



Aalto-yliopisto
Perustieteiden
korkeakoulu

ARIMA- ja neuroverkkomallien vertailua osakeindeksejä ennustettaessa (Valmiin työn esittely)

Matti Staudinger

02.12.2021

Ohjaaja: *prof. Pauliina Ilmonen*

Valvoja: *prof. Pauliina Ilmonen*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

Tausta

- Osakekursseja ja indeksejä pyritään ennustamaan monin menetelmin
- Sijoittajille on merkittävää rahallista hyötyä kyvystä ennakoida indeksin kehitystä
- Sijoituspäätöksen tueksi tarvitaan tietoa
- Ennustemalleja hyödynnetään automaattisessa kaupankäynnissä

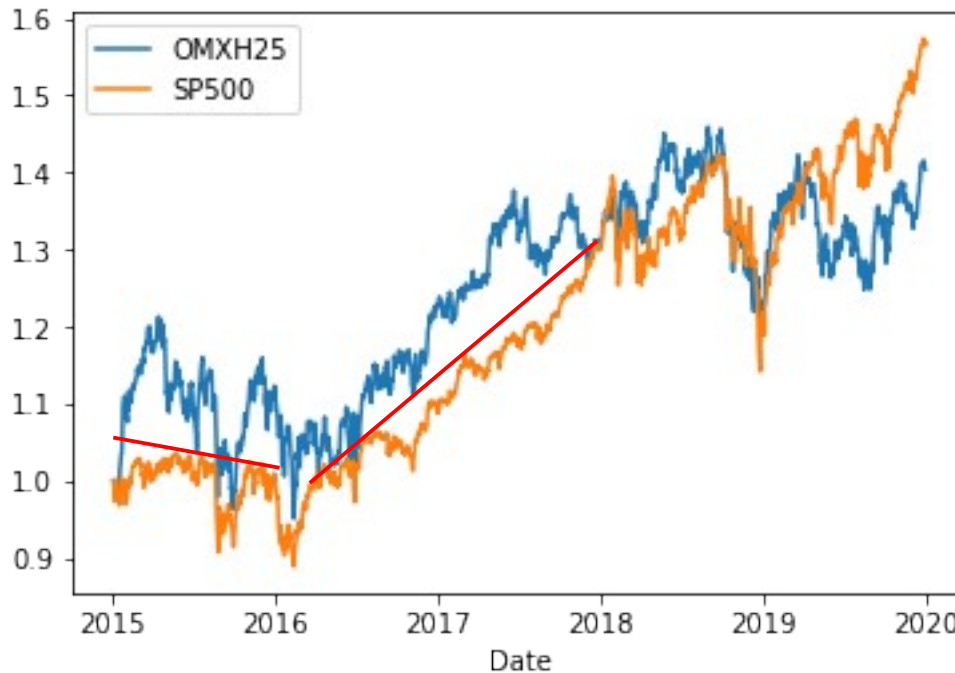
Tavoitteet

- Muodostaa ja verrata ARIMA- ja RNN-malleja osakeindeksien ennustamisessa
- Lyhyt aikaväli: ennusteet yhdestä viiden askeen päähän
- Arvioida ennusteiden mielekkyyttä ja hyödyntämismahdollisuuksia sijoituspäätöksen tukena

Rajaukset

- Valitut indeksit: S&P 500 ja OMXH25
- Vältetään käännekohtia, jotta saataisiin stabiilina ajankohtana toimivat ennusteet
- Valitaan 2 väliä ja koko aikasarja:
 - 1.1.2015 – 31.12.2015
 - 1.3.2016 – 31.12.2017
 - 1.1.2015 – 31.12.2019

Aikasarjat: S&P 500 ja OMXH25



Skaalattu: vuoden ensimmäinen päätösarvo on 1

Menetelmät ja työkalut

- Aikasarjojen mallinnus: ARIMA
- Neuroverkkojen mallinnus: RNN
- Ohjelmointi: Python
- Valmiit kirjastot: TensorFlow, Keras ja Statsmodels
- Kuviot: Draw.io, Matplotlib

ARIMA ja keskiarvo

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + \sum_{i=1}^P \Phi_i x_{t-is} + \sum_{i=1}^q \theta_i \epsilon_{t-q} + \sum_{i=1}^Q \Theta_i \epsilon_{t-Qs} + \epsilon_t$$

