EXPONENTIAL SMOOTHING METHODS

Contents

[I. Introduction 1](#_Toc103803498)

[II. First-Order Exponential Smoothing 5](#_Toc103803499)

[1. The value of **𝜆** 6](#_Toc103803500)

[2. The initial value, 7](#_Toc103803501)

[III. Modeling Time Series Data 8](#_Toc103803502)

[IV. Forcasting 9](#_Toc103803503)

[1. Constant Process 9](#_Toc103803504)

[2. Estimation of 12](#_Toc103803505)

[3. Adaptive Updating of the **Discount Factor** 13](#_Toc103803506)

1. Introduction

* Data set = **signal** + **noise**
* The **constant process**

 (1)

𝜇: the underlying constant level of system respone (mức độ phản ứng)

𝜀t: the noise at time t (uncorrelated with mean 0 and constant variance )

* **Smoothing** can be seen as a technique to separate the **signal** and the **noise**

Diagram

Description automatically generated

The **smoothers** will achieve this by simply relating the current observation to the previous ones.

* Use **backward looking smoothers**

That is, at any given **T**, the observation **yT** will be replaced by a combination of observations at and before **T**. Then, use some sort of an “**average**” of the current and the previous observations to smooth the data. An obvious choice is to replace the current observation with **the average of the observations at T, T − 1, ..., 1**.

* This is the “best” choice in **the least squares** sense for a **constant process** given in Eq (1).
* Bình phương nhỏ nhất (**least squares**) là khái niệm chung để chỉ một nhóm các phương pháp kinh tế lượng phục vụ cho việc ước lượng tham số trong phương trình hồi quy với mục đích tối thiểu hoá tổng bình phương của các khoảng cách theo phương thẳng đứng giữa số liệu thu thập được và đường hồi quy
* A **constant process** can be smoothed by replacing the current observation with the best estimate for **𝜇**. Using **the least squares** criterion, we define the **sum of squares error**, SSE, for the **constant process** as

A picture containing text, watch

Description automatically generated

* The **least squares** estimate of **𝜇** can be found by setting the derivative of SS with respect to 𝜇 to 0:

A picture containing clock

Description automatically generated (2)

* This is the average of observations up to time T

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

* As time goes on, the **smoother** in Eq (2) accumulates more and more data points and gains some sort of “**inertia**”. So, when there is a change in the process, it becomes increasingly more difficult for this smoother to react to it.
* The **smoother** in Eq (2) was that it reacted too slowly to process changes because of all the observations are weighted equally and having the same amount of influence on the average. Thus, if the weights of each observation are changed so that earlier observations are weighted less, a faster reacting **smoother** should be obtained.
* A common solution is to use the **simple moving average**:

A picture containing text, antenna

Description automatically generated

The most crucial issue is the choice of the **span**, N. A **simple moving average** will react faster to the changes if **N** is small, but the **variance of the moving average** gets bigger. This creates a dilemma in the choice of **N**.

Logo, company name

Description automatically generated

* If the process is expected to be **constant**, a **large N** can be used whereas a **small N** is preferred if the process is **changing**.

Chart, diagram, scatter chart

Description automatically generated

1. First-Order Exponential Smoothing

An **exponentially weighted smoother** (|𝜃| < 1):

Chart

Description automatically generated with medium confidence (3)

A picture containing text, clock, watch

Description automatically generated (4)

Because the **smoother** in Eq. (3) is not an average, we can adjust the **smoother** in Eq. (3) by multiplying it by (1 − 𝜃) / (1 − ). However, for large T values, goes to 0 and so **the exponentially weighted average** will have the following form:

Text, letter

Description automatically generated (5)

* This is called a **simple** or **first-order exponential smoother.**

An alternate expression in a recursive form for **simple exponential smoothing** is given by

Text, letter

Description automatically generated (6)

**First-order exponential smoothing** can also be seen as the linear combination of the current observation and the smoothed observation at the previous time unit.

is a **discount factor** represents the weight put on the last observation and **(1−𝜆)** represents the weight put on the smoothed value of the previous observations

 (7)

## 1. The value of **𝜆**

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated



* The latter the smoothed values follow the original observations more closely. In general, as **𝜆** gets closer to 1, and more emphasis is put on the last observation, the smoothed values will approach the original observations.

**𝜆** values between 0.1 and 0.4 are often recommended and do indeed perform well in practice.

- Mean squared deviation (MSD) is the average squared difference between the predicted and the true values, given as:

A picture containing text, clock

Description automatically generated

## 2. The initial value,

Take the average of the available data or a subset of the available data, , and set = . If the process is at least at the beginning locally constant, this starting value may be preferred.

As an alternative estimate for the initial value, we can also use the average of the data between June 1999 and June 2001

Chart, scatter chart

Description automatically generated

1. Modeling Time Series Data

The general class of models can be represented as:

 (9)

**𝛽** is the vector of unknown parameters

represents the uncorrelated errors

* the **constant process** can be represented as

 (10)

1. Forcasting

* Exponential smoothing techniques
* Point out the underlying patterns in the time series data
* Estimate the model parameters for the class of models given in Eq. (9)
* We will denote the **𝜏-step-ahead forecast** made at time T as

1. Constant Process

* **First-order exponential smoothing** for the constant process: 
* The constant level in Eq. (9), , can be estimated by **.** The **constant model** consists of 2 parts:
  + The random error that cannot be predicted
* Our forecast for the future observation is simply equal to the current value of the **exponential smoother**

 (11)

* The forecast in Eq. (11) is the same for all future values.
* We start accumulating more observations, we can update our forecast. If the data at **T+1** become available, our forecast for the future observations becomes:



*  (12)
* ****

A picture containing diagram

Description automatically generated (13)

(1) = − (T) is called the **one-step-ahead forecast** or **prediction error**

* Choice of 𝝀 We will define **the sum of the squared one-step-ahead forecast errors** as:

Text

Description automatically generated with medium confidence (14)

we can in general calculate **SSE** values for various values of **𝜆** and pick the value of 𝜆 that gives the smallest **sum of the squared forecast errors**

A picture containing table

Description automatically generated

Text

Description automatically generated

Chart

Description automatically generated

1. Estimation of

In the estimation of the variance of the forecast errors,, it is often assumed that the model (e.g., constant, linear trend) is correct and constant in time. With these assumptions, we have a way of estimating :

* Define **the mean absolute deviation Δ** as:

 (15)

* Calculate its estimate:

A picture containing diagram

Description automatically generated (16)

* Then the estimate of the is given by

 (17)

1. Adaptive Updating of the **Discount Factor**

* Estimation of the “best” discount factor, , by minimizing the sum of **the squared** **one-step-ahead forecasts errors**
* The changes in the underlying time series model will make it difficult for the **exponential smoother** with fixed **discount factor** to follow these changes
* modify the **discount factor** arises, so the **discount factor** will adapt to the changes in the time series model

 (18)

is given as a function of time and hence it is allowed to adapt to changes in the time series model

* *Smoothed error:*

 (19)

**𝛿** is a smoothing parameter

* Define **the tracking signal**:

 (20)

 in (16)

This ratio is expected to be close to 0 when the forecasting system performs well and to approach ±1 as it starts to fail. In fact, Trigg and Leach (1967) suggest setting the discount factor to:

A picture containing chart

Description automatically generated (21)

Chart, scatter chart

Description automatically generated

Graphical user interface, application, table, Excel

Description automatically generated

This plot shows that a better smoother can be obtained by making automatic updates to the **discount factor.** The adaptive smoothing procedure suggested by Trigg and Leach is a useful technique