Методи прогнозування

ЄВГЕН ОЛІЙНИК СНм - 51

Вступ

- Методи прогнозування застосовуються:
  - в економіці;
  - в медицині;
  - в техніці;
  - промисловості;
  - інших галузях, що пов’язані із числовими рядами
- Загальновизнанані методи прогнозування тимчасових рядів:
  - економетричні;
  - регресійні;
  - методи Бокса-дженкінса (ARIMA, ARMA);
  - нейромережні методи прогнозування.

1 Методи прогнозування, засновані на згладжуванні, експоненційному згладжуванні і ковзному середньому

1.1 "Наївні" моделі прогнозування

- основне припущення “завтра буде як сьогодні”;
- деякий основний період прогнозованого тимчасового ряду краще всього описує майбутнє цього прогнозованого ряду;
- найпростіша модель такого ряду **Y(t+1)=Y(t)**;
- моделі, що намагачться врахувати можливі тенденції

$$Y(t+1)=Y(t)+[Y(t)-Y(t-1)],$$

$$Y(t+1)=Y(t)*[Y(t)/Y(t-1)].$$

- модель, що намагається врахувати сезонні коливання

$$Y(t+1)=Y(t-s)$$

- недоліки:
  - Невелика точність;
  - Не враховують механізми, що визначають прогнозовані дані;
  - не захищені від випадкових коливань;
  - не враховують сезонні коливання і тенденції.

1 Методи прогнозування, засновані на згладжуванні, експоненційному згладжуванні і ковзному середньому

1.2 Середні і ковзаючі середні

- основне припущення “завтра буде як було в середньому за останній час”;
- значення тимчасового ряду з недалекого минулого краще описують прогноз, ніж усі попередні значення цього ж ряду;
- найпростіша модель на простому усередненюванні

Y(t+1)=(1/(t))*[Y(t)+Y(t-1)+...+Y(1)];

- Модель прогнозування заснована на ковзному середньому:

Y(t+1)=(1/(T+1))*[Y(t)+Y(t-1)+...+Y(t-T)].

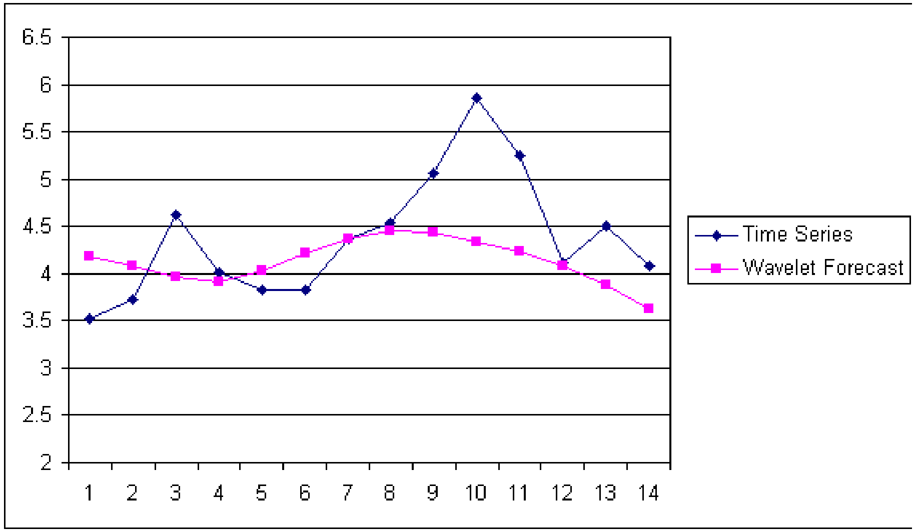
модель бачить тільки найближче минуле, T відліків за часом в глибину, грунтуючись на цих даних будує прогноз.

