

**Aki Taanila**

**AIKASARJAENNUSTAMINEN**

**30.5.2009**

# SISÄLLYS

JOHDANTO .....	1
Aikasarjaennustaminen .....	1
AIKASARJASSA EI OLE SYSTEMAATTISTA VAIHTELUA .....	2
Liukuvan keskiarvon menetelmä .....	2
Eksponentiaalinen tasoitus .....	3
AIKASARJASSA ON TRENDI.....	4
AIKASARJASSA ON TRENDI JA KAUSIVAIHTELU .....	5
Aikasarjaennusteissa huomioitavaa .....	7

## JOHDANTO

Tässä monisteessa käsitellään aikasarjaennustamista. Mukana ovat tarvittavat Excel-ohjeet (sopivat sekä Excel2007 että vanhempiin versioihin). Viimeisin versio tästä monisteesta ja siihen liittyvästä materiaalista löytyy osoitteesta

<http://myy.haaga-helia.fi/~taaak/m>

Yllä mainitusta osoitteesta löytyvät myös tähän monisteeseen liittyvät Excel esimerkit.

### Akin muita oppimateriaaleja

- Määrällisen tutkimuksen suunnittelu <http://myy.haaga-helia.fi/~taaak/t/>
- Aineiston esittäminen ja kuvailu <http://myy.haaga-helia.fi/~taaak/k/>
- Tilastollinen päättely <http://myy.haaga-helia.fi/~taaak/p/>

### Kommentit ja parannusehdotukset

Otan mielelläni vastaan kommentteja ja parannusehdotuksia sähköpostitse osoitteeseen [aki.taanila\(at\)haaga-helia.fi](mailto:aki.taanila(at)haaga-helia.fi).

### Aikasarjaennustaminen

Aikasarjaennustamisessa oletetaan, että toteutuneiden havaintojen muodostama aikasarja sisältää informaatiota, joka auttaa tulevien havaintojen ennustamisessa. Käytettävä ennustusmenetelmä riippuu siitä, minkälaista systemaattista vaihtelua aikasarjassa havaitaan. Seuraavassa tarkasteltavat tapaukset ovat:

- aikasarja ilman systemaattista vaihtelua
- aikasarja, jossa on trendi (pitkän aikavälin kehityssuunta)
- aikasarja, jossa on trendi ja kausivaihtelu (yleensä vuodenaikoihin liittyvä).

Olipa käytetty menetelmä mikä tahansa, niin ennusteet poikkeavat enemmän tai vähemmän toteutuvasta. Ennustusmenetelmän valinnassa pyritään siihen, että ennusteiden keskimääräinen virhe saadaan mahdollisimman pieneksi. Virhe voidaan yksinkertaisesti laskea ennusteen ja toteutuneen erotuksena. Tällöin keskimääräinen virhe saadaan virheiden itseisarvojen keskiarvona. Vaihtoehtoisten ennustusmenetelmien vertailu voidaan toteuttaa seuraavasti:

1. lasketaan menetelmän mukaisia ennusteita jo toteutuneita havaintoja vastaaville periodeille
2. lasketaan kuhunkin periodiin liittyvä virhe
3. lasketaan keskimääräinen virhe
4. valitaan ennustusmenetelmäksi se, joka tuottaa pienimmän keskimääräisen virheen.

## AIKASARJASSA EI OLE SYSTEMAATTISTA VAIHTELUA

Aikasarja ei ole ennustamisen kannalta hyödytön vaikka siinä esiintyykin ainoastaan satunnaista vaihtelua.

Esim. Jos tuotteen kysyntä vaihtelee satunnaisesti, niin viimeisin toteutunut kysyntä ei ole yleensä paras mahdollinen ennuste seuraavan periodin kysynnälle. Viimeisin toteutunut kysyntään voi sattumalta poiketa paljonkin kysynnän keskimääräisestä tasosta. Parempi ennuste saadaan laskemalla keskiarvo useamman periodin kysynnästä. Menetelmää kutsutaan liukuvan keskiarvon menetelmäksi.

### Liukuvan keskiarvon menetelmä

Jos ennuste lasketaan toteutuneiden havaintojen keskiarvona, niin laskijan on valittava, kuinka monen havainnon keskiarvoa käytetään. Mitä useamman havainnon keskiarvoa lasketaan sitä enemmän aikasarjassa esiintyvää vaihtelua tasoitetaan. Valinta tehdään yleensä siten että keskimääräinen ennustusvirhe saadaan mahdollisimman pieneksi.

