

ПОСТРОЕНИЕ РЕГРЕССИОННОЙ МОДЕЛИ ДЛЯ ОЦЕНИВАНИЯ РАЗМЕРА АРЕНДНОЙ ПЛАТЫ ЗА ИСПОЛЬЗОВАНИЕ СЕРВЕРА В РЕСПУБЛИКЕ БЕЛАРУСЬ

Е. С. Расолько, А. В. Соболевский

ГУО «Институт бизнеса
Белорусского государственного университета», г. Минск;
hellenra2001@gmail.com; sobolevsky.a@mail.ru;
науч. рук. – Г. А. Хацкевич, д-р экон. наук, проф.

В работе приведено построение модели для оценивания размера арендной платы за использование сервера в Республике Беларусь.

Ключевые слова: модель; арендная плата; регрессионная модель; факторы; параметры

Многие организации имеют дело с обработкой большого количество данных, которые требуют расширения пространства для их хранения. Большинство бумажных носителей требует немало времени для обработки, а при их увеличении и огромных средств для хранения, из-за чего встает вопрос о том, как снизит затраты и увеличить эффективность обработки информации. Решением для данной проблемы будет создание корпоративной серверной базы данных.

Несмотря на то, что все данные будут храниться в облачном хранилище сервера, для поддержания самого сервера требуются немалые ресурсы компьютера, которые как раз и отвечают за быстродействие сервера и его стабильную работу.

Для построения регрессионной модели были выбраны основные факторы, которые будут влиять на стоимость аренды сервера, а именно:

- m – объем ОЗУ (оперативная память)
- s – объем дискового хранилища (SSD-накопитель)
- v – количество ядер процессора
- f – тактовая частота процессора

Для оценивания размера арендной платы за использование сервера была выбрана степенная модель:

$$p = a_0 * m^{a_1} * s^{a_2} * c^{a_3} * f^{a_4}$$

Математическое выражение данной степенной зависимости представлено ниже:

$$\tilde{y}_t = b_0 * x_1^{b_1} * x_2^{b_2} * x_3^{b_3} * x_4^{b_4} + \varepsilon_t$$

Оценку неизвестных параметров b_0, b_1, b_2, b_3, b_4 будем производить по МНК (метод наименьших квадратов). Чтобы провести оценку неизвестных параметров, линеаризуем нашу модель прологарифмировав ее:

$$\ln p = \ln a_0 + a_1 \ln m + a_2 \ln s + a_3 \ln c + a_4 \ln f$$

Для проведения исследования был использован пакет прикладных программ EViews 8, который применяется при моделировании, оценке, прогнозировании экономических показателей [1, 2, 3].

На рис. 1 представлены конфигурации серверов (Т) по покупательским ценам (Р) в Республике Беларусь, количество серверов – 20, единица измерения – бел. руб., все данные взяты с платформы по аренде серверов IdeaHost [4].

View	Proc	Object	Name	Freeze	Default	Sort	Edit+/-	Smpl+/-	Compare+/-	Transpose+/-	T	M	V	F	S	P
1											1	8	8	2.40	300	294.0
2											2	16	4	3.60	480	353.5
3											3	32	4	3.30	480	388.5
4											4	32	12	2.66	480	406.0
5											5	32	4	3.40	960	528.5
6											6	32	10	3.70	960	1001.0
7											7	128	20	2.20	960	1043.0
8											8	32	14	3.30	960	1141.0
9											9	64	14	3.30	1920	1277.5
10											10	64	24	2.40	1920	1281.0
11											11	64	12	3.50	1920	1295.0
12											12	64	12	3.50	3840	1435.0
13											13	64	20	2.20	1920	1599.5
14											14	128	20	2.20	1920	1823.5
15											15	128	24	2.40	3840	2201.5
16											16	64	32	2.40	1920	2268.0
17											17	128	40	2.30	3840	2919.0
18											18	256	52	2.20	7680	4196.5
19											19	256	48	3.00	3840	4462.5
20											20	512	32	2.90	15360	6226.5

Рис. 1. Конфигурации серверов и стоимость их аренды

Для того, чтобы оценить неизвестные параметры, нужно прологарифмировать все значения этой таблицы. Полученные результаты представлены на рис. 2.

	T	LN_M	LN_V	LN_F	LN_S	LN_P
1	1	2.079442	2.079442	0.875469	5.703782	5.683580
2	2	2.772589	1.386294	1.280934	6.173786	5.867883
3	3	3.465736	1.386294	1.193922	6.173786	5.962293
4	4	3.465736	2.484907	0.978326	6.173786	6.006353
5	5	3.465736	1.386294	1.223775	6.866933	6.270043
6	6	3.465736	2.302585	1.308333	6.866933	6.908755
7	7	4.852030	2.995732	0.788457	6.866933	6.949856
8	8	3.465736	2.639057	1.193922	6.866933	7.039660
9	9	4.158883	2.639057	1.193922	7.560080	7.152660
10	10	4.158883	3.178054	0.875469	7.560080	7.155396
11	11	4.158883	2.484907	1.252763	7.560080	7.166266
12	12	4.158883	2.484907	1.252763	8.253228	7.268920
13	13	4.158883	2.995732	0.788457	7.560080	7.377446
14	14	4.852030	2.995732	0.788457	7.560080	7.508513
15	15	4.852030	3.178054	0.875469	8.253228	7.696894
16	16	4.158883	3.465736	0.875469	7.560080	7.726654
17	17	4.852030	3.688879	0.832909	8.253228	7.978996
18	18	5.545177	3.951244	0.788457	8.946375	8.342006
19	19	5.545177	3.871201	1.098612	8.253228	8.403464
20	20	6.238325	3.465736	1.064711	9.639522	8.736570

Рис. 2. Конфигурации серверов и стоимость их аренды
в прологарифмированном виде

После этого была построена линейная модель и рассчитаны неизвестные параметры. Результат регрессионного анализа представлены на рис. 3.