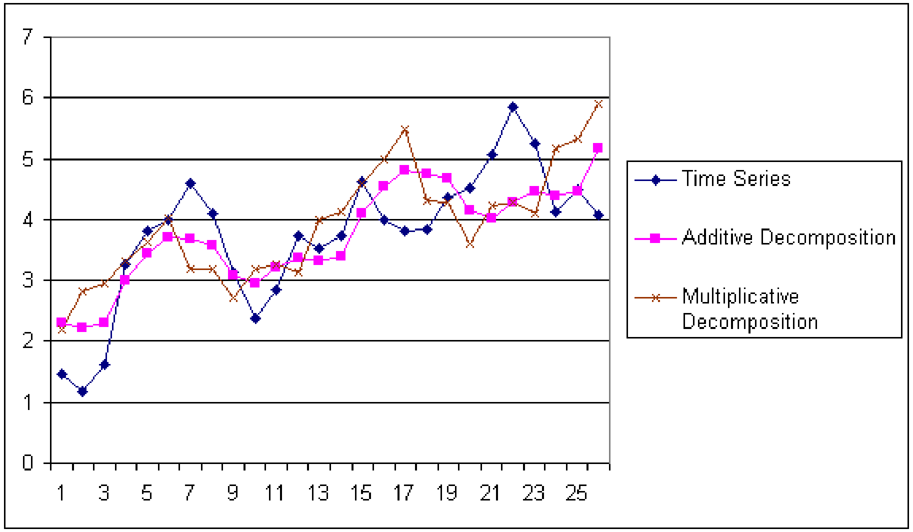
- модель, заснована на методі експоненційних середніх, яка постійно адаптується до нових значень:

Y(t+1)=*aY(t)+(1-a)*^Y(t),**

де **Y(t+1)** – прогноз на наступний період часу, **Y(t)** – реальне значення у момент часу t, **^Y(t)** – минулий прогноз на момент часу t, ***a*** – постійна згладжування **(0<=a<=1).**

Приклад





Прогнози при різних методах прогнозування

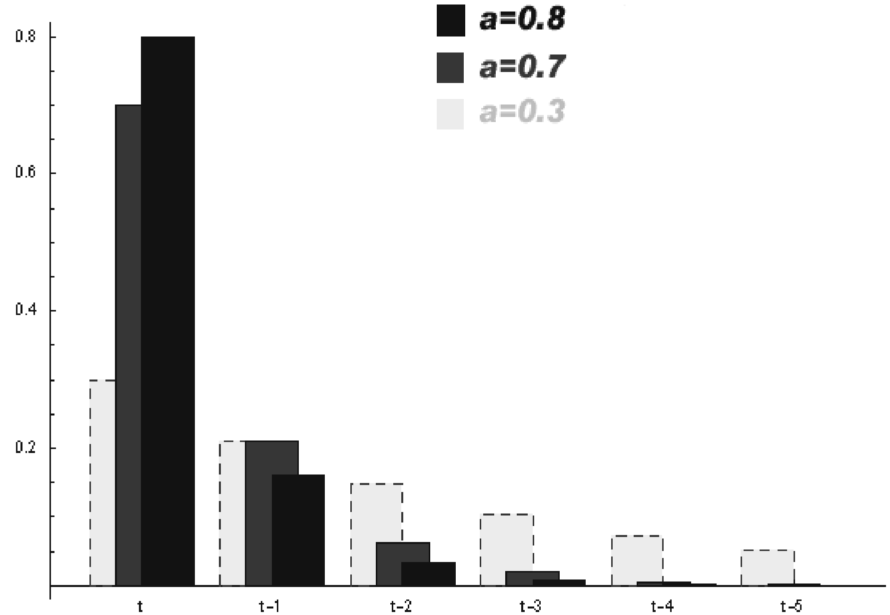
Приклад

Внутрішній параметр *a,* визначає залежність прогнозу від усіх розглянутих даних, причому вплив даних на прогноз експоненціально зменшується із "віком" даних.

При *a→1*, експоненціальна модель прямує до найпростішої "наївної" моделі.

При *a→0*, прогнозована величина стає рівною попередньому прогнозу.

На тестовому наборі будуються прогнози *a=[*0.01, 0.02 ..., 0.98, 0.99] і відстежується, при якому *a* точність прогнозування вища. Це значення *a* використовується при подальшому прогнозуванні .



Залежність впливу даних на прогноз при різних коефіцієнтах *a.*

2 Методи Хольта і Брауна

В даному алгоритмі значення рівня і тенденції згладжуються за допомогою експоненціального згладжування.

Параметри згладжування рівня і тенденції різні.