Seuraavassa taulukossa on käytetty 5 viikon keskiarvoja.

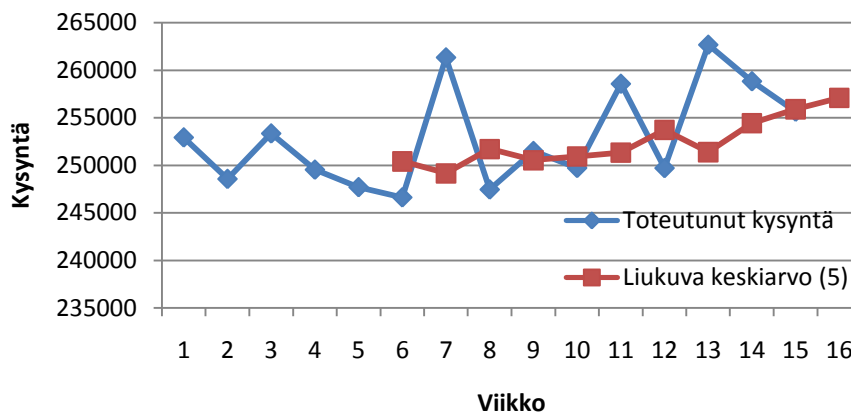
	A	B	C	D
	Viikko	Toteutunut kysyntä	5 viikon liukuva keskiarvo	Virheen itseisarvo
4				
5	1	252922		
6	2	248559		
7	3	253342		
8	4	249532		
9	5	247693		
10	6	246618	250410	3792
11	7	261333	249149	12184
12	8	247447	251704	4257
13	9	251492	250525	967
14	10	249713	250917	1204
15	11	258563	251321	7242
16	12	249702	253710	4008
17	13	262667	251383	11284
18	14	258829	254427	4402
19	15	255631	255895	264
20	16		257078	
21	Virheen itseisarvojen keskiarvo			4960

Solussa C10 on kaava  
 $=\text{AVERAGE}(\text{B5}:\text{B9})$  (suom.  
**KESKIARVO**)

Solussa D10 on kaava  $=\text{ABS}(\text{C10}-\text{B10})$  (suom. **ITSEISARVO**)

Solussa D21 on kaava  
 $=\text{AVERAGE}(\text{D10}:\text{D19})$  (suom.  
**KESKIARVO**)

Kuviosta nähdään, miten keskiarvojen käyttö on tasoittanut aikasarjassa esiintynyttä epäsäännöllistä vaihtelua.



## Ekspontiaalinen tasoitus

Vaihtoehtoinen ennustusmenetelmä on eksponentiaalinen tasoitus. Eksponentiaalisessa tasoituksessa ennuste lasketaan kaavalla:

$$\text{alfa} * \text{edellinen havainto} + (1 - \text{alfa}) * \text{edellinen ennuste}$$

Ennuste saadaan viimeisimmän tunnetun havainnon ja siihen liittyneen ennusteen painotettuna summana. Painokerroin alfa on välillä 0-1 oleva luku, joka ilmaisee, kuinka suurella painolla edellistä havaintoa painotetaan ennustetta laskettaessa. Jos alfa on 0, niin ennuste on sama kuin edellinen ennuste. Jos alfa on 1, niin ennuste on sama kuin edellinen toteutunut havainto. Suuret alfan arvot antavat ennusteita, jotka reagoivat herkästi aikasarjassa esiintyviin muutoksiin, koska viimeisimmillä havainnoilla on suurempi paino. Pienet alfan arvot tasoittavat voimakkaasti aikasarjan vaihtelua. Alfa arvo valitaan yleensä siten että keskimääräinen ennustusvirhe saadaan mahdollisimman pieneksi.

