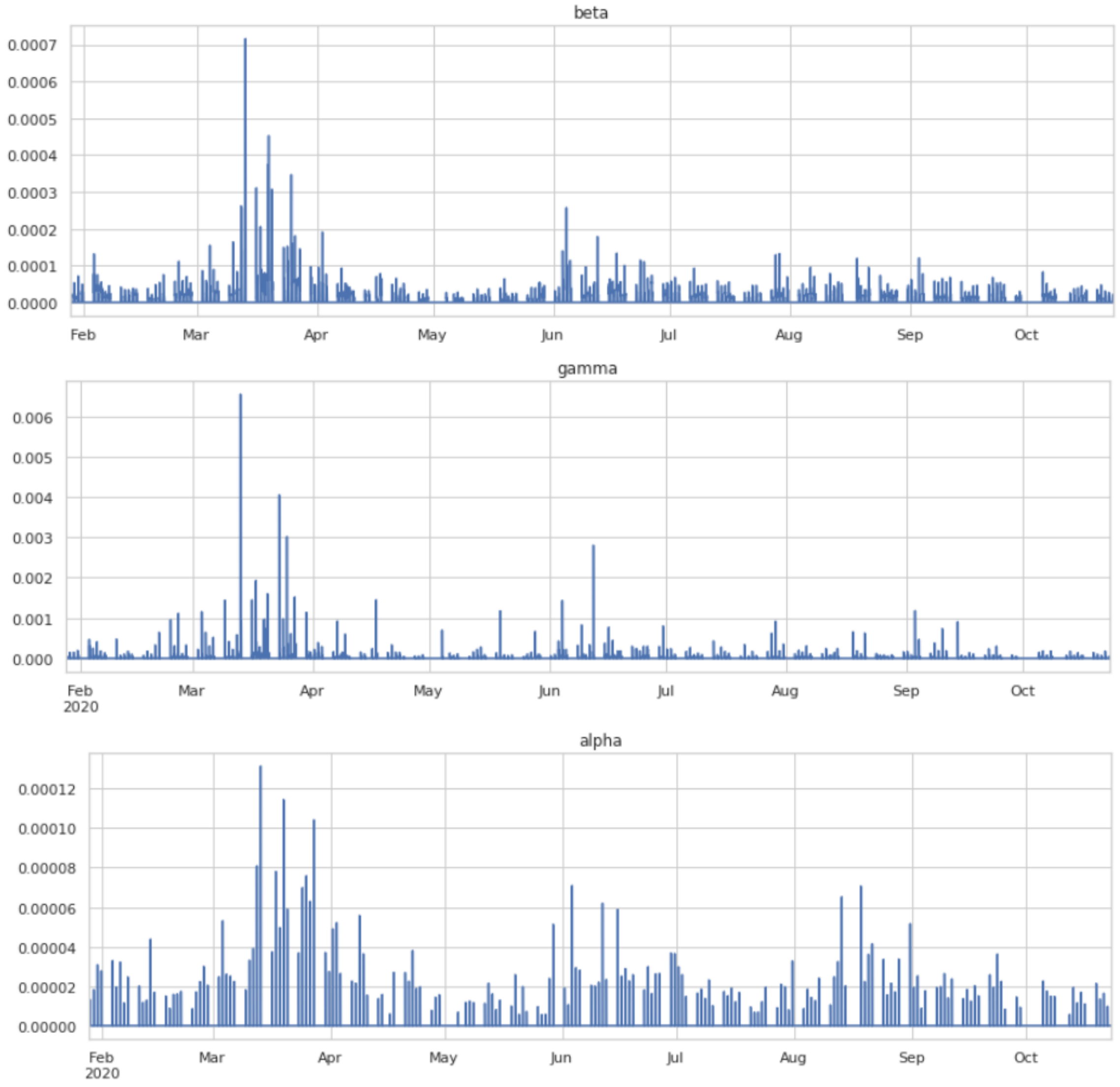
- **코윈과 슐츠**
  - **매매가 차이 추정기**
    - **원칙**
      - 고가는 항상 판매 호가에 매치, 저가는 구매 호가에 매치
        - 고가 대비 저가 비율은 근본적인 변동성과 매매가 차이 반영
      - 변동성에 기인한 고가-대비-저가 비율의 구성요소는 두 관측값 사이에 경과한 시간에 비례해 증가함
    - **수식**
      - 매매가 차이를 가격의 퍼센티지로 계산 가능
        - $S_t = \frac{2(e^{\alpha_t}-1)}{1+e^{\alpha_t}}$
          - $\alpha_t = \frac{\sqrt{2\beta_t}-\sqrt{\beta_t}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma_t}{3-2\sqrt{2}}}$
          - $\beta_t = E\left[\sum\left[\log\left(\frac{H_{t-j}}{L_{t-j}}\right)\right]^2\right]$
          - $\gamma_t = \left[\log\left(\frac{H_{t-1,t}}{L_{t-1,t}}\right)\right]^2$
            - $H_{t-1,t}$: 두 바 $t-1, t$에 대한 고가
            - $L_{t-1,t}$: 두 바 $t-1, t$에 대한 저가

```python
def getBeta(series,sl):
    h1=series[['High','Low']].values
    h1=np.log(h1[:,0]/h1[:,1])**2
    h1=pd.Series(h1,index=series.index)
    beta=h1.rolling(2).sum()
    beta=beta.rolling(sl).mean()
    return beta.dropna()
#--------------------------------

def getGamma(series):
    h2=series['High'].rolling(2).max()
    l2=series['Low'].rolling(2).min()
    gamma=np.log(h2.values/l2.values)**2
    gamma=pd.Series(gamma,index=h2.index)
    return gamma.dropna()

#--------------------------------
def getAlpha(beta,gamma):
    den=3-2*2**.5
    alpha=(2**.5-1)*(beta*.5)/den
    alpha-=(gamma/den)**.5
    alpha[alpha<0]=0 # set netative alphas to 0
    return alpha.dropna()

#--------------------------------
def corwinSchultz(series,sl=1):
    #Note:S<0 iff alpha<0
    beta=getBeta(series,sl)
    gamma=getGamma(series)
    alpha=getAlpha(beta,gamma)
    spread=2*(np.exp(alpha)-1)/(1+np.exp(alpha))
    startTime=pd.Series(series.index[0:spread.shape[0]],index=spread.index)
    spread=pd.concat([spread,startTime],axis=1)
    spread.columns=['Spread','Start_time']# 1st loc used to compute beta
    return spread
```

