Mean Reversion with a Variance Threshold (ICML 2013)

Marco Cuturi, Alexandre d'Aspremont

Arthur Gretton's notes

July 26, 2013

Problem setup

We are given samples $x_t \in \mathbb{R}^n$.
We want to find a projection of these samples, $y^\top x$, which:

- has high variance,
- is likely to "revert to its mean" (*tricky to define*)

Application: in finance, each dimension of x_t can be a time varying signal (eg a stock).
If a linear combination of the signals has both the above properties, you can profit when the signal is far from its mean.

Problem setup

We are given samples $x_t \in \mathbb{R}^n$.
We want to find a projection of these samples, $y^\top x$, which:

- has high variance,
- is likely to "revert to its mean" (*tricky to define*)

Application: in finance, each dimension of x_t can be a time varying signal (eg a stock).
If a linear combination of the signals has both the above properties, you can profit when the signal is far from its mean.
Real application: some nice math.

First proxy for mean reversion: predictability

First proxy for mean reversion: "predictability".
Univariate case

$$x_t = \hat{x}_t + \epsilon_t,$$

where $\hat{x}_t$ is the prediction, and noise is Gaussian i.i.d.

$$\underbrace{\mathbf{E}(x_t^2)}_{\sigma^2} = \underbrace{\mathbf{E}(\hat{x}_t^2)}_{\hat{\sigma}^2} + \mathbf{E}(\epsilon_t^2).$$

Define

$$\lambda = \frac{\hat{\sigma}^2}{\sigma^2}.$$

When this is close to zero, the observations are dominated by Gaussian noise.

First proxy for mean reversion: predictability

Predictability for multivariate case: define k-lag autocovariance

$$\mathcal{A}_k = \mathbf{E}(x_t x_{t+k}^\top),$$

with empirical estimate A_k. We want the projection $y^\top x$ with lowest predictability (closest to white noise).

First proxy for mean reversion: predictability

Predictability for multivariate case: define k-lag autocovariance

$$\mathcal{A}_k = \mathbf{E}(x_t x_{t+k}^\top),$$

with empirical estimate A_k. We want the projection $y^\top x$ with lowest predictability (closest to white noise).
Solve

$$y^* = \operatorname{argmin} \lambda(y) = \operatorname{argmin} \frac{y^\top \widehat{\mathcal{A}}_0 y}{y^\top \mathcal{A}_0 y}.$$

How do we compute the prediction covariance $\widehat{\mathcal{A}}_0$?

First proxy for mean reversion: predictability

Assume a p-th order autoregressive process (model),

$$\hat{x}_t = \sum_{k=1}^{p} \mathcal{H}_k x_{t-k}.$$

For the $p = 1$ case (for $p > 1$, just reparametrize),

$$\widehat{\mathcal{A}}_0 = \mathcal{H}_1 \mathcal{A}_0 \mathcal{H}_1^\top \qquad \mathcal{A}_1 = \mathcal{A}_0 \mathcal{H}_1.$$

By Yule-Walker, *empirical estimate* H_1 is

$$H_1 = A_0^{-1} A_1.$$

Making this substitution,

$$y^* = \operatorname{argmin} \lambda(y) = \operatorname{argmin} \frac{y^\top \left(A_1 A_0^{-1} A_1^\top \right) y}{y^\top A_0 y}.$$

Second proxy: portmanteau criterion

A second proxy for mean reversion is the portmanteau criterion,

$$\phi_p(y) = \frac{1}{p} \sum_{i=1}^{p} \left(\frac{y^\top \mathcal{A}_i y}{y^\top \mathcal{A}_0 y} \right)^2 .$$

This is zero for white noise. Hence we try to minimise this statistic over y.

Third proxy: crossing statistics

A third proxy for mean reversion is the expected frequency of the time series crossing the zero axis,

$$\gamma(x) = \mathbf{E}\left(\frac{\sum_{t=2}^{T} \mathbb{I}_{x_t x_{t-1} \leq 0}}{T-1}\right).$$

Given

$$x_t = a x_{t-1} + \epsilon_t$$

then

$$\gamma(x) = \frac{\mathrm{arcos}(a)}{\pi}.$$

Thus: *minimize* first order autocorrelation,

$$y^\top \mathcal{A}_1 y,$$

while ensuring all remaining $\left|y^\top \mathcal{A}_k y\right|$, $k > 1$ are small (so first order approximation is valid).

Optimization problem for predictability (1st)

To minimize predictability, an optimization problem is

$$\begin{array}{ll} \text{minimize} & y^\top A_1 A_0^{-1} A_1^\top y \\ \text{subject to} & y^\top A_0 y \geq \nu \\ & \|y\|_2 = 1 \end{array}$$

Second constraint imposes minimum variance. Third constraint is to avoid effects of scaling.

