BACHELOR THESIS

Sandra Martinková

Parameter estimation for Ornstein-Uhlenbeck process

Department of Probability and Mathematical Statistics

Supervisor of the bachelor thesis: Mgr. Ing. Pavel Kříž, Ph.D.
Study programme: General mathematics
Study branch: Stochastics

Prague 2019
I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

Prague, 10th July 2019  

signature of the author
This work is dedicated to all of my friends, coworkers, and family whose love and support was crucial to me throughout the entirety of this year. A special thank you goes out to my parents, Zlata Martinková and Bohuslav Pavlík, for enabling my university education. I would also like to express my sincere gratitude to Veronika Bauer for all her support, understanding and motivation. Last but not least, I would like to extend a warm thank you to my thesis supervisor Pavel Kríž, whose passion for the subject, kind nature, and willingness to help made writing this thesis so much more enjoyable.
Title: Parameter estimation for Ornstein-Uhlenbeck process

Author: Sandra Martinková

Department: Department of Probability and Mathematical Statistics

Supervisor: Mgr. Ing. Pavel Kříž, Ph.D., Department of mathematics, University of chemistry and technology Prague

Abstract: The Ornstein-Uhlenbeck process has countless practical applications most of which rely on having previously estimated the drift parameter. The literature offers two basic estimators - the least-squares estimator, which coincides with the maximum-likelihood estimator for Ornstein-Uhlenbeck process, and the method-of-moments estimator. However, the similarity in asymptotic properties of these estimators means that choosing which one to use is more of a random guess than an educated decision. This thesis focuses on finding differences between the two estimators when applied to the Ornstein-Uhlenbeck trajectories generated in R. The simulation study performed suggests that the method-of-moments is better suited when the initial condition is close to zero even if the observations are collected sparsely. On the other hand, the precision of the least-squares estimator is better when the initial condition is further away from zero, but it still requires having dense data points. Under the conditions of this study, the least-squares estimator performs better compared to the method-of-moments if the absolute value of the initial condition is large. On the other hand, the method-of-moments is superior in cases where we have infrequent observations and long time horizon.

Keywords: Ornstein-Uhlenbeck process, Drift parameter estimation, Method of moments, Least squares estimator
Contents

Introduction 2

1 Stochastic analysis 3
   1.1 Preliminaries .............................................. 3
   1.2 Itô integral ............................................. 3

2 Parameter estimators for Ornstein-Uhlenbeck process 6
   2.1 Least squares estimator ............................... 6
   2.2 Method of moments .................................. 7

3 Simulating Ornstein-Uhlenbeck process 8
   3.1 General setting of the experiment ................. 8
      3.1.1 Important variables ......................... 8
   3.2 Modelling trajectories ............................. 8
   3.3 Implementation of estimators .................. 9

4 Simulation results 11
   4.1 Stationary Ornstein-Uhlenbeck process .......... 11
   4.2 Deterministic initial condition ................. 13
      4.2.1 Zero initial condition ..................... 13
      4.2.2 Initial condition far from zero .......... 14
   4.3 The evolution of both estimators across our scenarios 16

Conclusion 17

Bibliography 18

List of Figures 19

List of Tables 20
Introduction

The Ornstein-Uhlenbeck process is one of the basic stochastic processes and can be defined by the following stochastic differential equation:

\[ dX_t = -\alpha X_t \, dt + \sigma \, dW_t, \quad X_0 = \eta. \]

The process was studied for the first time by Ornstein and Uhlenbeck [1930]. Since then, it has been frequently used in financial mathematics, for example interest rates (Okunev et al. [1994]), currency exchange rates (da Fonseca et al. [2015]) and commodity prices (Chaiyapo and Phewchean [2017]) can all be modelled using Ornstein-Uhlenbeck process. Furthermore, it has its application in physics including quantum mechanics as researched in Defendi and Roncadelli [1992].

If one measures real-life phenomena in order to provide a model for data and finds themselves with a pattern and properties painstakingly similar to the Ornstein-Uhlenbeck process it is important to be able to estimate its parameters. The data collected usually model a trajectory which can be used for this purpose. Having estimated the parameters one is able to try forecasting the future trajectory of the process. In this work, it is assumed that the parameter \( \sigma \) is known and the focus is on estimating the drift parameter \( \alpha \). If necessary, \( \sigma \) can be estimated separately and subsequently used as an input into the estimators of \( \alpha \). The impact of replacing the parameter \( \sigma \) with an estimate is outside of the scope of this work.

