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 μ, θ, σ is characterized by the stochastic differential equation

$$dX_t = \theta\ (\mu - X_t)\, dt + \sigma\, dB(t) \tag{1}$$

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

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

and

$$\operatorname{var} X_t = \frac{\sigma^2}{2\,\theta}\left(1 - e^{-2\,\theta\, t}\right).$$

2 Maximum likelihood-estimation

Given $n+1$ samples $\mathbf{x} = \{x_0, x_1, \ldots, x_n\}$ 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

$$\mathcal{L}(\Theta, \mathbf{x}) = -\frac{n}{2} \log\left(\frac{\sigma^2}{2\theta}\right) - \frac{1}{2}\sum_{i=1}^{n} \log\left(1 - e^{-2\,\theta\,\Delta_{t_i}}\right) - \frac{\theta}{\sigma^2}\sum_{i=1}^{n} \frac{\left(x_{t_i} - \mu - \left(x_{t_{i-1}} - \mu\right)\, e^{-\theta\,\Delta_{t_i}}\right)^2}{1 - e^{-2\,\theta\,\Delta_{t_i}}}. \quad (2)$$

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

$$\max_{\Theta} \mathcal{L}(\Theta, \mathbf{x}). \quad (3)$$

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

$$\mu = \sum_{i=1}^{n} \frac{x_{t_i} - x_{t_{i-1}}\, e^{-\theta\,\Delta_{t_i}}}{1 + e^{-\theta\,\Delta_{t_i}}} \left(\sum_{i=1}^{n} \frac{1 - e^{-\theta\,\Delta_{t_i}}}{1 + e^{-\theta\,\Delta_{t_i}}}\right)^{-1}. \quad (4)$$

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

$$\sigma^2 = \frac{2\theta}{n} \sum_{i=1}^{n} \frac{\left(x_{t_i} - \mu - (x_{t_{i-1}} - \mu) e^{-\theta\,\Delta_i}\right)^2}{1 - e^{-2\theta\,\Delta_i}}. \quad (5)$$

Lastly, the condition $\frac{\partial\,\mathcal{L}}{\partial\,\theta}(\Theta, \mathbf{x}) = 0$ must also hold at the optimal, however, the expression for $\frac{\partial\,\mathcal{L}}{\partial\,\theta}(\Theta, \mathbf{x})$ is a bit more complicated (see Appendix C) and does not lead to an explicit solution for θ as was the case for parameters μ and σ^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 θ_{MLE}. Then plug back that solution into (5) and (4) to obtain σ_{MLE} and μ_{MLE}.

	Solver	Options
Multi	`scipy.optimize.minimize`	`method = 'L-BFGS-B', bounds = ((None, None), (0.05, None), (0.05, None)) jac = supplied`
Scalar	`scipy.optimize.minimize_scalar`	`method = 'bounded', bounds = (0, 10)`
Scalar root	`scipy.root_scalar`	

Table 1: Optimization software and settings

3. Plug the first order conditions (4) and (5) into $\frac{\partial \mathcal{L}}{\partial \theta}(\Theta, \mathbf{x}) = 0$ and use a root finder to find θ_{MLE}. Then plug back that solution into (5) and (4) to obtain σ_{MLE} and μ_{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 $\frac{\partial \mathcal{L}}{\partial \theta}(\theta) = 0$ 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.

	μ	θ	σ^2
true value	20.00	1.20	16.00
Multi: 55.83 seconds			
mean	19.9932	1.2756	16.6517
median	19.9947	1.2332	16.2070
minimum	18.9605	0.6272	8.7389
maximum	20.9328	8.9435	103.5889
standard deviation	0.2709	0.3010	3.5388
Scalar: 18.12 seconds			
mean	19.9932	1.2758	16.6532
median	19.9947	1.2330	16.2070
minimum	18.9605	0.6272	8.7388
maximum	20.9328	9.6353	106.3378
standard deviation	0.2709	0.3050	3.5720
Scalar root: 31.05 seconds			
mean	19.9932	1.2758	16.6532
median	19.9947	1.2330	16.2070
minimum	18.9605	0.6272	8.7389
maximum	20.9328	9.6354	106.3380
standard deviation	0.2709	0.3050	3.5720

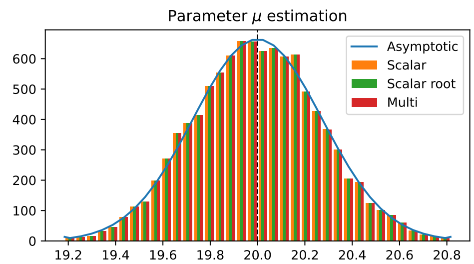
Table 2: Parameter estimate statistics and runtimes by method

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.

