

# Low Frequency Spread Measurements and Price Impact Proxies

Giulio Surya Lo Verde

November 20, 2024

## 1 Low Frequency Spread Measurements

### 1.1 Roll

Nota

la covarianza va bene come formula?

$$S = \begin{cases} 2\sqrt{-\text{Cov}(\Delta P_t, \Delta P_{t-1})} & \text{se } \text{Cov}(\Delta P_t, \Delta P_{t-1}) < 0 \\ 0 & \text{se } \text{Cov}(\Delta P_t, \Delta P_{t-1}) \geq 0 \end{cases}$$

**Variabili:**

- $\Delta P_t$ : variazione di prezzo

**Implementazione della formula di Roll:**

1. Raccogli i dati sui prezzi delle transazioni per un titolo in un determinato periodo.
2. Calcola le variazioni di prezzo giornaliere  $\Delta P_t = P_t - P_{t-1}$ .
3. Calcola la covarianza seriale delle variazioni di prezzo utilizzando la formula:

$$\text{Cov}(\Delta P_t, \Delta P_{t-1}) = \frac{1}{N-1} \sum_{t=1}^N (\Delta P_t - \bar{\Delta P})(\Delta P_{t-1} - \bar{\Delta P})$$

dove  $\bar{\Delta P}$  è la media delle variazioni di prezzo nel periodo di riferimento (nel nostro caso un mese).

4. Sostituisci il valore della covarianza nella formula di Roll per ottenere lo spread stimato  $S$ .

## 1.2 Effective Tick

$$\text{Effective Tick} = \frac{\sum_{j=1}^J \hat{g}_j s_j}{\bar{P}}$$

**Variabili:**

- $\hat{g}_j$ : probabilità vincolata di osservare uno spread effettivo di dimensione  $s_j$
- $s_j$ : dimensione dello spread effettivo per il cluster  $j$
- $\bar{P}$ : prezzo medio nel periodo considerato

### Implementazione dell'Effective Tick

L'Effective Tick è una misura dello spread effettivo basata sulla frequenza con cui i prezzi si raggruppano su determinati livelli (*price clustering*). L'idea principale è che i prezzi di negoziazione tendono a concentrarsi su incrementi prefissati, e la frequenza con cui si osservano questi prezzi fornisce un'indicazione delle dimensioni dello spread.

#### 1. Step 1: Definire i Cluster di Prezzo

Per ogni transazione, classifichiamo il prezzo in uno dei seguenti cluster di prezzo, a seconda della granularità del sistema di prezzi. Questi cluster riflettono possibili dimensioni dello spread nel mercato.

- $s_1 = \$0,01$  (Penny)
- $s_2 = \$0,05$  (Nickel)
- $s_3 = \$0,10$  (Dime)
- $s_4 = \$0,25$  (Quarter)
- $s_5 = \$1,00$  (Dollar)

#### 2. Step 2: Calcolare la Frequenza dei Cluster

Per ogni prezzo osservato, assegnarlo al cluster corrispondente in base alla dimensione dello spread che rappresenta. Calcolare il numero di transazioni  $N_j$  per ciascun cluster  $j$ . Successivamente, determinare la probabilità empirica  $F_j$  per ogni cluster:

$$F_j = \frac{N_j}{\sum_{k=1}^J N_k}$$

dove  $J$  è il numero totale di cluster.

#### 3. Step 3: Calcolare le Probabilità Non Vincolate

Le probabilità non vincolate  $U_j$  per ciascun cluster sono calcolate come segue:

$$U_j = \begin{cases} 2F_j, & \text{se } j = 1 \\ 2F_j - F_{j-1}, & \text{se } 2 \leq j \leq J-1 \\ F_j - F_{j-1}, & \text{se } j = J \end{cases}$$

#### 4. Step 4: Vincolare le Probabilità

Per garantire che le probabilità siano comprese tra 0 e 1, applichiamo il vincolo:

$$\hat{g}_j = \begin{cases} \min(\max(U_j, 0), 1), & \text{per } j = 1 \\ \min(\max(U_j, 0), 1 - \sum_{k=1}^{j-1} \hat{g}_k), & \text{per } 2 \leq j \leq J \end{cases}$$

Questo vincolo garantisce che le probabilità siano valide e la loro somma non ecceda 1.

#### 5. Step 5: Calcolare lo Spread Effettivo Stimato

Infine, lo spread effettivo stimato è una media ponderata degli spread dei vari cluster:

$$\text{Effective Tick} = \frac{\sum_{j=1}^J \hat{g}_j s_j}{\bar{P}}$$

dove  $\bar{P}$  è il prezzo medio del titolo nel periodo di osservazione.

#### Giorni di Calcolo dell'Effective Tick

L'Effective Tick viene calcolato utilizzando i dati dei giorni in cui si sono verificate transazioni (**giorni di trading con volume positivo**). Ciò significa che consideriamo solo i giorni in cui il titolo è stato scambiato almeno una volta.

#### Esempio:

Supponiamo di voler calcolare l'Effective Tick per un titolo durante una settimana di trading. Se durante quella settimana il titolo è stato scambiato nei giorni Lunedì, Martedì e Giovedì, ma non nei giorni Mercoledì e Venerdì, utilizzeremo solo i dati di Lunedì, Martedì e Giovedì per il calcolo.

### 1.3 Effective Tick 2

L'Effective Tick 2 segue gli stessi passaggi dell'Effective Tick, ma include **tutti i giorni** nel periodo di analisi, sia quelli con volume di negoziazione positivo che quelli senza transazioni. Nei giorni senza transazioni, i prezzi di riferimento (ad esempio, i prezzi bid-ask o i prezzi di chiusura del giorno precedente) vengono utilizzati per determinare le informazioni sullo spread.

#### Giorni di Calcolo dell'Effective Tick 2

L'Effective Tick 2 viene calcolato utilizzando i dati di **tutti i giorni di trading** nel periodo considerato, inclusi i giorni in cui non si sono verificate

transazioni per il titolo. Questo approccio permette di incorporare informazioni anche nei giorni di inattività del titolo.

**Esempio:**

Riprendendo l'esempio precedente, per calcolare l'Effective Tick 2 durante la settimana di trading, considereremo i dati di tutti i cinque giorni (Lunedì, Martedì, Mercoledì, Giovedì e Venerdì), anche se il titolo non è stato scambiato nei giorni Mercoledì e Venerdì. Nei giorni senza transazioni, utilizzeremo informazioni disponibili come i prezzi bid-ask o i prezzi di chiusura per stimare le frequenze dei cluster di prezzo.

