Pairs trading / Avellaneda & Lee paper

Let P_t & Q_t be time series for two stocks that are assumed to be correlated.

The price evolution relations b/w P & Q are:

(1) $$\log \frac{P_t}{P_{t_0}} = \alpha(t-t_0) + \beta \log \frac{Q_t}{Q_{t_0}} + \underbrace{X_t}_{\text{stationary/mean-reverting}}$$

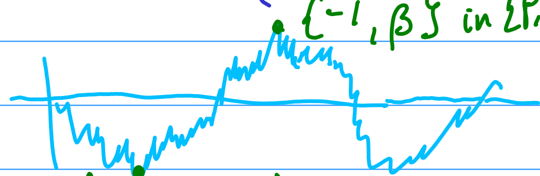
(2) $$\frac{dP_t}{P_t} = \underbrace{\alpha\, dt}_{\text{small}} + \beta \frac{dQ_t}{Q_t} + dX_t$$

$$\frac{\log P_{t_0} + \Delta P_t}{\log P_{t_0}} = \frac{\log(P_{t_0}+\Delta P_t) - \log P_{t_0}}{}$$

$$= \log P_{t_0}\left(1+\frac{\Delta P_t}{P_{t_0}}\right) - \log P_{t_0}$$

$$= \log P_{t_0} + \underbrace{\log\left(1+\frac{\Delta P_t}{P_{t_0}}\right)}_{\approx \frac{dP_t}{P_{t_0}}} - \log P_{t_0}$$

$\{-1, \beta\}$ in $\{P, Q\}$



Want to invest: $\{1, -\beta\}$ in $\{P, Q\}$

$$dX_t = \frac{dP_t}{P_t} - \beta \frac{dQ_t}{Q_t}$$

$(\alpha\, dt \ll dX_t)$

Model dX_t as an Ornstein-Uhlenbeck process:

only non-trivial process that's
- stationary
- Gaussian
- Markovian

$$dX_t = k(m - X_t)\,dt + \sigma\, dW_t$$

To solve OU SDE, use the following trick:

$$Y_t := X_t - m \Rightarrow dY_t = -k Y_t\, dt + \sigma\, dW_t$$

$$Z_t := e^{kt} Y_t \Rightarrow dZ_t = k e^{kt} Y_t\, dt + e^{kt} dY_t$$

$$= k e^{kt} Y_t\, dt + e^{kt}(-k Y_t\, dt + \sigma\, dW_t)$$

$$= e^{kt} \sigma\, dW_t$$

Integrating, we get $Z_t = Z_0 + \sigma\int_0^t e^{ks}\,dW_s$

(3) $\rightarrow$ $\Rightarrow X_{t_0+\Delta t} = m + e^{-k\Delta t}(X_{t_0} - m) + \sigma\int_{t_0}^{t_0+\Delta t} e^{-k(t_0+\Delta t - s)}\,dW_s$

$I := \int f(s)\,dW_s \sim \mathcal{N}\left(0, \underbrace{\mathbb{E}\left(\int f(s)\,dW_s\right)^2}\right) \rightarrow \int \mathbb{E}(f(s))^2\,ds$

Here, $\operatorname{Var} I = \sigma^2\int_0^{\Delta t} \mathbb{E}\, e^{-2k(\Delta t - s)}\,ds = \sigma^2 e^{-2k\Delta t}\int_0^{\Delta t} e^{2ks}\,ds$

$= \frac{1 - e^{-2k\Delta t}}{2k}\sigma^2$

* letting $\Delta t \to \infty$ we see that the equilibrium distn of $X_t \sim \mathcal{N}(m, \underbrace{\frac{\sigma^2}{2k}}_{:=\sigma_{eq}})$

Thus, the exact OU solution is

(4) $X_t = m + \underbrace{e^{-k\Delta t}}_{:=b}(X_{t_0} - m) + \xi_t$ $\quad \xi_t \sim \mathcal{N}\left(0, \frac{1 - e^{-2k\Delta t}}{2k}\sigma^2\right)$

