Maximum Likelihood Estimation of a Mean Reverting Process

José Carlos García Franco

May 14, 2023

THIS IS A RE-WRITE OF A NOTE ORIGINALLY POSTED IN 2003

Abstract

Mean reverting processes are often used to model commodity prices in real options or to price financial derivatives. Given a specified model, estimating model parameters, based on on observations collected at arbitrary times, can be done using a variety of methods including regression and moment matching. When parametric specification is highly trusted, Maximum-likelihood is the method of choice, especially if likelihood functions are explicitly available. The Ornstein-Uhlenbeck mean-reverting process is a continuous time process that has an explicit likelihood function that can be maximized to obtain maximum likelihood estimates. While the theory assures us of asymptotic convergence, numerical estimation is sensitive to solver techniques. This work compares a couple ways of obtaining maximum-likelihood estimates for an Ornstein-Uhlenbeck mean-reverting model based on samples obtained at arbitrary (even random) points in time.

Keywords — Mean reversion, Ornstein-Uhlenbeck, maximum likelihood, scipy

1 Introduction

A mean-reverting Ornstein-Uhlenbeck process $X_t$ with parameters $\mu, \theta, \sigma$ is characterized by the stochastic differential equation

$$dX_t = \theta \left( \mu - X_t \right) dt + \sigma dB(t)$$

(1)

where $B_t$ is standard Brownian motion and $X_0 = x_0$. It can be shown that $X_t$ is normally distributed (see appendix A) with

$$E X_t = \mu + (x_0 - \mu) e^{-\theta t}$$

and

$$\text{var} X_t = \frac{\sigma^2}{2\theta} \left( 1 - e^{-2\theta t} \right).$$
2 Maximum likelihood-estimation

Given \( n + 1 \) samples \( x = \{ x_0, x_1, \ldots, x_n \} \) corresponding to times \( t_0, t_1, \ldots, t_n \), the unknown parameter vector \( \Theta = [\theta, \mu, \sigma] \) can be estimated using maximum likelihood. Note that the samples can be collected at arbitrary times in the sense that it is not required that they be collected at evenly spaced time intervals. Let \( \Delta t_i = t_i - t_{i-1} \) for \( i = 1, \ldots, n \), since \( X_t|x_{t-1} \) is normally distributed the log-likelihood function is

\[
\mathcal{L}(\Theta, x) = -\frac{n}{2} \log \left( \frac{\sigma^2}{2\theta} \right) - \frac{1}{2} \sum_{i=1}^{n} \log \left( 1 - e^{-2\theta \Delta t_i} \right) - \frac{\theta}{\sigma^2} \sum_{i=1}^{n} \frac{(x_{t_i} - \mu - (x_{t_{i-1}} - \mu) e^{-\theta \Delta t_i})^2}{1 - e^{-2\theta \Delta t_i}}. \tag{2}
\]

Maximum likelihood estimates (MLE’s) are found by solving

\[
\max_{\Theta} \mathcal{L}(\Theta, x). \tag{3}
\]

From optimization theory it is known that the solution to (3) must satisfy the first order necessary condition \( \frac{\partial \mathcal{L}}{\partial \mu}(\Theta, x) = 0 \) which implies that at the optimal

\[
\mu = \frac{1}{n} \sum_{i=1}^{n} \frac{x_{t_i} - x_{t_{i-1}} e^{-\theta \Delta t_i}}{1 + e^{-\theta \Delta t_i}} \left( \sum_{i=1}^{n} \frac{1 - e^{-\theta \Delta t_i}}{1 + e^{-\theta \Delta t_i}} \right)^{-1}. \tag{4}
\]

Similarly, the first order condition \( \frac{\partial \mathcal{L}}{\partial \sigma^2}(\Theta, x) = 0 \) must be satisfied, implying that

\[
\sigma^2 = \frac{2\theta}{n} \sum_{i=1}^{n} \frac{(x_{t_i} - \mu - (x_{t_{i-1}} - \mu) e^{-\theta \Delta t_i})^2}{1 - e^{-2\theta \Delta t_i}}. \tag{5}
\]

