Low Frequency Spread Measurements and Price Impact Proxies

Giulio Surya Lo Verde

November 14, 2024

1 Low Frequency Spread Measurements

1.1 Roll

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

Variables:

- ΔP_t: price change

Roll Formula Implementation:

1. Collect transaction price data for a security over a given period.
2. Calculate daily price changes $\Delta P_t = P_t - P_{t-1}$.
3. Calculate the serial covariance of price changes using:

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

   where $\bar{\Delta P}$ is the mean of the price changes.

4. Substitute the covariance value into Roll's formula to obtain the estimated spread S.

1.2 Effective Tick

$$\text{Effective Tick} = \frac{\sum_{j=1}^{J} g_j s_j}{P}$$

Variables:

- g_j: probability of observing a price corresponding to the spread s_j
- s_j: effective spread

- P: average price over the period

Effective Tick Implementation

The Effective Tick is an effective spread measure based on the frequency at which prices cluster at certain levels (*price clustering*). The main idea is that trading prices tend to cluster at predetermined increments, and the frequency of these price clusters provides an indication of spread size.

1. Step 1: Define Price Clusters For each transaction, classify the price into one of the following price clusters, based on the price system granularity.

   > **Note**
   >
   > This should be s_j.

   - Penny: 0.01
   - Nickel: 0.05
   - Dime: 0.10
   - Quarter: 0.25
   - Dollar: 1.00

   These clusters reflect possible spread sizes in the market.

2. Step 2: Calculate Cluster Frequency For each observed price, assign it to the corresponding cluster and calculate the number of transactions for each cluster N_j. Next, determine the empirical probability F_j for each cluster j:

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

   where J is the total number of clusters.

3. Step 3: Calculate Unconstrained Probabilities Unconstrained probabilities U_j for each cluster are calculated as:

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

4. Step 4: Constrain Probabilities To ensure probabilities are between 0 and 1, apply a constraint:

   $$\hat{g}_j = \begin{cases} \text{Min}\left(\max(U_j, 0), 1\right), & \text{for } j = 1 \\ \text{Min}\left(\max(U_j, 0), 1 - \sum_{k=1}^{j-1} \hat{g}_k\right), & \text{for } 2 \le j \le J \end{cases}$$

   This constraint ensures valid probabilities.

5. Step 5: Calculate Estimated Effective Spread Finally, the estimated effective spread is a weighted average of the spreads for various clusters:

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

where s_j is the cluster size j (e.g., 0.01 for penny, 0.05 for nickel, etc.), and $\overline{P}$ is the average price of the security during the observation period.

1.3 Effective Tick 2

Identical to *Effective Tick*, but calculated using data from all days, not only those with positive volume.

Note

I assume this refers to days when the market is closed.

1.4 Holden

The Holden estimator combines serial correlation and price clustering to provide an accurate measure of effective spread.

The Holden proxy is based on the following general formula for spread:

$$S_H = \sum_{j=1}^{J} g_j s_j$$

where:

- S_H is the estimated spread,
- g_j is the probability associated with possible spread s_j,
- s_j represents the various possible spreads (1/8, 1/4, 1/2, etc.).

1. Define Observed and Unobserved Variables:
    - P_t: observed price of the last trade on day t.
    - S_t: effective spread realized in the last trade of the day, which is to be estimated.
    - Q_t: buy (+1) or sell (-1) indicator for the last trade of the day.
    - C_t: observed price cluster on day t. For example, $C_t = 1$ for a spread of 1/8, $C_t = 2$ for 1/4, etc.
    - ϵ_t: public information shock, normally distributed $\epsilon_t \sim N(0, \sigma_\epsilon^2)$.
2. Determine Price Cluster: Each day, the observed price P_t is mapped to a specific price cluster C_t, associated with one of the possible spreads. For example:

- If the closing price P_t is on a number like 25.125 (an odd eighth), then $C_t = 1$ corresponds to a spread of 1/8.
- If the closing price P_t is a number like 25.25 (quarters), then $C_t = 2$ corresponds to a spread of 1/4.