Ennusteen kaava voidaan kirjoittaa myös toiseen muotoon:

$$\text{edellinen ennuste} + \text{alfa} * (\text{edellinen havainto} - \text{edellinen ennuste}) = \\ \text{edellinen ennuste} + \text{alfa} * \text{virhe}$$

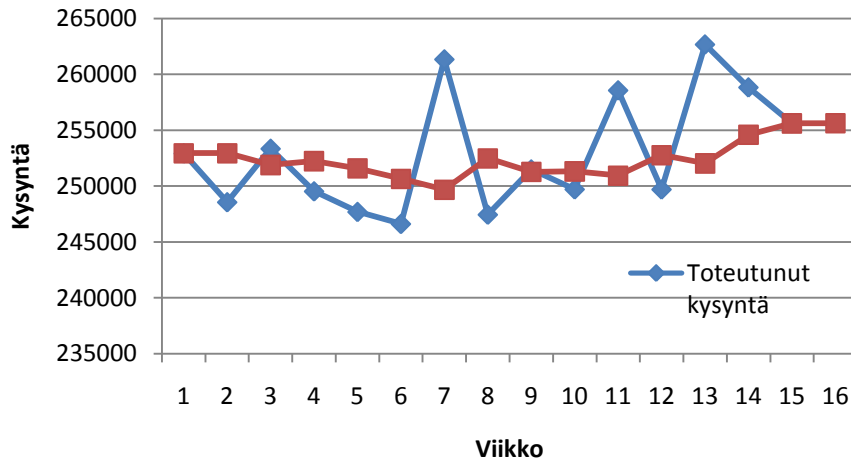
Ennustetta siis korjataan jokaisen toteutuneen havainnon jälkeen korjaustermillä  $\text{alfa} * \text{virhe}$ .

Seuraavan taulukossa ensimmäinen ennuste (viikko 1) on keskiarvo viikkojen 1-15 toteutuneista kysynnöistä. Solussa C5 on siis kaava  $=\text{AVERAGE}(B5:B19)$  (suom. **KESKIARVO**). Solussa C6 on kaava  $=\$F\$4 * B5 + (1 - \$F\$4) * C5$ .

	A	B	C	D	E	F
1	Ennustaminen eksponentiaalisella tasoituksella					
2	Aikasarjassa ei ole trendiä eikä kausivaihtelua					
3						
4	Viikko	Toteutunut kysyntä	Ennuste	Virheen itseisarvo	alfa = 0,24	
5	1	252922	252936			
6	2	248559	252933	4374		
7	3	253342	251883	1459		
8	4	249532	252233	2701		
9	5	247693	251585	3892		
10	6	246618	250651	4033		
11	7	261333	249683	11650		
12	8	247447	252479	5032		
13	9	251492	251271	221		
14	10	249713	251324	1611		
15	11	258563	250938	7625		
16	12	249702	252768	3066		
17	13	262667	252032	10635		
18	14	258829	254584	4245		
19	15	255631	255603	28		
20	16		255610			
21	Virheen itseisarvojen keskiarvo			4327		

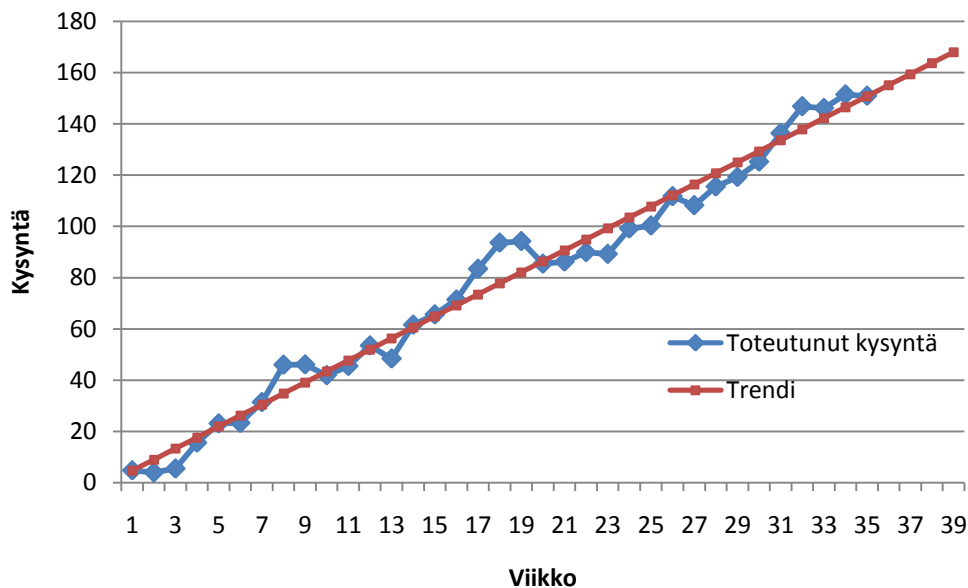
Sopiva alfan arvo, joka tuottaa pienimmän keskimääräisen ennustevirheen, löydetään helposti kokeilemalla.