Lastly, the condition \( \frac{\partial \mathcal{L}}{\partial \theta}(\Theta, x) = 0 \) must also hold at the optimal, however, the expression for \( \frac{\partial \mathcal{L}}{\partial \theta}(\Theta, x) \) is a bit more complicated (see Appendix C) and does not lead to an explicit solution for \( \theta \) as was the case for parameters \( \mu \) and \( \sigma^2 \). Nevertheless, the relationships outlined above will prove useful in designing an algorithm to solve (3). Accordingly, a few strategies to find our MLE’s will be investigated:

1. Find the parameter set that minimizes the log-likelihood function (2) with a multi-dimensional numerical solver. The gradient of the log-likelihood function (shown in Appendix C) helps speed things up.

2. Plug the first order conditions (4) and (5) into the log-likelihood function (2) and solve a 1-dimensional maximization problem to obtain the MLE estimate \( \theta_{\text{MLE}} \). Then plug back that solution into (5) and (4) to obtain \( \sigma_{\text{MLE}} \) and \( \mu_{\text{MLE}} \).
3. Plug the first order conditions (4) and (5) into \( \frac{\partial L}{\partial \theta}(\Theta, x) = 0 \) and use a root finder to find \( \theta_{\text{MLE}} \). Then plug back that solution into (5) and (4) to obtain \( \sigma_{\text{MLE}} \) and \( \mu_{\text{MLE}} \).

While these solution strategies are theoretically equivalent not all of them are equally practical from the implementation point of view.

**Example**

Consider a mean-reverting process \( X \) as defined in (1) with \( \theta = 1.2, \mu = 20, \sigma = 4 \) and \( x_0 = 12 \). We compare three ML estimation methods for the parameter vector \( \Theta = [\mu, \theta, \sigma] \):

- Multi: maximizes (2) doing a 3-dimensional search with gradient supplied to the solver.\(^1\)
- Scalar: combines (4), (5) and (2) to perform the log-likelihood maximization using a scalar solver.
- Scalar root: combines (4), (5) to find the root of \( \frac{\partial L}{\partial \theta}(\theta) = 0 \) using a scalar root finder.

Monte Carlo simulation is used to compare the statistical behavior of the estimators and the performance of each method. Specifically, 10,000 sample paths of the process \( X \) were generated each with 168 steps at intervals \( \Delta_i = 1, i = 1, \ldots, 168 \). MLE estimates are obtained for each sample path using each one of the tested methods. The experiment was performed in Python 3.9.6 using the solvers of the `scipy` package. Table 1 shows the solver configuration options utilized.

The results of the experiment are summarized in Table 2 where it can be seen that all three methods are essentially equivalent in terms of the statistical properties of their

\(^1\)A configuration without passing the gradient was also tested. The results are identical to the gradient supplied search but the run time doubles.
corresponding estimators. In fact, for almost every sample path, they all arrive at the same solution. On the performance side, however, the scalar minimization method is faster than the others and the easiest to implement.

Panels (a), (b) and (c) in Figure 1 show the empirical estimator distributions corresponding to each parameter. We can see that the mean reversion level $\mu$ is the easiest to estimate and that its empirical distribution is consistent with its asymptotic distribution (which is essentially the best possible estimation scenario). Given the size of our sample paths, the estimators for parameters $\theta$ and $\sigma^2$ are not as well behaved as the estimator for $\mu$, their variances are patently greater than their asymptotic lower bound. Nevertheless, we can see that on average all estimators tend to be centered around the true value of the