Holden Implementation

3. Modeling Price Change: The price change between consecutive days t and $t+1$ is described by the following equation, accounting for the informational component and spread:

$$\Delta P_t = \frac{1}{2} S_t Q_t - (1-\lambda)\frac{1}{2} S_{t-1} Q_{t-1} + \epsilon_t$$

where:

- S_t is the effective spread on day t,
- Q_t is the buy (+1) or sell (-1) indicator,
- λ represents the proportion of the spread attributed to adverse selection and inventory costs,
- ϵ_t is the information shock.

4. Calculate the Likelihood Function: Once you have the price change ΔP_t, calculate the likelihood function for the model parameters. In this case, the likelihood is based on the normal distribution of the information shocks ϵ_t, with zero mean and variance σ_ϵ^2. The likelihood function for a triplet of observed prices (P_t, P_{t+1}, P_{t+2}) is:

$$L(P_t, P_{t+1}, P_{t+2} \mid \mu, g_1, g_2, S_H, \sigma_\epsilon, \lambda) = \prod_{(H_t, H_{t+1}, H_{t+2})} \Pr(C_t) \cdot \Pr(H_t \mid C_t) \cdot \mathcal{N}(P_{t+1} \mid P_t, \epsilon_t)$$

where $\mathcal{N}$ denotes the normal distribution density.

5. Maximize the Likelihood Function: Maximize the likelihood function with respect to parameters $\mu, g_1, g_2, S_H, \sigma_\epsilon, \lambda$, subject to constraints $0 \leq g_j \leq 1$ (probabilities must be valid) and $0 \leq \lambda \leq 1$ (for adverse selection).

6. Interpret Optimized Parameters: After estimating the parameters via maximum likelihood, interpret the results:

- S_H: represents the estimated average effective spread.
- λ: indicates the portion of the spread due to adverse selection and inventory costs. A value near 1 suggests that most of the spread is due to these factors.
- g_j: probabilities associated with different observed spreads in price clusters.

Note

My question is: in calculating this likelihood, am I primarily interested in calculating the probabilities associated with the spread levels, right? We don't really need the other parameters? It seems that I have various spread levels, and I weight them by the associated probabilities, just as in Effective Tick, but the probabilities are calculated differently here.

7. Calculate the Final Estimator: The final estimator result is S_H, representing the average effective spread. This value is calculated as a weighted sum of possible spreads s_j, using optimized probabilities g_j:

$$S_H = \sum_{j=1}^{J} g_j s_j$$

1.5 Gibbs Sampler for Spread Estimation

The Gibbs sampler is a Markov Chain Monte Carlo (MCMC) method used to estimate model parameters when some variables are latent, i.e., not directly observed. Here, we estimate the effective spread of a security using observed price data, considering the fundamental price and order direction as latent variables.

Note

The article information isn't enough, so I've loaded Hasbrouck's 2004 paper explaining the Gibbs sampler. The following section is a draft, and everything here needs verification. I wonder if we might need additional papers for other proxies.

Variables:

- P_t: observed price of the security on day t.
- V_t: fundamental (latent) price on day t.
- Q_t: order direction (+1 for buy, -1 for sell).
- S: effective spread (to be estimated).
- ϵ_t: information shock, normally distributed $N(0, \sigma^2)$.

Gibbs Sampler Implementation Steps:

1. **Define Latent Variables**
   Latent variables are the fundamental price V_t and order direction Q_t, which are not directly observable and must be sampled conditionally on the observed data.

2. **Model Observed Price**
   The observed price P_t relates to the fundamental price V_t and spread S as follows:
   $$P_t = V_t + \frac{1}{2} S Q_t$$
   where S is the spread to estimate, and Q_t indicates order direction (buy or sell).

3. **Fundamental Price Evolution**
   The fundamental price V_t evolves as a stochastic process:
   $$V_t = V_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2)$$
   This implies V_t follows a random walk with shock ϵ_t.

4. **Sampling Using the Gibbs Sampler**
   The Gibbs sampler iteratively updates parameters and latent variables by sampling from conditional distributions.