Optimization problem for predictability (1st)

To minimize predictability, an optimization problem is

$$\begin{array}{ll} \text{minimize} & y^\top A_1 A_0^{-1} A_1^\top y \\ \text{subject to} & y^\top A_0 y \geq \nu \\ & \|y\|_2 = 1 \end{array}$$

Second constraint imposes minimum variance. Third constraint is to avoid effects of scaling.
Matrix version: define $yy^\top = Y$. Then solve

$$\begin{array}{ll} \text{minimize} & \mathrm{tr}(A_1 A_0^{-1} A_1^\top Y) \\ \text{subject to} & \mathrm{tr}(A_0 Y) \geq \nu \\ & \mathrm{tr}(Y) = 1,\ \mathrm{rank}(Y) = 1,\ Y \succeq 0 \end{array}$$

Optimization problem for predictability (1st)

Semidefinite relaxation: for $Y \in S_n$ (positive definite cone),

$$\begin{array}{ll} \text{minimize} & \mathrm{tr}(A_1 A_0^{-1} A_1^\top Y) \\ \text{subject to} & \mathrm{tr}(A_0 Y) \geq \nu \\ & \mathrm{tr}(Y) = 1, \; Y \succeq 0. \end{array}$$

Optimization problem for predictability (1st)

Semidefinite relaxation: for $Y \in S_n$ (positive definite cone),

$$\begin{array}{ll} \text{minimize} & \mathrm{tr}(A_1 A_0^{-1} A_1^\top Y) \\ \text{subject to} & \mathrm{tr}(A_0 Y) \geq \nu \\ & \mathrm{tr}(Y) = 1, \; Y \succeq 0. \end{array}$$

From Brickman (1961),

$$\left\{ \left(y^\top A y, y^\top B y \right) \; : \; y \in \mathbb{R}^n, \|y_2\| = 1 \right\}$$
$$= \{(\mathrm{tr}(AY), \mathrm{tr}(BY)) \; : \; Y \in S_n, \; \mathrm{tr}(Y) = 1\}$$

Hence solution Y^* of the semidefinite relaxation can be written $y^* y^{*\top}$.

Optimization problem for portmanteau (2nd)

For portmanteau statitic, an optimization problem is

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{p} (y^\top A_i y)^2 \\ \text{subject to} & y^\top A_0 y \geq \nu \\ & \|y\|_2 = 1 \end{array}$$

Optimization problem for portmanteau (2nd)

For portmanteau statitic, an optimization problem is

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{p} (y^\top A_i y)^2 \\ \text{subject to} & y^\top A_0 y \geq \nu \\ & \|y\|_2 = 1 \end{array}$$

Consider simple case $p = 1$. Replace objective by $\left|y^\top A_1 y\right|$.
Then by Brickman, semidefinite relaxation gives exact solution,

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \mathrm{tr}(A_1 Y) \leq t \\ & \mathrm{tr}(A_1 Y) \geq -t \\ & \mathrm{tr}(A_0 Y) \geq \nu \\ & \mathrm{tr}(Y) = 1,\ Y \succeq 0. \end{array}$$

Optimization problem for portmanteau (2nd)

Portmanteau statitic for $p > 1$ is

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{p} (y^\top A_i y)^2 \\ \text{subject to} & y^\top A_0 y \geq \nu \\ & \|y\|_2 = 1 \end{array}$$

We have the following semidefinite prorgram:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{p} \mathrm{tr}(A_i Y)^2 \\ \text{subject to} & \mathrm{tr}(A_0 Y) \geq \nu \\ & \mathrm{tr}(Y) = 1, \; Y \succeq 0. \end{array}$$

Optimization problem for portmanteau (2nd)

Portmanteau statitic for $p > 1$ is

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{p} (y^\top A_i y)^2 \\ \text{subject to} \quad & y^\top A_0 y \geq \nu \\ & \|y\|_2 = 1 \end{aligned}$$

We have the following semidefinite prorgram:

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{p} \mathrm{tr}(A_i Y)^2 \\ \text{subject to} \quad & \mathrm{tr}(A_0 Y) \geq \nu \\ & \mathrm{tr}(Y) = 1, \; Y \succeq 0. \end{aligned}$$

By e.g. Ben-Tal et al (2009),

$$\mathrm{SDP} \leq \mathrm{OPT} \leq \mathrm{SDP}\, c \log p,$$

Optimization problem for crossing (3rd)

Crossing statistic gives optimization problem

$$\begin{array}{ll} \text{minimize} & y^\top A_1 y + \mu \sum_{i=2}^{p} (y^\top A_i y)^2 \\ \text{subject to} & y^\top A_0 y \geq \nu \\ & \|y\|_2 = 1 \end{array}$$

and semidefinite program

$$\begin{array}{ll} \text{minimize} & \mathrm{tr}(A_1 Y) + \sum_{i=2}^{p} \mathrm{tr}(A_i Y)^2 \\ \text{subject to} & \mathrm{tr}(A_0 Y) \geq \nu \\ & \mathrm{tr}(Y) = 1,\ Y \succeq 0. \end{array}$$

Same upper and lower bound-style guarantees as Portmanteau.

"If this works so well then why aren't you rich?"

"If this works so well then why aren't you rich?"

Experiments compare against three methods that don't maximize variance.