3

1세대 가격 시퀀스

- 코윈과 슐츠
  - $S_t = \frac{2(e^{a_t}-1)}{1+e^{a_t}}$
  - $a_t = \frac{\sqrt{2\beta_t}-\sqrt{\beta_t}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma_t}{3-2\sqrt{2}}}$
  - $\beta_t = E\left[\sum\left[log\left(\frac{H_{t-j}}{L_{t-j}}\right)\right]^2\right]$
  - $\gamma_t = \left[log\left(\frac{H_{t-1,t}}{L_{t-1,t}}\right)\right]^2$



4

2세대: 전략적 거래 모델

- **목표**
  - **유동성의 이해와 측정**
    - **저유동성**
      - 초과 구매력과 연계된 리스크
    - 거래를 정보-기반 투자가와 그렇지 못한 투자가 사이의 전략적 상호작용으로 설명
    - 거래량의 부호와 주문 흐름의 불균형
- **방법론**
  - 회귀를 통해 계산

1세대 2세대

	Date and Time	Price	Volume
0	2020-01-28 09:00:15	59400	1130369
1	2020-01-28 09:00:15	59400	100
2	2020-01-28 09:00:15	59400	10
3	2020-01-28 09:00:15	59400	8333
4	2020-01-28 09:00:15	59400	100
...	...	...	...
17484738	2020-10-23 15:19:58	60100	100
17484739	2020-10-23 15:19:58	60100	3
17484740	2020-10-23 15:19:59	60100	10
17484741	2020-10-23 15:19:59	60000	1
17484742	2020-10-23 15:30:21	60200	846611

17484743 rows × 3 columns

4

2세대: 전략적 거래 모델

- 카일의 람다
  - 모델
    - 자산
      - 종료가치가 $v \sim N[p_0, \Sigma_0]$
    - 거래자
      - v와 독립적으로 $u = N[0, \sigma_u^2]$만큼 수량을 거래하는 비합리적 투자가
      - v를 알고 수량 x를 시장 주문을 통해 요청한 정보-기반 투자가
  - 전문 투자가
    - 전체 주문 흐름 $y = x + u$ 관찰 후 가격 p설정
      - 현 상황에서는 비합리적 투자가의 주문과 정보-기반 투자가의 주문 구분 못함
    - 가격을 주문 흐름 불균형의 함수로 조정
      - 이 경우 정보-기반 거래자의 존재 확인 가능
      - → 가격 변화와 주문 흐름 불균형 사이에는 양의 관계가 존재