MALLI 1

SARIMA(p,d,q)(P,D,Q)_s

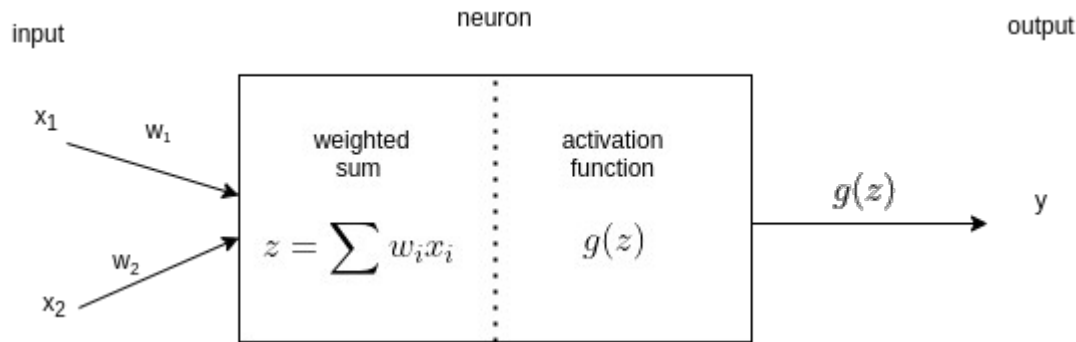
- Autoregressio, p, P
- Liukuva keskiarvo, q, Q
- Erotukset, d, D
- Kauden pituus, s

MALLI 2

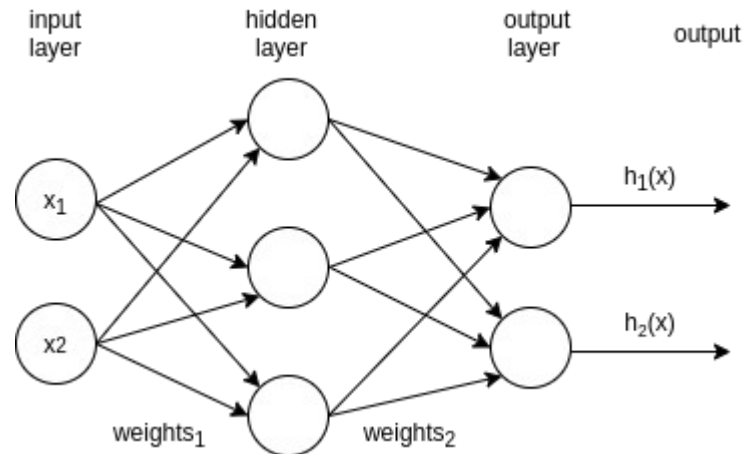
ARMA(0,0)

- differensioidun aikasarjan keskiarvo

Neuroverkko

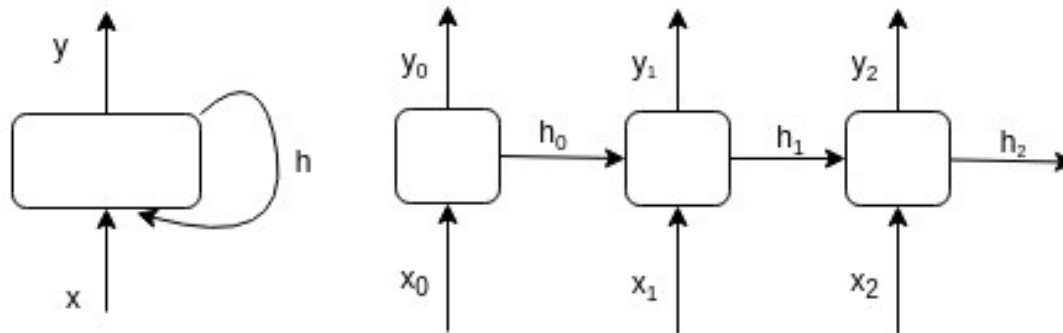


Neuroni laskee painotetun summan syötesignaaleista ja laskee summasta kuvauksen aktivaatiofunktiolla.



Neuronit asetetaan kerroksiksi ja painot valitaan siten, että tappiofunktio minimoidaan.

