FE570 Financial Markets and Trading
Lecture 3. The Roll Model of Trade Prices
(Ref. Joel Hasbrouck - Empirical Market Microstructure)

Steve Yang
Stevens Institute of Technology
9/11/2012
Outline

1. Asset Returns and Distributional Properties of Returns
2. Spread and Liquidity
3. The Random-Walk Model of Security Prices
4. Statistical Analysis of Price Series
5. The Roll Model of Bid, Ask, and Transaction Prices
Asset Returns:

Let $P_t$ be the price of an asset at time $t$, and assume no dividend.

One-period simple return: Gross return

$$1 + R_t = \frac{P_t}{P_{t-1}}$$

or

$$P_t = P_{t-1}(1 + R_t) \quad (1)$$

Simple return:

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (2)$$

Multiperiod simple return: Gross return

$$1 + R_t(k) = \frac{P_t}{P_{t-k}}$$

$$= \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \ldots \times \frac{P_{t-k+1}}{P_{t-k}}$$

$$= (1 + R_t)(1 + R_{t-1})\ldots(1 + R_{t-k+1})$$
**Example**

**Asset Returns Example:** Suppose the closing prices of an asset are

<table>
<thead>
<tr>
<th>Period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>37.84</td>
<td>38.49</td>
<td>37.12</td>
<td>37.60</td>
<td>36.30</td>
</tr>
</tbody>
</table>

- What is the simple return from period 1 to period 2?
  
  Ans: \( R_2 = \frac{38.49 - 37.84}{37.84} = 0.017. \)

- What is the simple return from period 1 to period 5?
  
  Ans: \( R_5(4) = \frac{36.30 - 37.84}{37.84} = -0.041. \)

- Verify that \( 1 + R_5(4) = (1 + R_2)(1 + R_3)\ldots(1 + R_5). \)
Continuously Compounded (or Log) Returns:

- **Log return:**

  \[ r_t = \ln(1 + R_t) = \ln \frac{P_t - P_{t-1}}{P_{t-1}} = p_t - p_{t-1} \]  

  where \( p_t = \ln(P_t) \).

- **Multiperiod log return**

  \[ r_t(k) = \ln[1 + R_t(k)] \]

  \[ = \ln[(1 + R_t)(1 + R_{t-1})...(1 + R_{t-k+1})] \]

  \[ = \ln(1 + R_t) + \ln(1 + R_{t-1}) + ... + \ln(1 + R_{t-k+1}) \]

  \[ = r_t + r_{t-1} + ... + r_{t-k+1}. \]
Example

Log Returns Example: Suppose the closing prices of an asset are

<table>
<thead>
<tr>
<th>Period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>37.84</td>
<td>38.49</td>
<td>37.12</td>
<td>37.60</td>
<td>36.30</td>
</tr>
</tbody>
</table>

- What is the log return from period 1 to period 2?
  Ans: \( r_2 = \ln(38.49) - \ln(37.84) = 0.017. \)

- What is the log return from period 1 to period 5?
  Ans: \( r_5(4) = \ln(36.30) - \ln(37.84) = -0.042. \)

- Is it easy to verify that \( r_5(4) = r_2 + r_3\ldots r_5? \)
Distributional Properties of Returns:

- **Moments of a random variable** $X$ with density $f(x)$: $\ell$ – th moment
  \[
m'_\ell = E(X^\ell) = \int_{-\infty}^{\infty} x^\ell f(x) dx \tag{4}
  \]

- $\ell$ – th central moment
  \[
m_\ell = E[(X - \mu_x)^\ell] = \int_{-\infty}^{\infty} (x - \mu_x)^\ell f(x) dx \tag{5}
  \]

- **First moment**: mean or expectation of $X$
- **Second moment**: variance of $X$
- **Skewness** (symmetry) and **Excess Kurtosis** (fat-tails)

  \[
  S(x) = E\left[\frac{(X - \mu_x)^3}{\sigma_x^3}\right] \tag{6}
  \]

  \[
  K(x) = E\left[\frac{(X - \mu_x)^4}{\sigma_x^4}\right] \tag{7}
  \]
Example

**Properties of Returns:**

- Why are mean and variance of returns important?
  
  Ans: They are concerned with long-term return and risk, respectively.

- Why is symmetry of interest in financial study?
  
  Ans: Symmetry has important implications in holding short or long financial positions and in risk management.

- Why is kurtosis important?
  