4

2세대: 전략적 거래 모델

- **카일의 람다**
  - **정보-기반 투자가**
    - 전문 투자가들이 선형 가격 조정 함수 $p = \lambda y + \mu$를 사용할 것이라고 추정
      - λ : 유동성 척도의 역
      - 이익 : $\pi = (v - p)x$
        - 이차 조건 $\lambda > 0$, $x = \frac{v-u}{2\lambda}$에서 최대화
  - **전문 투자가**
    - 정보-기반 투자가들의 수요는 선형 함수일 것이라고 추정
      - $v: x = \alpha + \beta v$
        - $\alpha = -\frac{\mu}{2\lambda}, \beta = \frac{1}{2\lambda}$
      - 더 낮은 유동성은 더 높은 λ를 의미
      - 이는 정보-기반 거래자로부터의 더 낮은 수요 의미

4

2세대: 전략적 거래 모델

- 카일의 람다
  - 목표
    - 전문 투자가는 수익 극대화와 시장 효율 사이의 균형을 찾아야 함
    - 선형 함수에 따른 유일한 해$(\mu, \alpha, \lambda, \beta)$
      - $\mu = p_0$
      - $\alpha = p_0\sqrt{\frac{\sigma_u^2}{\Sigma_0}}$
      - $\lambda = \frac{1}{2}\sqrt{\frac{\Sigma_0}{\sigma_u^2}}$
      - $\beta = \sqrt{\frac{\sigma_u^2}{\Sigma_0}}$
    - 정보-기반 투자가의 기대 수익
      - $\boldsymbol{E[\pi] = \frac{(v-p_0)^2}{2}\sqrt{\frac{\sigma_u^2}{\Sigma_0}} = \frac{1}{4\lambda}(v - p_0)^2}$

2세대: 전략적 거래 모델

- 카일의 람다
  - 정보-기반 투자가의 기대 수익
    - 수익의 원천
      - 증권의 가격 오류
      - 비합리적 투자가의 순주문 흐름의 분산
        - 비합리적 투자가들이 많을수록 정보-기반 거래자들이 자신의 의도를 감추기가 쉬움
      - 증권 분산 종단가의 역, 변동성이 낮을수록 가격 오류로부터 수익을 얻기 쉬움
    - λ
      - 가격 영향 포착
      - 저유동성은 v의 불확실성에 따라 증가하고, 잡음의 양에 따라 감소
      - $\Delta \mathrm{p_t} = \lambda(b_t V_t) + \epsilon_t$
        - p_t: 가격 시계열
        - b_t: 어그레서 신호의 시계열
        - V_t: 거래량의 시계열
        - $b_t V_t$: 거래량의 부호, 또는 순주문 흐름의 시계열

```python
from sklearn.linear_model import LinearRegression
def kyles_lambda(tick_prices, tick_volumes, tick_signs, regressor=LinearRegression()):
    price_change = tick_prices.diff()
    net_order_flow = tick_signs * tick_volumes
    x_val   = net_order_flow.values[1:].reshape(-1, 1)
    y_val = price_change.dropna().values
    lambda_ = regressor.fit(x_val  , y_val)
    return lambda_.coef_[0]
kyles_lambda(df["Price"], df["Volume"], aggressor)
```

```
0.00025116805392242964
```