Various estimators can be used to estimate \( \alpha \). Two very popular and quite simple ones are the least-squares estimator and the method-of-moments estimator, which were chosen for implementation in this work. Their asymptotic properties are well known and they also happen to be very similar - both are consistent and asymptotically normal with the same speed of convergence as proved in Kutoyants [2004]. This means that there is very little distinction between the two estimators analytically. Such a situation makes deciding which one to use rather difficult.

In practice, research is done over a finite period of time and measurements are taken at discrete moments, rarely does one find themselves with the number of observations and time tending to infinity. That is why there is a need for a work that would compare these estimators by running simulations on trajectories generated under various conditions. Mean-square error (MSE) and squared bias are used to compare the quality of these estimators in various scenarios. The scenarios differ in the initial condition of the process, length of time-window and the length of time-step.
1. Stochastic analysis

The majority of the theory and results presented in this chapter come from [Oksendal 2003].

1.1 Preliminaries

In order to be able to work with the Ornstein-Uhlenbeck process, one needs to be familiar with the Wiener process.

Definition 1. Let \((W_t, t \geq 0)\) be a stochastic process with the following properties:

(i) the distribution of \(W_t - W_s\) is normal, \(E[W_t - W_s] = 0\) and \(E[(W_t - W_s)^2] = (t - s)\) for \(s < t\);

(ii) when \(s < t\) the random variable \(W_t - W_s\) is independent of \(W_r\) for \(r \leq s\);

(iii) function \(t \to W_t\) is a continuous function almost surely and \(W_0 = 0\).

Such a process shall be called the Wiener process.

Furthermore, we need to be sure of its existence. For the Wiener process to exist it is necessary that we work on a probability space that is rich enough to contain it.

Theorem 1. There exists a probability space \((\Omega, \mathcal{A}, P)\) such that the Wiener process defined in Definition 1 exists on this space.

For proof see [Oksendal 2003], section 2.2.

In addition to this, the adaptedness of a process also needs to be understood. The Itô integral with respect to the Wiener process cannot be constructed without this property.

Definition 2. Let \((W_s, s \geq 0)\) be a Wiener process and \(\mathcal{W}_t = \sigma(W_s, 0 \leq s \leq t)\) be the increasing family of \(\sigma\)-algebras of subsets of \(\Omega\) generated by \(W_s\). A random process \(X: [0, \infty) \times \Omega \to \mathbb{R}^n\) is \(\mathcal{W}_t\)-adapted if for each \(t \geq 0\) the function \(\omega \to X(t, \omega)\) is \(\mathcal{W}_t\)-measurable.

1.2 Itô integral

From here on, it is assumed that we are working on a complete probability space \((\Omega, \mathcal{A}, P)\) with a Wiener process \(W\) and an interval \([r, s] \subset [0, +\infty)\). To begin with, a class of functions for which we will be able to construct the Itô integral must be defined. Such a class of functions will be denoted by \(\mathcal{V}(r, s)\).

Definition 3. Functions \(f \in \mathcal{V}(r, s), f : [0, \infty) \times \Omega \to \mathbb{R}\) are functions that fulfill the following conditions:

(i) \((t, \omega) \to f(t, \omega)\) is \(\mathcal{B} \times \mathcal{A}\)-measurable, where \(\mathcal{B}\) is the Borel \(\sigma\)-algebra on \([0, \infty)\);

(ii) the random process \(f\) is \(\mathcal{W}_t\)-adapted;

(iii) \(E[f_r^2(t, \omega)dt] < \infty\) for \(0 \leq r \leq s < \infty\).
The Itô integral for elementary functions \( \phi \), i.e. functions such that \( \phi(t, \omega) = \sum_j e_j(\omega) \cdot \chi_{[t_j, t_{j+1})}(t) \) where \( e_j \) is a random variable measurable with respect to \( \mathcal{W}_{t_j} \) and \( 0 < r = t_1 < t_2 < \ldots < t_n = s \), can be defined easily as:

\[
\int_r^s \phi \, d\mathcal{W}_t(\omega) = \sum_{j=0}^{n-1} e_j(\omega)[\mathcal{W}_{t_{j+1}}(\omega) - \mathcal{W}_{t_j}(\omega)]
\]  

(1.1)

The following lemma holds true for elementary functions and it is the first step towards obtaining the Itô isometry.