Kuviosta nähdään, miten eksponentiaalinen tasoitus alfan arvolla 0,24 on tasoittanut alkuperäistä aikasarjaa.



## AIKASARJASSA ON TRENDI

Jos aikasarjassa on trendi, niin ennusteita voidaan laskea tapaukseen sopivaa regressiomallia käyttäen. Jos trendi on lineaarinen, niin ennusteena käytetään lineaarisen regressiomallin antamalta suoralta laskettuja pisteitä.



Excelissä on helppo laskea regressiosuoran pisteitä funktiolla

**=FORECAST(x; tunnetut x;tunnetut y) (suom. ENNUSTE)**

Funktion ensimmäinen lähtötieto **x** tarkoittaa sen periodin järjestysnumeroa, jolle lasketaan ennuste. Funktion toinen lähtötieto **tunnetut x** tarkoittaa niiden periodien järjestysnumeroita, joista on käytettävissä havainnot. Funktion kolmas lähtötieto **tunnetut y** tarkoittaa käytettävissä olevia havaintoja.

Tarkempia ennusteita saadaan yhdistämällä eksponentiaaliseen tasoitukseen trendi, jota korjataan jokaisen toteutuneen havainnon jälkeen. Mallia kutsutaan kehittäjänsä mukaan Holtin malliksi. Excel esimerkeistä löytyy laskettu esimerkki myös Holtin mallista.

## AIKASARJASSA ON TRENDI JA KAUSIVAIHTELU

Trendin ja kausivaihtelun huomiointi voidaan toteuttaa seuraavassa kuvattavien vaiheiden kautta:

### Lasketaan kausivaihtelusta puhdistettu aikasarja

Lasketaan kausivaihtelusta puhdistettu aikasarja liukuvia keskiarvoja käyttäen. Jos havainnot ovat vuosineljänneksittäin, niin kausivaihtelu puhdistuu neljän vuosineljänneksen keskiarvoilla (vastaavasti kuukausittaisista havainnoista puhdistetaan kausivaihtelu 12 kuukauden keskiarvoilla). Seuraavassa solun D7 liukuva keskiarvo on laskettu kaavalla

=**AVERAGE(C5:C8)** (suom. **KESKIARVO**)

ja solun D8 liukuva keskiarvo kaavalla

=**AVERAGE(C6:C9)** (suom. **KESKIARVO**)

	A	B	C	D	E
4	Neljännes		Toteutunut kysyntä	Liukuva ka.	Keskistetty liukuva ka.
5	1	1	8		
6	2	2	13		
7	3	3	23	19,50	19,75
8	4	4	34	20,00	20,63
9	1	5	10	21,25	21,25
10	2	6	18	21,25	21,75
11	3	7	23	22,25	22,50
12	4	8	38	22,75	22,13
13	1	9	12	21,50	22,63
14	2	10	13	23,75	24,13
15	3	11	32	24,50	
16	4	12	41		

Koska soluun D7 sijoitettu keskiarvo on neljännesten 1-4 keskiarvo, niin sen oikea paikka olisi neljännesten 2 ja 3 puolivälissä. Koska soluun D8 sijoitettu keskiarvo on neljännesten 2-5 keskiarvo, niin sen oikea paikka olisi neljännesten 3 ja 4 puolivälissä. Jotta keskiarvot voidaan sijoittaa taulukkoon asianmukaisesti, lasketaan vielä kahden liukuvan keskiarvon keskiarvot (keskistetyt liukuvat keskiarvot). Solujen D7 ja D8 keskiarvon oikea paikka on neljänneksen 3 kohdalla. Näin ollen soluun E7 on laskettu keskistetty liukuva keskiarvo kaavalla

=**AVERAGE(D7:D8)** (suom. **KESKIARVO**)

Keskistettyjen liukuvien keskiarvojen muodostamaa aikasarjaa voidaan pitää kausivaihtelusta puhdistettuna aikasarjana.