## 1.4 Holden

### Parte Teorica

Il modello di Holden fornisce un metodo per stimare l'effettivo bid-ask spread utilizzando i prezzi di scambio osservati, tenendo conto di fattori come la selezione avversa, gli effetti di inventario e il clustering dei prezzi. L'effettivo spread è una misura chiave dei costi di transazione e della liquidità di mercato. Di seguito, scomponiamo il quadro teorico passo dopo passo.

### 1. Evoluzione del Valore Fondamentale

Sia  $V_t$  il valore fondamentale non osservabile di un'azione alla fine del giorno  $t$ . Il valore fondamentale evolve secondo:

$$V_t = V_{t-1} + \frac{1}{2}\alpha S_{t-1}Q_{t-1} + \epsilon_t \quad (1)$$

- $\alpha$ : Porzione della metà dello spread attribuibile alla selezione avversa, la selezione avversa si riferisce al rischio che i market maker o i dealer affrontano quando scambiano con trader che possiedono informazioni private superiori o privilegiate sul valore fondamentale di un'azione. Questo fenomeno si verifica perché i trader informati possono sfruttare le loro informazioni per trarre vantaggio nelle transazioni, mentre i trader non informati (o meno informati) non possono distinguere se stanno scambiando con qualcuno che possiede tali informazioni.

La selezione avversa porta i market maker ad ampliare lo spread bid-ask per compensare il rischio di perdite potenziali derivanti dal trading con individui informati. In altre parole, una parte dello spread è utilizzata come protezione contro la possibilità che il market maker stia acquistando (vendendo) un'azione a un prezzo superiore (inferiore) rispetto al suo valore fondamentale, a causa dell'asimmetria informativa.

- $S_{t-1}$ : Spread effettivo nel giorno  $t - 1$ .
- $Q_{t-1}$ : Indicatore di acquisto/vendita/assenza di scambio nel giorno  $t - 1$ .

- $\epsilon_t$ : Shock di informazione pubblica non correlato in serie nel giorno  $t$ .

Questa equazione modella come il valore fondamentale cambia a causa di informazioni private (selezione avversa) e shock di informazione pubblica.

## 2. Indicatore di Acquisto/Vendita/No-Trade

L'indicatore  $Q_t$  cattura la direzione degli scambi:

$$Q_t = \begin{cases} 1 & \text{se il prezzo di chiusura è un acquisto (probabilità } \mu/2) \\ 0 & \text{se non c'è scambio; il prezzo riportato è il midpoint di chiusura (probabilità } 1 - \mu) \\ -1 & \text{se il prezzo di chiusura è una vendita (probabilità } \mu/2) \end{cases} \quad (2)$$

- $\mu/2$ : Probabilità di una transazione di chiusura al prezzo ask (acquisto) o bid (vendita).
- $1 - \mu$ : Probabilità di un giorno senza scambi.

## 3. Midpoint Bid-Ask Non Osservato

Il midpoint bid-ask non osservato  $M_t$  è dato da:

$$M_t = V_t + \omega_t \quad (3)$$

- $\omega_t$ : Effetto cumulativo dell'inventario da scambi precedenti, riflettendo gli squilibri di inventario del dealer. Quando un dealer o un market maker facilita le transazioni tra acquirenti e venditori, accumula posizioni in azioni (positive o negative) a seconda che abbia venduto o acquistato azioni. Questo accumulo di posizioni può portare a:
  - Rischio di Inventario: Se un dealer accumula un grande inventario (lungo o corto), è esposto al rischio che il valore dell'azione cambi sfavorevolmente.
  - Aggiustamento dei Prezzi: Per gestire questo rischio, il dealer potrebbe aggiustare i prezzi bid e ask per incentivare le controparti di mercato a prendere l'opposto della sua posizione, aiutandolo a riportare l'inventario a un livello desiderato.

## 4. Variazione Giornaliera del Midpoint

Differenziando l'Equazione (3) e sostituendo nell'Equazione (1):

$$\Delta M_t = \frac{1}{2} \alpha S_{t-1} Q_{t-1} + \epsilon_t + \Delta \omega_t \quad (4)$$

- $\Delta$ : Operatore di differenza ( $\Delta M_t = M_t - M_{t-1}$ ).

## 5. Prezzo di Scambio Osservato

Il prezzo di scambio osservato  $P_t$  è:

$$P_t = M_t + \frac{1}{2}S_tQ_t + \eta_t(S_t) \quad (5)$$

- $\eta_t(S_t)$ : Termine di errore dovuto all'arrotondamento del prezzo a cluster di prezzo discreti, con  $E[\eta_t(S_t)] = 0$ .

## 6. Processo Chiave del Cambiamento di Prezzo

Combinando le Equazioni (4) e (5):

$$\Delta P_t = \frac{1}{2}S_tQ_t - (1 - \alpha)\frac{1}{2}S_{t-1}Q_{t-1} + e_t \quad (6)$$

- $e_t = \epsilon_t + \Delta\eta_t + \Delta\omega_t$ : Termine di errore assunto normalmente distribuito con media  $\bar{e}$  e deviazione standard  $\sigma_e$ .

## 7. Definizione della Metà dello Spread

Definiamo la metà dello spread  $H_t$  nel giorno  $t$ :

$$H_t = \frac{1}{2}S_tQ_t \quad (7)$$

## 8. Espressione del Termine di Errore

Risolvendo per il termine di errore  $e_t$ :

$$e_t = \Delta P_t - (H_t - (1 - \alpha)H_{t-1}) \quad (8)$$

Questa equazione esprime  $e_t$  in termini di quantità osservabili e del parametro  $\alpha$ .

## 9. Clustering dei Prezzi e Spread Effettivo

Il clustering dei prezzi si verifica a causa della preferenza dei trader per certi incrementi di prezzo per minimizzare i costi di negoziazione. Su una griglia di prezzi frazionaria, i trader scelgono cluster di prezzo come ottavi, quarti, metà o interi dollari. Lo spread effettivo  $S_t$  è assunto uguale all'incremento del cluster di prezzo scelto nel giorno  $t$ .

Si modellano possibili spread effettivi  $s_j$  con probabilità  $\gamma_j$ , dove  $s_j$  sono ordinati dal più piccolo al più grande (ad esempio,  $\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, \$1$ ).