Основна система рівнянь:

$$\begin{cases} \Omega_t = \alpha Y_t + (1-\alpha)(\Omega_{t-1} - T_{t-1}), \\ T_t = \beta(\Omega_t - \Omega_{t-1}) + (1-\beta)T_{t-1}, \\ \hat{Y}_{t+p} = \Omega_t + pT_t \end{cases}$$

- Перше рівняння описує згладжений ряд загального рівня.
- Друге рівняння служить для оцінки тенденції.
- Третє рівняння визначає прогноз на *p* відліків за часом вперед.

Завжди можна підібрати таку пару параметрів, яка дає велику точність на тестовому наборі і потім використовувати цю пару, при реальному прогнозуванні.

Окремим випадком методу Хольта є метод Брауна, коли $\alpha = \beta$.

3 Метод Вінтерса

В даному методі робиться спроба враховувати сезонні складові даних, на відміну від попереднього, який не міг “бачити” такі складові.

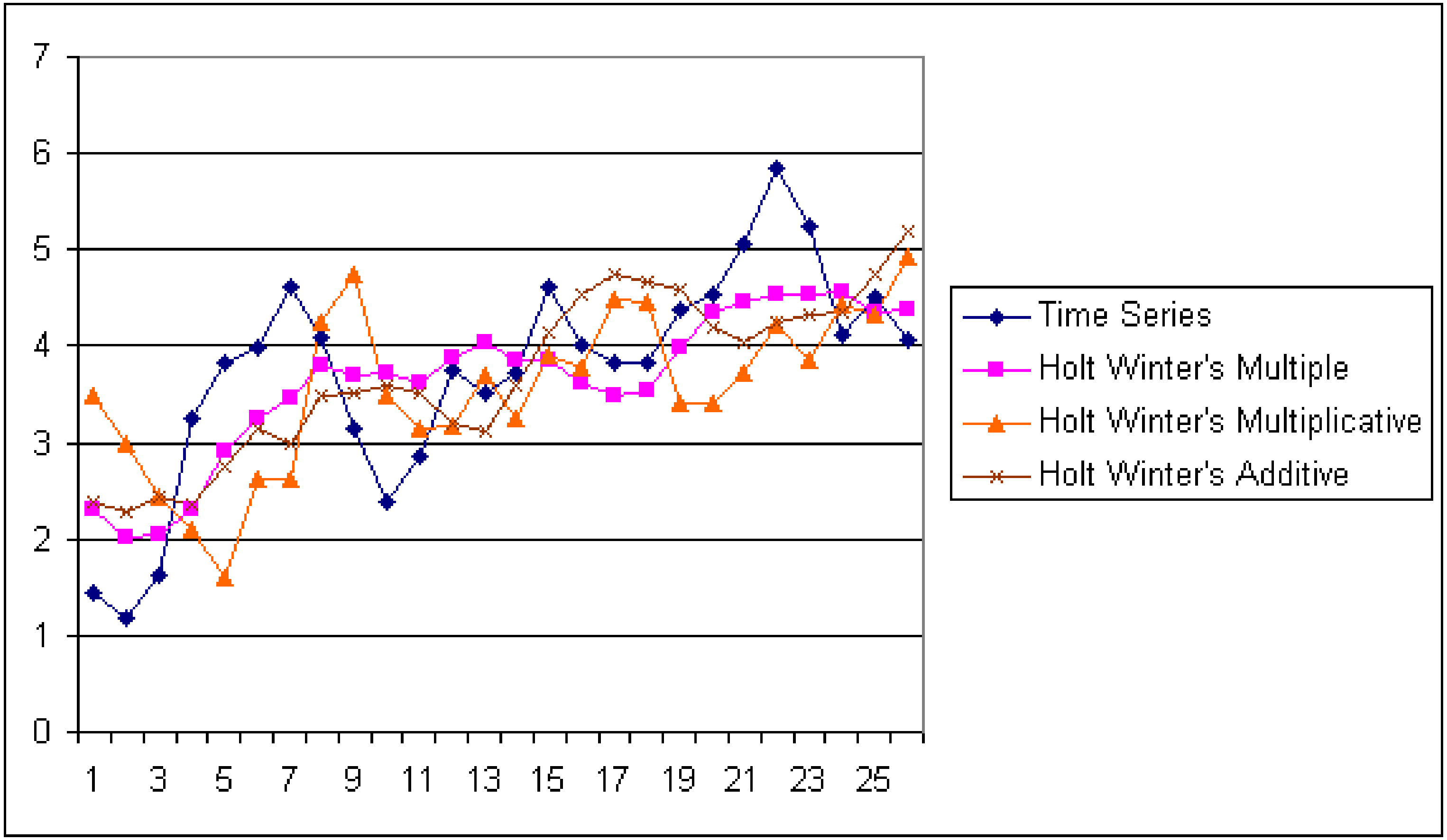
Даний метод фактично є розширенням попереднього.