Takaisinkytketty neuroverkko, RNN



Toistuvissa neuroverkoissa neuronit muistavat aikaisemman aika-
askeleen tilansa ja välittävät sen syötteenä seuraavaan aika-askeleeseen.

Mallien toteukseen liittyviä valintoja

- Aikasarjat stationarisoitu ottamalla peräkkäisistä termeistä erotus
- Mallin harjoittamiseen käytettiin 200 havaintoa lyhyemmissä sarjoissa ja 800 havaintoa koko sarjaa tarkastellessa
- Mallien hyvyyslukuina käytettiin ennusteiden keskimääräisiä neliövirheitä

Mallit

ARIMA-mallit saatiin poistamalla trendi ja arvioimalla autokorrelaatio- sekä osittaisautokorrelaatiofunktioita.

Keskiarvoa, ARMA(0, 0), käytettiin kilpaileva mallina.

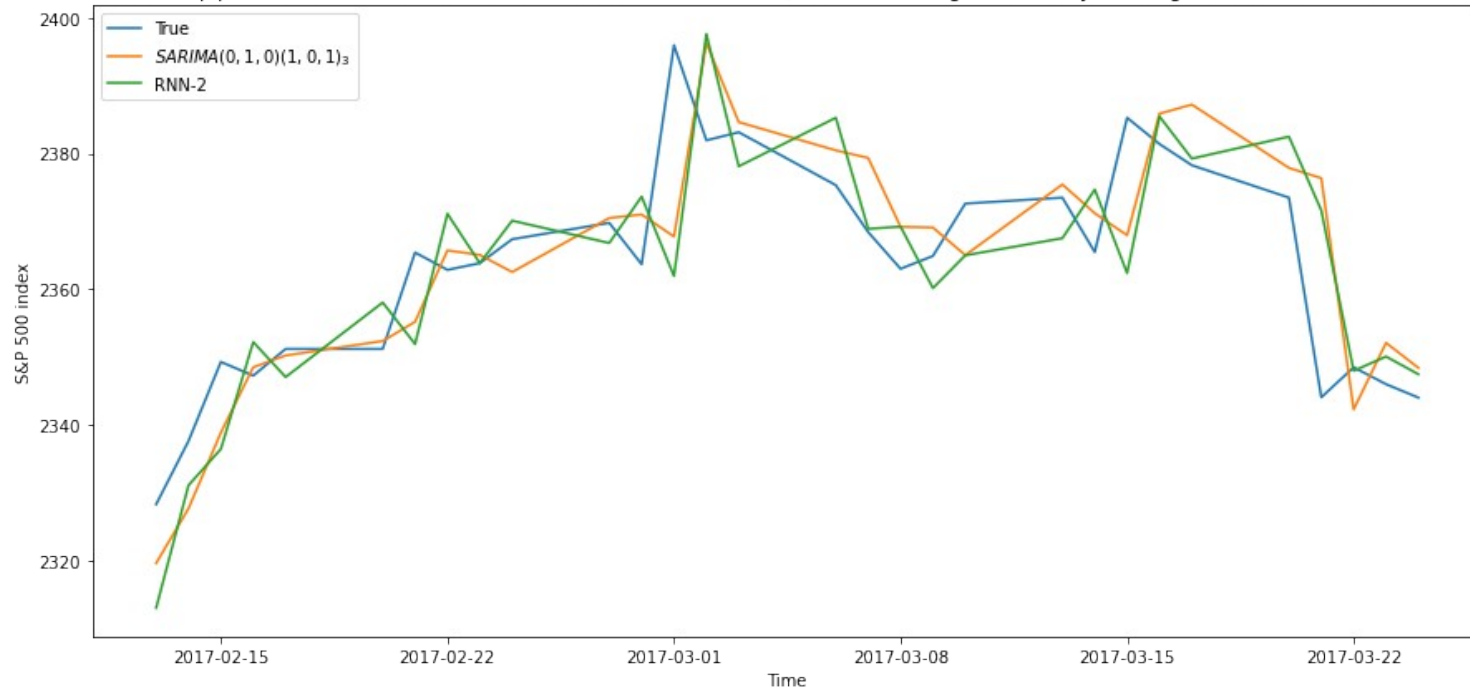
time series	model
S&P 500 2015	SARIMA(0,1,0)(1,0,1) ₄
OMXH25 2015	SARIMA(0,1,0)(1,0,1) ₄
S&P 500 2016-2017	SARIMA(0,1,0)(1,0,1) ₃
OMH25 2016-2017	SARIMA(0,1,0)(1,0,1) ₅
S&P 500 2015-2019	SARIMA(0,1,0)(1,0,1) ₆
OMXH25 2015-2019	SARIMA(0,1,0)(1,0,1) ₅