## 10. Funzione di Verosimiglianza

L'obiettivo è stimare la proxy dello spread, che è lo spread percentuale effettivo ponderato:

$$\text{Holden} = \frac{\sum_{j=1}^J \hat{\gamma}_j s_j}{\bar{P}}$$

- $\hat{\gamma}_j$ : Probabilità stimate degli spread effettivi.
- $\bar{P}$ : Prezzo medio di scambio nel periodo di aggregazione.

Vincoli:

$$\sum_{j=1}^J \hat{\gamma}_j = 1$$

### Funzione di Verosimiglianza per Triplette di Prezzi

Utilizzando triple di prezzi osservati  $(P_t, P_{t+1}, P_{t+2})$ , la verosimiglianza è:

$$\mathcal{L} = \sum_{(H_t, H_{t+1}, H_{t+2}) \in H} \left\{ \prod_{i=t}^{t+2} \Pr(C_i) \Pr(H_i | C_i) \right\} \times \left\{ \prod_{i=t+1}^{t+2} n(e_i) \right\}$$

- $H$ : Insieme di triple di metà spread fattibili date le cluster di prezzo osservate.
- $\Pr(C_i)$ : Probabilità della cluster di prezzo  $C_i$ .
- $\Pr(H_i | C_i)$ : Probabilità condizionale della metà dello spread  $H_i$  data la cluster  $C_i$ .
- $n(e_i)$ : Funzione di densità normale del termine di errore  $e_i$ .

I termini di errore  $e_{t+1}$  ed  $e_{t+2}$  sono calcolati usando l'Equazione (8):

$$\begin{aligned} e_{t+1} &= \Delta P_{t+1} - (H_{t+1} - (1 - \alpha)H_t) \\ e_{t+2} &= \Delta P_{t+2} - (H_{t+2} - (1 - \alpha)H_{t+1}) \end{aligned}$$

### Parametri da Stimare

- $\mu$ : Probabilità di un giorno di scambio.
- $\gamma_j$ : Probabilità degli spread effettivi  $s_j$ .
- $\bar{e}$ : Media del termine di errore.
- $\sigma_e$ : Deviazione standard del termine di errore.

- $\alpha$ : Porzione dello spread dovuta alla selezione avversa.

Vincoli:

- **Probabilità del Giorno di Scambio:**

$$0 \leq \mu \leq 1$$

- **Probabilità degli Spread Effettivi:**

$$0 \leq \gamma_j \leq 1 \quad \text{per ogni } j = 1, 2, \dots, J$$

$$\sum_{j=1}^J \gamma_j = 1$$

- **Parametro di Selezione Avversa:**

$$0 \leq \alpha \leq 1$$

- **Termine di Errore:**

$$\sigma_e > 0$$

$$\bar{e} \in R$$

## 11. Probabilità di Cluster di Prezzo e Metà dello Spread

Le probabilità  $\Pr(C_t)$  e  $\Pr(H_t|C_t)$  sono derivate in base ai cluster di prezzo scelti e agli spread effettivi.

### Probabilità dei Cluster di Prezzo

Per i cluster di prezzo di scambio:

$$\Pr(C_t = j) = \frac{\sum_{k=1}^j \gamma_k \mu^{(j-k+1)} \left(\frac{1}{2}\right)^k}{\sum_{k=1}^J \gamma_k \mu^{(J-k)} \left(\frac{1}{2}\right)^k}$$

Per i cluster di midpoint (giorni senza scambio):

$$\Pr(C_t = J + j) = \gamma_j (1 - \mu)$$

### Probabilità Condizionali della Metà dello Spread

Per i cluster di prezzo di scambio:

$$\Pr(H_t = h_k | C_t = j) = \frac{\gamma_k \left(\frac{1}{2}\right)^{j-k+1}}{\Pr(C_t = j)}$$

Per i cluster di midpoint:

$$\Pr(H_t = 0 | C_t = J + j) = 1$$



## 12. Massimizzazione della Funzione di Verosimiglianza

La funzione di log-verosimiglianza è massimizzata su tutte le triple di prezzi osservate per stimare i parametri.

I vincoli sono applicati per assicurare che le probabilità sommino a uno e siano entro intervalli validi.

## Parte Pratica

Per implementare la proxy di Holden utilizzando i dati dell'S&P 500, seguire questi passaggi:

### 1. Requisiti e Preparazione dei Dati

- **Dati Necessari:**

- Prezzi di chiusura giornalieri  $P_t$  per le azioni dell'S&P 500.
- Date corrispondenti a ciascun prezzo di chiusura.
- Indicatore se una transazione è avvenuta al bid, ask o midpoint (potrebbe essere necessario inferire  $Q_t$ ).

- **Preprocessing dei Dati:**

- Inferire  $Q_t$  se non direttamente osservato (ad esempio, utilizzando il test del tick).

### 2. Calcolo delle Variabili

- **Calcolare le Variazioni di Prezzo:**

$$\Delta P_t = P_t - P_{t-1}$$

- **Determinare i Cluster di Prezzo  $C_t$ :**

- Identificare gli incrementi di prezzo utilizzati ogni giorno (es.  $\frac{1}{8}, \frac{1}{4}$ ).
- Assegnare  $C_t$  in base al clustering di prezzo osservato.

- **Assegnare le Metà degli Spread  $H_t$ :**

- Basandosi su  $C_t$ , generare i possibili valori di  $H_t$ .
- Utilizzare le probabilità condizionali  $\Pr(H_t|C_t)$  per considerare i possibili  $H_t$ .

### 3. Impostazione dell'Estrazione

- **Definire i Parametri da Stimare:**

- $\gamma_j$  per ciascuno spread effettivo  $s_j$ .
- $\alpha, \bar{e}, \sigma_e$ .
- $\mu$ , se si modella la probabilità di giorni di scambio.

- **Valori Iniziali:**

- Fornire valori iniziali ragionevoli per l'ottimizzatore.
- Assicurarsi che  $\gamma_j$  sommino a uno.