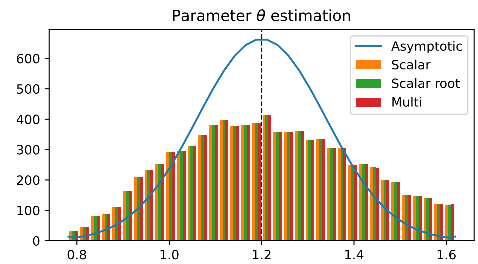
Panels (a), (b) and (c) in Figure 1 show the empirical estimator distributions corresponding to each parameter. We can see that the mean reversion level μ is the easiest to estimate and that its empirical distribution is consistent with its asymptotic distribution[2] (which is essentially the best possible estimation scenario). Given the size of our sample paths, the estimators for parameters θ and σ^2 are not as well behaved as the estimator for μ, their variances are patently greater than their asymptotic lower bound.[3] Nevertheless, we can see that on average all estimators tend to be centered around the true value of the

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

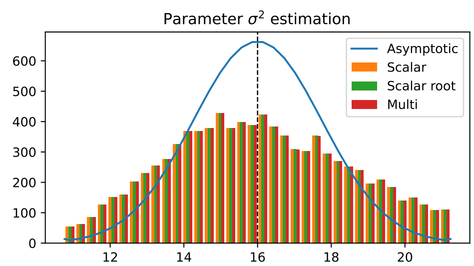
[3]Note that we run a slightly restricted likelihood estimation (see Table 1). It is left to the reader to evaluate the convenience of doing this and its effects on small sample estimations.



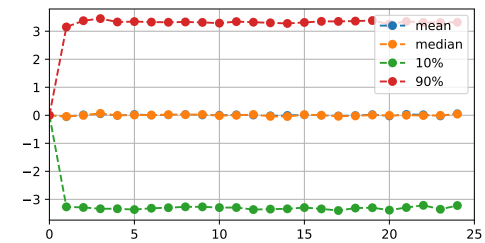
(a) Parameter μ estimation



(b) Parameter θ estimation



(c) Parameter σ^2 estimation



(d) Forecast error by forecast horizon

Figure 1: Parameter estimate histograms and forecast statistics

parameter - as they should! 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

$$dX_t = \theta\ (\mu - X_t)\, dt + \sigma\, dB_t \tag{6}$$

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

$$d\left(e^{\theta\, t}X_t\right) = \frac{\partial f}{\partial t}(t,X_t)dt + \frac{\partial f}{\partial x}(t,X_t) = \theta\, e^{\theta\, t}X_t dt + e^{\theta\, t}dX_t. \tag{7}$$

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

$$d\left(e^{\theta\, t}X_t\right) = \theta\, e^{\theta\, t}\mu\, dt + e^{\theta\, t}\sigma dB_t \tag{8}$$

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

$$X_t = X_0 e^{-\theta\, t} + \int_0^t \theta\, e^{-\theta\,(t-s)}\mu\, ds + \int_0^t e^{-\theta\,(t-s)}\sigma dB_s. \tag{9}$$

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

$$\begin{aligned}
\mathrm{E}\left(\int_0^t e^{-\theta\,(t-s)}\sigma dB_s\right)^2 &= \int_0^t \left(e^{-\theta\,(t-s)}\sigma\right)^2 ds = \int_0^t e^{-2\,\theta\,(t-s)}\sigma^2 ds \\
&= \frac{\sigma^2}{2\,\theta}\left(1 - e^{-2\,\theta\, t}\right).
\end{aligned}$$

Therefore, given $X_0 = x_0$, X_t is normally distributed with

$$\mathrm{E}\, X_t = \mu + (x_0 - \mu)\ e^{-\theta\, t} \tag{10}$$

and

$$\operatorname{var} X_t = \frac{\sigma^2}{2\,\theta}\left(1 - e^{-2\,\theta\,t}\right). \tag{11}$$