Основна система рівнянь:

$$\begin{cases} \Omega_t = \alpha \dfrac{Y_t}{S_{t-s}} + \ 1-\alpha \quad \Omega_{t-1} - T_{t-1} \ , \\ T_t = \beta \ \ \Omega_t - \Omega_{t-1} \ + \ 1-\beta \ T_{t-1}, \\ S_t = \Upsilon \dfrac{Y_t}{\Omega_t} + \ 1-\Upsilon \ S_{t-s}, \\ \hat{Y}_{t+p} = \ \Omega_t + pT_t \ \ S_{t-s+p} \end{cases}$$

Дріб в першому рівнянні служить для виключення сезонності з Y(t). Після виключення сезонності алгоритм працює з "чистими" даними, в яких немає сезонних коливань. З'являються вони вже в самому фінальному прогнозі, коли "чистий" прогноз, порахований майже по методу Хольта умножається на сезонний коефіцієнт.

Приклад



Методи Вінтерса

4 Регресійні методи прогнозування

Суть методів регресійного прогнозування:

- Існує прогнозована змінна Y (залежна змінна) і відібраний заздалегідь комплект змінних, від яких вона залежить, X1, X2 ..., XN (незалежні змінні).
- Природа незалежних змінних різна.
- Модель множинної регресії в загальному випадку описується виразом: $Y = F\left(X_1, X_2, \ldots, X_N\right) + \varepsilon$
- У варіанті лінійної регресійної моделі залежність залежної змінної від незалежних має вигляд:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_N X_N + \varepsilon$$

- $\beta_1, \beta_2, \ldots, \beta_N$ – підбирані коефіцієнти регресії
- ε – компонента помилки. Передбачається, що всі помилки незалежні і нормально розподілені.
- Знайдені моделі обов’язково потрібно перевіряти на адекватність.

Приклад

Якщо припустити, що Y - рівень попиту на деякий продукт в наступному місяці, то незалежними змінними можуть бути:

- рівень попиту на цей же продукт в минулий і позаминулий місяці,
- витрати на рекламу,
- рівень платоспроможності населення,
- економічна обстановка,
- діяльність конкурентів і т.д.

Для побудови регресійних моделей необхідно мати базу даних спостережень приблизно такого вигляду:

	змінні				
	незалежні				**залежна**
№	***X1***	***X2***	**...**	***XN***	***Y***
1	x_11	x_12	...	x_1N	Y_1
2	x_21	x_22	...	x_2N	Y_2
...	...	...	...	...	...
m	x_M1	x_M2	...	x_MN	Y_m

5 Методи Бокса-дженкинса (ARIMA)

- ***A**uto**r**egressive **I**ntegrated **M**oving **A**verage*
  - Заснування класу методів: середина 90-х років минулого століття.
  - *Засновники* : G.E.P. Box, G.M. Jenkins.
  - *Дані методи широко вбудовуються в спеціалізовані пакети для прогнозування.*
  - *Моделі спираються тільки на інформацію, що міститься в передісторії прогнозованих рядів.*
  - *У класичному варіанті ARIMA не використовує незалежні змінні.*
  - *В методології ARIMA не передбачається якої-небудь чіткої моделі для прогнозування даної тимчасової серії.*
  - *Задається загальний клас моделей, що описують часовий ряд, які дозволяють виражати поточне значення змінної через її попередні значення.*
  - *Алгоритм, підстроюючи внутрішні параметри, сам вибирає найбільш відповідну модель прогнозування.*
- Ієрархія моделей Бокса-дженкінса:

AR(p)+MA(q)→ARMA(p,q)→ARMA(p,q)(P,Q)→ARIMA(p,q,r)(P,Q,R)→...