### 4. Implementazione della Funzione di Verosimiglianza

- **Calcolare la Verosimiglianza per Ogni Tripletta:**

- Per ciascuna tripletta di prezzi sovrapposta  $(P_t, P_{t+1}, P_{t+2})$ :
  - \* Generare tutte le combinazioni possibili di  $(H_t, H_{t+1}, H_{t+2})$  date  $C_t$ .
  - \* Calcolare  $e_{t+1}$  ed  $e_{t+2}$  usando l'Equazione (8).
  - \* Calcolare la densità normale  $n(e_i)$  per  $e_{t+1}$  ed  $e_{t+2}$ .
  - \* Moltiplicare le probabilità e le densità secondo la funzione di verosimiglianza.

- **Somma dei Log-Verosimiglianze:**

- Sommare il log delle verosimiglianze su tutte le triplette.

### 5. Stima dei Parametri

- **Routine di Ottimizzazione:**

- Utilizzare metodi di ottimizzazione numerica (es. Newton-Raphson, BFGS) per massimizzare la funzione di log-verosimiglianza.
- Applicare i vincoli (es. utilizzando metodi di barriera o algoritmi di ottimizzazione vincolati).

- **Gestione dei Vincoli:**

- Assicurarsi che  $\gamma_j \geq 0$  e  $\gamma_j \leq 1$ .
- Applicare  $\sum_{j=1}^J \gamma_j = 1$ .
- Mantenere  $\sigma_e > 0$  e  $0 \leq \alpha \leq 1$ .

## 6. Calcolo della Proxy di Holden

- **Calcolare la Stima dello Spread Effettivo:**

$$\text{Holden} = \frac{\sum_{j=1}^J \hat{\gamma}_j s_j}{\bar{P}}$$

- **Interpretazione:**

- La proxy di Holden stimata rappresenta lo spread percentuale effettivo medio nel periodo.

### 1.5 Gibbs Sampler per la stima dello spread

### 1.6 Implementazione del Modello LOT (Lesmond, Ogden, Trzcinka)

Il modello LOT (Lesmond, Ogden e Trzcinka) utilizza il modello a variabile dipendente limitata (LDV) per stimare i costi di transazione (round-trip) dei titoli azionari. Questo modello presuppone che i rendimenti dei titoli siano funzione del rendimento di mercato, ma che i costi di transazione influenzino la capacità degli investitori di trarre vantaggio dalle informazioni, creando una zona di non-trading dove i rendimenti misurati sono zero.

#### 1. Raccolta dei Dati

- **Rendimenti dei Titoli ( $R_{jt}$ ):** Raccogliere i rendimenti giornalieri (o periodici) per ciascun titolo  $j$  nel periodo  $t$ .
- **Rendimenti del Mercato ( $R_{mt}$ ):** Ottenere i rendimenti dell'indice di mercato corrispondente per lo stesso periodo  $t$ .

#### 2. Definizione dei Rendimenti

I rendimenti sono calcolati come le variazioni percentuali dei prezzi tra due periodi consecutivi. Matematicamente:

- **Rendimento del Titolo  $j$  nel periodo  $t$ :**

$$R_{jt} = \frac{P_{jt} - P_{j,t-1}}{P_{j,t-1}} = \frac{\Delta P_{jt}}{P_{j,t-1}}$$

Dove:

- $P_{jt}$  è il prezzo del titolo  $j$  al tempo  $t$ .
- $P_{j,t-1}$  è il prezzo del titolo  $j$  al tempo  $t - 1$ .
- $\Delta P_{jt} = P_{jt} - P_{j,t-1}$  è la variazione del prezzo del titolo tra  $t - 1$  e  $t$ .

- **Rendimento del Mercato nel periodo  $t$ :**

$$R_{mt} = \frac{I_t - I_{t-1}}{I_{t-1}} = \frac{\Delta I_t}{I_{t-1}}$$

Dove:

- $I_t$  è il valore dell'indice di mercato al tempo  $t$ .
- $I_{t-1}$  è il valore dell'indice di mercato al tempo  $t - 1$ .
- $\Delta I_t = I_t - I_{t-1}$  è la variazione dell'indice tra  $t - 1$  e  $t$ .

### Specificazione del Modello LDV

Il modello LDV considera che i costi di transazione creano una zona di non-trading dove i rendimenti misurati sono zero.

#### 1. Rendimento "Vero":

$$R_{jt}^* = \beta_j R_{mt} + \epsilon_{jt}$$

dove:

- $R_{jt}^*$ : Rendimento "vero" non osservato del titolo  $j$  al tempo  $t$ .
- $\beta_j$ : Coefficiente di sensibilità al mercato.
- $\epsilon_{jt}$ : Termine di errore (assunto normalmente distribuito con media zero e varianza  $\sigma_j^2$ ).

#### 2. Rendimento Misurato (Osservato):

##### 1.6.1 Le Soglie dei Costi di Transazione nel Modello LOT

Nel modello LOT (Lesmond, Ogden e Trzcinka), le soglie  $\alpha_{1j}$  e  $\alpha_{2j}$  sono parametri fondamentali che rappresentano i costi di transazione proporzionali associati alla vendita e all'acquisto del titolo  $j$ , rispettivamente.

#### Definizione delle Soglie

- $\alpha_{1j}$ : **Soglia inferiore dei costi di transazione**, associata alla **vendita** del titolo  $j$ . È un valore negativo, cioè  $\alpha_{1j} < 0$ .
- $\alpha_{2j}$ : **Soglia superiore dei costi di transazione**, associata all'**acquisto** del titolo  $j$ . È un valore positivo, cioè  $\alpha_{2j} > 0$ .

**Equazioni del Modello** Il modello LDV specifica il rendimento osservato  $R_{jt}$  in funzione del rendimento "vero" non osservato  $R_{jt}^*$  e delle soglie di costo di transazione:

$$R_{jt} = \begin{cases} R_{jt}^* - \alpha_{1j} & \text{se } R_{jt}^* < \alpha_{1j} \\ 0 & \text{se } \alpha_{1j} \leq R_{jt}^* \leq \alpha_{2j} \\ R_{jt}^* - \alpha_{2j} & \text{se } R_{jt}^* > \alpha_{2j} \end{cases}$$

dove  $R_{jt}^*$  è il rendimento "vero" non osservato del titolo  $j$  al tempo  $t$ , determinato come:

$$R_{jt}^* = \beta_j R_{mt} + \epsilon_{jt}$$

con:

- $\beta_j$ : Coefficiente di sensibilità al mercato del titolo  $j$ .
- $R_{mt}$ : Rendimento del mercato al tempo  $t$ .
- $\epsilon_{jt}$ : Termine di errore (assunto normalmente distribuito con media zero e varianza  $\sigma_j^2$ ).