**Lemma 2.** For \( \phi \) bounded and elementary the following holds true

\[
E[(\int_r^s \phi(t, \omega) d\mathcal{W}_t)(\omega))^2] = E[\int_r^s \phi^2(t, \omega) dt]
\]  

(1.2)

For proof see [Oksendal 2003], Lemma 3.1.5

Furthermore, for any function \( f \in \mathcal{V}(r, s) \) there exist elementary functions \( \phi_n \), such that

\[
E[\int_r^s (f - \phi_n)^2 dt] \to 0 \text{ as } n \to \infty.
\]  

(1.4)

Let us now demonstrate that the limit in (1.3) exists. The sequence of elementary functions is a Cauchy sequence with regard to \( L_2(\Omega \times [r, s]) \) metrics defined as \( d(\phi_m, \phi_n) = \sqrt{E \int_r^s (\phi_n(t, \omega) - \phi_m(t, \omega))^2 dt} \). Therefore, if one takes \( m, n \geq n_r \), the following holds true

\[
E[\int_r^s (\phi_m - \phi_n)^2 dt] < \epsilon.
\]

By applying the Itô isometry and using linearity (which is obvious) we get

\[
E(\int_r^s \phi_m d\mathcal{W}_t - \int_r^s \phi_n d\mathcal{W}_t)^2 < \epsilon.
\]

Therefore, the sequence \( (\int_r^s \phi_n d\mathcal{W}_t)_{n \in N} \) is also Cauchy in \( L_2(\Omega) \). Moreover, the \( L_2(\Omega) \) space is complete; hence the limit from Definition 4 (Equation 1.3) exists.

Furthermore, let us also prove that the limit does not depend on the approximating sequence. Take two sequences of elementary functions \( \phi_n \) and \( \theta_n \) which fulfill \( E[\int_r^s (f - \phi_n)^2 dt] \to 0 \) and \( E[\int_r^s (f - \theta_n)^2 dt] \to 0 \) as \( n \to \infty \). Once again due to the Itô isometry

\[
E(\int_r^s \theta_n d\mathcal{W}_t - \int_r^s \phi_n d\mathcal{W}_t)^2 = E(\int_r^s (\theta_n - \phi_n)^2 dt).
\]
In addition to this, the following also holds true
\[
E\left(\int_r^s (\theta_n - \phi_n)^2 dt\right) \leq 2E\left(\int_r^s (f - \phi_n)^2 dt\right) + 2E\left(\int_r^s (f - \theta_n)^2 dt\right).
\]
Therefore, we have
\[
E\left(\int_r^s \theta_n dW_t - \int_r^s \phi_n dW_t\right)^2 \to 0 \quad \text{as} \quad n \to \infty.
\]
Hence the expected result is obtained
\[
\lim_{n \to \infty} \int_r^s \phi_n(t, \omega)dW_t = \lim_{n \to \infty} \int_r^s \theta_n(t, \omega)dW_t \quad \text{a.s.}
\]
For all functions \( f \in \mathcal{V}(r, s) \) the following version of Itô isometry holds true.

**Theorem 3.** The Itô isometry holds true for all \( f \in \mathcal{V}(r, s) \), i.e.
\[
E[(\int_r^s f(t, \omega)dW_t(\omega))^2] = E[\int_r^s f^2(t, \omega)dt] \quad (1.5)
\]
Proof is a combination of Lemma 2 and Definition 4. For details see Oksendal 2003, Corrolary 3.1.8.

Having established all of these results it is finally possible to define the Ornstein-Uhlenbeck process. Throughout the rest of this work, the Ornstein-Uhlenbeck process will be referred to via its differential equation rather than its integral form.

**Definition 5.** Ornstein-Uhlenbeck process (OU process) is the solution of differential equation
\[
dX_t = -\alpha X_t dt + \sigma dW_t, \quad X_0 = \eta \quad (1.6)
\]
where \( \sigma \) and \( \alpha \) are positive constants and \( \eta \) is a random initial condition independent of \( W_t \). This means that OU process is a continuous process that fulfills
\[
X_t(\omega) = \eta - \int_0^t \alpha X_s(\omega) ds + \int_0^t \sigma dW_s(\omega) \quad \text{a.s.} \quad (1.7)
\]
The right-hand side of the equation is well defined because \( X_s \) has continuous trajectories. Hence, the first integral exists in the Riemann sense and the second integral is simply a constant integrated with respect to \( W_t \) in the Itô sense.

**Theorem 4.** A unique solution to equation (1.6) exists and can be written explicitly as
\[
X_t = e^{-\alpha t} \eta + \int_0^t e^{-\alpha(t-s)} \sigma dW_s \quad (1.8)
\]
Proof can be found in Karatzas and Shreve 1991, example 6.8.

In this work, the Ornstein-Uhlenbeck process with a deterministic initial condition as well as its stationary version will be simulated. For this reason, it is necessary to establish the existence of a stationary solution to the Ornstein-Uhlenbeck equation.

**Theorem 5.** \( X_t \) as a solution to equation (1.6) is strictly stationary for \( \eta \sim N(0, \sigma^2/2\alpha) \) independent of \( W \). Such a \((X_t, t \geq 0)\) will be referred to as stationary Ornstein-Uhlenbeck process.