  Ans: Related to volatility forecasting, efficiency in estimation and tests, etc. High kurtosis implies heavy (or long) tails in distribution.
Estimation for a given sample data \( \{x_1, ..., x_T\} \):

- **Sample mean:**
  \[
  \hat{\mu}_x = \frac{1}{T} \sum_{t=1}^{T} x_t, \tag{8}
  \]

- **Sample variance:**
  \[
  \hat{\sigma}_x^2 = \frac{1}{T-1} \sum_{t=1}^{T} (x_t - \hat{\mu}_x)^2, \tag{9}
  \]

- **Sample skewness:**
  \[
  \hat{S}(x) = \frac{1}{(T-1)\hat{\sigma}_x^3} \sum_{t=1}^{T} (x_t - \hat{\mu}_x)^3, \tag{10}
  \]

- **Sample kurtosis:**
  \[
  \hat{K}(x) = \frac{1}{(T-1)\hat{\sigma}_x^4} \sum_{t=1}^{T} (x_t - \hat{\mu}_x)^4, \tag{11}
  \]
Normality Test: Under the normality assumption, we have:

\[ \hat{S}(x) \sim N(0, \frac{6}{T}), \hat{K}(x) - 3 \sim N(0, \frac{24}{T}) \]  \hspace{1cm} (12)

- **Test for symmetry:**

\[ S^* = \frac{\hat{S}(x)}{\sqrt{6/T}} \sim N(0, 1), \]  \hspace{1cm} (13)

Description rule: Reject $H_0$ of a symmetric distribution if $|S^*| > Z_{\alpha/2}$ or p-value is less than $\alpha$.

- **Test for tail thickness:**

\[ K^* = \frac{\hat{K}(x) - 3}{\sqrt{24/T}} \sim N(0, 1), \]  \hspace{1cm} (14)

Description rule: Reject $H_0$ of a symmetric distribution if $|K^*| > Z_{\alpha/2}$ or p-value is less than $\alpha$. 
Figure: Simulated E-mini Market Return Distribution.
**Spread:** The size of the bid/ask spread is an important object of the microstructure theory.

- *The quoted spread* between ask $A_t$ and bid $B_t$ that is averaged over $T$ periods equals:

$$S^Q = \frac{1}{T} \sum_{t=1}^{T} (A_t - B_t)$$

- *The average spread* in terms of the asset fundamental price $P^*_t$ is defined as:

$$S = \frac{1}{T} \sum_{t=1}^{T} 2q_t (P_t - P^*_t)$$

where $q_t$ is 1 for buy orders and -1 for sell orders. Since the value of $P^*_t$ is not observable, the *effective spread* in terms of mid-price $M_t = \frac{1}{2}(A_t + B_t)$ is usually used:

$$S^E = \frac{1}{T} \sum_{t=1}^{T} 2q_t (P_t - M_t)$$

- *The realized spread* is applied in post-trade analysis:

$$S^R = \frac{1}{T} \sum_{t=1}^{T} 2q_t (P_t - M_{t+1})$$
**Spread Components:** The bid/ask spread is the price of immediacy of trading.

- The spread incorporates the dealers’ operational costs, such as trading system development and maintenance, clearing and settlement, etc. If dealers are not compensated for their expenses, there is no rationale for them to stay in the business.

- Dealers’ inventory costs contribute to the bid/ask spread, because they must recover their potential losses by widening the spread. Since deals must satisfy order flows on both sides of the market, they maintain inventories of risky instruments (and sometimes undesirable).

- Spread covers the dealers’ risk of trading with counterparts who have superior information about true security value. Informed traders trade at one side of the market and may profit from trading with dealers. This component of the bid/ask spread is called the *adverse-selection* component since dealers confront one-sided selection of their order flow.
Liquidity: is a notion that is widely used in finance, yet it has no strict definition and in fact may have different meanings. Generally, the term liquid asset implies that it can be quickly and cheaply sold for cash. A popular notion defines liquidity as the market’s breadth, depth, and resiliency.

- It means that buying price and selling price of a liquid instrument are close, that is, the bid/ask spread is small. In a deep market, there are many orders from multiple market makers, so that order cancellations and transactions do not affect notably the total order inventory available for trading.

- Market resiliency means that if some liquidity loss does occur, it is quickly replenished by market makers. In other words, market impact has only temporary effect.