$a := m(1 - b) \qquad \operatorname{Var}\xi = \sigma^2\frac{1 - b^2}{2k}$

$\Rightarrow X_t = a + bX_{t-1} + \xi_t \quad \leftarrow$ AR(1) model

$\xi_t \sim \mathcal{N}\left(0, \sigma^2\frac{1 - b^2}{2k}\right)$

We want to estimate:

(a) Correlation parameter β

(b) AR(1) model parameters

For (a), observe that in Eq. 2,

$$\frac{dP_t}{P_t} = \frac{P_t - P_{t-1}}{P_t} \approx \frac{P_t - P_{t-1}}{P_{t-1}} := R_t \quad \text{stock return}$$

Rewrite (2) as $R_n^S = \beta_0 + \beta R_n^I + \epsilon_n, \quad n = 1, \dots, 60$

(Will run this regression over a 60-day window)

$$\beta_0 = \alpha \Delta t \Rightarrow \alpha = \beta_0 \frac{1}{\Delta t} = \beta_0 \cdot 252 \quad (t \text{ measured in years})$$

The process $X_t = \int_0^t dX_s$. Here, in our discretization scheme, we write

(5) $$X_t = \sum_{j=1}^{k} \epsilon_j \, , \; k = 1, \dots, 60$$

* Note that by construction, $X_{60} = 0$ since the sum of the residuals $= 0$ over the whole 60-day regr. window.

For (b), we use the AR(1) result above:

$$b := e^{-k\Delta t} \qquad a := m(1-b)$$

$$X_t = a + b X_{t-1} + \zeta_t$$

$$\zeta_t \sim N\left(0, \sigma^2 \frac{1-b^2}{2k}\right)$$

From the AR(1) parameters $\{a, b, \operatorname{Var} \zeta\}$ we can get the OU param. estimates like so:

(6)
$$b := e^{-k\Delta t}$$

$$a := m(1-b)$$

$$\operatorname{Var} \zeta_t = \sigma^2 \frac{1-b^2}{2k}$$

$$\operatorname{Var} \zeta_t \Big|_{b=0} := \sigma_{eq}^2 = \frac{\sigma^2}{2k}$$

$$k = 252 \cdot (-\log b)$$

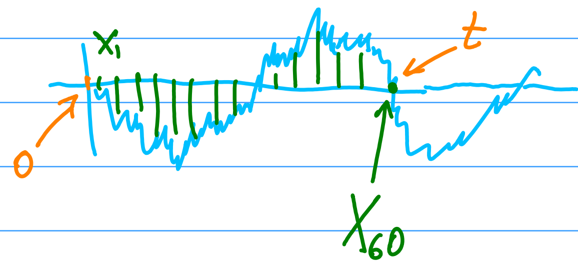
$$\mu = \frac{a}{1-b}$$

$$\sigma = \sqrt{\frac{\operatorname{Var} \zeta \cdot 2k}{1-b^2}}$$

$$\sigma_{eq} = \sqrt{\frac{\operatorname{Var} \zeta}{1-b}}$$

Trading signals are based on the normalized deviation of the estimated O-U process X_t from its estimated mean; the normalization is by the equilibrium sdev σ_{eq}:

$$s := \frac{X_t - m}{\sigma_{eq}} = -\frac{m}{\sigma_{eq}}$$



$$\downarrow$$

$$X_t = X_{60} = 0$$

(7) $$\Rightarrow s = -\frac{a\sqrt{1-b^2}}{(1-b)\sqrt{\mathrm{Var}\,\zeta}}$$

* Avellaneda & Lee recommend using "centered means" (averaged over stocks), so s becomes

$$s = -\frac{\bar{m}}{\sigma_{eq}}, \quad \text{where } \bar{m} = m - \langle m \rangle$$

avg. over stocks

$$= \underbrace{\frac{a}{1-b} - \left\langle \frac{a}{1-b} \right\rangle}$$

(8) $$s = \left(\quad \right)\left(-\sqrt{\frac{1-b^2}{\mathrm{Var}\,\zeta}} \right)$$

To do pairs trading based on sector ETFs, the process is:

→ For every stock, pick its sector ETF, get the β & residuals.

→ For the 60-term residual series, fit AR(1)

→ compute the s-score from (8).