For proof see Karatzas and Shreve 1991, example 6.8.


2. Parameter estimators for Ornstein-Uhlenbeck process

In this section it will be discussed how to estimate the drift parameter $\alpha$ from a single trajectory of the Ornstein-Uhlenbeck process using the least-squares and method-of-moments estimators. Both of these approaches can be used for a deterministic as well as stationary initial conditions. Most of the results in this chapter are taken from Kutoyants [2004].

2.1 Least squares estimator

In the first place, the motivation behind this estimator will be examined. It is impossible to calculate the derivative of the Wiener process with respect to $t$ because its trajectories are not differentiable - for further information see Oksendal [2003]. However, let us at least for a while imagine that we were able to do so. If this were the case the Equation 1.6 could be rewritten as

$$X'_{t}=\alpha X_{t}+\sigma W'_{t}$$

(2.1)

where $W'_{t}$ and $X'_{t}$ would be the time-derivatives of the Wiener process and the Ornstein-Uhlenbeck process respectively.

It would now be possible to proceed in a manner similar to the construction of the traditional least-squares estimate by looking for such $\alpha$ that best fulfills the following condition

$$\min_{\alpha \in \mathbb{R}} \int_{0}^{t} (X_{s}' - \alpha X_{s})^{2} \, ds.$$ 

By denoting the function which is being minimized as $F(\alpha)$ and setting the derivative with respect to $\alpha$ equal to zero, one gets

$$F'(\alpha) = \left. \int_{0}^{t} 2(X_{s}' - \alpha X_{s})(-X_{s}) \, ds \right|_{\alpha = \alpha_{LSE}} = 0.$$ 

Now by rearranging this to express $\alpha$ a formula for the estimate is obtained

$$\alpha_{LSE} = \frac{\int_{0}^{t} X_{s} \cdot X_{s}' \, ds}{\int_{0}^{t} X_{s}^{2} \, ds}.$$ 

Even though $X'_{s}$ does not exist, one could formally write $X'_{s} \, ds = dX_{s}$. Therefore, one procures a well-defined formula in terms of the Itô integral

$$\alpha_{LSE} = \frac{\int_{0}^{t} X_{s} \, dX_{s}}{\int_{0}^{t} X_{s}^{2} \, ds}.$$ 

Even though this is only a heuristic inference it provides a motivation for the definition of this estimator.
Definition 6. Let $X_s$ be an Ornstein-Uhlenbeck process with $X_0 = \eta$ and parameters $\alpha$ and $\sigma$, where $\sigma$ is given. The least-squares estimator of $\alpha$ is

$$
\alpha_{LSE}(t) = \frac{\int_0^t X_s dX_s}{\int_0^t X_s^2 ds}
$$

(2.2)

where $\int_0^t X_s dX_s$ can be defined as $\int_0^t -\alpha X_s^2 ds + \int_0^t X_s \sigma dW_s$.

The known asymptotic properties of this estimator are the following:

Theorem 6. The least-squares estimator is strongly consistent and asymptotically normal, i.e. the following holds true:

$$
\alpha_{LSE}(t) \overset{a.s.}{\rightarrow} \alpha \quad \text{when} \quad t \to \infty
$$

$$
\sqrt{t}(\alpha_{LSE}(t) - \alpha) \overset{d}{\rightarrow} N(0, \sigma_{LSE}^2)
$$

For proof see [Kutoyants 2004], example 1.35.

Remark. Least-squares estimator coincides with the maximum-likelihood estimator in the case of estimating the drift parameter $\alpha$ of the Ornstein-Uhlenbeck process.

2.2 Method of moments

It is possible to infer the method-of-moments estimator using an approach similar to the classical method-of-moments. Let $X_t$ be a stationary Ornstein-Uhlenbeck process. It is known that $EX_t^2 = \frac{\sigma^2}{2\alpha}$ (see Theorem 5), which can be used to express the method-of-moments estimator. Simply rearranging this relation with $\alpha$ on one side gives

$$
\alpha = \frac{\sigma^2}{2EX_t^2}
$$

Here the unknown $EX_t^2$ can be estimated by the sample moment of $\frac{1}{t} \int_0^t X_s^2 ds$.