### Lasketaan trendi

Trendi lasketaan kausivaihtelusta puhdistetun aikasarjan avulla. Esimerkissämme periodiin 1 liittyvä trendi saadaan Excelissä kaavalla

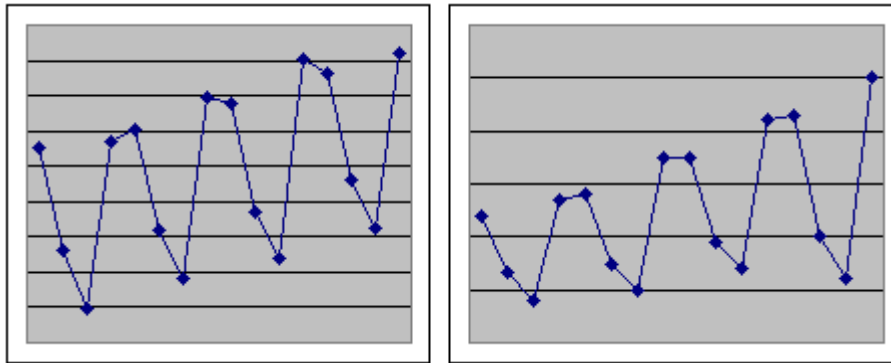
=**FORECAST(B5,\$B\$7:\$B\$14,\$E\$7:\$E\$14)** (suom. **ENNUSTE**)

Kaavaa alaspäin kopioimalla saadaan muihin periodeihin liittyvät trendit.

## Lasketaan kausivaihtelun suuruus

Kausivaihtelun suuruuden laskemisessa voidaan soveltaa summamallia tai tulomallia.

Summamalli soveltuu hyvin tilanteisiin, joissa kausivaihtelun absoluuttinen suuruus on ennustettavan muuttujan arvoista riippumaton. Tulomalli taas soveltuu tilanteisiin, joissa kausivaihtelun absoluuttinen suuruus on sitä suurempi, mitä suurempi on ennustettavan muuttujan arvo. Seuraavista kuvista vasemmanpuoleiseen sopii summamalli, kun taas oikeanpuoleiseen tulomalli on sopivampi.



Summamallissa lasketaan kutakin havaintoa vastaava kausivaihtelu toteutuneen havainnon ja trendin erotuksena. Tulomallissa lasketaan kausivaihtelu toteutuneen havainnon ja trendin osamääränä. Seuraavassa taulukossa solun G5 kausivaihtelu on laskettu kaavalla  $=C5/F5$  (tulomalli).

Jos samasta periodista (esim. samasta vuosineljänneksestä) on käytettävissä useampia arvoja, niin lopullinen kausivaihtelu saadaan keskiarvona. Seuraavassa taulukossa solun K5 ensimmäisen neljänneksen kausivaihtelu on laskettu kaavalla  $=AVERAGE(G5;G9;G13)$  (suom. **KESKIARVO**).

## Korjataan trendiä kausivaihtelun verran

Lopulliset ennusteet saadaan korjaamalla trendiä kausivaihtelun verran. Summamallissa trendiin lisätään kyseessä olevan periodin kausivaihtelu. Tulomallissa trendi kerrotaan kyseessä olevan periodin kausivaihtelulla. Seuraavassa taulukossa solun H17 ennuste on laskettu kaavalla  $=F17*K5$  (tulomalli).

	A	B	C	D	E	F	G	H	I	J	K
4	Neljännes		Toteutunut kysyntä	Liukuva ka.	Keskistetty liukuva ka.	Trendi	Kausivaihtelu	Ennuste	Virheen itseisarvo	Neljännes	Kausivaihtelu
5	1	1	8			18,96	0,42	8,94	0,94	1	0,47
6	2	2	13			19,49	0,67	13,32	0,32	2	0,68
7	3	3	23	19,50	19,75	20,01	1,15	23,43	0,43	3	1,17
8	4	4	34	20,00	20,63	20,53	1,66	34,18	0,18	4	1,66
9	1	5	10	21,25	21,25	21,06	0,47	9,93	0,07		
10	2	6	18	21,25	21,75	21,58	0,83	14,75	3,25		
11	3	7	23	22,25	22,50	22,11	1,04	25,88	2,88		
12	4	8	38	22,75	22,13	22,63	1,68	37,66	0,34		
13	1	9	12	21,50	22,63	23,15	0,52	10,92	1,08		
14	2	10	13	23,75	24,13	23,68	0,55	16,18	3,18		
15	3	11	32	24,50		24,20	1,32	28,33	3,67		
16	4	12	41			24,72	1,66	41,15	0,15		
17	5	13				25,25		11,91			
18	6	14				25,77		17,61			
19	7	15				26,30		30,79			
20	8	16				26,82		44,64			
21	Virheen itseisarvojen keskiarvo:								1,37		