Let $\Theta = [\mu, \theta, \sigma^2]$, under regularity conditions the vector of MLE estimates $\hat{\Theta}$ is asymptotically normally distributed with mean $\Theta$ and covariance $(I(\Theta))^{-1}$ where $I(\Theta) = -E \frac{\partial^2 L}{\partial \Theta \partial \Theta'}$, see [Greene, 1997]. Appendix C includes expressions for the main diagonal of $\frac{\partial^2 L}{\partial \Theta \partial \Theta'}$ and the corresponding expectations can be calculated with the results included in Appendix B. For example, we have $E \frac{\partial^2 L}{\partial \mu^2} = -2n/\sigma^2(1-e^{-\theta})(1-e^{-2\theta})$ and $E \frac{\partial^2 L}{\partial \sigma^2} = -n/(2\sigma^4)$. An expression for $E \frac{\partial^2 L}{\partial \theta^2}$ can also be obtained but it is too long to show here.

---

3Note that we run a slightly restricted likelihood estimation (see Table 1). It is left to the reader to evaluate the convenience of doing this and its effects on small sample estimations.

<table>
<thead>
<tr>
<th></th>
<th>$\mu$</th>
<th>$\theta$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true value</td>
<td>20.00</td>
<td>1.20</td>
<td>16.00</td>
</tr>
<tr>
<td><strong>Multi: 55.83 seconds</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>19.9932</td>
<td>1.2756</td>
<td>16.6517</td>
</tr>
<tr>
<td>median</td>
<td>19.9947</td>
<td>1.2332</td>
<td>16.2070</td>
</tr>
<tr>
<td>minimum</td>
<td>18.9605</td>
<td>0.6272</td>
<td>8.7389</td>
</tr>
<tr>
<td>maximum</td>
<td>20.9328</td>
<td>8.9435</td>
<td>103.5889</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.2709</td>
<td>0.3010</td>
<td>3.5388</td>
</tr>
<tr>
<td><strong>Scalar: 18.12 seconds</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>19.9932</td>
<td>1.2758</td>
<td>16.6532</td>
</tr>
<tr>
<td>median</td>
<td>19.9947</td>
<td>1.2330</td>
<td>16.2070</td>
</tr>
<tr>
<td>minimum</td>
<td>18.9605</td>
<td>0.6272</td>
<td>8.7388</td>
</tr>
<tr>
<td>maximum</td>
<td>20.9328</td>
<td>9.6353</td>
<td>106.3378</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.2709</td>
<td>0.3050</td>
<td>3.5720</td>
</tr>
<tr>
<td><strong>Scalar root: 31.05 seconds</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>19.9932</td>
<td>1.2758</td>
<td>16.6532</td>
</tr>
<tr>
<td>median</td>
<td>19.9947</td>
<td>1.2330</td>
<td>16.2070</td>
</tr>
<tr>
<td>minimum</td>
<td>18.9605</td>
<td>0.6272</td>
<td>8.7388</td>
</tr>
<tr>
<td>maximum</td>
<td>20.9328</td>
<td>9.6354</td>
<td>106.3380</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.2709</td>
<td>0.3050</td>
<td>3.5720</td>
</tr>
</tbody>
</table>

Table 2: Parameter estimate statistics and runtimes by method
Figure 1: Parameter estimate histograms and forecast statistics

Panel (d) in Figure 1 shows the average and median error when forecasting $X_t$ 1 to 24 periods beyond the last observation of each sample path as well as the corresponding 10 and 90 percentile levels.