Definition 7. Let $X_s$ be an Ornstein-Uhlenbeck process with $X_0 = \eta$ and parameters $\alpha$ and $\sigma$, where $\sigma$ is given. The method-of-moments estimator of $\alpha$ is $\alpha_{MoM}(t)$ defined as follows

$$
\alpha_{MoM}(t) = \frac{\sigma^2}{2\frac{1}{t} \int_0^t X_s^2 ds}
$$

(2.3)

Now the asymptotic properties of the method-of-moments estimator can be formulated:

Theorem 7. The method-of-moments estimate is strongly consistent and asymptotically normal, i.e.

$$
\alpha_{MoM}(t) \overset{a.s.}{\rightarrow} \alpha \quad \text{when} \quad t \to \infty
$$

$$
\sqrt{t}(\alpha_{MoM}(t) - \alpha) \overset{d}{\rightarrow} N(0, \sigma_{MoM}^2)
$$

For proof see [Kutoyants 2004], example 1.45.
3. Simulating Ornstein-Uhlenbeck process

3.1 General setting of the experiment

In order to be able to measure the performance of the two estimators trajectories of the Ornstein-Uhlenbeck process with a given parameter $\alpha$ were generated. Thereafter, these trajectories were used as an input for different types of parameter estimation and retrospectively the estimates were compared to the original value of $\alpha$.

First, a function in statistical software R which generates trajectories of the Ornstein-Uhlenbeck process with input parameters described below was implemented.

3.1.1 Important variables

The model uses the following parameters:

- $\alpha$ is the drift parameter of the Ornstein-Uhlenbeck process
- $\sigma$ is the volatility of the Ornstein-Uhlenbeck process
- $npaths$ is the number of trajectories generated
- $M$ is the total number of observations in one trajectory
- $T$ is the time of the last observation of the process, i.e. the time horizon
- $firstval$ is the value of the initial condition $\eta$
- $\delta = \frac{T}{M}$ is the time step
- $\Delta_t$ are independent random variables generated from $N(0, \sqrt{\frac{{\sigma^2(1-e^{-2\alpha \delta})}}{2\alpha}})$ distribution

3.2 Modelling trajectories

To be able to test different methods of parameter estimation the trajectories of the Ornstein-Uhlenbeck process needed to be generated. Since this work is examining stationary as well as a non-stationary process both types of trajectories were generated. An example of a stationary trajectory can be seen in Figure 3.1.

Subsequently, the computed values of $X_t$ were saved into a matrix $M$, each trajectory being one row, with columns representing the time instant at which the value of $X_t$ was calculated. Hence the value in $m_{ij}$ represented the value of process number $i$ at time $j$.

Initial condition $\eta$ was either a fixed value for the non-stationary process or a number from $N(0, \sqrt{\frac{T}{2\alpha}})$ distribution for the stationary process. With $X_t$
To ensure that the correct process was generated the Pearson’s chi-square goodness of fit test was performed. This test verifies the hypothesis that if a fixed time $T_1$ is chosen, the values of generated stationary processes are normally distributed with mean 0 and variance $\frac{\sigma^2}{2\alpha}$.

To perform this test 10 000 trajectories were modelled and $T_1=54$ was chosen. Figure 3.2 illustrates the distribution of values of X at $T_1$, with the curve showing the normal distribution. Applying the goodness of fit test gave a p-value of 0.772. Therefore, the hypothesis that the values of generated stationary processes are normally distributed with mean 0 and variance $\frac{\sigma^2}{2\alpha}$ cannot be rejected.

### 3.3 Implementation of estimators

The algorithm takes the matrix M with values of $X_t$ as an input into the parameter estimation part. In this part, the parameter estimate was calculated for each of the trajectories separately.

The effectiveness of estimators in this work was measured by mean square error (MSE), i.e.

$$\sum_{i \in \{1,2,\ldots,\text{npaths}\}} \frac{(\alpha_i - \alpha)^2}{\text{npaths}}$$

and squared bias, i.e.

$$\left( \sum_{i \in \{1,2,\ldots,\text{npaths}\}} \frac{\alpha_i}{\text{npaths}} - \alpha \right)^2$$
Figure 3.2: Values of X at time $T_1 = 54$

where $\alpha_i$ is the calculated estimate of $\alpha$ from the $i$-th trajectory.

Box plots of these estimates were generated showing the effectiveness of a given estimator in further detail.

Given that the trajectories were generated at discrete times with time step $\delta$, the integral in the formulas for the two estimators was substituted with sums. Therefore the implementation of the estimators was as follows:

\[
\alpha_{MoM} = \frac{\sigma^2}{2 \cdot \frac{1}{M} \sum_{s=0}^{M} X_{s \delta}^2}
\]

\[
\alpha_{LSE} = \frac{\sum_{s=0}^{M} X_{s \delta} (X_{(s+1)\delta} - X_{s \delta})}{\sum_{s=0}^{M} X_{s \delta}^2 \delta}
\]
4. Simulation results

The interval lengths that were chosen to be examined are $T \in \{10, 1000\}$. These numbers seem reasonable considering that they provide a fine time difference. The density of partitioning frequency $\delta$ is either 1 or 0.001, which helps identify differences for high frequency and low-frequency data.