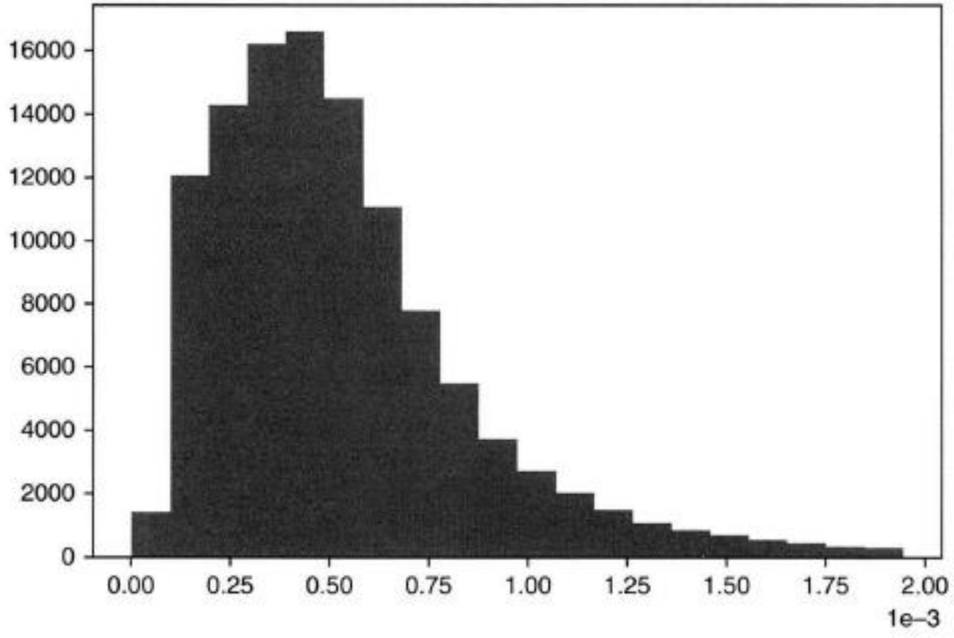


그림 19-1 E-mini S&P 500 선물 계열에 계산한 카일의 람다

4

2세대: 전략적 거래 모델

- 아미후드의 람다
  - **목표**
    - **수익률의 절댓값과 저유동성 사이의 양의관계**
    - $|\Delta \log [\widetilde{p_t}]| = \lambda \sum_{t \in B_\tau} (p_t V_t) + \epsilon_\tau$
      - B_τ : τ에 포함된 거래의 집합
      - $\tau, \overline{p_t}$ 의 종가
      - $p_t V_t$는 거래 $t \in B_\tau$에 관여된 달러 거래량