Similarly, for $t > s$

$$\begin{aligned}
\operatorname{cov}[X_t, X_s] &= \mathrm{E}\left((X_t - \mathrm{E}X_t)(X_s - \mathrm{E}X_s)\right) \\
&= \mathrm{E}\left(\left(\int_0^t e^{-\theta\,(t-u)}\sigma dB_u\right)\left(\int_0^s e^{-\theta\,(s-u)}\sigma dB_u\right)\right) \\
&= \sigma^2 e^{-\theta(t+s)}\mathrm{E}\left(\left(\int_0^t e^{\theta\,u} dB_u\right)\left(\int_0^s e^{\theta\,u} dB_u\right)\right) \\
&= \sigma^2 e^{-\theta(t+s)}\left(\mathrm{E}\left(\left(\int_0^s e^{\theta\,u} dB_u\right)^2\right) + \mathrm{E}\left(\left(\int_s^t e^{\theta\,u} dB_u\right)\left(\int_0^s e^{\theta\,u} dB_u\right)\right)\right) \\
&= \sigma^2 e^{-\theta(t+s)}\int_0^s \left(e^{\theta\,u}\right)^2 du \\
&= \frac{\sigma^2}{2\,\theta}\left(e^{-\theta(t-s)} - e^{-\theta(t+s)}\right) = e^{-\theta(t-s)}\operatorname{var} X_s. \qquad (12)
\end{aligned}$$

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

$$Y_{t_i} = X_{t_i} - \mu - (X_{t_{i-1}} - \mu)e^{-\theta\,\Delta_i} \tag{13}$$

where $\Delta_i = t_i - t_{i-1}$. Note that

$$\begin{aligned}
\mathrm{E}\,Y_{t_i} &= \mathrm{E}\,(X_{t_i} - \mu) - e^{-\theta\,\Delta_i}\mathrm{E}\,(X_{t-1} - \mu) \\
&= \left(\mu + (x_0 - \mu)\,e^{-\theta\,t_i} - \mu\right) - \left(\mu + (x_0 - \mu)\,e^{-\theta\,t_{i-1}} - \mu\right)e^{-\theta\,\Delta_i} \\
&= (x_0 - \mu)\,e^{-\theta\,t_i} - (x_0 - \mu)\,e^{-\theta\,(t_{i-1} + \Delta_i)} \\
&= 0 \qquad (14)
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{var} Y_{t_i} &= \operatorname{var}\left((X_{t_i} - \mu) - e^{-\theta\,\Delta_i}\left(X_{t_{i-1}} - \mu\right)\right) \\
&= \operatorname{var} X_{t_i} - 2e^{-\theta\Delta_i}\operatorname{cov}\left(X_{t_i}, X_{t_{i-1}}\right) + e^{-2\theta\Delta_i}\operatorname{var}\,X_{t_{i-1}} \\
&= \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta t_i}\right) - 2e^{-\theta\Delta_i}\frac{\sigma^2}{2\theta}\left(e^{-\theta\Delta_i} - e^{-\theta(t_i + t_{i-1})}\right) + e^{-2\theta\Delta_i}\frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta t_{i-1}}\right) \\
&= \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta\,\Delta_i}\right), \qquad (15)
\end{aligned}$$

where $\mathrm{E}\,X_t$, $\mathrm{var}\,X_t$ and $\mathrm{cov}\,(X_t, X_s)$ come from (10), (11) and (12). Since $\mathrm{E}\,Y_{t_i} = 0$ we have $\mathrm{E}\,Y_{t_i}^2 = \mathrm{var}\,Y_{t_i} + (\mathrm{E}\,Y_{t_i})^2 = \mathrm{var}\,Y_{t_i}$, hence

$$\mathrm{E}\,Y_{t_i}^2 = \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta\,\Delta_i}\right). \tag{16}$$

Let $Z_i = X_{t_i} - \mu$ then $\mathrm{E}Z_i = (x_0 - \mu)e^{-\theta t_i}$ and

$$\begin{aligned}
\mathrm{E}Z_i^2 &= \mathrm{E}\,(X_{t_i} - \mu)^2 \\
&= \mathrm{E}X_{t_i}^2 - 2\mu\mathrm{E}X_{t_i} + \mu^2 \\
&= \mathrm{var}X_{t_i} + (\mathrm{E}X_{t_i})^2 - 2\mu\mathrm{E}X_{t_i} + \mu^2 \\
&= \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta t_i}\right) + (x_0 - \mu)^2\, e^{-2\theta t_i} \qquad (17)
\end{aligned}$$

and similarly

$$\begin{aligned}
\mathrm{E}\,(Z_{i-1}Y_{t_i}) &= \mathrm{E}\left(\left(X_{t_{i-1}} - \mu\right)\left(X_{t_i} - \mu - \left(X_{t_{i-1}} - \mu\right)\gamma_i\right)\right) \\
&= \mathrm{E}\left(\left(X_{t_{i-1}} - \mu\right)(X_{t_i} - \mu)\right) - \gamma_i \mathrm{E}\left(X_{t_{i-1}} - \mu\right)^2 \\
&= \mathrm{cov}\left(X_{t_{i-1}}, X_{t_i}\right) + \mathrm{E}Z_{i-1}\mathrm{E}Z_i - \gamma_i \mathrm{var}X_{t_{i-1}} - \gamma_i\left(\mathrm{E}Z_{i-1}\right)^2 \\
&= (x_0 - \mu)^2\, e^{-\theta(t_i + t_{i-1})} - (x_0 - \mu)^2\, e^{-2\theta t_{i-1}}\gamma_i \\
&= 0 \qquad (18)
\end{aligned}$$

where $\gamma_i = e^{-\theta\Delta_i}$.

