Notes on Mean Reverting Process

Mean reversion models and stochastic differential equations

Equity models are often based on geometric Brownian motion, but this model tends to be inadequate for modeling prices of commodities (like oil) because, while they exhibit random fluctuations, they seem to fluctuate around a long-term price, a phenomenon that has been explained using economic principles. For this reason, mean-reverting processes have been found to be useful. These models have applications beyond finance to the the physical sciences.

A standard model for mean-reversion is a random process X_t that starts at the value X_s at time s and whose evolution is governed by the a stochastic differential equation of the form

$$dX_t = \theta(\mu - X_t)dt + \sigma dB_t,$$

where B is standard Brownian motion and μ, $\theta > 0$ and $\sigma > 0$ are parameters that may or may not be known.

To see why this process is mean-reverting (with μ as the long-run mean) observe that as X_t becomes larger than μ we tend to get dX_t negative, which pulls the process down, while X_t becoming smaller than μ makes dX_t positive, which pulls the process up.

Generating realizations

For the purpose of doing Monte-Carlo simulation of such a process, we need a method of generating realizations. A simple and direct way of simulating a realization of this process at equally spaced time points is to pick some small time increment $\Delta > 0$, initialize X_s and then define $X_{s+\Delta}$, $X_{s+2\Delta}$, $X_{s+3\Delta}, \ldots$ iteratively by taking

$$X_{s+(j+1)\Delta} = X_{s+j\Delta} + \theta(\mu - X_{s+j\Delta})\Delta + \sigma\delta_j\sqrt{\Delta}$$

where the δ_j are independent random variables, each distributed $N(0, 1)$.

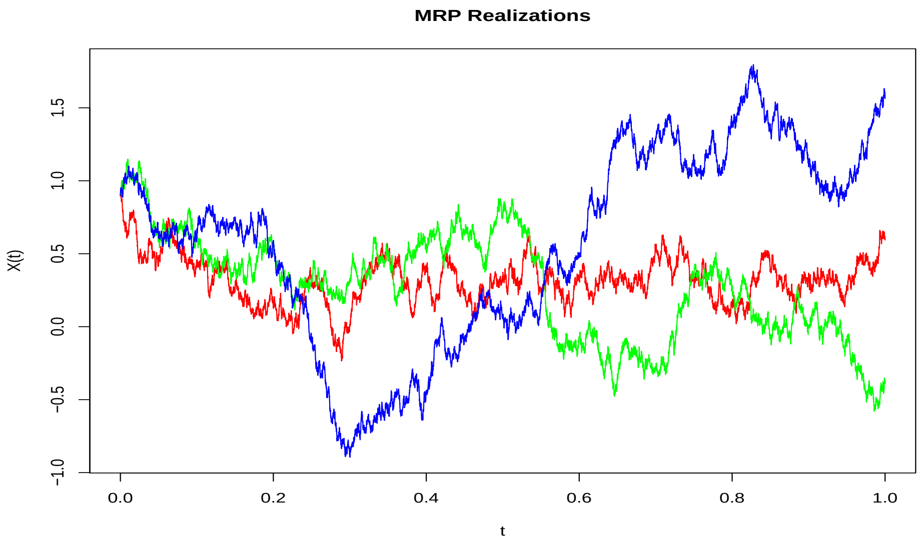

Figure 1: Three realizations of the MRP starting at 0.9 at time 0 with $\mu = 1.$, $\theta = 0.8$ and $\sigma = 1.3$.

Deriving the conditional distribution of X_t given X_s for some $s < t$

There is another way to generate a realization in which we only need to consider time points we're interested. We can derive an interesting and important property of this process for the purpose of inference as follows.