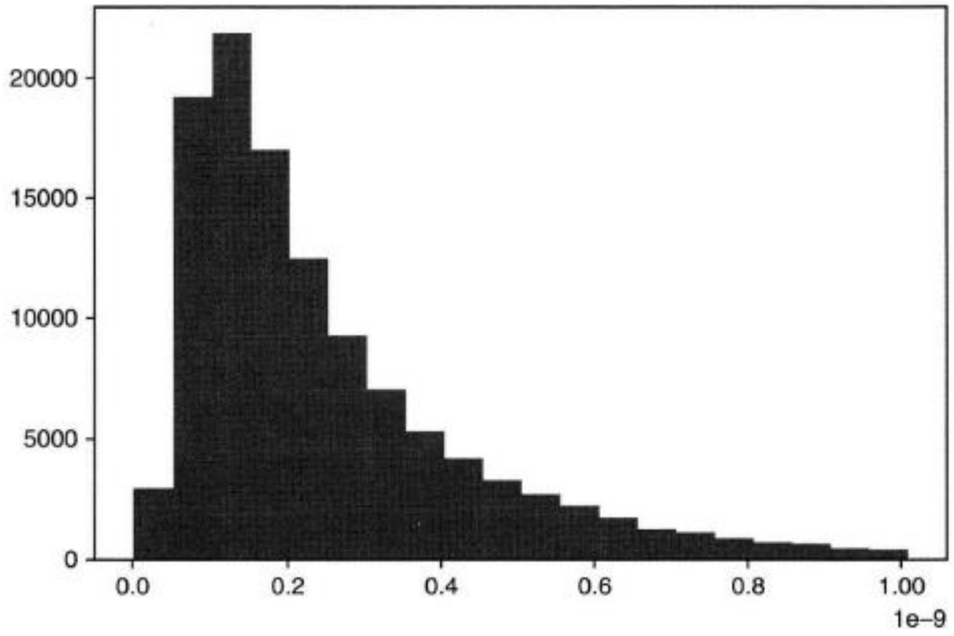


그림 19–2 E–mini S&P 500 선물 계열에 계산한 아미후드의 람다

4

2세대: 전략적 거래 모델

- 하스브룩의 람다
  - **목표**
    - **일거래 데이터에 기반을 둔 거래 영향 계수 계산**
      - **깁스 표본 기법**을 사용해 회귀 식의 베이즈 추정 만듬
    - $\log[\widetilde{p_{i,\tau}}] - \log[\tilde{p}_{i,\tau-1}] = \lambda_i \sum\left(b_{i,t}\sqrt{p_{i,t}V_{i,t}}\right) + \epsilon_{i,\tau}$
      - $B_{i,\tau}$는 증권 i에 대한 바 τ에 속한 거래의 집합
      - $i = 1, \cdots, I$
      - $\tilde{\boldsymbol{p}}_{\boldsymbol{i},\boldsymbol{\tau}}$: 증권 i의 바 τ의 종가
      - $\boldsymbol{b}_{\boldsymbol{i},\boldsymbol{t}} \in \{-\mathbf{1}, \mathbf{1}\}$: 거래
      - $\boldsymbol{t} \in \boldsymbol{B}_{\boldsymbol{i},\boldsymbol{\tau}}$: 구매자 개시인지 판매자 개시인지 알려줌
      - $\boldsymbol{p}_{\boldsymbol{i},\boldsymbol{t}}\boldsymbol{V}_{\boldsymbol{i},\boldsymbol{t}}$: 거래 $t \in B_{i,\tau}$에 개입된 달러의 거래량
      - 모든 증권 i에 대해 λ_i 계산
      - 거래의 효율적 비용을 근사하는데 특징으로 사용
  - **팁**
    - 5분 시간- 바 추천

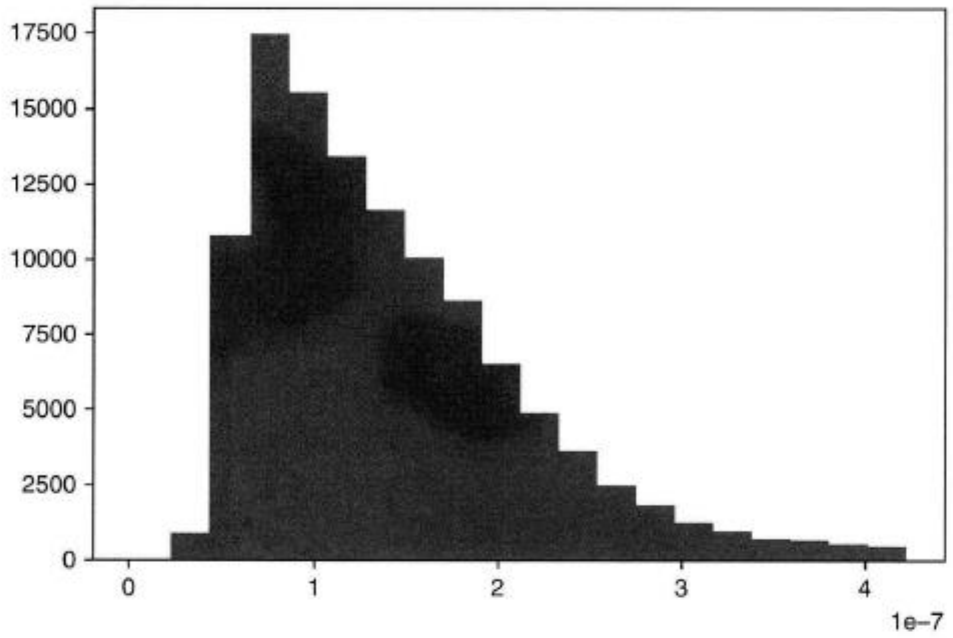


그림 19–3 E-mini S&P 500 선물 계열에 계산한 하스브룩의 람다

5

3세대: 순차적 거래 모델

- **앞선 모델은 단일 정보-기반 투자자가 다수의 시간대에 거래**
- **이 절에서는 랜덤하게 선택된 투자자가 순차적이며 독립적으로 시장에 등장**
- **이 모델은 유동성 공급자들이 직면하는 불확실성의 원인 포함**
  - 정보성 이벤트가 발생할 확률
  - 이런 이벤트가 부정적일 확률
  - 비이성적 투자가들이 등장할 확률
  - 정보-기반 투자가들이 등장할 확률