- Sometimes inverse liquidity - illiquidity based on the price impact caused by trading volume, is used:

$$ILLIQ = \frac{1}{N} \sum_{k=1}^{N} \frac{|r_k|}{V_k},$$

where $r_k$ and $V_k$ are the return and trading volume at time $k$. 
The Random-Walk Model of Security Prices

- **1900 - Louis Bachelier** The French mathematician Louis Bachelier first documented the idea and provided insights about stock market prices in his Ph.D. dissertation titled "The Theory of Speculation".

- **1953 - Maurice Kendall** A British statistician proposed the Random-Walk hypothesis in his paper titled "The Analytics of Economic Time Series, Part 1: Prices".

- **1964 - Paul Cootner** MIT Sloan professor developed the same idea in his book titled "The Random Character of Stock Market Prices".

- **1965 - Eugene Fama** Eugene Fama at University Chicago further developed the idea in a paper titled "Random Walks In Stock Market Prices", and eventually he infused the idea into the Efficient-Market hypothesis.

**The Random-Walk Model** is no longer considered to be a complete and valid description of a short-term price dynamics, but it nevertheless retains an important role as a model for the fundamental security value. (Martin Weber, Andrew Lo, etc.)
The Random-Walk Model - General Model

- Let $p_t$ denote the transaction price at time $t$, where $t$ indexes regular points of real ("calendar" or "wall-clock") time, for example, end-of-day, end-of-hour, end-of-minute, end-of-second, etc. Because it is unlikely that trades occur exactly at these times, we will approximate these observations by using the prices of the last (most recent) trade, for example, the day’s closing price.

- **The Random-Walk model** (with drift) is defined as:
  
  $p_t = p_{t-1} + \mu + u_t,$

  - where $u_t$, $t = 0, 1, \ldots$ are independently and identically distributed random variables. Intuitively, they arise from new information that bears on the security value.
  - $\mu$ is the expected price change (the drift).
  - The units of $p_t$ are either levels (e.g. dollars per share) or logarithms. The log form is sometimes more convenient because price changes can be interpreted as continuously compounded returns.
The Random-Walk Model - Martingale

- The drift can be dropped in this model in most of the microstructure analysis. When $\mu = 0$, $p_t$ cannot be forecast beyond its most recent value:

$$E[p_{t+1}|p_t, p_{t-1}, ...] = p_t,$$

A process with this property is generally described as a martingale.

One definition of a martingale is a discrete stochastic process $x_t$ where $E|x_t| < \infty$ for all $t$, and $E(x_{t+1}|x_t, x_{t-1}, ...) = x_t$

- Note that expectation in this formulation is conditioned on lagged $p_t$ or $x_t$ that is the history of the process.
- A more general definition of a martingale process involves conditioning on broader information set. The process $x_t$ is a martingale with respect to another (possibly multidimensional) process $z_t$ if $E|x_t| < \infty$ for all $t$ and $E(x_{t+1}|z_t, z_{t-1}, ...) = x_t$
- When the conditioning information is ”all public information”, the conditional expectation is sometimes called the fundamental value or the efficient price of the security.
The Random-Walk Model - Observations

- To define a Random-Walk formally, take independent random variables $Z_1, Z_2, ..., Z_n$, where each variable is either 1 or -1, with a 50% probability for either value, and $S_0 = 0$ and $S_n = \sum_{j=1}^{n} Z_j$.

- The series $S_n$ is called the **simple random walk on $\mathbb{Z}$**.

  - This series (the sum of the sequence of -1s and 1s) gives the distance walked, if each part of the walk is of length one. The expectation $E(S_n)$ of $S_n$ is zero.

  - This follows by the finite additivity property of expectation: $E(S_n) = \sum_{j=1}^{n} E(Z_j) = 0$.

- A Random-Walk is a process constructed as the sum of independently and identically distributed (i.i.d) zero-mean random variables - a special case of martingale.

- In microstructure analysis, transaction prices are usually not martingales, but by imposing economic and statistical structure it is often possible to identify a martingale component of the price.
Suppose we have a sample \( \{ p_1, p_2, ..., p_T \} \) generated from a random walk process.
2D Random Walks

Across time \(0, 1, \ldots, 100\) with initial value of 10

\[1.96 \times \sqrt{t}\]

2D Random Walk (for fun!!!).
Statistical Analysis of Price Series - Empirical Evidence

- By the Random-Walk model (with drift):
  \[ p_t = p_{t-1} + \mu + u_t, \]
  Price changes \( \Delta p_t = p_t - p_{t-1} \) should be i.i.d with mean
  \[ E(u_t) = \mu \]
  and variance \( Var(u_t) = \sigma^2_u \), for which we can calculate the usual estimates. However the empirical evidence shows the following features:

  - **Near-Zero Mean Returns.** In microstructure data samples \( \mu \) is usually small relative to the estimation error of its usual estimate, the arithmetic mean. Zero is, of course, a biased estimate of \( \mu \), but its estimation error will generally be lower than that of the arithmetic mean.