5 Методи Бокса-дженкинса (ARIMA)

- **5.1 AR(p) -авторегресивна модель порядку p.**
  - Модель має вигляд:

$$Y(t) = f_0 + f_1 \cdot Y(t-1) + f_2 \cdot Y(t-2) + \ldots + f_p \cdot Y(t-p) + E(t)$$

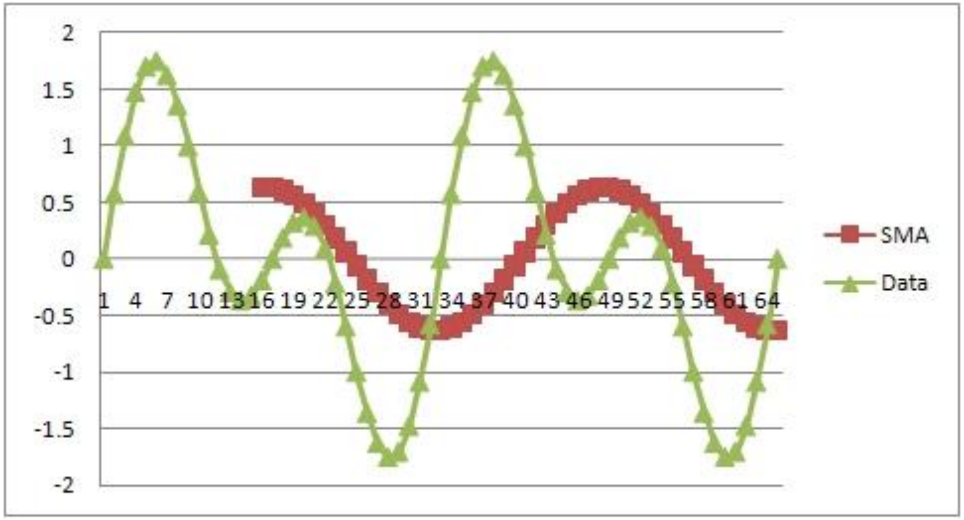
Y(t) – залежна змінна у момент часу t;

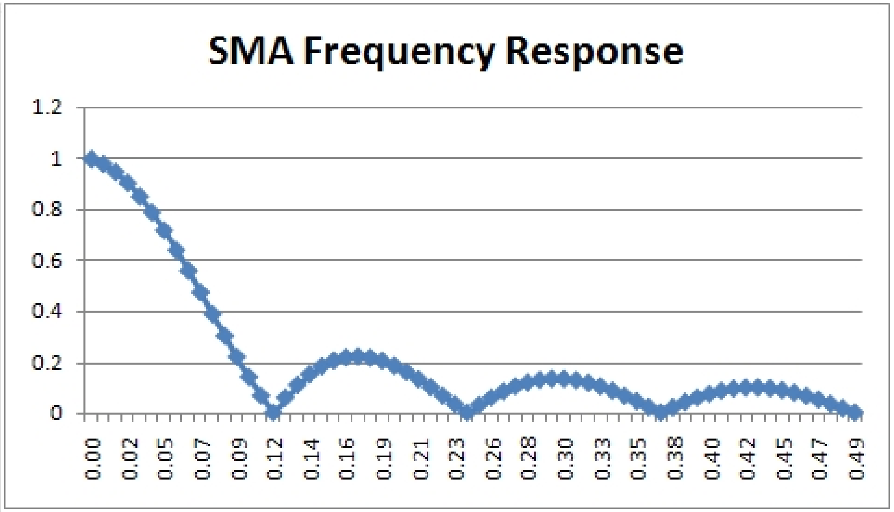
$f_0, f_1, f_2 \ldots, f_p$ - оцінювані параметри;

E(t) - помилка від впливу змінних, які не враховуються в даній моделі.

  - Завдання визначити $f_0, f_1, f_2 \ldots, f_p$
    - через систему рівнянь Юла-уолкера(потребує розрахунку значень автокореляційної функції);
    - методом найменших квадратів.

Приклад





Моделювання із ковзаючим середнім, спад адекватності моделі при ковзному середньому.