5

3세대: 순차적 거래 모델

- 정보-기반 거래의 확률
  - **개별 증권의 정보-기반 거래 확률 계산 위해 거래 데이터 사용**
    - 거래를 전문 투자가와 전략적 투자가 사이에 다수의 거래 주기에 걸쳐 반복되는 게임으로 인식
    - $\boldsymbol{S}$: 증권 가격
      - 특정 양의 새로운 정보가 가격 S에 포함될 경우
        - $\boldsymbol{S_B}$: 나쁜 뉴스일때의 가격
        - $\boldsymbol{S_G}$: 좋은 뉴스일때의 가격
    - $\boldsymbol{S_0}$: 현재 가격
    - $\boldsymbol{\alpha}$: 분석 중인 시간 프레임 내 새로운 정보가 도달할 확률
    - $\boldsymbol{\delta}$: 해당 뉴스가 나쁜 뉴스일 확률
    - $\boldsymbol{1 - \delta}$: 해당 뉴스가 좋은 뉴스일 확률
  - **시간 t에서 증권 가격의 기댓값**
    - $E[S_t] = (1 - \alpha_t)S_0 + \alpha_t[\delta_t S_B + (1 - \delta_t)S_g]$
  - **정보-기반 투자자들** : 푸아송 분포를 따라 μ의 비율로 등장
  - **비합리적 투자가** : ϵ의 비율로 등장

5 3세대: 순차적 거래 모델

- **정보-기반 거래의 확률**
  - **시간 t에서 증권 가격의 기댓값**
    - $E[S_t] = (1-\alpha_t)S_0 + \alpha_t[\delta_t S_B + (1-\delta_t)S_g]$
  - **정보-기반 투자자들** : 푸아송 분포를 따라 μ의 비율로 등장
  - **비합리적 투자가** : 푸아송 분포를 따라 ϵ의 비율로 등장
  - 정보 기반 전문 투자가 손익분기점
    - 매수가 레벨 B_t에서 손익분기점
      - $E[B_t] = E[S_t] - \frac{\mu\alpha_t\delta_t}{\epsilon + \mu\alpha_t\delta_t}(E[S_t] - S_B)$
    - 매도가 레벨 A_t의 손익분기점
      - $E[A_t] = E[S_t] + \frac{\mu\alpha(1-\delta_t)}{\epsilon + \mu\alpha_t(1-\delta_t)}(S_G - E[S_t])$
  - **손익 분기 매매가의 차이**
    - $E[A_t - B_t] = \frac{\mu\alpha_t(1-\delta_t)}{\epsilon + \mu\alpha(1-\delta_t)}(S_G - E[S_t]) + \frac{\mu\alpha_t\delta_t}{\epsilon + \mu\alpha_t\delta_t}(E[S_t] - S_B)$

5

3세대: 순차적 거래 모델

- **정보-기반 거래의 확률**
  - **손익 분기 매매가의 차이**
    - $E[A_t - B_t] = \frac{\mu\alpha_t(1-\delta_t)}{\epsilon+\mu\alpha(1-\delta_t)}(S_G - E[S_t]) + \frac{\mu\alpha_t\delta_t}{\epsilon+\mu\alpha_t\delta_t}(E[S_t] - S_B)$
  - If $\delta_t = \frac{1}{2}$
    - $\frac{\alpha_t\mu}{\alpha_t\mu+2\epsilon}(S_G - S_B)$
    - 유동성을 공급하는 가격의 범위를 결정하는 중요 요인은 다음과 같다
      - $PIN_t = \frac{\alpha_t\mu}{\alpha_t\mu+2\epsilon}$
        - 이를 추정하기 위해서는 $\alpha, \delta, \mu, \epsilon$ 추정 필요
        - 최대-우도 기법은 세 가지 푸아송 분포의 혼합 적합화
          - $P[V^B, V^S] = (1-\alpha)P[V^B, \epsilon]P[V^S, \epsilon] + \alpha(\delta P[V^B, \epsilon]P[V^S, \mu+\epsilon] + (1-\delta)P[V^B, \mu+\epsilon]P[V^S, \epsilon])$
          - V^B : 매도가(구매자 개시 거래)거래에 대한 거래량(거래량 바 τ에 포함된 매수 틱으로부터의 거래량 합)
          - V^S : 매수가(판매자 개시 거래) 거래에 대한 거래량(거래량 바 τ에 포함된 매도 틱으로부터의 거래량 합)
        - 전체 주문 흐름에 상대적인 정보를 가진 거래자의 주문 비율

