Statistical correlations in high-frequency financial trading

Yang WU
Lady Margaret Hall (2013)
Why traders need correlations?

**Strategy 1:** Toss a coin?

Good strategy if
1. lady luck is with you; *(well... hmm....)*
2. the market is trending upwards.

In midst of the financial crisis right now, I think ... ... I need to be **better than being a gambler**!

**Q:** Up or down?

**Strategy 2:** Analyze and make predictions.

**Fundamental analysis**
1. look out for related news, reports, ... ;
2. hope that future price will correlate/anti-correlate accordingly;

**Statistical/ Technical analysis**
1. look at historical pricing data ... ;
2. hope that future price will correlate according to some models in mind...
Statistical/ Technical analysis

Essential Concepts

1) What determines the **price** and **return**?

---

*Mid-pricing scheme:*

\[ m_X(t) = \frac{A_X(1, t) + B_X(1, t)}{2} \]

*Weighted-pricing scheme:*

\[ w_X(t) = \frac{A_X(1, t)B_X^{q}(1, t) + B_X(1, t)A_X^{q}(1, t)}{A_X(1, t) + B_X^{q}(1, t)} \]

Price denoted in general by \( x(t) \)

Return

\[ r(t) = x(t) - x(t-1) \]

Time unit is important!

1-second return?
1-minute return?
1-day return?
Statistical correlations in high-frequency financial trading

Statistical/ Technical analysis

Essential Concepts

2) An example of a 2-hr USD/JPY exchange rate.

volatility $\sigma_x(t)$:
- standard deviation of $\{x(t) \mid t \text{ domain}\}$;
- definitely dependent on time-domain;
- time unit is important
- more specifically to be realized/ integrated/ historical volatility

recent data have more relevance:
- history should have more weighting in near past;
- moving average schemes (EMA, GMA, or ... )
- see thesis appendix.

The FX market is an typical example of a high frequency market, i.e.
1. large nr. of changes in order book over small period of time;
2. very liquid such that trades modify minimally on price;
3. huge volume (lots of transactions, “honest market”)
Statistical/ Technical analysis

Essential Concepts

3) Stochastic pricing model:

\[ dX_t = \mu_t dt + \sigma_t dW_t \]

- i. instantaneous volatility \( \sigma_t \);  
- ii. many models in estimating \( \sigma_t \), e.g. ARCH, GARCH, ...  
- iii. this is geometric brownian motion (GBM)

4) Covariance matrix (2 assets):

\[ \Sigma = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} \]

Objective: Estimate the correlation \( \rho = \frac{\sigma_{ab}}{\sigma_a \sigma_b} \).

Correlation gives traders a strategic sense of direction!
5) Single derivative:

1. Before its expiry date, e.g. 14th Feb 2014, this derivative (contract) must have a value

\[ V = V(S,t) \]

(This derivative can be traded in markets!)

2. Ito’s lemma:

\[ dV = \sigma S \frac{\partial V}{\partial S} dW_t + \left( \mu S \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t} \right) dt \]

3. Delta hedging:

\[ \Pi = V - \Delta S \quad \left( \Delta = \frac{\partial V}{\partial S} \right) \]

By hedging this derivative with delta \( \Delta \) units of DJI, risks due to randomness vanish!

4. Black-Scholes Model:

\[ \frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0 \]

\textit{implied volatility} best interest rate in the market (no arbitrage principle)
6) Many derivatives:

Suppose I own 3 derivatives with correlated underlying ...

\[
V_1 = V_1(S_1, S_2, S_3, t) \\
V_2 = V_2(S_1, S_2, S_3, t) \\
V_3 = V_3(S_1, S_2, S_3, t)
\]

Similarly,

\[
dV_i = \sum_{\alpha} \frac{\partial V_i}{\partial S_\alpha} dS_\alpha + \frac{1}{2} \sum_{\alpha\beta} \frac{\partial^2 V_i}{\partial S_\alpha \partial S_\beta} dS_\alpha dS_\beta
\]

where (asset pricing model):

\[
\frac{dS_\alpha}{S_\alpha} = \sum_k \sigma_{\alpha k} dW_k + \mu_{\alpha} dt
\]

Hedging portfolio:

\[
\Pi = \sum_i V_i - \sum_{ij} \Delta_{ij} S_j - \sum_{ijk} \Gamma_{ijk} S_j S_k
\]

Delta hedging:

\[
\Delta_{ij} = \frac{\partial V_i}{\partial S_j}
\]

Gamma hedging:

\[
\Gamma_{ijk} = \frac{\partial^2 V_i}{\partial S_j \partial S_k}
\]

Hedging is an another way to minimize risks.
**Statistical/ Technical analysis**

**Volatility estimate**

USD/ JPY Hourly Av. Volatility (100ms tick)

- August 2009
- September 2009
- October 2009

Market is imperfect:

\[ Y_i = X(t_i) + \epsilon_i \]