   (a) **Sampling V_t (Fundamental Price)**
   Conditional on observed P_t and Q_t values, sample the fundamental price from the conditional normal distribution:
   $$V_t \sim N\left(P_t - \frac{1}{2} S Q_t, \sigma^2\right)$$

   (b) **Sampling Q_t (Order Direction)**
   Conditional on P_t and V_t, sample the order direction Q_t as follows:
   $$Q_t = \begin{cases} +1, & \text{if } P_t > V_t \quad \text{(buy)} \\ -1, & \text{if } P_t < V_t \quad \text{(sell)} \end{cases}$$
   In case of equality, sample Q_t randomly.

   (c) **Sampling the Spread S**
   After sampling V_t and Q_t, update the spread S using the relationship between observed prices and latent variables:
   $$S \sim N\left(\frac{2(P_t - V_t)}{Q_t}, \tau^2\right)$$
   where τ^2 is a hyperparameter reflecting spread estimation uncertainty.

5. **Iterate Gibbs Sampler**
   Repeat the above steps iteratively for N iterations. During initial iterations (warm-up), discard samples to allow the chain to converge.

6. **Estimate Final Spread**

   After Gibbs sampler convergence, estimate the spread S as the mean of the samples:

   $$\hat{S} = \frac{1}{N} \sum_{i=1}^{N} S_i$$

   where N is the effective number of iterations (excluding initial burn-in iterations).

1.6 LOT Model (Lesmond, Ogden, Trzcinka) Implementation

The LOT model (Lesmond, Ogden, Trzcinka) uses a Limited Dependent Variable (LDV) model to estimate round-trip transaction costs of equity securities. This model posits that security returns depend on market returns but that transaction costs affect the ability of investors to capitalize on information, creating a non-trading range where observed returns are zero.

1. **Data Collection**

   - **Stock Returns (R_{jt})**: Gather daily (or periodic) returns for each stock j over period t.
   - **Market Returns (R_{mt})**: Obtain the corresponding market index returns over the same period t.
   - **Other Variables (if necessary)**: Trading volume, market capitalization, etc., for additional analysis or control.

2. **LDV Model Specification**

   The LDV model assumes that transaction costs create a non-trading zone where observed returns are zero.

   (a) **"True" Return:**

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

   where:

   - R_{jt}^*: Unobserved "true" return of stock j at time t.
   - β_j: Market sensitivity coefficient.
   - ϵ_{jt}: Error term (assumed normally distributed with zero mean and variance σ_j^2).

   (b) **Observed Return:**

1.6.1 Transaction Cost Thresholds in the LOT Model

In the LOT model (Lesmond, Ogden, Trzcinka), the thresholds α_{1j} and α_{2j} represent proportional transaction costs associated with selling and buying stock j, respectively.

Threshold Definition

- α_{1j}: **Lower transaction cost threshold**, associated with **selling** stock j. This is a negative value, i.e., $\alpha_{1j} < 0$.
- α_{2j}: **Upper transaction cost threshold**, associated with **buying** stock j. This is a positive value, i.e., $\alpha_{2j} > 0$.

Model Equations The LDV model specifies the observed return R_{jt} as a function of the unobserved "true" return R^*_{jt} and the transaction cost thresholds:

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

where R^*_{jt} is the unobserved "true" return of stock j at time t, defined as:

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

with:

- β_j: Market sensitivity coefficient for stock j.
- R_{mt}: Market return at time t.
- ϵ_{jt}: Error term (assumed normally distributed with zero mean and variance σ_j^2).

Trading Region Interpretation The thresholds α_{1j} and α_{2j} define three distinct trading regions:

i. **Selling Region** ($R^*_{jt} < \alpha_{1j}$):
   - *Region 2*: $R_{jt} \neq 0$ and $R_{mt} < 0$ (indicating $R^*_{jt} < 0$, the observed return is non-zero, so a trade occurred, and the market is down).
   - *Condition*: The "true" return is sufficiently negative to surpass (in absolute value) the selling transaction cost threshold.
   - *Incentive to Sell*: The investor is incentivized to sell, expecting a significant price drop.
   - *Observed Return*: $R_{jt} = R^*_{jt} - \alpha_{1j}$. Since $\alpha_{1j} < 0$, subtracting α_{1j} makes the observed return less negative than the "true" return.

ii. **No-Trade Region** ($\alpha_{1j} \leq R^*_{jt} \leq \alpha_{2j}$):
   - *Region 0*: $R_{jt} = 0$, no transaction occurs.
   - *Condition*: The "true" return is not large enough (in absolute value) to surpass the transaction cost thresholds for buying or selling.