Observe that an application of Ito's lemma gives

$$d(e^{\theta t} X_t) = \theta e^{\theta t} X_t dt + e^{\theta t} dX_t$$

and substituting the expression for dX_t from above we obtain

$$d(e^{\theta t} X_t) = \theta e^{\theta t} X_t dt + e^{\theta t} \left\{ \theta(\mu - X_t) dt + \sigma dB_t \right\}.$$

Simplifying the right-hand side gives

$$d(e^{\theta t} X_t) = \mu \theta e^{\theta t} dt + \sigma e^{\theta t} dB_t$$

Consequently,

$$e^{\theta t} X_t - e^{\theta s} X_s = \int_{u=s}^{t} d(e^{\theta u} X_u)$$
$$= \int_{u=s}^{t} e^{\theta u} \theta \mu du + \int_{u=s}^{t} \sigma e^{\theta u} dW_u.$$

$$= \mu(e^{\theta t} - e^{\theta s}) + \sigma \int_{u=s}^{t} e^{\theta u} dW_u,$$

and so, solving for X_t we can write

$$X_t = e^{-\theta(t-s)} X_s + \mu(1 - e^{\theta-(t-s)}) + \sigma e^{-\theta t} \left\{ \int_{u=s}^{t} e^{\theta u} dW_u \right\},$$

The term $U = \int_{u=s}^{t} e^{\theta u} dW_u$ has a normal distribution with mean 0 and to determine its variance we can use the Ito isometry, which states that for a large class of processes Y_t we have

$$E\left[\int_{u=s}^{t} Y_u dW_u\right]^2 = E\left[\int_{u=s}^{t} Y_u^2 du\right].$$

For the special case in which $Y_u = e^{\theta u}$ we can conclude that the variance of U is

$$E[U^2] = E\left[\int_{u=s}^{t} Y_u^2 du\right] = E\left[\int_{u=s}^{t} e^{2\theta u} du\right] = (e^{2\theta t} - e^{2\theta s})/(2\theta).$$ [1]

Thus, the term

$$\sigma e^{-\theta t} \int_{u=s}^{t} e^{\theta u} dW_u,$$

has a normal distribution with mean 0 and variance

$$(\sigma e^{-\theta t})^2 (e^{2\theta t} - e^{2\theta s})/(2\theta).$$

So we conclude by stating the following useful result.

Conditional distribution: For $t > s$ the conditional distribution of X_t given X_s is normal with mean

$$m = e^{-\theta(t-s)} X_s + \mu(1 - e^{-\theta(t-s)})$$

and variance

$$v = \sigma^2 (1 - e^{-2\theta(t-s)})/(2\theta).$$

This result leads to the following.

Generating a realization at specific time points

To generate a realization of the mean reverting process at time points $0 = t_0 < t_1 < \cdots < t_N$ starting at X_0 we can define $X_{t_0} = 0$ then for $i = 1, \ldots, N$ iteratively by taking X_{t_i} to be normally distributed with mean

$$e^{-\theta \Delta t_i} X_{t_{i-1}} + \mu(1 - e^{-\theta \Delta t_i})$$

[1] In special cases, like the current one, when the Y_u is deterministic this has a very intuitive explanation.

and variance

$$v = \sigma^2(1 - e^{-2\theta\Delta t_i})/(2\theta).$$

where $\Delta t_i = t_i - t_{i-1}$.

Covariance:

Using the conditional distribution, we find that the conditional covariance between X_s and X_t given X_0 takes the form:

$$\frac{1}{2}e^{-\theta(s+t)}\left(e^{2s\theta} - 1\right)\sigma^2/\theta = \frac{1}{2}\left(e^{-\theta(t-s)} - e^{\theta(t+s)}\right)\sigma^2/\theta \text{ for } t > s.$$

Long-run Behavior

The just-stated fact has an immediate consequence. The limiting expected value of X_t given X_0 is μ and the limiting variance is $\sigma^2/(2\theta)$.

Likelihood function

Suppose we don't know the parameters μ, θ and σ but we have observations $(X_{t_0}, X_{t_1}, \ldots, X_{t_N})$ at (known) times $t_0, t_1, \ldots, t_N$. In principle, we can estimate the parameters by maximizing the likelihood function, which is obtained by first coming up with the joint pdf

$$f_{X_{t_0},\ldots,X_{t_N}}(x_{t_0}, x_{t_1}, \ldots, x_{t_N}|\mu, \theta, \sigma),$$

plugging in the observed values for the x_{t_i} to get the likelihood function

$$f_{X_{t_0},\ldots,X_{t_N}}(X_{t_0}, X_{t_1}, \ldots, X_{t_N}|\mu, \theta, \sigma),$$

which is a function of the unknown parameters μ, θ, and σ, and maximizing this function in the parameters to get the estimates.