Instantaneous return:

\[ R_i = Y_i - Y_{i-1} = X(t_i) - X(t_{i-1}) + \epsilon_i - \epsilon_{i-1} \]

Variance:

\[ \text{Var}(R_i) = \sigma^2 T(i) + 2\eta^2 \]

Unbiased estimate:

\[ \hat{\sigma}^2 = n^{-1} \sum_i \left( \frac{R_i^2}{T(i) + 2\eta^2/\sigma^2} \right) \]

Volatility estimate is dependent on volume effects!
Statistical/ Technical analysis

Correlation for non-synchronous data

(Dots indicate trading events.)

\[ X_A(t_i) \]

\[ X_B(s_j) \]

Asynchronous trading

Return:

\[ Y_a(i) = X_a(t_i) - X_a(t_{i-1}), \quad i = 1, \ldots, n \]
\[ Y_b(j) = X_b(s_j) - X_b(s_{j-1}), \quad j = 1, \ldots, m \]

Motivations:

1. Treat each trading event individually
2. Assume no lagged correlations among returns Y.

A simpler illustration:

1. Asset A: 3 trading events at time t1, t2, t3
   Asset B: 3 trading events at time s1, s2, s3
2. 6 individual returns, i.e. \{Y_A(t1), Y_A(t2), Y_A(t3), Y_B(s1), Y_B(s2), Y_B(s3)\}

\[ V = \begin{pmatrix}
\sigma_A^2 T(t_1) & 0 & 0 \\
0 & \sigma_A^2 T(t_2) & 0 \\
0 & 0 & \sigma_A^2 T(t_3)
\end{pmatrix} \quad \begin{pmatrix}
\sigma_{AB} \tau_{11} & \sigma_{AB} \tau_{12} & \sigma_{AB} \tau_{13} \\
\sigma_{AB} \tau_{21} & \sigma_{AB} \tau_{22} & \sigma_{AB} \tau_{23} \\
\sigma_{AB} \tau_{31} & \sigma_{AB} \tau_{32} & \sigma_{AB} \tau_{33}
\end{pmatrix}
\]

\[ Y = \begin{pmatrix}
Y_A(t_1) \\
Y_A(t_2) \\
Y_A(t_3) \\
Y_B(s_1) \\
Y_B(s_2) \\
Y_B(s_3)
\end{pmatrix} \]

Time interval overlap \( \tau_{33} \)

\[ \tau_{13} = \tau_{21} = \tau_{23} = \tau_{31} = 0 \]

Of course, in this example:
Statistical/ Technical analysis

Correlation for non-synchronous data

Using the same example:

Method-of-moment estimate:

\[
\hat{\sigma}_A^2 = \frac{1}{n} \sum_{n=1}^{3} \frac{Y_A^2(t_n)}{T(t_n)}
\]

\[
\hat{\sigma}_B^2 = \frac{1}{m} \sum_{m=1}^{3} \frac{Y_B^2(s_m)}{T(s_m)}
\]

\[
\hat{\sigma}_{AB} = \frac{\sum_{n,m} w_{nm} Y_A(t_n) Y_B(s_m)}{\sum_{n,m} w_{nm} T(t_n)}
\]

\[
\hat{\rho} = \frac{\hat{\sigma}_{ab}}{\hat{\sigma}_a \hat{\sigma}_b}
\]

1. What is the weight function \( w_{nm} \)?
2. Best determined by maximum likelihood estimation ...

Maximum likelihood estimate (MLE):

\[
L(\theta) = \frac{\exp \left( -\frac{1}{2} y^T V^{-1} y \right)}{(2\pi)^{(m+n)/2} \sqrt{\det(V)}}
\]

Optimize \( L(\theta) \) w.r.t. \( \theta \), such that \( \theta = (\sigma_a^2, \sigma_b^2, \sigma_{ab}) \)
either numerically or analytically (see thesis)

Our ultimate objective is to estimate correlation.
Statistical/Technical analysis

Trending or mean-reverting?

If we were God, we would see the future to be trending or mean-reverting ...

But (un)fortunately we are not, we need to guess the future ...

Q: Can the nr. of local maxima tell me whether it trends?
A: Yes ... within statistical confidence

\[ E(T_n) = \frac{2(n - 2)}{3} \quad \text{Var}(T_n) = \frac{16n - 29}{90} \]

1. Analytical proof - see thesis.
2. Numerically confirmed by Monte Carlo simulations

\[ n : \text{nr. of tick points} \quad T_n : \text{nr. of local maxima} \]
Jumps affect hedging strategies

An example of hedging strategy model:

\[ \Pi = \sum_i V_i - \sum_{ij} \Delta_{ij} S_j - \sum_{ijk} \Gamma_{ijk} S_j S_k \]

1. This hedging strategy model assumes continuity in price.
   - how to handle jumps?

2. How long after the jump can we trust this continuous model again?

3. How to hedge in close vicinity to the jump?
Thank you for your attention!