**Interpretazione delle Soglie e delle Regioni di Trading** Le soglie  $\alpha_{1j}$  e  $\alpha_{2j}$  definiscono tre regioni distinte di comportamento di trading:

(a) **Regione di Vendita** ( $R_{jt}^* < \alpha_{1j}$ ):

- *regione 2*:  $R_{jt} \neq 0$  e  $R_{mt} < 0$  il rendimento osservato è diverso da 0 quindi è avvenuta una transazione e il mercato scende
- *Condizione*: Il rendimento "vero" è sufficientemente negativo da superare (in valore assoluto) il costo di transazione per la vendita.
- *Incentivo a Vendere*: L'investitore è incentivato a vendere poiché si aspetta che il prezzo del titolo diminuisca significativamente.
- *Rendimento Osservato*:  $R_{jt} = R_{jt}^* - \alpha_{1j}$ . Poiché  $\alpha_{1j} < 0$ , sottrarre  $\alpha_{1j}$  equivale ad aggiungere il costo di transazione (rendendo il rendimento osservato meno negativo rispetto al rendimento "vero").

(b) **Regione di Nessun Trading** ( $\alpha_{1j} \leq R_{jt}^* \leq \alpha_{2j}$ ):

- *regione 0*:  $R_{jt} = 0$  nessuna transazione avvenuta
- *Condizione*: Il rendimento "vero" non è sufficientemente grande (in valore assoluto) da superare i costi di transazione per l'acquisto o la vendita.
- *Nessun Incentivo a Fare Trading*: L'investitore non è incentivato né a comprare né a vendere.
- *Rendimento Osservato*:  $R_{jt} = 0$ .

(c) **Regione di Acquisto** ( $R_{jt}^* > \alpha_{2j}$ ):

- *regione 1*:  $R_{jt} \neq 0$  e  $R_{mt} > 0$  il rendimento osservato è diverso da 0 quindi è avvenuta una transazione e il mercato sale
- *Condizione*: Il rendimento "vero" è sufficientemente positivo da superare il costo di transazione per l'acquisto.
- *Incentivo ad Acquistare*: L'investitore è incentivato ad acquistare poiché si aspetta che il prezzo del titolo aumenti significativamente.
- *Rendimento Osservato*:  $R_{jt} = R_{jt}^* - \alpha_{2j}$ , riflettendo il costo di transazione sostenuto per l'acquisto.

### Valori Teorici delle Soglie

- $\alpha_{1j} < 0$ : Essendo associata alla vendita, la soglia è negativa perché riflette un rendimento "vero" negativo necessario per superare il costo di transazione della vendita.
- $\alpha_{2j} > 0$ : Essendo associata all'acquisto, la soglia è positiva perché riflette un rendimento "vero" positivo necessario per superare il costo di transazione dell'acquisto.

**Calcolo del Costo di Transazione Totale** Il costo di transazione totale (*round-trip*) per il titolo  $j$  è dato dalla differenza tra le soglie:

$$\text{Costo totale di transazione} = \alpha_{2j} - \alpha_{1j}$$

Questo valore rappresenta il costo complessivo che un investitore sostiene per acquistare e successivamente vendere il titolo  $j$ .

**Costruzione della Funzione di Verosimiglianza** La funzione di verosimiglianza  $L$  per il titolo  $j$  è data dal prodotto delle probabilità associate alle osservazioni in ciascuna regione:

$$L = \left( \prod_{t \in \mathcal{R}_1} \frac{1}{\sigma_j} \phi \left( \frac{R_{jt} + \alpha_{1j} - \beta_j R_{mt}}{\sigma_j} \right) \right) \times \left( \prod_{t \in \mathcal{R}_0} \left[ \Phi \left( \frac{\alpha_{2j} - \beta_j R_{mt}}{\sigma_j} \right) - \Phi \left( \frac{\alpha_{1j} - \beta_j R_{mt}}{\sigma_j} \right) \right] \right) \times \left( \prod_{t \in \mathcal{R}_2} \frac{1}{\sigma_j} \phi \left( \frac{R_{jt} - \alpha_{2j} - \beta_j R_{mt}}{\sigma_j} \right) \right)$$

dove:

- $\phi(\cdot)$ : Funzione di densità della normale standard.
- $\Phi(\cdot)$ : Funzione di distribuzione cumulativa della normale standard.
- $\sigma_j$ : Deviazione standard dei residui (stimata utilizzando solo i rendimenti misurati non nulli).

#### Nota

qua da vedere bene cosa sono queste funzioni di densità, l'articolo di lesmond a pag 11 dice "The terms  $\phi_1$  and  $\phi_2$  refer to standard normal density functions for decreases and increases in the measured return, respectively." mentre per  $\Phi$  "is the distribution function of the standard normal distribution" a me sembra la cumulata invece.

**Funzione di Log-Verosimiglianza** Prendendo il logaritmo naturale della funzione di verosimiglianza, otteniamo la funzione di log-verosimiglianza  $\ln L$ :

$$\begin{aligned}\ln L = & \sum_1 \ln \left( \frac{1}{(2\pi\sigma_j^2)^{1/2}} \right) - \sum_1 \frac{1}{2\sigma_j^2} (R_{jt} + \alpha_{1j} - \beta_j \cdot R_{mt})^2 \\ & + \sum_2 \ln \left( \frac{1}{(2\pi\sigma_j^2)^{1/2}} \right) - \sum_2 \frac{1}{2\sigma_j^2} (R_{jt} + \alpha_{2j} - \beta_j \cdot R_{mt})^2 \\ & + \sum_0 \ln(\Phi_{j2} - \Phi_{j1})\end{aligned}$$

**Stima dei Parametri** I parametri  $\alpha_{1j}$ ,  $\alpha_{2j}$ ,  $\beta_j$  e  $\sigma_j$  sono stimati massimizzando la funzione di log-verosimiglianza  $\ln L$  rispetto a tali parametri.

### 3. Calcolo dei Costi di Transazione

- **Costo di Round-Trip:**

$$\text{Costo totale} = \alpha_{2j} - \alpha_{1j}$$

- **Interpretazione:** Questo valore rappresenta il costo totale per eseguire un acquisto e una vendita sul titolo  $j$ .