To make the notation simpler we leave out the parameters μ, θ, and σ in the various pdf's and conditional pdf's that appear below. Generally, we have

$$f_{X_{t_0},X_{t_1},\ldots,X_{t_N}}(x_{t_0}, x_{t_1}, \ldots, x_{t_N})$$
$$= f_{X_{t_0}}(t_0) f_{X_{t_1}|X_{t_0}}(x_{t_1}|x_{t_0}) f_{X_{t_2}|X_{t_0},X_{t_1}}(x_{t_2}|x_{t_0}, x_{t_1}) \cdots f_{X_{t_n}|X_{t_0},X_{t_1},\ldots,X_{t_{N-1}}}(x_{t_N}|x_{t_0}, x_{t_1}, \ldots, x_{t_{N-1}}).$$

However, since our process has the **Markov property**[2] this simplifies enourmously to

$$f_{X_{t_0}}(x_{t_0}) f_{X_{t_1}|X_{t_0}}(x_{t_1}|x_{t_0}) f_{X_{t_2}|X_{t_1}}(x_{t_2}|x_{t_1}) \cdots f_{X_{t_N}|X_{t_{N-1}}}(x_{t_N}|x_{t_{N-1}})$$

[2]For $t_0 < t_1 < \ldots < t_N$ the conditional distribution of X_{t_n} given $X_{t_0}, \ldots, X_{t_{n-1}}$ is the same as the conditional distribution of X_{t_n} given $X_{t_{n-1}}$ so we can build up the joint pdf using conditional pdf's only involving pairs of random variables.

Log-likelihood

As usual, working with the log-likelihood is usually numerically better behaved than the likelihood, and maximizing the log-likelihood is equivalent to maximizing the likelihood since the log function is monotone increasing. Here, the log-likelihood is given by

$$\log(f_{X_{t_0}}(X_{t_0})) + \sum_{i=1}^{N} \log(f_{X_{t_i}|X_{t_{i-1}}}(X_{t_i}|X_{t_{i-1}}).$$

Here, we assume the starting point is known so the first time is some constant and can be ignored when we maximize.

Now, as derived above, the conditional distribution of X_{t_i} given $X_{t_{i-1}}$ is normal with mean

$$m_i = e^{-\theta \Delta t_i} X_{t_{i-1}} + \mu(1 - e^{-\theta \Delta t_i})$$

and variance

$$v_i = \sigma^2(1 - e^{-2\theta \Delta t_i})/(2\theta).$$

The notation can be simplified a bit by defining

$$w_i = e^{-\theta \Delta t_i}$$

so that

$$m_i = w_i X_{t_{i-1}} + \mu(1 - w_i),$$

and

$$v_i = \sigma^2(1 - w_i^2)/(2\theta).$$

Then it follows that the contribution to the i-th term of the log-likelihood sum above is

$$\log(f_{X_{t_i}|X_{t_{i-1}}}) = \log\left(\frac{1}{\sqrt{2\pi v_i}} \exp\left\{-\frac{1}{2}(X_{t_i} - m_i)^2/v_i\right\}\right)$$

$$= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log v_i - \frac{1}{2}(X_{t_i} - m_i)^2/v_i$$

and the full log-likelihood is given by

$$\ell(\mu, \sigma, \theta) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\log v_i - \frac{1}{2}\sum_{i=1}^{N}(X_{t_i} - m_i)^2/v_i$$

Maximizing the log-likelihood

On the surface, it appears that maximization of the log-likelihood $\ell(\mu, \sigma, \theta)$ would involve a search in the 3-dimensional parameter space for the optimum. In fact, the situation is considerably simpler. For any fixed value of θ we attempt to maximize the log-likelihood in μ and we find that the maximizing value $\hat{\mu} = \hat{\mu}(\theta)$ does not depend on σ. Then we can find σ to maximize $\ell(\hat{\mu}(\theta), \sigma, \theta)$ to produce $\hat{\sigma} = \hat{\sigma}(\hat{\mu}, \theta)$. Finally, we need only maximize the function

$$\ell(\hat{\mu}(\theta), \hat{\sigma}(\hat{\mu}(\theta), \theta), \theta)$$

which is a function of a single variable θ using some iterative scheme to produce $\hat{\theta}$.

