Тема лекции №7

Авторегрессионная модель

Цель лекции - изучить особенности авторегрессионого моделирования

План лекции.

1. Авторегрессонная модель и её виды
2. Операторное представление
3. Автокорреляционная функция

1. Авторегрессонная модель и её виды

Авторегрессионная (**AR-**) **модель** — модель временных рядов, в которой значения временного ряда в данный момент линейно зависят от предыдущих значений этого же ряда.

Общий вид модели авторегрессии:

$$Y_i = a_0 + \Sigma a_i * Y_{i-1} + \mathcal{E}_i$$

где a_0 — постоянная - коэффициент описывающий ситуацию прохождение влияющих факторов через начало координат, то есть показывает каким будет итог модели в случае, когда влияющие факторы равны нулю;

a_i — коэффициенты, которые описывают степень зависимости итогового Y от влияющих факторов, в данном случае, от того каким был Y в прошлом периоде регрессии;

Y_{i-1} — влияющие факторы, которые в данном случае и есть итоговый Y, но тот, каким он был раньше;

$\mathcal{E}_i$ — случайная компонента или как еще ее принято называть погрешность модели (по сути, это разница между расчетным значением модели за известные периоды и между самими известными значениями, то есть ***Yрасч. - Y***).

Виды моделей

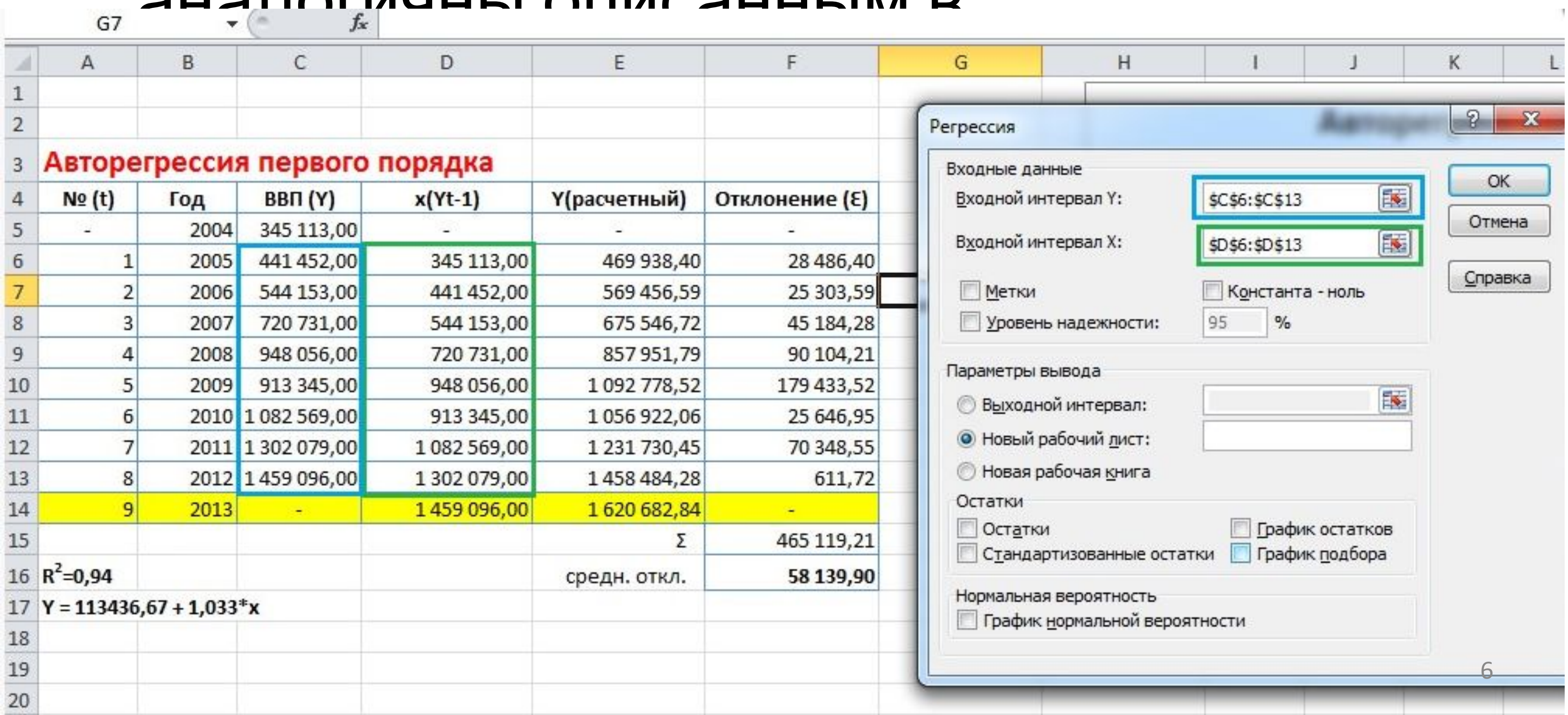
Авторегрессия первого порядка (AR I -)