5

3세대: 순차적 거래 모델

- **정보-기반 거래의 거래량-동기화 확률**
  - $E[V^B - V^S] = (1-\alpha)(\epsilon - \epsilon) + \alpha(1-\delta)\big(\epsilon - (\mu + \epsilon)\big) + \alpha\delta(\mu + \epsilon - \epsilon) = \alpha\mu(1-2\delta)$
  - 충분히 큰 μ에 대해서 $E\Big[[V^B - V^S]\Big] \approx \alpha\mu$
  - PIN의 고빈도 추정
    - 정보-기반 거래의 거래량-동기화 확률(VPIN)
    - 데이터 추출을 거래량에 의해 포착된 시장 활동에 동기화된 거래량 시계에 도입
      - $\frac{1}{\mathrm{n}}\sum|V_\tau^B - V_\tau^S| \approx \alpha\mu$
        - V_τ^B : 거래량 바 τ 이내에 매수-개시로부터 거래된 거래량의 합
        - V_τ^S : 거래량 바 τ 이내에 매도-개시로부터 거래된 거래량의 합
        - N : 이 계산을 위해 사용된 바의 개수
      - $\frac{1}{n}\sum(V_\tau^B + V_\tau^S) = V = \alpha\mu + 2\epsilon$
        - 총 거래량
    - **고빈도에서의 계산**
      - $VPIN_\tau = \frac{\sum|V_\tau^B - V_\tau^S|}{\sum(V_\tau^B + V_\tau^S)} = \frac{\sum|V_\tau^B - V_\tau^S|}{nV}$

6 미시 구조적 데이터 세트로부터의 추가 특징

- **주문 크기의 분포**
  - **주문 크기당 주문 빈도수를 연구**
    - 반올림된 크기의 거래가 빈번함
    - GUI 투자가들에 의함
    - 이는 시장에서 족적을 감추기 위해 랜덤화된 거래를 프로그램화 하는 실리콘 투자가들의 패턴과는 거리가 멂
  - **유용한 특징**
    - 반올림된 크기의 거래의 정규 빈도를 확인, 기댓값으로부터의 편차 모니터링
      - 평소보다 많은 반올림 크기의 거래가 특정 경향을 갖는지 확인
        - 수동 투자시에는 근본적 시각, 믿음, 확신을 갖고 투자하는 경향이 많기 때문
      - 평소보다 반올림 크기의 거래가 줄어들었다면 가격이 횡보할 가능성이 높아짐
        - 실리콘 투자가들은 대개 장기적인 관점을 갖고 있지 않기 때문

6 미시 구조적 데이터 세트로부터의 추가 특징

- **취소율, 지정가 주문, 시장가 주문**
  - 소형주와 대형주가 다르다는 것 발견
    - 이런 크기 모델링이 매매가 차이의 동적 모델링과 연계되어 있음
  - 대규모 호가 취소율은 낮은 유동성 신호
    - 참여자들이 실행되길 원하지 않은 호가를 제시한 것이기 때문
      - **호가 스터퍼**
        - 잠재 차익 거래를 함
        - 경쟁 알고리즘의 속도를 저하시키려는 의도로 메시지를 통한 거래소 혼란 수반
      - **호가 댄글러**
        - 스퀴즈된 투자가를 자신의 이익에 반하게 행동하게 되는 상황으로 몰아부치는 호가
      - **유동성 스퀴저**
        - 고통받던 대형 투자가들이 현재 포지션을 해제하려 할 때, 약탈적 알고리즘은 동일한 방향으로 거래해 유동성을 최대한 없애버림
        - 가격은 치솟고 이득을 봄
      - **팩 헌터**
        - 독립적으로 사냥하던 약탈자들이 서로의 행위를 알아차리고 연쇄적 효과를 발발시키기 위해 무리 형성

6 미시 구조적 데이터 세트로부터의 추가 특징

- **시간-가중 평균 가격 실행 알고리즘**
  - **정의**
    - TWAP 알고리즘 : 대규모 주문을 작을 것으로 분할
    - 일정한 시간 간격으로 제출해 미리 정의된 TWAP 달성
  - **실험**
    - **기간**
      - 2010년 11월 7일 ~ 2011년 11월 7일
    - **종목**
      - E-mini S&P 500 선물 거래
    - **방법**
      - 하루를 24시간으로 나누고, 각 시간별로 분과 관계없이 초당 거래량 덧붙임
      - x축을 초당 거래량, y축은 일별 시간, z축은 종합된 거래량으로 정의
      - 날짜의 진행에 따른 각 분당 거래량의 분포 → 저빈도 투자가들의 대량 주문 실행을 시간 흐름 공간 상에서 찾을 수 있게 함
    - **결과**
      - 1분 이내의 최대 밀집 거래량은 대부분 일 중 매 시간의 처음 몇 초 사이에 일어남
      - 아시아 증권 시장의 장 시작 시간, 영국과 유럽 증권 시장의 장 시간, 미국 증권 시장의 장 시간, 미국 증권의 장 종료 시간에 큼