model	hidden layers	neurons in hidden layer	activation function	optimizer	loss	batch size	epochs
RNN-2	2	10	tanh	Adam	mse	5	5
RNN-3	3	20	ReLU	Adam	mse	5	10

Neuroverkkomallit saatiin kokeilemalla eri parametrien arvoja. Haluttiin kaksi hyvin toimivaa mallia.

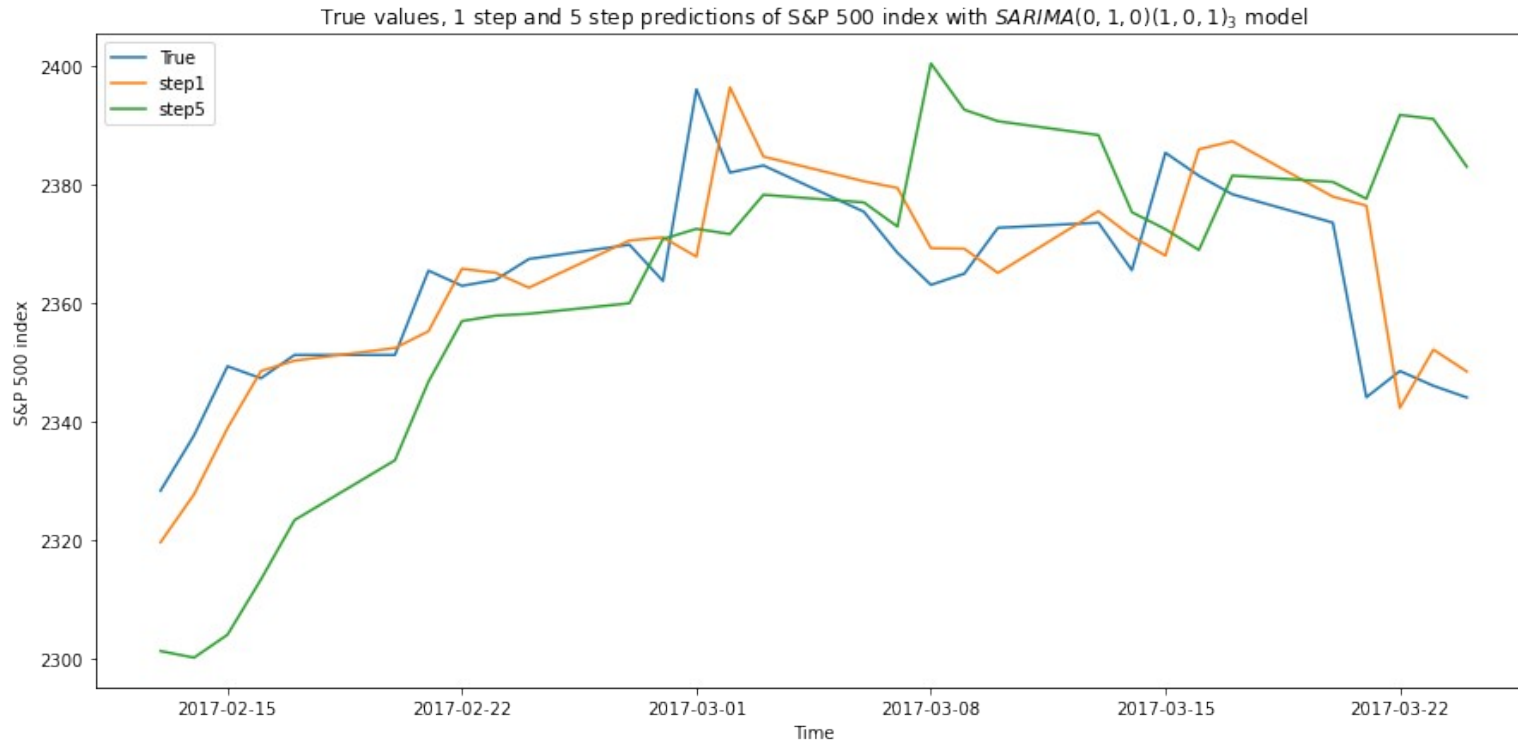
Ennusteet: ARIMA ja RNN

True values and 1 step predictions of S&P 500 index with $SARIMA(0, 1, 0)(1, 0, 1)_3$ and RNN having 2 hidden layers using time series from 1.3.2016 to 31.12.2017