5 Методи Бокса-дженкинса (ARIMA)

- **5.2 MA(q) -модель з ковзаючим середнім порядку q.**
  - Модель має вигляд:

$$Y(t) = m + e(t) - w_1 \cdot e(t-1) - w_2 \cdot e(t-2) - \ldots - w_p \cdot e(t-p)$$

Y(t) – залежна змінна у момент часу t;

w_0, w_1, w_2 ..., w_p - оцінювані параметри;

- **5.3 Авторегресійне ковзне середнє ARMA(*p,q*).**
  - модель, *p* авторегресійних складових, *що* містить *q,* ковзаючих середніх. Модель ARMA(*p,q*) включає моделі AR(*p*) і MA(*q*)
  - Модель має вигляд:

$$X_t = c + e_t + \sum_{i=1}^{q} \theta_i e_{t-i} + \sum_{i=1}^{p} \varphi_i X_{t-i},$$

ε_t – значення помилки, вважають, що вона нормально розподілена із нульовим середнім $e_t \sim N(0, \sigma^2)$,

5 Методи Бокса-дженкинса (ARIMA)

- **5.4 ARIMA (p,d,q).**
  - Залежно від аналізованого ряду модель *ARIMA (p,d,q)* може трансформуватися до авторегресійної моделі *AR(p),* моделі ковзного середнього *MA(q)* або змішаній моделі *ARMA (p,q)*.
  - При переході від нестаціонарного ряду до стаціонарного значення параметра *d*, що визначає порядок різниці, приймається рівним 0 або 1.
  - З поля зору дослідників випадає ситуація, коли параметр *d* може приймати дробові значення.

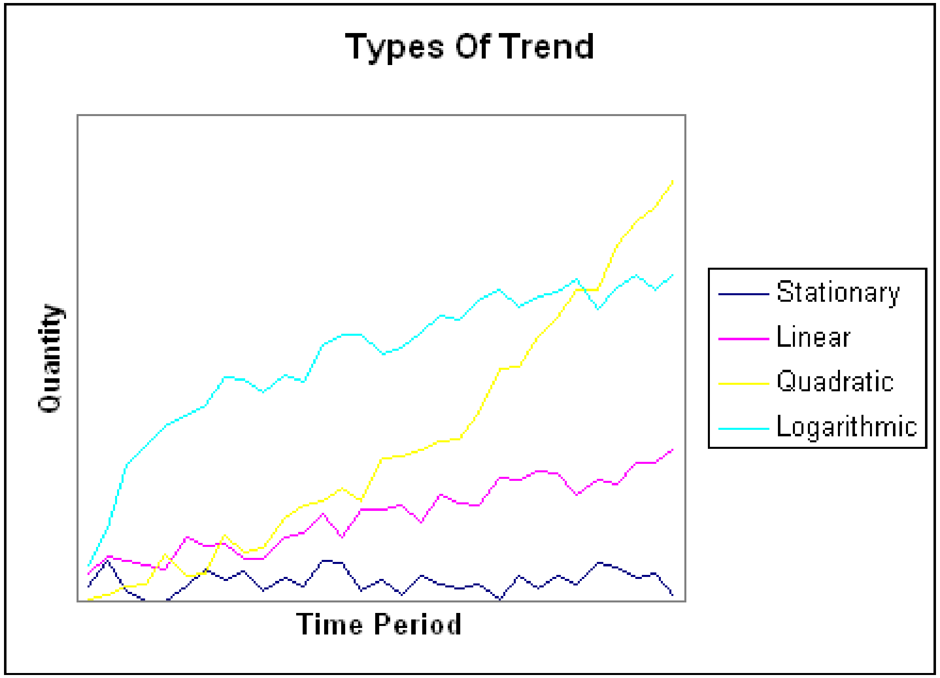
5 Методи Бокса-дженкинса (ARIMA)