6 미시 구조적 데이터 세트로부터의 추가 특징

- 옵션 시장
  - **무라브예프**
    - **목적**
      - 미국 주식과 옵션으로부터의 미시 구조적 정보를 사용해 두 시장이 불일치되는 이벤트 연구
        - 풋-콜 패리티 호가에 내재된 기저 매매가 차이 범위 도출 후, 실제 주식의 매매가 차이 범위와 비교
    - **결과**
      - 불일치가 주식 호가를 향해 해소되는 경향이 있음
      - 옵션 호가는 경제적으로 주요한 정보를 갖지 않는다는 것을 의미한다는 결론
      - 다만 옵션 호가는 경제적으로 주식가에 포함되지 않은 정보를 포함하고 있다는 것 발견
  - **크레머 와인바움**
    - 비싼 콜을 가진 주식이 상대적으로 비싼 풋을 가진 주식보다 주당 50 베이시스 포인트만큼 더 높은 성능을 냄
    - 옵션 유동성이 높고, 주식 유동성이 낮으면 더 커진다

6 미시 구조적 데이터 세트로부터의 추가 특징

- **부호가 있는 주문 흐름의 계열 상관관계**
  - **토스**
    - 주문 부호가 여러 날동안 양의 자기 상관관계를 가짐
    - 무리짓기와 주문 나누기로 표현
      - 수시간 이하의 시간 규모에서 주문 흐름이 일관적인 것은 분할 때문
    - 시장 미시 구조 이론이 일관된 주문 흐름 불균형을 정보-기반 투자가의 존재 때문으로 설명
      - 부호가 있는 거래량의 계열 상관관계를 통해 이런 일관성의 강도 측정해보는 것 의미

7

미시 구조적 정보란 무엇인가?

- **문제의식**
  - 거래라는 맥락에서 정보란 어떻게 정의되는가?
    - 선행 연구들은 이 개념에 대해 형식적인 것이 아니라 놀라우리만큼 느슨하게 사용
- **정보란?**
  - 전문 투자가들이 특정 레벨에서 유동성을 공급할 것인지, 수동적 호가를 취소할 것인지를 결정하는 상황
  - 특징 행렬 X 고려
    - VPIN, 카일의 람다, 취소율 등
    - 각 결정 지점마다 하나의 행을 가짐
      - Ex) 10,000건의 계량 거래될 때, 혹은 가격에 심각한 변동이 있을 때 action
  - **구성**
    - 1) 시장 조성 수익을 낸 경우 1, 손실을 낸 경우 0
    - 2) 훈련 집합 (X,y)에 분류기 적합화
    - 3) 새로운 OOS 관측값($\tau > T$)이 도착하면 레이블 관측을 위해 적합화된 분류기 사용($\hat{y}_\tau = E_\tau\left[y_\tau[X]\right]$)
    - 4) 관측에 대한 교차-엔트로피 손실 L_τ
    - 5) 커널밀도추정기(KDE)를 음의 교차-엔트로피 손실에 적합해 F 도출
    - 6) 시각 t에서 미시 구조적 정보를 $\phi_\tau = F[-L_\tau]$로 계산

7

미시 구조적 정보란 무엇인가?

- **해석**
  - ϕ_τ는 전문 투자가의 결정 모델에 의해 맞닥뜨린 복잡도로 이해할 수 있음
  - 정상적인 시장 조건하에서 전문 투자가는 정보-기반 예측을 낮은 교차-엔트로피로 생성
  - 전략적 투자가들에게 유동성을 공급함으로써 수익
  - 비대칭적인 정보-기반의 투자가가 있다면 전문 투자가는 높은 교차-엔트로피 손실에 의해 측정된 비합리적 예측 생성
  - 따라서 미시구조적 정보는 오직 전문 투자가의 예측력에 관련해서만 정의되거나 측정될 수 있음
- **예시**
  - 2010년 5월 6일 플래시 크래시
    - 전문 투자가들은 매수에 걸려 있는 수동적 호가 주문이 실행될 것으로 잘못 예측하여 더 높은 가격에 팜
    - 폭락은 부정확한 하나의 예측 때문이 아니라 수천개의 추정 오류가 누적돼 발생한 것
  - 전문 투자가들이 자신들의 예측에서 교차-엔트로피 손실이 증가되는 것을 보았다면 역선택 확률에 대해 인식할 수 있음

Thank you