$$Y_i = a_0 + a_i * Y_{i-1} + \varepsilon_i$$

Линейная модель авторегрессии первого порядка состоит только из одного влияющего фактора, а именно из *Y-1*, то есть изучается наиболее тесная зависимость только от того каким был итоговый показатель периодом с шагом назад.

Пример

Рассмотрим построение модели с помощью "пакета анализа" в эксель (вся процедура и поочередность шагов аналогичны описанным в



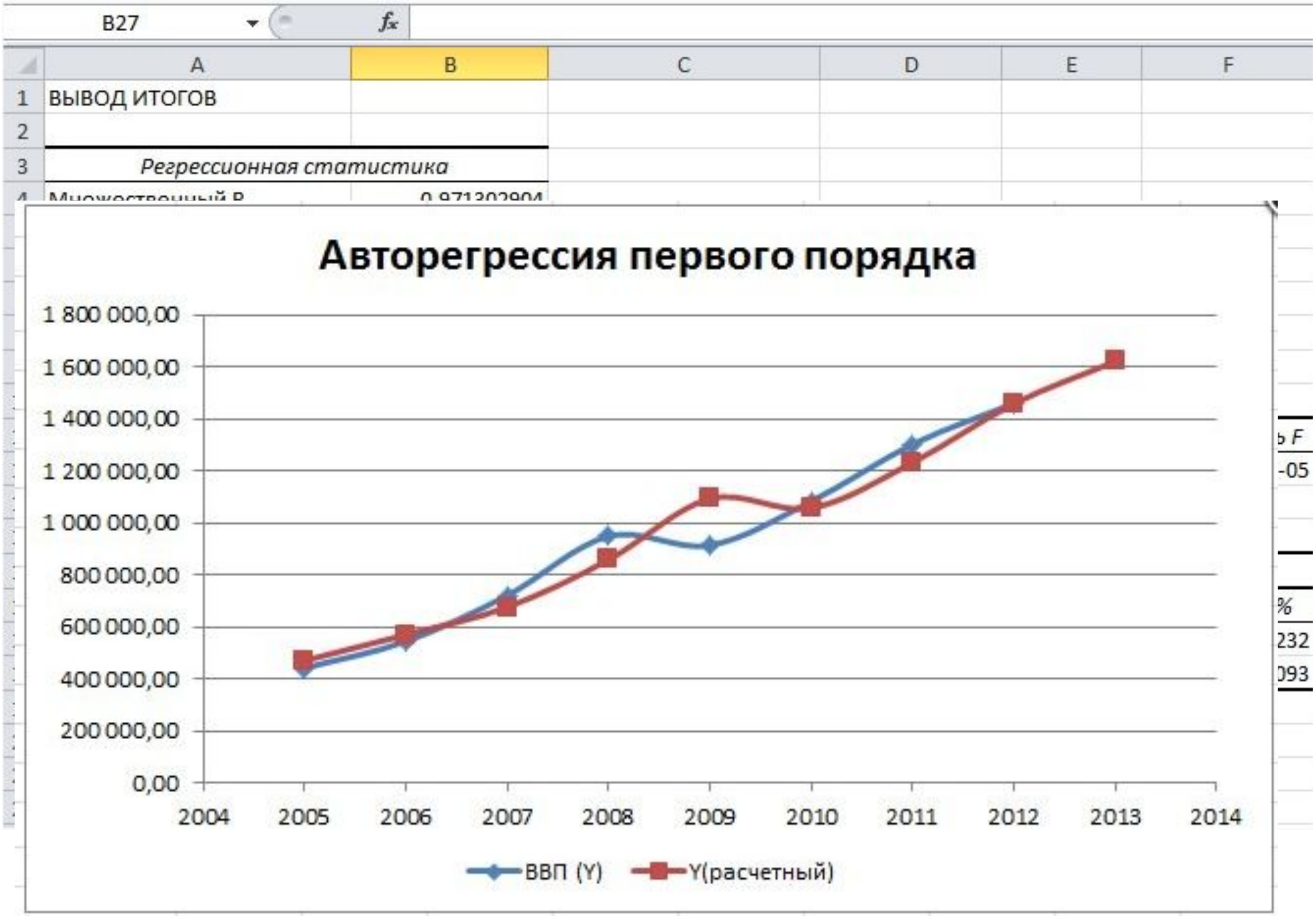
Авторегрессия первого порядка

№ (t)	Год	ВВП (Y)	x(Yt-1)	Y(расчетный)	Отклонение (Ɛ)
-	2004	345 113,00	-	-	-
1	2005	441 452,00	345 113,00	469 938,40	28 486,40
2	2006	544 153,00	441 452,00	569 456,59	25 303,59
3	2007	720 731,00	544 153,00	675 546,72	45 184,28
4	2008	948 056,00	720 731,00	857 951,79	90 104,21
5	2009	913 345,00	948 056,00	1 092 778,52	179 433,52
6	2010	1 082 569,00	913 345,00	1 056 922,06	25 646,95
7	2011	1 302 079,00	1 082 569,00	1 231 730,45	70 348,55
8	2012	1 459 096,00	1 302 079,00	1 458 484,28	611,72
9	2013	-	1 459 096,00	1 620 682,84	-
				Σ	465 119,21
				средн. откл.	**58 139,90**

$R^2=0{,}94$

$Y = 113436{,}67 + 1{,}033*x$

Результат регрессии



Авторегрессия второго порядка (AR II -)

$$Y_i = a_0 + a_i{*}Y_{i-1} + a_i{*}Y_{i-2} + \varepsilon_i$$

Модель авторегрессии второго порядка отличается от первой тем, что она включает в себя еще один влияющий фактор Y_{i-2}, то есть показывается зависимость от того каким был Y не только один период назад, но и от того каким он был два периода назад. Порой это позволяет выявить большую взаимосвязь и соответственно построить более точный прогноз.