### 4. Considerazioni sulle Specifiche del Modello

- **Intercetta Soppressa:** Nel modello, l'intercetta del modello di mercato è soppressa per evitare problemi di identificazione con i costi di transazione.

- **Efficienza dell'Indice di Mercato:** Assicurarsi che l'indice di mercato utilizzato sia appropriato e rappresentativo per evitare bias nelle stime.

## Passi Pratici per l'Implementazione

### 1. Preparazione dei Dati

- Importare i dati dei prezzi dei titoli e dell'indice di mercato.
- Calcolare i rendimenti giornalieri per i titoli e l'indice di mercato

### 2. Classificazione delle Osservazioni

- Per ogni osservazione  $t$ , calcolare  $R_{jt}$  e  $R_{mt}$ .
- Classificare le osservazioni in:
  - $\mathcal{R}_1$ :  $R_{mt} < 0$  e  $R_{jt} \neq 0$  (rendimenti negativi).
  - $\mathcal{R}_0$ :  $R_{jt} = 0$  (rendimenti zero).
  - $\mathcal{R}_2$ :  $R_{mt} > 0$  e  $R_{jt} \neq 0$  (rendimenti positivi).

### 3. Stima Iniziale di $\beta_j$

- Eseguire una regressione lineare senza intercetta dei rendimenti dei titoli sui rendimenti di mercato utilizzando solo le osservazioni con  $R_{jt} \neq 0$  per ottenere una stima iniziale di  $\beta_j$ .

questa può essere una buona inizializzazione per beta?

### 4. Costruzione della Funzione di Verosimiglianza

- Implementare la funzione di verosimiglianza  $L$  come definita sopra.
- Utilizzare le stime iniziali dei parametri per avviare l'ottimizzazione.

### 5. Ottimizzazione Numerica

- Utilizzare un algoritmo di ottimizzazione (ad esempio, il metodo di Newton-Raphson, algoritmi di massima verosimiglianza) per massimizzare  $L$  rispetto a  $\alpha_{1j}$ ,  $\alpha_{2j}$ ,  $\beta_j$  e  $\sigma_j$ .
- Impostare vincoli appropriati  $\alpha_{j1} \leq 0$ ,  $\alpha_{j2} \geq 0$ ,  $\beta_j \geq 0$ ,  $\sigma_j \geq 0$  e  $\alpha_{j2} > \alpha_{j1}$ .

### 6. Verifica dei Risultati

- Controllare la convergenza dell'algoritmo e la significatività statistica dei parametri stimati.
- Analizzare i residui per verificare l'adeguatezza del modello.



## 7. Calcolo del Costo di Transazione

- Calcolare Costo di Transazione =  $\alpha_{2j} - \alpha_{1j}$ .
- Interpretare il risultato come stima del costo di transazione round-trip per il titolo  $j$ .

## 8. Ripetizione per Tutti i Titoli

- Ripetere i passi precedenti per ogni titolo dell'S&P 500.
- Confrontare i costi di transazione stimati tra i vari titoli.

### 1.7 Lot Y-split

Simile a *Lot Mixed*, simile al LOT ma differente definizione delle regioni

#### Definizione delle Regioni di Trading

- **Regione 0:**  $R_{jt} = 0$   
Nessuna transazione avviene; il rendimento osservato è zero.
- **Regione 1 (Acquisto):**  $R_{jt} > 0$   
L'investitore è incentivato ad acquistare.
- **Regione 2 (Vendita):**  $R_{jt} < 0$   
L'investitore è incentivato a vendere.

### 1.8 Proxy di Liquidità basate sui Giorni a Rendimento Zero

Lesmond, Ogden e Trzcinka (1999) introducono la proporzione di giorni con rendimenti zero come proxy per la liquidità di un titolo. Questa misura si basa su due argomentazioni principali:

1. **Bassa Liquidità e Giorni a Volume Zero:** I titoli con bassa liquidità sono più propensi ad avere giorni con volume di scambio zero e, di conseguenza, sono più propensi ad avere giorni con rendimento zero.
2. **Alti Costi di Transazione e Acquisizione di Informazioni:** I titoli con alti costi di transazione hanno minore acquisizione di informazioni private (poiché è più difficile superare costi di transazione più elevati) e, pertanto, anche in giorni con volume positivo, sono più propensi ad avere giorni a rendimento zero senza rivelazione di nuove informazioni.

#### 1.8.1 Definizione delle Misure

Le proxy basate sui giorni a rendimento zero sono definite come segue:

- **Zeros:**

$$\text{Zeros} = \frac{\text{Numero di giorni con rendimento zero}}{T}$$

dove  $T$  è il numero totale di giorni di trading in un mese.

- **Zeros2:**

$$\text{Zeros2} = \frac{\text{Numero di giorni con volume positivo e rendimento zero}}{T}$$

### 1.8.2 Implementazione delle Misure

#### 1. Raccolta dei Dati:

- Ottenere i dati giornalieri dei rendimenti ( $R_{jt}$ ) e dei volumi di scambio per ciascun titolo  $j$  nel periodo di interesse.
- Determinare il numero totale di giorni di trading  $T$  nel periodo considerato (ad esempio, un mese).

#### 2. Calcolo di Zeros:

- Contare il numero di giorni in cui il rendimento del titolo è esattamente zero ( $R_{jt} = 0$ ).
- Applicare la formula:

$$\text{Zeros} = \frac{\text{Numero di giorni con } R_{jt} = 0}{T}$$

#### 3. Calcolo di Zeros2:

- Identificare i giorni in cui il volume di scambio è positivo.
- Tra questi, contare il numero di giorni in cui il rendimento è zero ( $R_{jt} = 0$ ).
- Applicare la formula:

$$\text{Zeros2} = \frac{\text{Numero di giorni con volume positivo e } R_{jt} = 0}{T}$$

## 2 Proxy per il Price Impact

In questa sezione, deriviamo e descriviamo varie misure proxy per il *price impact*, come presentato nell'articolo. Forniamo le formule matematiche, spieghiamo come vengono calcolate e discutiamo la loro implementazione.

## 2.1 1. Misura di Illiquidità di Amihud (2002)

### 2.1.1 Descrizione

La misura di illiquidità di Amihud cattura il *price impact* medio per unità di volume scambiato. Essa rappresenta la risposta media del prezzo associata a un dollaro di volume di trading.