- *No Incentive to Trade*: The investor is not motivated to buy or sell.
- *Observed Return*: $R_{jt} = 0$.

iii. **Buying Region** ($R_{jt}^* > \alpha_{2j}$):
   - *Region 1*: $R_{jt} \neq 0$ and $R_{mt} > 0$ (indicating $R_{jt}^* > 0$, the observed return is non-zero, so a trade occurred, and the market is up).
   - *Condition*: The "true" return is sufficiently positive to surpass the buying transaction cost threshold.
   - *Incentive to Buy*: The investor is incentivized to buy, expecting a significant price increase.
   - *Observed Return*: $R_{jt} = R_{jt}^* - \alpha_{2j}$, reflecting the transaction cost for buying.

Theoretical Values of Thresholds

- $\alpha_{1j} < 0$: Associated with selling, the threshold is negative because a sufficiently negative "true" return is required to cover the transaction cost of selling.
- $\alpha_{2j} > 0$: Associated with buying, the threshold is positive as it requires a sufficiently positive "true" return to cover the transaction cost of buying.

Total Transaction Cost Calculation The total (round-trip) transaction cost for stock j is given by the difference between the thresholds:

$$\text{Total Transaction Cost} = \alpha_{2j} - \alpha_{1j}$$

This value represents the overall cost for executing a buy and subsequent sale on stock j.

Likelihood Function Construction The likelihood function L for stock j is the product of probabilities associated with observations in each region:

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

where:

- $\phi(\cdot)$: Standard normal density function.
- $\Phi(\cdot)$: Standard normal cumulative distribution function.
- σ_j: Standard deviation of residuals (estimated using only non-zero observed returns).

Note

According to page 11 of the Lesmond paper, "The terms ϕ_1 and ϕ_2 refer to standard normal density functions for decreases and increases in the measured return, respectively," while Φ "is the cumulative distribution function of the standard normal distribution." It seems that Φ is indeed the cumulative function.

Interpretation of Terms

- **First Term** ($t \in \mathcal{R}_1$): Corresponds to observations with negative measured returns (sales), where $R_{jt} \neq 0$.
- **Second Term** ($t \in \mathcal{R}_0$): Represents the probability of a zero measured return (no trading).
- **Third Term** ($t \in \mathcal{R}_2$): Corresponds to observations with positive measured returns (purchases), where $R_{jt} \neq 0$.

Log-Likelihood Function Taking the natural logarithm of the likelihood function yields the log-likelihood function $\ln L$:

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

Parameter Estimation The parameters α_{1j}, α_{2j}, β_j, and σ_j are estimated by maximizing the log-likelihood function $\ln L$ with respect to these parameters.

—

This completes the implementation and description of the LOT model.

i. **Data Collection**

- **Stock Returns (R_{jt})**: Gather daily (or periodic) returns for each stock j over period t.
- **Market Returns (R_{mt})**: Obtain the corresponding market index returns over the same period t.
- **Other Variables (if necessary)**: Trading volume, market capitalization, etc., for additional analysis or control.

1.7 5. Roll Impact

1.7.1 Description

The Roll Impact extends the Roll measure (1984) to estimate the *price impact* based on the serial covariance of price changes.

1.7.2 Derivation

The Roll spread measure is given by:

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

when the covariance is negative.
The Roll Impact is defined as:

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

1.7.3 Implementation

A. **Calculate Covariance**:
   - Compute daily price changes $\Delta P_t = P_t - P_{t-1}$.
   - Calculate the covariance $\text{Cov}(\Delta P_t, \Delta P_{t-1})$.