3 Conclusion

This write-up provides a practical introduction to the Ornstein-Uhlenbeck mean reverting process and the estimation of its parameters via maximum-likelihood. It contains all the information required to implement parameter estimation algorithms and to build forecasts. A key feature of the approach utilized is that it allows for the usage of sample data collected at arbitrary times without the requirement sampling at equally spaced time intervals. At a more didactic level, the document includes sufficient math background and appendices to allow the practitioner to develop model extensions, or refinements to the estimation algorithms. Finally, it also serves as a good starting point into the study of stochastic process and their applications and may serve as a stepping stone towards the study of more general modeling and estimation frameworks.
Appendix

A Solution to the mean-reverting Ornstein-Uhlenbeck stochastic differential equation (SDE)

Let \( X_t \) be defined by the SDE
\[
dX_t = \theta (\mu - X_t) \, dt + \sigma \, dB_t
\]
(6)
where \( B_t \) is brownian motion and \( X_0 = x_0 \). Let \( f(t, x) = e^{\theta t} x \), it follows that \( d \left( e^{\theta t} x \right) = d f(t, x) \). Note that \( \frac{\partial f}{\partial t}(t, x) = \theta e^{\theta t} x \), \( \frac{\partial f}{\partial x}(t, x) = e^{\theta t} \) and \( \frac{\partial^2 f}{\partial x^2}(t, x) = 0 \). Applying the Ito formula (See [Øksendal, 1995, Chapter 4]) we get
\[
d \left( e^{\theta t} X_t \right) = \frac{\partial f}{\partial t}(t, X_t) \, dt + \frac{\partial f}{\partial x}(t, X_t) \, dB_t = \theta e^{\theta t} X_t \, dt + e^{\theta t} dB_t.
\]
(7)

Multiplying (6) by \( e^{\theta t} \) we get
\[
e^{\theta t} \, dX_t = \theta e^{\theta t} (\mu - X_t) \, dt + e^{\theta t} \sigma \, dB_t
\]
which together with (7) implies
\[
d \left( e^{\theta t} X_t \right) = \theta e^{\theta t} \mu \, dt + e^{\theta t} \sigma \, dB_t
\]
(8)

and we can write
\[
e^{\theta t} X_t = X_0 + \int_0^t \theta e^{\theta s} \mu \, ds + \int_0^t e^{\theta s} \sigma \, dB_s
\]
or equivalently
\[
X_t = X_0 e^{-\theta t} + \int_0^t \theta e^{-\theta (t-s)} \mu \, ds + \int_0^t e^{-\theta (t-s)} \sigma \, dB_s.
\]
(9)

The first integral on the right hand side of (9) evaluates to \( \mu \left(1 - e^{-\theta t}\right) \) and since \( B_t \) is brownian motions, the second integral is normally distributed with mean zero and variance equal to \( E \left( \int_0^t e^{-\theta (t-s)} \sigma \, dB_s \right)^2 \). By Ito isometry (See [Øksendal, 1995, Chapter 3]) we have that
\[
E \left( \int_0^t e^{-\theta (t-s)} \sigma \, dB_s \right)^2 = \int_0^t \left( e^{-\theta (t-s)} \sigma \right)^2 \, ds = \int_0^t e^{-2\theta (t-s)} \sigma^2 \, ds = \frac{\sigma^2}{2 \theta} \left(1 - e^{-2\theta t}\right).
\]

Therefore, given \( X_0 = x_0 \), \( X_t \) is normally distributed with
\[
E X_t = \mu + (x_0 - \mu) \, e^{-\theta t}
\]
(10)
and
\[ \text{var } X_t = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t}). \]  
\hspace{1cm} (11)

Similarly, for \( t > s \)
\[
\text{cov} [X_t, X_s] = \mathbb{E} \left( (X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s) \right)
= \mathbb{E} \left( \left( \int_0^t e^{-\theta(t-u)}\sigma dB_u \right) \left( \int_0^s e^{-\theta(s-u)}\sigma dB_u \right) \right)
= \sigma^2e^{-\theta(t+s)} \mathbb{E} \left( \left( \int_0^t e^{\theta u} dB_u \right)^2 \right) + \mathbb{E} \left( \left( \int_0^s e^{\theta u} dB_u \right)^2 \right) + \mathbb{E} \left( \left( \int_s^t e^{\theta u} dB_u \right)(\int_0^s e^{\theta u} dB_u) \right)
= \sigma^2e^{-\theta(t+s)} \int_0^s (e^{\theta u})^2 \, du
= \frac{\sigma^2}{2\theta}(e^{-\theta(t-s)} - e^{-\theta(t+s)}) = e^{-\theta(t-s)} \text{var } X_s.
\hspace{1cm} (12)

where the equality on the second to last line follows from Ito isometry and the independence of the random variables \( B_{t_0}, B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}} \) for \( t_0 < t_1 < t_2 < \ldots t_{n-1} < t_n. \)