### 2.1.2 Formula

La misura è definita come:

$$\text{Illiquidity} = \frac{1}{T} \sum_{t=1}^T \left( \frac{|R_t|}{\text{Volume}_t} \right),$$

dove:

- $R_t$  è il rendimento del titolo nel giorno  $t$ ,
- $\text{Volume}_t$  è il volume in dollari scambiato nel giorno  $t$ ,
- $T$  è il numero totale di giorni di trading nel periodo di interesse.

### 2.1.3 Implementazione

#### 1. Raccolta dei Dati:

- Ottenere i dati giornalieri dei rendimenti ( $R_t$ ) e dei volumi scambiati ( $\text{Volume}_t$ ) per il titolo in esame.

#### 2. Calcolo della Misura:

- Calcolare il valore assoluto del rendimento giornaliero  $|R_t|$ .
- Dividere  $|R_t|$  per  $\text{Volume}_t$  per ogni giorno.
- Calcolare la media dei valori ottenuti su tutti i giorni di trading  $T$ .
- porre attenzione: il calcolo viene fatto solo sui giorni con volume positivo (altrimenti non ha senso fare il rapporto con 0 a denominatore)

### 2.1.4 Interpretazione

Un valore elevato della misura di illiquidità di Amihud indica che piccoli volumi di trading hanno un impatto significativo sul prezzo, suggerendo bassa liquidità e alto *price impact*.

—

## 2.2 2. Proxies Estese di Amihud

### 2.2.1 Descrizione

Le proxies estese di Amihud sono una classe di misure sviluppate estendendo la misura originale di Amihud. Queste misure cercano di isolare la componente di liquidità del rendimento, eliminando il rumore non correlato alla liquidità.

### 2.2.2 Formula della Proxy Estesa

La proxy estesa di Amihud è definita come:

$$\text{Extended Amihud Proxy} = \frac{\text{Spread Proxy}}{\text{Average Daily Dollar Volume}},$$

dove:

- Spread Proxy può essere una qualsiasi proxy per lo spread (ad esempio, Roll, Effective Tick, ecc.),
- Average Daily Dollar Volume è il volume medio giornaliero in dollari scambiato nel periodo considerato.

### 2.2.3 Implementazione

#### 1. Calcolo della Spread Proxy:

- Scegliere una misura di spread appropriata (vedi sezioni successive per le diverse proxy).
- Calcolare lo spread proxy per il periodo di interesse.

#### 2. Calcolo del Volume Medio:

- Calcolare il volume medio giornaliero in dollari scambiato nel periodo.

#### 3. Calcolo della Proxy Estesa:

- Dividere lo spread proxy per il volume medio giornaliero.

### 2.2.4 Interpretazione

Questa misura stima il *price impact* associato a un'unità di volume scambiato, tenendo conto dello spread effettivo.

—

## 2.3 3. Misura Gamma di Pastor e Stambaugh (2003)

### 2.3.1 Descrizione

La misura Gamma di Pastor e Stambaugh cattura l'elasticità del rendimento rispetto al flusso di ordini precedenti. Essa quantifica l'entità con cui i rendimenti si invertano dopo giorni con volumi elevati.

### 2.3.2 Formula

La misura Gamma è stimata tramite la seguente regressione:

$$r_{t+1}^e = \theta + \phi r_t + (\text{Gamma}) \cdot \text{sign}(r_t^e) \cdot \text{Volume}_t + \eta_t,$$

dove:

- $r_t^e$  è il rendimento in eccesso del titolo rispetto al mercato nel giorno  $t$ ,
- $\text{Volume}_t$  è il volume in dollari scambiato nel giorno  $t$ ,
- $\text{sign}(r_t^e)$  è il segno del rendimento in eccesso nel giorno  $t$ ,
- $\theta$  e  $\phi$  sono parametri della regressione,
- $\eta_t$  è il termine di errore.

### 2.3.3 Implementazione

#### 1. Calcolo dei Rendimenti in Eccesso:

- Calcolare i rendimenti giornalieri del titolo  $r_t$ .
- Calcolare i rendimenti giornalieri del mercato  $r_{mt}$ .
- Ottenere i rendimenti in eccesso:  $r_t^e = r_t - r_{mt}$ .

#### 2. Preparazione dei Dati per la Regressione:

- Calcolare  $\text{sign}(r_t^e) \cdot \text{Volume}_t$  per ogni giorno.

#### 3. Stima della Regressione:

- Eseguire la regressione lineare di  $r_{t+1}^e$  su  $r_t$  e  $\text{sign}(r_t^e) \cdot \text{Volume}_t$ .
- Ottenere la stima del coefficiente Gamma.

### 2.3.4 Interpretazione

Un valore negativo di Gamma indica che i prezzi tendono a invertire la direzione dopo giorni con volumi elevati, suggerendo un alto *price impact*.

## 2.4 4. Rapporto di Liquidità di Amivest

### 2.4.1 Descrizione

Il rapporto di liquidità di Amivest misura la liquidità come il volume scambiato per unità di variazione di prezzo. Un valore elevato indica alta liquidità.

### 2.4.2 Formula

$$\text{Liquidity} = \frac{1}{T} \sum_{t=1}^T \left( \frac{\text{Volume}_t}{|R_t|} \right).$$

### 2.4.3 Implementazione

#### 1. Raccolta dei Dati:

- Ottenere i dati giornalieri dei rendimenti  $R_t$  e dei volumi  $\text{Volume}_t$ .

#### 2. Calcolo del Rapporto di Liquidità:

- Calcolare  $\frac{\text{Volume}_t}{|R_t|}$  per ogni giorno con rendimento diverso da zero.
- Calcolare la media dei valori ottenuti su tutti i giorni considerati.

### 2.4.4 Interpretazione

Un valore elevato del rapporto indica che grandi volumi sono scambiati con piccole variazioni di prezzo, suggerendo alta liquidità e basso *price impact*.

## 2.5 5. Roll Impact

### 2.5.1 Descrizione

Il Roll Impact estende la misura di Roll (1984) per stimare il *price impact* basandosi sulla covarianza seriale dei cambiamenti di prezzo.

### 2.5.2 Derivazione

La misura di Roll per lo spread è data da:

$$\text{Roll Spread} = 2\sqrt{-\text{Cov}(\Delta P_t, \Delta P_{t-1})},$$

quando la covarianza è negativa.