B. **Calculate Roll Spread**:
   - Apply the Roll spread formula.

C. **Calculate Roll Impact**:
   - Calculate the average daily dollar volume.
   - Divide the Roll spread by the average daily volume.

1.7.4 Interpretation

The Roll Impact estimates the *price impact* based on the relationship between price changes and traded volume.
—

1.8 6. Effective Tick Impact

1.8.1 Description

The Effective Tick Impact extends the Effective Tick spread measure to estimate the *price impact*.

1.8.2 Formula

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

where the Effective Tick Spread is calculated as in the Effective Tick spread measure.

1.8.3 Implementation

A. **Calculate Effective Tick Spread**:
   - Follow the steps described in the Effective Tick spread measure section.

B. **Calculate Effective Tick Impact**:
   - Divide the Effective Tick Spread by the average daily dollar volume.

1.8.4 Interpretation

This measure estimates the *price impact* while accounting for the distribution of price changes.
—

1.9 7. Effective Tick2 Impact

1.9.1 Description

Similar to Effective Tick Impact, but uses all days, including those with zero volume.

1.9.2 Formula

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

where the Effective Tick2 Spread is calculated as in the Effective Tick2 spread measure.

1.9.3 Implementation

A. **Calculate Effective Tick2 Spread**:
   - Follow the steps described in the Effective Tick2 spread measure section.

B. **Calculate Effective Tick2 Impact**:
   - Divide the Effective Tick2 Spread by the average daily dollar volume.

—

1.10 8. Holden Impact

1.10.1 Description

The Holden Impact extends the Holden spread measure to estimate the *price impact*.

1.10.2 Formula

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

where the Holden Spread is calculated according to the Holden measure.

1.10.3 Implementation

A. **Calculate Holden Spread**:
   - Follow the steps described in the Holden spread measure section.

B. **Calculate Holden Impact**:
   - Divide the Holden Spread by the average daily dollar volume.

—

1.11 9. Gibbs Impact

1.11.1 Description

The Gibbs Impact extends the Gibbs spread measure to estimate the *price impact*.

1.11.2 Formula

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

where the Gibbs Spread is calculated using the methodology of Hasbrouck (2004).

1.11.3 Implementation

A. **Calculate Gibbs Spread**:
   - Use the Gibbs sampler to estimate model parameters and obtain the spread.

B. **Calculate Gibbs Impact**:
   - Divide the Gibbs Spread by the average daily dollar volume.

—

1.12 10. LOT Mixed Impact

1.12.1 Description

The LOT Mixed Impact extends the LOT Mixed spread measure to estimate the *price impact*.

1.12.2 Formula

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

where the LOT Mixed Spread is calculated using the methodology of Lesmond, Ogden, and Trzcinka (1999).

1.12.3 Implementation

A. **Calculate LOT Mixed Spread**:
   - Follow the steps described in the LOT Mixed spread measure section.

B. **Calculate LOT Mixed Impact**:
   - Divide the LOT Mixed Spread by the average daily dollar volume.

—

1.13 11. LOT Y-split Impact

1.13.1 Description

The LOT Y-split Impact extends the LOT Y-split spread measure to estimate the *price impact*.

1.13.2 Formula

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

where the LOT Y-split Spread is calculated using the methodology developed in the article.

1.13.3 Implementation

A. **Calculate LOT Y-split Spread**:
   - Follow the steps described in the LOT Y-split spread measure section.

B. **Calculate LOT Y-split Impact**:
   - Divide the LOT Y-split Spread by the average daily dollar volume.

—

1.14 12. Zeros Impact and Zeros2 Impact

1.14.1 Description

The Zeros Impact and Zeros2 Impact measures extend the Zeros and Zeros2 measures to estimate the *price impact*.

1.14.2 Formula

For Zeros Impact:

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

For Zeros2 Impact:

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

1.14.3 Implementation

A. **Calculate Zeros and Zeros2 Measures**:
   - Follow the steps described in the Zeros and Zeros2 measures section.

B. **Calculate Zeros Impact**:
   - Divide the Zeros measure by the average daily dollar volume.

C. **Calculate Zeros2 Impact**:
   - Divide the Zeros2 measure by the average daily dollar volume.

—