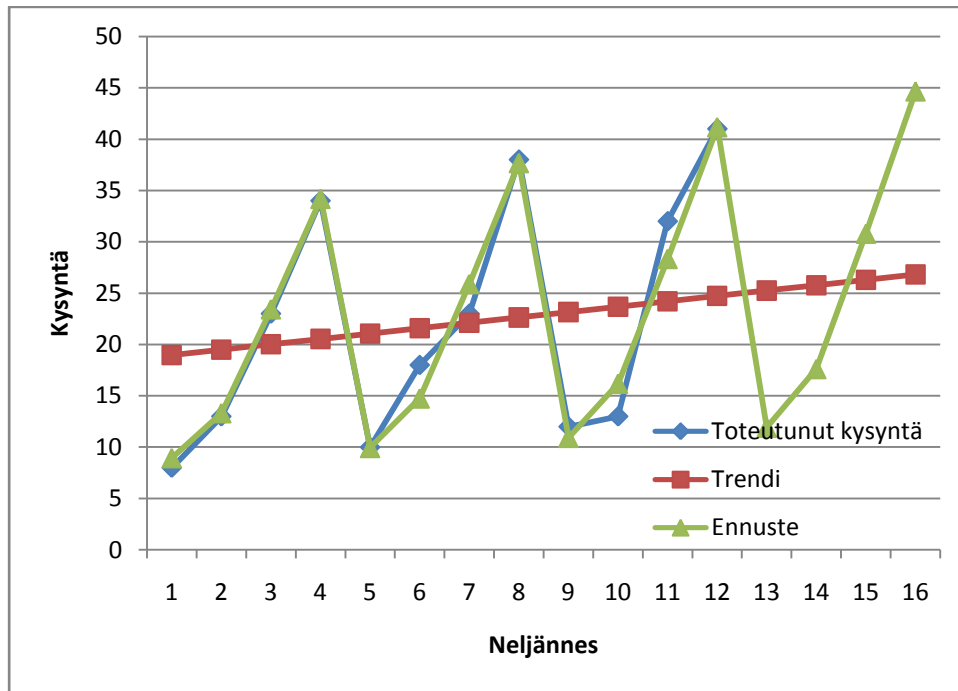
Tarkempia ennusteita saadaan laajentamalla Holtin malli Holt-Winterin malliksi, jossa kausivaihtelua korjataan uusien havaintojen myötä. Excel esimerkeistä löytyy laskettu esimerkki myös Holt-Winterin mallista.

## Aikasarjaennusteissa huomioitavaa

Edellä kuvatun kaltaiset ennusteet ovat usein vasta lähtökohta lopulliselle ennusteelle. Ennustetta voidaan korjata esimerkiksi

- ennustettavaan ilmiöön liittyvän kokemustiedon pohjalta
- tutkimuslaitoksilta saatavien suhdanne-ennusteiden pohjalta
- suunniteltujen kampanjoiden tai hinnanmuutosten johdosta.

Ennustusvirheeseen on kiinnitettävä huomiota muutoinkin kuin ennustusmenetelmää valittaessa. Keskimääräinen ennustusvirhe antaa hyvän kuvan siitä kuinka tarkkoja ennusteita menetelmän voidaan odottaa antavan. Ennusteiden tarkkuutta kannattaa aina arvioida myös kuvion avulla. Seuraava kuvio kuvaa aiemmin esitettyyn esimerkkilaskelmaan liittyviä ennusteita.



Kuvion mukaan ennusteet näyttävät seuraavan hyvin toteutunutta kysyntää ja näin ollen voidaan odottaa ennusteiden toteutuvan hyvin myös jatkossa (olettaen, että toimintaympäristössä ei tapahdu olennaisia muutoksia).

### Valmisohjelmat

Aikasarjaennustamiseen on olemassa valmiita tietokoneohjelmia. Esimerkiksi moniin toiminnanohjausjärjestelmiin liittyy aikasarjaennustamisen osio. Vaikka ohjelmat suorittavatkin tarvittavat laskutoimitukset, niin käyttäjän vastuulle jää tilanteeseen sopivan mallin valinta ja määrittely.