ARIMA ja RNN ennusteet ovat hyvin lähellä toisiaan.

Ennusteet eivät ennako



Ennusteet eivät kykene ennakoimaan tulevia muutoksia, vaan ne näkyvät viiveellä.

Ennusteiden keskimääräiset neliövirheet

S&P 500 indeksi 1.3.2016 – 31.12.2017

models	1 step	2 step	3 step	4 step	5 step
SARIMA (0,1,0)(1,0,1) ₃	102,38	185,26	261,28	323,55	366,91
ARMA(0, 0)	103,67	185,90	264,40	325,47	369,61
RNN 3-layers	110,40	185,08	243,96	310,56	374,62
RNN 2-layers	109,48	192,72	265,39	337,28	400,89

OMXH25 indeksi 1.3.2016 – 31.12.2017

models	1 step	2 step	3 step	4 step	5 step
SARIMA (0,1,0)(1,0,1) ₅	688,74	1341,74	1911,97	2514,91	2976,72
ARMA(0, 0)	663,85	1290,58	1848,58	2440,88	2929,24
RNN 3-layers	678,81	1329,41	1892,12	2502,50	2972,88
RNN 2-layers	678,33	1383,60	1829,09	2451,42	2880,24

Yhden ennustemallin ei voida sanoa olevan paras. Huonoimman ja parhaan ennusteen välinen ero oli enimmillään 9,3 %.

Yhteenveto tuloksista

- ARIMA ja RNN tuottavat samankaltaisia ennusteita
- Monimutkaisempi malli ei paranna tai huononna ennustetta
- Tasaisimmassa sarjassa S&P 500 2016-2017 1 päivän ennusteen keskimääräinen neliö virhe vastaa 10 pisteen eli 0,4 - 0,5 % muutosta indeksissä.
- Huonoimmillaan koko OMXH25 indeksiä ennusteessa virhe vastaa 83 pisteen eli 2 - 3 % muutosta indeksissä
- Osakeindeksien ennustaminen on äärimmäisen vaikeaa

Aineistot

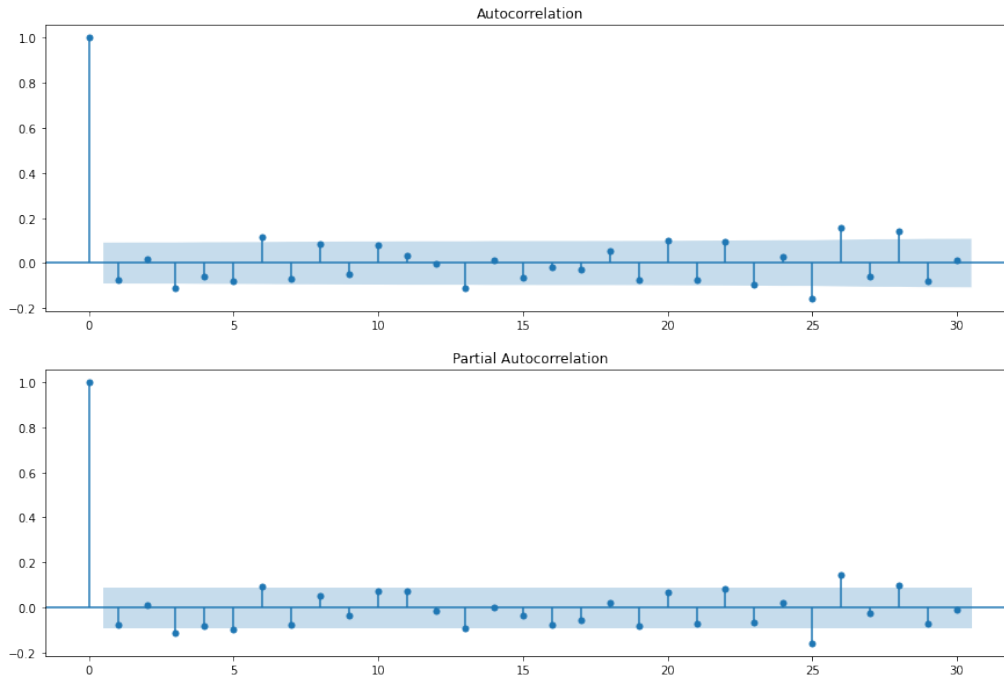
- S&P 500:
<https://www.nasdaq.com/market-activity/index/spx/historical>
- OMXH25:
<https://indexes.nasdaqomx.com/Index/History/OMXH25>