Пример

21	**Авторегрессия второго порядка**						
22	**№ (t)**	**Год**	**ВВП (Y)**	**x_1(Yt-1)**	**x_2(Yt-2)**	**Y(расчетный)**	**Отклонение (Ɛ)**
23	-	2004	345 113,00	-	-	-	-
24	-	2005	441 452,00	345 113,00	-	-	-
25	1	2006	544 153,00	441 452,00	345 113,00	581 443,40	37 290,40
26	2	2007	720 731,00	544 153,00	441 452,00	686 627,40	34 103,60
27	3	2008	948 056,00	720 731,00	544 153,00	847 334,19	100 721,81
28	4	2009	913 345,00	948 056,00	720 731,00	1 068 422,45	155 077,45
29	5	2…					33 466,69
30	6	2…					74 632,66
31	7	2…					18 572,74
32	8	2…					-
33							453 865,34
34	R^2=0,927						**64 837,91**
35	Y=151395,987+0,7…						
36							

Авторегрессия третьего порядка (AR III -)

$$Y_i = a_0 + a_i * Y_{i-1} + a_i * Y_{i-2} + a_i * Y_{i-3} + \mathcal{E}_i$$

Модель авторегрессии третьего порядка наиболее тесно описывает зависимость от того каким был итоговый показатель раньше, так как в качестве влияющих факторов используется три отправные точки - каким Y был 1 период назад, 2 периода назад и 3 периода назад.

Требования к размаху исследуемого динамического ряда у этой модели выше - так как диапазон исходных данных сокращается на три периода, то чтобы не пострадало качество модели, необходимо расширять исследуемый период.