Thus, for any choice of parameter values (μ, σ, θ) we have

$$\ell(\mu, \sigma, \theta) \leq \ell(\hat{\mu}(\theta), \sigma, \theta) \leq \ell(\hat{\mu}(\theta), \hat{\sigma}(\hat{\mu}(\theta), \theta), \theta) \leq \ell(\hat{\mu}(\hat{\theta}), \hat{\sigma}(\hat{\mu}(\hat{\theta}), \hat{\theta}), \hat{\theta})$$

and as a result we will have determined our MLE's:

Maximization in μ for fixed θ

We can write the portion of the log-likelihood to be maximized (we ignore the non-random term) as

$$\ell(\mu, \sigma, \theta) = \sum_{i=1}^{N} \left[-\frac{1}{2} \log(v_i) - \frac{1}{2} (X_{t_i} - m_i)^2 / v_i \right]$$

This expression is quadratic and concave in μ so we can express the MLE $\hat{\mu}$ by differentiating with respect to μ and equating to zero. Here

$$\frac{\partial}{\partial \mu} \ell = \sum_{i=1}^{N} (X_{t_i} - m_i)/v_i \frac{\partial m_i}{\partial \mu} = \sum_{i=1}^{N} (X_{t_i} - m_i)(1 - w_i)/v_i.$$

$$= \sum_{i=1}^{N} (X_{t_i} - w_i X_{t_{i-1}} - \mu(1 - w_i))(1 - w_i)/v_i.$$

Equating this to zero gives

$$\hat{\mu} = \frac{\sum_{i=1}^{N} (X_{t_i} - w_i X_{t_{i-1}})(1 - w_i)/v_i}{\sum_{i=1}^{N} (1 - w_i)^2 / v_i}$$

and using the fact that $v_i = \sigma^2 (1 - w_i^2)/(2\theta)$ we obtain

$$\hat{\mu} = \frac{\sum_{i=1}^{N} (X_{t_i} - w_i X_{t_{i-1}})(1 - w_i)/(1 - w_i^2)}{\sum_{i=1}^{N} (1 - w_i)^2 / (1 - w_i^2)}$$

$$= \frac{\sum_{i=1}^{N} (X_{t_i} - X_{t_{i-1}} w_i)/(1 + w_i)}{\sum_{i=1}^{N} (1 - w_i)/(1 + w_i)}$$

Note that this expression does not depend on σ but it does depend on θ so we can refer to this expession as $\hat{\mu}(\theta)$.

Maximization in σ for fixed θ

Once again, fixing θ and having maximized in μ, we can substitute $\hat{\mu}(\theta)$ for μ and maximize in σ. Letting

$$\hat{m}_i = w_i X_{t_{i-1}} + \hat{\mu}(\theta)(1 - w_i)$$

we see that

$$\begin{aligned}\ell(\hat{\mu}(\theta), \sigma, \theta) &= -\frac{1}{2}\sum_{i=1}^{N} \log\left[\frac{\sigma^2(1-w_i^2)}{2\theta}\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{N} \frac{(X_{t_i} - \hat{m}_i)^2(2\theta)}{(1-w_i^2)} \\ &= -n\log(\sigma) - \frac{1}{2}\sum_{i=1}^{N} \log\left[\frac{(1-w_i^2)}{2\theta}\right] - \frac{1}{2}S/\sigma^2\end{aligned}$$

where

$$S = 2\theta \sum_{i=1}^{N} (X_{t_i} - \hat{m}_i)^2/(1-w_i^2).$$

Differentiating with respect to σ and equating to zero gives

$$\hat{\sigma} = \sqrt{S/N}.$$

This still depends on θ so we ought to refer to this as $\hat{\sigma}(\theta)$.