C Log-likelihood derivatives

We have the log-Likelihood function $\mathcal{L}(\Theta, \mathbf{x}, \boldsymbol{\Delta})$, $\Theta = [\mu, \theta, \sigma]$, $\mathbf{x} = \{X_{t_0}, \ldots, X_{t_n}\}$, $\boldsymbol{\Delta} = \{\Delta_1, \ldots, \Delta_n\}$, $\Delta_i = t_i - t_{i-1}$

$$\mathcal{L}(\Theta, \mathbf{x}) = -\frac{n}{2}\log\left(\frac{\sigma^2}{2\theta}\right) - \frac{1}{2}\sum_{i=1}^{n}\log\left(1 - e^{-2\,\theta\,\Delta_i}\right) - \frac{\theta}{\sigma^2}\sum_{i=1}^{n}\frac{Y_{t_i}^2}{1 - e^{-2\,\theta\,\Delta_i}}. \tag{19}$$

with Y_t as defined in (13). Then

$$\frac{\partial\mathcal{L}}{\partial\,\mu} = \frac{2\theta}{\sigma^2}\sum_{i=1}^{n}\frac{Y_{t_i}(1 - \gamma_i)}{1 - \gamma_i^2} \tag{20}$$

$$\frac{\partial\mathcal{L}}{\partial\,\theta} = \frac{n}{2\theta} - \sum_{i=1}^{n}\frac{\Delta_i\gamma_i^2}{1 - \gamma_i^2} + \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{Y_{t_i}\left(Y_{t_i}\left(\gamma_i^2 - 1\right) + 2\theta\Delta_i\gamma_i\left(Z_i\gamma_i - Z_{i-1}\right)\right)}{\left(1 - \gamma_i^2\right)^2} \tag{21}$$

$$\frac{\partial\mathcal{L}}{\partial\,(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{\theta}{\sigma^4}\sum_{i=1}^{n}\frac{Y_{t_i}^2}{1 - \gamma_i^2} \tag{22}$$

and

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \mu^2} &= -\frac{2\,\theta}{\sigma^2} \sum_{i=1}^{n} \frac{(1-\gamma_i)^2}{1-\gamma_i^2} && (23) \\
\frac{\partial^2 \mathcal{L}}{\partial \theta^2} &= -\frac{n}{2\theta^2} + 2 \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i^2}{\left(1-\gamma_i^2\right)^2} - \frac{4\theta}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i^2 Y_{t_i}^2}{\left(1-\gamma_i^2\right)^2} - \frac{8\theta}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i^4 Y_{t_i}^2}{\left(1-\gamma_i^2\right)^3} \\
&\quad + \frac{2\theta}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i Z_{i-1} Y_{t_i}}{1-\gamma_i^2} + \frac{8\theta}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i^3 Z_{i-1} Y_{t_i}}{\left(1-\gamma_i^2\right)^2} - \frac{2\theta}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i^2 \gamma_i^2 Z_{i-1}^2}{1-\gamma_i^2} \\
&\quad + \frac{4}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i \gamma_i^2 Y_{t_i}^2}{\left(1-\gamma_i^2\right)^2} - \frac{4}{\sigma^2} \sum_{i=1}^{n} \frac{\Delta_i \gamma_i Z_{i-1} Y_{t_i}}{1-\gamma_i^2} && (24) \\
\frac{\partial^2 \mathcal{L}}{\partial \left(\sigma^2\right)^2} &= \frac{n}{2\sigma^4} - \frac{2\theta}{\sigma^6} \sum_{i=1}^{n} \frac{Y_{t_i}^2}{1-\gamma_i^2} && (25)
\end{aligned}
$$

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

References

[Greene, 1997] Greene, W. H. (1997). *Econometric Analysis, Third Edition*. Prentice-Hall.

[Øksendal, 1995] Øksendal, B. (1995). *Stochastic Differential Equations, An Introduction with Applications, Fourth Edition*. Springer.