Пример

43	Авторегрессия третьего порядка							
44	№ (t)	Год	ВВП (Y)	x_1(Yt-1)	x_2(Yt-2)	x_3(Yt-3)	Y(расчетный)	Отклонение (Ɛ)
45	-	2004	345 113,00	-	-	-	-	-
46	-	2005	441 452,00	345 113,00	-	-	-	-
47	-	2006	544 153,00	441 452,00	345 113,00	-	-	-
48	1	2007	720 731,00	544 153,00	441 452,00	345 113,00	720 950,11	219,11
49	2	2008	948 056,00	720 731,00	544 153,00	441 452,00	880 804,39	67 251,61
50	3	2009	913 345,00	948 056,00	720 731,00	544 153,00	1 102 475,60	189 130,60
51	4	2010	1 082 569,00	913 345,00	948 056,00	720 731,00	1 141 391,63	58 822,63
52	5	2011	1 302 079,00	1 082 569,00	913 345,00	948 056,00	1 242 824,03	59 254,97
53	6	2012	1 459 096,00	1 302 079,00	1 082 569,00	913 345,00	1 464 216,14	5 120,14
54	7	2013	-	1 459 096,00	1 302 079,00	1 082 569,00	1 643 817,11	-
55							Σ	379 799,05
56	R^2=0,888						средн. откл.	63 299,84
57	Y=189893,52+0,745*x1+0,33*x2-0,058*x3							

ПЛЮСЫ:

1. Получение высококачественной модели с адекватным прогнозом при минимуме временных затрат и требований к исходным данным.

МИНУСЫ:

1. Прогноз по исходным данным возможен только на один период вперед. Если нужно сделать прогноз на более длительный срок, то в качестве влияющих факторов для расчета придется брать не реально существующий Y, а тот который рассчитан по модели, что в итоге даст прогноз на прогнозе, а значит адекватность такого прогноза, как минимум, в два раза меньше.

2. С увеличением разрядности авторегрессии возникает необходимость расширять диапазон исходных данных.

2. Операторное представление

Если ввести лаговый оператор *L: LY$_t$=Y$_{t-1}$* то авторегрессионную модель можно представить следующим образом

$$Y_t = a_0 + \sum_{i=1}^{p} a_i L^i Y_t + \varepsilon_t$$

или

$$a(L)Y_t = (1 - \sum_{i=1}^{p} a_i L^i)Y_t = a_0 + \varepsilon_t .$$

Ла́говый оператор — оператор смещения, позволяющий получить значения элементов временного ряда на основании ряда предыдущих значений.

Стационарность авторегрессионного процесса зависит от корней характеристического полинома

$$a(z) = 1 - \sum_{i=1}^{P} a_i z^i .$$

Для того чтобы процесс был стационарным, достаточно, чтобы все корни характеристического полинома лежали вне единичного круга в комплексной плоскости $|z|>1$.

В частности, для AR I - процесса $a(z)=1-rz$, следовательно корень этого «полинома» $z=1/r$, поэтому условие стационарности можно записать в виде $|r|<1$, то есть коэффициент авторегрессии (он же в данном случае коэффициент автокорреляции) должен быть строго меньше 1 по модулю.

Для AR(2)-процесса можно показать, что условия стационарности имеют вид: $|a_2|<1$, $a_2 \pm a_1 < 1$.

3. Автокорреляционная функция

Автоковариационная и автокорреляционная функции AR(p)-процесса удовлетворяют рекуррентным соотношениям:

$$\gamma(k)=\sum_{j=1}^{p} a_j\gamma(k-j)$$

или

$$r(k)=\sum_{j=1}^{p} a_j r(k-j)$$

В простейшем случае AR(1)-процесса, математическое ожидание равно

$$\mu=\frac{a_0}{1-a}$$

а дисперсия

$$\gamma(0)-\frac{\sigma_{\varepsilon}^{2}}{1-a^{2}}.$$

В общем случае выражение для математического ожидания через параметры модели было указано выше, однако, выражение для дисперсии временного ряда - существенно усложняется. Можно показать, что дисперсия ряда *γ(0)* и вектор автокорреляций *γ*

выражаются через параметры следующим образом:

$$\gamma(0) = (1 + a^T (C - a \cdot a^T)^{-1} a)\sigma_\varepsilon^2, \;\; \gamma = \sigma_\varepsilon^2 (C - a \cdot a^T)^{-1} a.$$

где *a* - вектор параметров,

C - матрица порядка *p* , элементы которой определяются следующим образом. Диагональные элементы равны $c_{ii}=1-a_{2i}$. Элементы выше диагонали равны $-a_{2i+j-1}$, а элементы ниже диагонали равны $-(a_j+a_{2i-j})$. Здесь подразумевается, что если индекс превышает порядок модели *p*, то соответствующая величина приравнивается к нулю.