Il Roll Impact è definito come:

$$\text{Roll Impact} = \frac{\text{Roll Spread}}{\text{Average Daily Dollar Volume}}.$$

### 2.5.3 Implementazione

#### 1. Calcolo della Covarianza:

- Calcolare i cambiamenti di prezzo giornalieri  $\Delta P_t = P_t - P_{t-1}$ .
- Calcolare la covarianza  $\text{Cov}(\Delta P_t, \Delta P_{t-1})$ .

#### 2. Calcolo dello Spread di Roll:

- Applicare la formula dello spread di Roll.

#### 3. Calcolo del Roll Impact:

- Calcolare il volume medio giornaliero in dollari.
- Dividere lo spread di Roll per il volume medio giornaliero.

### 2.5.4 Interpretazione

Il Roll Impact stima il *price impact* basandosi sulla relazione tra variazioni di prezzo e volume scambiato.

## 2.6 6. Effective Tick Impact

### 2.6.1 Descrizione

L'Effective Tick Impact estende la misura dello spread Effective Tick per stimare il *price impact*.

### 2.6.2 Formula

$$\text{Effective Tick Impact} = \frac{\text{Effective Tick Spread}}{\text{Average Daily Dollar Volume}},$$

dove l'Effective Tick Spread è calcolato come nella misura dello spread Effective Tick.

### 2.6.3 Implementazione

#### 1. Calcolo dell'Effective Tick Spread:

- Seguire i passaggi descritti nella misura dello spread Effective Tick.

#### 2. Calcolo dell'Effective Tick Impact:

- Dividere l'Effective Tick Spread per il volume medio giornaliero in dollari.

#### 2.6.4 Interpretazione

Questa misura stima il *price impact* tenendo conto della distribuzione delle variazioni di prezzo.

---

### 2.7 7. Effective Tick2 Impact

#### 2.7.1 Descrizione

Simile all'Effective Tick Impact, ma utilizza tutti i giorni, inclusi quelli senza volume.

#### 2.7.2 Formula

$$\text{Effective Tick2 Impact} = \frac{\text{Effective Tick2 Spread}}{\text{Average Daily Dollar Volume}},$$

dove l'Effective Tick2 Spread è calcolato come nella misura dello spread Effective Tick2.

#### 2.7.3 Implementazione

1. Calcolo dell'Effective Tick2 Spread:

- Seguire i passaggi descritti nella misura dello spread Effective Tick2.

2. Calcolo dell'Effective Tick2 Impact:

- Dividere l'Effective Tick2 Spread per il volume medio giornaliero in dollari.
- 

### 2.8 8. Holden Impact

#### 2.8.1 Descrizione

L'Holden Impact estende la misura dello spread di Holden per stimare il *price impact*.

#### 2.8.2 Formula

$$\text{Holden Impact} = \frac{\text{Holden Spread}}{\text{Average Daily Dollar Volume}},$$

dove l'Holden Spread è calcolato secondo la misura di Holden.



### 2.8.3 Implementazione

#### 1. Calcolo dell'Holden Spread:

- Seguire i passaggi descritti nella misura dello spread di Holden.

#### 2. Calcolo dell'Holden Impact:

- Dividere l'Holden Spread per il volume medio giornaliero in dollari.
- 

## 2.9 9. Gibbs Impact

### 2.9.1 Descrizione

Il Gibbs Impact estende la misura dello spread Gibbs per stimare il *price impact*.

### 2.9.2 Formula

$$\text{Gibbs Impact} = \frac{\text{Gibbs Spread}}{\text{Average Daily Dollar Volume}},$$

dove lo spread Gibbs è calcolato secondo la metodologia di Hasbrouck (2004).

### 2.9.3 Implementazione

#### 1. Calcolo dello Spread Gibbs:

- Utilizzare il campionatore di Gibbs per stimare i parametri del modello e ottenere lo spread.

#### 2. Calcolo del Gibbs Impact:

- Dividere lo spread Gibbs per il volume medio giornaliero in dollari.
- 

## 2.10 10. LOT Mixed Impact

### 2.10.1 Descrizione

Il LOT Mixed Impact estende la misura dello spread LOT Mixed per stimare il *price impact*.

### 2.10.2 Formula

$$\text{LOT Mixed Impact} = \frac{\text{LOT Mixed Spread}}{\text{Average Daily Dollar Volume}},$$

dove lo spread LOT Mixed è calcolato secondo la metodologia di Lesmond, Ogden e Trzcinka (1999).

### 2.10.3 Implementazione

#### 1. Calcolo dello Spread LOT Mixed:

- Seguire i passaggi descritti nella misura dello spread LOT Mixed.

#### 2. Calcolo del LOT Mixed Impact:

- Dividere lo spread LOT Mixed per il volume medio giornaliero in dollari.
- 

## 2.11 11. LOT Y-split Impact

### 2.11.1 Descrizione

Il LOT Y-split Impact estende la misura dello spread LOT Y-split per stimare il *price impact*.

### 2.11.2 Formula

$$\text{LOT Y-split Impact} = \frac{\text{LOT Y-split Spread}}{\text{Average Daily Dollar Volume}},$$

dove lo spread LOT Y-split è calcolato secondo la metodologia sviluppata nell'articolo.

### 2.11.3 Implementazione

#### 1. Calcolo dello Spread LOT Y-split:

- Seguire i passaggi descritti nella misura dello spread LOT Y-split.

#### 2. Calcolo del LOT Y-split Impact:

- Dividere lo spread LOT Y-split per il volume medio giornaliero in dollari.
- 

## 2.12 12. Zeros Impact e Zeros2 Impact

### 2.12.1 Descrizione

Le misure Zeros Impact e Zeros2 Impact estendono le misure Zeros e Zeros2 per stimare il *price impact*.

### 2.12.2 Formula

Per Zeros Impact:

$$\text{Zeros Impact} = \frac{\text{Zeros}}{\text{Average Daily Dollar Volume}}.$$

Per Zeros2 Impact:

$$\text{Zeros2 Impact} = \frac{\text{Zeros2}}{\text{Average Daily Dollar Volume}}.$$

### 2.12.3 Implementazione

#### 1. Calcolo delle Misure Zeros e Zeros2:

- Seguire i passaggi descritti nelle misure Zeros e Zeros2.

#### 2. Calcolo dello Zeros Impact:

- Dividere la misura Zeros per il volume medio giornaliero in dollari.

#### 3. Calcolo dello Zeros2 Impact:

- Dividere la misura Zeros2 per il volume medio giornaliero in dollari.