Tietolähteet

- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer Science & Business Media, 2009.
- Seabold, Skipper, and Josef Perktold. “*statsmodels: Econometric and statistical modeling with python.*” Proceedings of the 9th Python in Science Conference. 2010.
- Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and Tensor-Flow: Concepts, tools, and techniques to build intelligent systems*. O’Reilly Media, 2019.

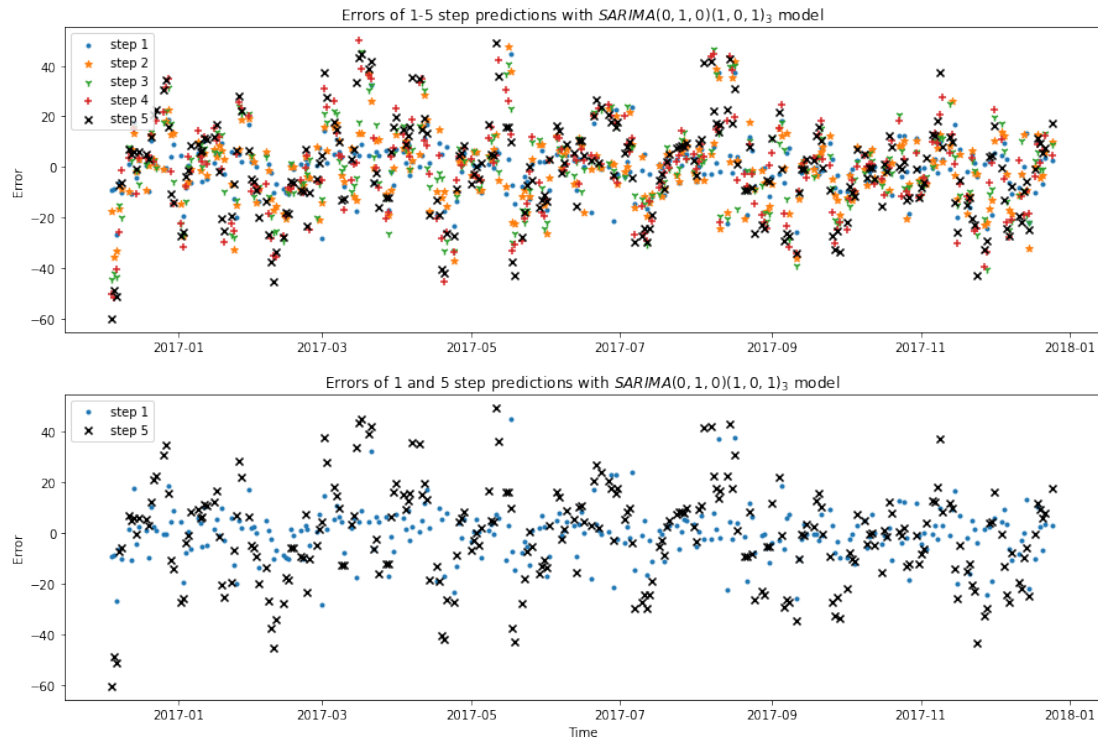
ARIMA mallin tunnistaminen

S&P 500 2016-2017



- Autoregressio (AR)
- Liukuva keskiarvo (MA)
- Differensioimalla trendi (I)
- Kausivaihtelu (S)

Ennustevirheet



Ennustevirheissä ei näy säännönmukaisuutta