4.1 Stationary Ornstein-Uhlenbeck process

Using the stationary trajectories we decided to take a look mainly at how the length and the density of partitioning of the interval over which data is measured influences these estimators.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$T$</th>
<th>$\delta$</th>
<th>LSE MSE</th>
<th>LSE bias$^2$</th>
<th>Method-of-moments MSE</th>
<th>Method-of-moments bias$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>10</td>
<td>1</td>
<td>0.075190</td>
<td>0.010599</td>
<td>0.2187190</td>
<td>0.0394935</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.073736</td>
<td>0.013657</td>
<td>0.0660257</td>
<td>0.017948</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>0.000187</td>
<td>0.000022</td>
<td>0.0001832</td>
<td>0.0000033</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.000200</td>
<td>0.000007</td>
<td>0.0002334</td>
<td>0.0000077</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>1</td>
<td>0.157928</td>
<td>0.077020</td>
<td>0.5773550</td>
<td>0.1260679</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.316861</td>
<td>0.043184</td>
<td>0.2502070</td>
<td>0.0224119</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>0.135052</td>
<td>0.134121</td>
<td>0.0027119</td>
<td>0.0000302</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.002059</td>
<td>0.000054</td>
<td>0.0018886</td>
<td>0.000100</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>1</td>
<td>81,122609</td>
<td>81,003646</td>
<td>80,175940</td>
<td>9,6114560</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>2,135854</td>
<td>0.089042</td>
<td>2,2513730</td>
<td>0.0808611</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>80,971343</td>
<td>80,970334</td>
<td>0.2183912</td>
<td>0.0050373</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.024037</td>
<td>0.004094</td>
<td>0.0017149</td>
<td>0.0000013</td>
</tr>
</tbody>
</table>

Table 4.1: Stationary Ornstein-Uhlenbeck process

With $\alpha = 0.1$ both estimators perform very well if given enough data. When the length of the interval is 1000, MSE and squared bias are smaller than $\frac{1}{1000}$. With the shorter interval where $T = 10$, MSE is smaller than $\frac{1}{10}$ for least-squares while it reaches more than $\frac{2}{10}$ for the method-of-moments. All of these numbers suggest that both estimators provide good estimates in this setting.

For $\alpha = 10$ the influence of a bigger drift parameter can clearly be seen. The least-squares estimator performs worse in comparison with method-of-moments for the longer interval. The MSE for the least-squares estimator reaches numbers as high as 80, thus making the estimator virtually useless. It can be seen that when $T = 10$, $\delta = 1$ the squared bias for the least-squares estimate is the same as its MSE. Therefore, it is clear that the least-squares estimate systematically underestimates the parameter. In this case method-of-moments does not perform better in terms of MSE. However, its squared bias is significantly lower. This can also be seen in Figure 4.1 where the horizontal line denotes the actual alpha.
The only setting for $\alpha = 10$ where the least-squares estimate performs reasonably is the one where $T = 1000$, $\delta = 0.001$. Compared to this, the method-of-moments performs reasonably well in both low and high density of partitioning when $T$ is 1000. It seems to be sufficient to lengthen the interval to get a good enough estimate with the method-of-moments, whereas with the least-squares estimator one needs both long interval and high density of partitioning. This means that the least-squares estimator requires substantially more data to offset the bigger drift parameter. Overall, it can be seen that the method-of-moments performs better with $\alpha$ large, however, it still needs sufficient data to be able to do this.
Since the least-squares estimator uses an integral with respect to $dX_s$ to calculate the estimate it is expected that it would perform worse when the density of partitioning is low or the interval short. Given that $dX_s$ is substituted with a simple difference of $X_{s+\delta} - X_s$ if two values are further apart that will result in a worse approximation of $dX_s$. On the other hand, method-of-moments shows lower sensitivity to change in length of time as well as density of partitioning because it only takes into account the values of $X_s$ and $\sigma$ to calculate the estimate and $dX_s$ is not needed.

Figure 4.2 illustrates how lengthening of time horizon helps the method-of-moments estimator to improve its accuracy. The mean of method-of-moments estimates is quite close to the actual $\alpha = 10$ represented by the horizontal line of the box plot. The same doesn’t do much for the least-squares estimator where the mean estimate of $\alpha$ close to 1. The least-squares estimator would also require a denser partitioning to provide a good result.