B Some useful quantities

For \( X_t \) where \( dX_t = \theta(\mu - X_t) \, dt + \sigma \, dB_t \), \( B_t \) is brownian motion and \( X_0 = x_0 \), let
\[ Y_t = X_t - \mu - (X_{t_{i-1}} - \mu)e^{-\theta \Delta_i} \]  
\hspace{1cm} (13)

where \( \Delta_i = t_i - t_{i-1} \). Note that
\[
\mathbb{E} Y_{t_i} = \mathbb{E} (X_{t_i} - \mu) - e^{-\theta \Delta_i} \mathbb{E} (X_{t_{i-1}} - \mu)
= \left( \mu + (x_0 - \mu) e^{-\theta t_i} - \mu \right) - \left( \mu + (x_0 - \mu) e^{-\theta t_{i-1}} - \mu \right) e^{-\theta \Delta_i}
= (x_0 - \mu) e^{-\theta t_i} - (x_0 - \mu) e^{-\theta(t_{i-1} + \Delta_i)}
= 0
\]  
\hspace{1cm} (14)

and
\[
\text{var} Y_{t_i} = \text{var} \left( (X_{t_i} - \mu) - e^{-\theta \Delta_i} (X_{t_{i-1}} - \mu) \right)
= \text{var} X_{t_i} - 2e^{-\theta \Delta_i} \text{cov} (X_{t_i}, X_{t_{i-1}}) + e^{-2\theta \Delta_i} \text{var} X_{t_{i-1}}
= \frac{\sigma^2}{2\theta} \left( 1 - e^{-2\theta t_i} \right) - 2e^{-\theta \Delta_i} \frac{\sigma^2}{2\theta} \left( e^{-\theta \Delta_i} - e^{-\theta(t_i + t_{i-1})} \right) + e^{-2\theta \Delta_i} \frac{\sigma^2}{2\theta} \left( 1 - e^{-2\theta t_{i-1}} \right)
= \frac{\sigma^2}{2\theta} \left( 1 - e^{-2\theta \Delta_i} \right),
\]  
\hspace{1cm} (15)
where $E X_t$, var $X_t$ and cov $(X_t, X_s)$ come from (10), (11) and (12). Since $E Y_{t_i} = 0$ we have $E Y_{t_i}^2 = \text{var} Y_{t_i} + (E Y_{t_i})^2 = \text{var} Y_{t_i}$, hence
\[
E Y_{t_i}^2 = \frac{\sigma^2}{2\theta} \left(1 - e^{-2\theta \Delta_i}\right). \tag{16}
\]
Let $Z_i = X_{t_i} - \mu$ then $EZ_i = (x_0 - \mu)e^{-\theta t_i}$ and
\[
EZ_i^2 = E (X_{t_i} - \mu)^2 = E X_{t_i}^2 - 2\mu E X_{t_i} + \mu^2 = \text{var} X_{t_i} + (E X_{t_i})^2 - 2\mu E X_{t_i} + \mu^2 = \frac{\sigma^2}{2\theta} \left(1 - e^{-2\theta t_i}\right) + (x_0 - \mu)^2 e^{-2\theta t_i} \tag{17}
\]
and similarly
\[
E (Z_{t_{i-1}}Y_{t_i}) = E \left( (X_{t_{i-1}} - \mu) (X_{t_i} - \mu - (X_{t_{i-1}} - \mu) \gamma_i) \right) = E \left( (X_{t_{i-1}} - \mu) (X_{t_i} - \mu) \right) - \gamma_i E (X_{t_{i-1}} - \mu)^2 = \text{cov} (X_{t_{i-1}}, X_{t_i}) + EZ_{i-1}EZ_i - \gamma_i \text{var} X_{t_{i-1}} - \gamma_i (EZ_{i-1})^2 = (x_0 - \mu)^2 e^{-\theta (t_i + t_{i-1})} - (x_0 - \mu)^2 e^{-2\theta t_{i-1}} \gamma_i = 0 \tag{18}
\]
where $\gamma_i = e^{-\theta \Delta_i}$.