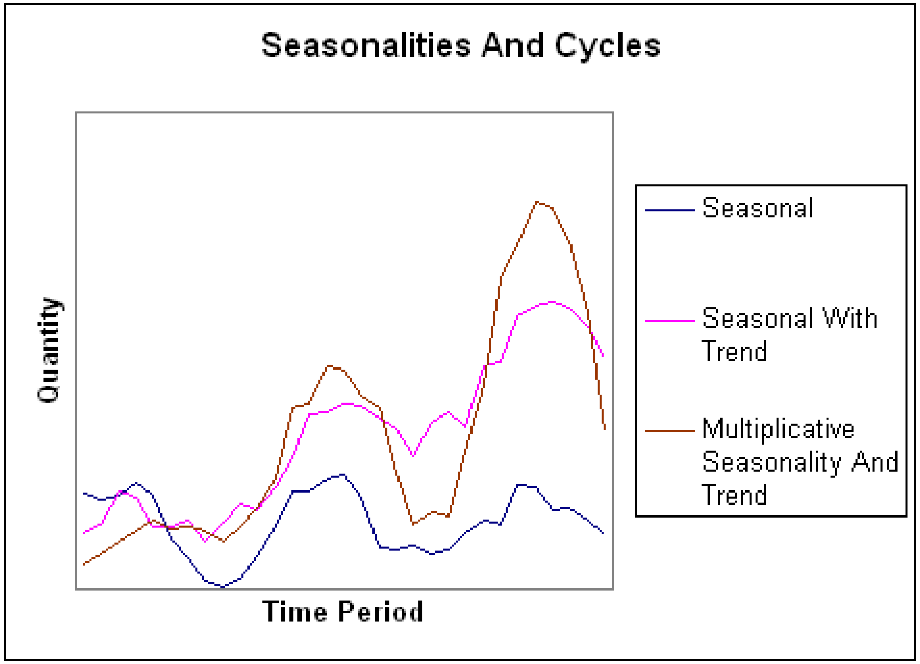
- **5.5 ARFIMA(p,d,q)** *F: fractional* - дріб
  - *C.W.Granger, J.R.Hosking, P.M.Robinson, R. Beran.*
  - допускає можливість нецілого параметра *d* авторегресійний дріб.
  - Інтегрований процес ковзного середнього.
  - Специфіка таких рядів:
    - самоподібність;
    - дробовою розмірність;
    - поволі спадаюча кореляція.
  - Прогнозування тимчасових рядів за допомогою моделі ARFIMA(p,d,q) відкриває ширші перспективи для підвищення точності прогнозу.
- **5.6 Модель вигляду ARIMA (p,d,q)(P,D,Q)S**
  - *p* - авторегресійні доданки;
  - *D* - різниці;
  - *q* - доданки ковзаючого середнього;
  - *P* – сезонні авторегресійні доданки;
  - *D* – сезонні різниці на інтервалі *S*;
  - *Q* – доданки сезонного ковзаючого середнього

6 Метод "Гусениці" SSA

- Може бути використаний для різних загальних завдань дослідження тимчасових рядів:
  - зокрема - для виділення сигналу і знаходження його ЛРФ.
- При його використанні по ряду F_N будується траєкторна матриця X заданого розміру L x K, 1 < L < N.
- Алгоритм:
  - 1. Вибір довжини вікна L і побудова траєкторної матриці $X \in \mathbb{R}^{L \times K}$ по ряду F_N;
  - 2. Сингулярне розкладання траєкторної матриці $X = \sum_i \sqrt{\lambda_i} U_i V_i^T$;
  - 3. Вибір сингулярних трійок, відповідних сигналу S_N;
  - 4. Побудова по власних векторах вибраних сингулярних трійок наближеної ЛРФ сигналу порядку L – 1;
  - 5. Знаходження кореня характеристичного полінома цієї ЛРФ;
  - 6. Пошук головного кореня серед всієї безлічі коренів;
  - 7. Отримання наближеною мінімальною ЛРФ (порядка 2) сигналу по головному кореню.

Приклад





Розклад на тренди, та сезонно циклічний розклад із прогнозуванням.

Список використаних джерел

- http://ipu-conf.ru/kmu/sbornik_VMKPU2008.pdf (Лютий 2010);
- http://www.guap.ru/guap/main/avtoref_krichevsky.doc (Лютий 2010);
- http://www.machinelearning.ru/wiki/index.php?title= Авторегрессионное_скользящее_среднее (Лютий 2010);
- www.neuroproject.ru/forecasting_tutorial.php#mlp (Лютий 2010);
- http://www.pdmi.ras.ru/~theo/autossa/files/SSAvsREGR--paper.pdf (Лютий 2010);
- http://www.nsu.ru/ef/tsy/ecmr/Forum/topic77.htm (Лютий 2010);

ДЯКУЮ ЗА УВАГУ