4.2 Deterministic initial condition

This section evaluates how the value of initial condition $\eta$ influences the estimates when the Ornstein-Uhlenbeck process with the deterministic initial condition is used. The conjecture was that due to the nature of our two estimators, method-of-moments should perform worse than the least-squares estimator when $\eta$ is far from zero. For $\eta$ close to zero the results should be quite similar to the stationary case.

4.2.1 Zero initial condition

In this case, it was chosen to have $\eta = 0$. For this initial condition, the results table is quite similar to the stationary case. The precision of both estimates for $\alpha = 0.1$ and partially for $\alpha = 1$ is higher with $\eta = 0$ than it is in the stationary case. It is unclear why this is the case and further investigation into this matter would be recommended.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$T$</th>
<th>$\delta$</th>
<th>LSE</th>
<th>Method-of-moments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MSE bias^2</td>
<td>MSE bias^2</td>
</tr>
<tr>
<td>0.1</td>
<td>100</td>
<td>1</td>
<td>0.004890</td>
<td>0.0000404</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.005027</td>
<td>0.0000611</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>0.000232</td>
<td>0.0000011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.000199</td>
<td>0.000002</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>1</td>
<td>0.144052</td>
<td>0.136470</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.019272</td>
<td>0.000519</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>0.132336</td>
<td>0.131511</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.001951</td>
<td>0.000103</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>1</td>
<td>81.254255</td>
<td>81.247199</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.208942</td>
<td>0.005554</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>80.962636</td>
<td>80.961679</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.017032</td>
<td>0.000511</td>
</tr>
</tbody>
</table>

Table 4.2: Initial condition $\eta = 0$
One possible, though not verified, theory is that for stationary process the starting point is random and there can be outliers that start farther away from zero, therefore decreasing the estimate of $\alpha$ mainly for the method-of-moments. However, this does nothing to explain the increased precision of the least-squares estimator in this case.

If the initial condition of the Ornstein-Uhlenbeck process is zero or close to zero, then both estimators perform well for $\alpha = 0.1$. As $\alpha$ increases to 1 it can be seen again that the method-of-moments estimator is more precise than the least-squares estimator. For the low density of partitioning the method-of-moments is about ten times as precise as the least-squares estimator.

Figure 4.3: Estimated values of $\alpha = 10$ for $\delta = 1$, $T=1000$ and $\eta = 0$

Furthermore, method-of-moments performs significantly better than least-squares for $\alpha = 10$ as can be seen in Figure 4.3. Once again, the actual alpha is denoted by the horizontal line here. From this, it can again be reasoned that the method-of-moments has lower sensitivity to changes in $\alpha$ than the least-squares estimator, particularly when the density of partitioning is low.

4.2.2 Initial condition far from zero

On the other hand, if one takes initial condition that is far away from zero, e.g. $\eta = 300$, it can be seen that the least-squares estimator does better than the method-of-moments because it is more precise in all scenarios.

With $\alpha = 10$ we see that MSE is practically equal to squared bias for method-of-moments meaning that the estimates are virtually equal to 0 in all cases. Therefore, method-of-moments is not a viable option in any combination of parameters $\delta$ and $T$. On the other hand, the least-squares estimator still works for $\delta = 0.001$. This can be clearly seen in Figure 4.4.
In this case, it can be expected that method-of-moments should perform worse than the least-squares estimator because it only takes into account the values of \( X_s \) and assumes stationarity of the process. A process with large initial condition \( \eta \) will have large values of \( X_s^2 \) in the beginning. After a while, the values will become smaller due to the effect of the drift parameter bringing the process into stationary phase. Nonetheless, these large values of \( X_s^2 \) at the start of the process decrease the method-of-moments estimate of the drift parameter. This happens because of the assumed stationarity, i.e. the estimator assumes that large values of \( X_s^2 \) are the result of small \( \alpha \). This results in the under-estimation of \( \alpha \), especially when the true value of \( \alpha \) is large.

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( T )</th>
<th>( \delta )</th>
<th>LSE ( \text{MSE} ) ( \text{bias}^2 )</th>
<th>Method-of-moments ( \text{MSE} ) ( \text{bias}^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>100</td>
<td>1</td>
<td>0.000025 0.000023</td>
<td>0.009980 0.009980</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,001</td>
<td>0.000002 0.000000</td>
<td>0.009978 0.009978</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>0.000027 0.000025</td>
<td>0.009801 0.009801</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,001</td>
<td>0.000002 0.000000</td>
<td>0.009782 0.009782</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>1</td>
<td>0.135422 0.135419</td>
<td>0.999040 0.999040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,001</td>
<td>0.000022 0.000001</td>
<td>0.997783 0.997783</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>0.135270 0.135266</td>
<td>0.990460 0.990460</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,001</td>
<td>0.000027 0.000001</td>
<td>0.978165 0.978165</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>1</td>
<td>81.001897 81.001897</td>
<td>99.988890 99.988890</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,001</td>
<td>0.002911 0.002700</td>
<td>99.780345 99.780345</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>81.001515 81.001514</td>
<td>99.888981 99.888981</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0,001</td>
<td>0.002654 0.002423</td>
<td>97.835118 97.835118</td>
</tr>
</tbody>
</table>