  - **Extreme Dispersion.** The convenient assumption that price changes are normally distributed is routinely violated. Statistical analysis of price changes at all horizons generally encounter sample distribution with fat tails.

  - **Dependence of Successive Observations.** The increments (changes) in random walk should be uncorrelated. In actual samples, the first order autocorrelations of short-run price changes are usually negative.
The Roll Model of Bid, Ask, and Transaction Prices

- Keep the random-walk assumption, and apply it to the efficient price instead of the actual transaction price.

- Denote efficient price by $m_t$, and we assume $m_t = m_{t-1} + u_t$, where $u_t$ are i.i.d. zero-mean random variables.

- All trades are conducted through dealers - the best dealer quotes are bid and ask prices, $b_t$ and $a_t$, i.e. if a customer wants to buy, he/she must pay the dealer’s ask price (therefore lifting the ask); if a customer wants to sell, he/she receives the dealer’s bid price (hitting the bid).

- Dealer incur a cost of $c$ per trade. This charge reflects costs like clearing fees and per trade allocations of fixed costs.

- If dealers compete to the point where the costs are just covered, the bid and ask are $m_t - c$ and $m_t + c$, respectively.
The Roll Model - Two Parameters

- The bid-ask spread is $a_t - b_t = 2c$, a constant.

- At time $t$, there is a trade at transaction price $p_t$, which may be expressed as: $p_t = m_t + q_t c$
  where $q_t$ is a trade direction indicator set to +1 if the customer is buying and -1 if the customer is selling.

- We also assume that buys and sells are equally likely, serially independent (a buy this period does not change the probability of a buy next period), and that agents buy or sell independently of $u_t$ (a customer buy or sell is unrelated to the evolution of $m_t$).

- The Roll model has two parameters, $c$ and $\sigma_u^2$. These are most conveniently estimated from the variance and first-order autocovariance of the price changes, $\Delta p_t$:
  - The variance $\gamma_0 \equiv F(c, \sigma_u^2)$
  - The first order autocovariance $\gamma_1 \equiv G(c, \sigma_u^2)$
The Roll Model - Variance and Auto-covariance

- The variance:

\[ \gamma_0 \equiv \text{Var}(\Delta p_t) = E[(\Delta p_t)^2] - (E[\Delta p_t])^2 \]

\[ = E[q_{t-1}^2 c^2 + q_t^2 c^2 - 2q_{t-1}q_t c^2 - 2q_{t-1}u_t c + 2q_t u_t c + u_t^2] \]

\[ = 2c^2 + \sigma_u^2. \]

In the expectation of this equation all of the cross-products vanish except for those involving \( q_t^2, q_{t-1}^2, \) and \( u_t^2. \)

- The first-order auto-covariance:

\[ \gamma_1 \equiv \text{Cov}(\Delta p_{t-1}, \Delta p_t)^2 = E(\Delta p_{t-1}\Delta p_t) \]

\[ = E[c^2(q_{t-2}q_{t-1} - q_{t-1}^2 - q_{t-2}q_t + q_{t-1}q_t) + c(q_t u_{t-1} - q_{t-1}u_{t-1} + u_t q_{t-1} - u_t q_{t-2})] = -c^2. \]

- It can be easily verified that all autocovariances of order 2 or higher are zero.
The Roll Model - Estimate of Model Parameters

- From the equations obtained earlier, we can compute the model parameters as:

\[ c = \sqrt{-\gamma_1} \]  \hspace{1cm} (15)
\[ \sigma_u^2 = \gamma_0 + 2\gamma_1 \]  \hspace{1cm} (16)

- Given a sample data, it is sensible to estimate \( \gamma_0 \) and \( \gamma_1 \) and apply these transformations to obtain estimate of the model parameters.

- Example: Based on all trades for PCO in October 2003, the estimated first-order autocovariance of the price changes is \( \hat{\gamma}_1 = -0.0000294 \). This implies \( c = $0.017 \) and a spread of \( 2c = $0.034 \).

The Roll model is often used in situations where we don’t possess bid and ask data. In this example, we do know that the time-weighted average NYSE spread in the sample is $0.032, so the Roll estimate is fairly close.