C Log-likelihood derivatives

We have the log-Likelihood function $L(\Theta, \mathbf{x}, \Delta)$, $\Theta = [\mu, \theta, \sigma], \mathbf{x} = \{X_{t_0}, \ldots, X_{t_n}\}, \Delta = \{\Delta_1, \ldots, \Delta_n\}, \Delta = t_i - t_{i-1}$
\[
L(\Theta, \mathbf{x}) = -n \frac{1}{2} \log \left(\frac{\sigma^2}{2\theta}\right) - \frac{1}{2} \sum_{i=1}^n \log \left(1 - e^{-2\theta \Delta_i}\right) - \frac{\theta}{\sigma^2} \sum_{i=1}^n \frac{Y_{t_i}^2}{1 - e^{-2\theta \Delta_i}}. \tag{19}
\]
with $Y_t$ as defined in (13). Then
\[
\frac{\partial L}{\partial \mu} = \frac{2\theta}{\sigma^2} \sum_{i=1}^n \frac{Y_{t_i} (1 - \gamma_i)}{1 - \gamma_i^2} \tag{20}
\]
\[
\frac{\partial L}{\partial \theta} = \frac{n}{2\theta} - \sum_{i=1}^n \frac{\Delta_i \gamma_i^2}{1 - \gamma_i^2} + \frac{1}{\sigma^2} \sum_{i=1}^n \frac{Y_{t_i} (Y_{t_i} \left(\gamma_i^2 - 1\right) + 2\theta \Delta_i \gamma_i (Z_i \gamma_i - Z_{i-1}))}{(1 - \gamma_i^2)^2} \tag{21}
\]
\[
\frac{\partial L}{\partial (\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{\theta}{\sigma^4} \sum_{i=1}^n \frac{Y_{t_i}^2}{1 - \gamma_i^2} \tag{22}
\]
and

\[
\frac{\partial^2 L}{\partial \mu^2} = -\frac{2 \theta}{\sigma^2} \sum_{i=1}^{n} \frac{(1 - \gamma_i)^2}{1 - \gamma_i} 
\]

(23)

\[
\frac{\partial^2 L}{\partial \theta^2} = -\frac{n}{2\theta^2} + 2 \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i^2}{1 - \gamma_i^2} - \frac{4\theta}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i^2 Y_{ti}^2}{(1 - \gamma_i^2)^2} - \frac{8\theta}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i^4 Y_{ti}^2}{(1 - \gamma_i^2)^3} 
\]

\[
+ \frac{2\theta}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i Z_{t_{i-1}} Y_{ti}}{1 - \gamma_i^2} \]

\[
+ \frac{8\theta}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i^3 Z_{t_{i-1}} Y_{ti}}{(1 - \gamma_i^2)^2} - \frac{2\theta}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i^2 Z_{t_{i-1}}^2}{1 - \gamma_i^2} 
\]

(24)

\[
\frac{\partial^2 L}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{2\theta}{\sigma^6} \sum_{i=1}^{n} \frac{Y_{ti}^2}{1 - \gamma_i^2} 
\]

(25)

where \(\gamma_i = e^{-\theta \Delta_i}\) and \(Z_i = X_{t_i} - \mu\).
References