Table 4.3: Initial condition \( \eta = 300 \)

Figure 4.4: Estimated values of \( \alpha = 10 \) for \( \delta = 0.001, T=1000 \) and \( \eta = 300 \)
4.3 The evolution of both estimators across our scenarios

The least-squares estimator performs better when the initial condition is farther away from zero. There is an increase in precision in almost all settings of $T$, $\delta$ and $\alpha$ when comparing the results for $\eta = 0$ and $\eta = 300$. In scenarios where $\eta = 300$ and $\alpha = 0.01$ as well as when $\alpha \in \{1, 10\}$ and $\delta = 0.001$ the precision of the estimates increases about hundredfold compared to when initial condition $\eta = 0$. In the rest of the cases, the precision stays at a similar level. This is because the least-squares estimator depends on $X_{s\delta}(X_{(s+1)\delta} - X_{s\delta})$. Thus it significantly benefits from trends observed in the non-stationary phase of the process.

On the other hand, the opposite is true for the method-of-moments estimator. Overall, this estimator performs worse when the initial condition $\eta$ is far from zero as opposed to $\eta$ being close to zero. Moreover, whilst the method-of-moments estimator handles larger $\alpha$ quite well for the stationary process or $\eta = 0$, the moment the initial condition $\eta$ is far from zero and $\alpha$ is larger the estimator is very imprecise.
Conclusion

From the simulation, several scenarios for which the estimators perform differently were identified. The first scenario is where $\eta$ is close to zero, $\alpha$ and $T$ are big, i.e. at least 10 and 1000 respectively, and the density of partitioning is low. In this case, the method-of-moments proves to be a better choice by being 400 times as precise as the least-squares estimator. This means that method-of-moments performs better for initial condition close to zero, large alpha and low density of partitioning.

Second main scenario is where both $\eta$ and $\alpha$ are big and the density of partitioning is high, i.e. $\delta = 0.001$. In this case for both longer and shorter intervals, the least-squares estimator is 5000 times more precise than the method-of-moments. Therefore, the least-squares estimator will perform better if we have initial condition farther away from zero even if $\alpha$ is large but needs a high density of partitioning to do so.

Furthermore, if we have a scenario where the initial condition is far from zero the least-squares estimator will always perform better than the method-of-moments. Its biggest limitation is for the low density of partitioning and large $\alpha$ where it performs very similarly to the method-of-moments. In this case, we would not recommend using either of these estimates.

The precision of the least-squares estimator increases as the initial condition moves away from zero. This holds true even for cases where the least-squares estimator was precise even with the initial condition being 0, i.e. for $\alpha$ small. So if one has reason to believe that $\alpha$ should be small in their model and the initial condition large, the least-squares estimator will provide a very precise estimate of $\alpha$. On the other hand, fo the stationary Ornstein-Uhlenbeck process observed with low frequency, method-of-moments is the preferred option, especially if the time-window is large.
Bibliography


# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Example of a stationary Ornstein-Uhlenbeck process trajectory</td>
<td>9</td>
</tr>
<tr>
<td>3.2</td>
<td>Values of $X$ at time $T_1 = 54$</td>
<td>10</td>
</tr>
<tr>
<td>4.1</td>
<td>Estimated values of $\alpha = 10$ for $\delta = 1$, $T=10$ in the stationary case</td>
<td>12</td>
</tr>
<tr>
<td>4.2</td>
<td>Estimated values of $\alpha = 10$ for $\delta = 1$ and $T=1000$ in stationary case</td>
<td>12</td>
</tr>
<tr>
<td>4.3</td>
<td>Estimated values of $\alpha = 10$ for $\delta = 1$, $T=1000$ and $\eta = 0$</td>
<td>14</td>
</tr>
<tr>
<td>4.4</td>
<td>Estimated values of $\alpha = 10$ for $\delta = 0.001$, $T=1000$ and $\eta = 300$</td>
<td>15</td>
</tr>
</tbody>
</table>
List of Tables

4.1 Stationary Ornstein-Uhlenbeck process ........................................... 11
4.2 Initial condition $\eta = 0$ ................................................................. 13
4.3 Initial condition $\eta = 300$ ................................................................. 15