Equation: UNTITLED Workfile: ПРИМЕР::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LN_P_
Method: Least Squares
Date: 04/24/22 Time: 21:28
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.486491	0.441489	3.366991	0.0042
LN_M_	0.149123	0.099055	1.505461	0.1530
LN_V_	0.555797	0.111530	4.983360	0.0002
LN_F_	0.629617	0.290697	2.165889	0.0468
LN_S_	0.386384	0.099645	3.877606	0.0015

R-squared	0.969404	Mean dependent var	7.160111
Adjusted R-squared	0.961246	S.D. dependent var	0.871125
S.E. of regression	0.171491	Akaike info criterion	-0.476256
Sum squared resid	0.441137	Schwarz criterion	-0.227322
Log likelihood	9.762555	Hannan-Quinn criter.	-0.427661
F-statistic	118.8167	Durbin-Watson stat	2.254688
Prob(F-statistic)	0.000000		

Рис. 3. Основные параметры регрессионного анализа модели

По полученным данным анализа можно дать оценку качества самой модели, важно проанализировать значение нормированного коэффициента детерминации R^2 , значения T -статистики параметров модели и значение F -критерия:

- Коэффициент детерминации показывает на тесноту связи параметров, и чем значение ближе к единице, тем модель будет считаться лучше. В нашем случае коэффициент детерминации равен 0,96. Поскольку его значение почти равно единице, можно сказать, что модель вполне адекватна.

- По критерию Стьюдента параметры a_0, a_2, a_3, a_4 статистически значимы для данной модели. Расчетные их значения больше критических. Параметр a_1 представляет статистически не значимую величину, его расчетное значение меньше критических значений для всех уровней надежности.

- Значение F -статистики равно 118,8, что больше его критического значения. Следовательно, коэффициент детерминации является статистически значимым.

Также важно отметить, что значение критерия Дарбина-Вотсона равно 2,25 и попадает в промежуток $d_u < d < 4 - d_u$, из-за чего можно сделать вывод о том, что в модели отсутствует автокорреляция.

Для проверки качества спецификации модели был использован критерий Рамсея. Результат теста представлен на рис. 4.

Ramsey RESET Test
Equation: EQ01
Specification: LN_P_ C LN_M_ LN_V_ LN_F_ LN_S_
Omitted Variables: Powers of fitted values from 2 to 4

	Value	df	Probability
F-statistic	0.147477	(3, 12)	0.9293
Likelihood ratio	0.724117	3	0.8675

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	0.015686	3	0.005229
Restricted SSR	0.441137	15	0.029409
Unrestricted SSR	0.425451	12	0.035454
Unrestricted SSR	0.425451	12	0.035454

LR test summary:

	Value	df
Restricted LogL	9.762555	15
Unrestricted LogL	10.12461	12

Unrestricted Test Equation:
Dependent Variable: LN_P_
Method: Least Squares
Date: 04/25/22 Time: 09:48
Sample: 1 20

Рис. 4. Основные параметры качества спецификации модели по тесту Рамсея

Из результатов тестирования видно, что F-статистика по критерию Рамсея равна 0,15, что намного меньше, чем критическое значение. Следовательно, можно сделать вывод, что модель хорошо специфицирована.

Была проведена проверка модели на наличие мультиколлинеарности с помощью метода инфляционных факторов. Результат проверки представлен на рис. 5.

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	0.194913	132.5527	NA
LN_M_	0.009812	123.5878	6.243508
LN_V_	0.012439	69.02684	4.913978
LN_F_	0.084505	62.66681	2.108861
LN_S_	0.009929	379.4347	6.406625

Рис. 5. Мультиколлинеарность модели методом инфляционных факторов

По данному отчету можно сделать вывод, что в модели присутствуют элементы мультиколлинеарности. Так как значение VIF у факторов m и s больше 5. Это указывает на то, что при увеличении мощности процессора повышается и объем оперативной памяти, и появляется зависимость между этими переменными. Однако при увеличении числа серверов (например, 6 ГБ., 10 ГБ., 13 ГБ), мультиколлинеарность модели можно устранить.

После выполнения основных проверок по качеству и спецификации модели, ее можно использовать для прогнозирования цены за аренду сервера в будущем. Итоговый вид формульной модели представлен ниже:

$$p = e^{1,49} * m^{0,15} + v^{0,56} + f^{0,63} + s^{0,39}$$

На основании построенной модели была определена цена серверов при увеличении мощности комплектующих. Результат расчетов представлен в таблице.

Прогноз цен за аренду сервера

№ Сервера	Объем оперативной памяти, ГБ	Количество ядер процессора	Тактовая частота процессора, ГГц	Объем хранилища, ГБ	Стоимость, бел.руб
1	256	48	3,4	7680	5956,58
2	512	52	3,2	15360	8688,49
3	1024	52	3,7	30720	13798,97
4	2048	40	2,9	61440	14828,77

В результате была построена модель для оценивания-размера арендной платы за использование сервера в Республике Беларусь, которую можно использовать для построения прогноза, используя желаемые значения факторов.

Библиографические ссылки

1. Русилко Т. В. Эконометрика : учеб. пособие / Т. В. Русилко, Г. А. Хацкевич. – Гродно: ГрГУ, 2014. – 268 с.
2. *Eviews*. User Guide. Version 8. Irvine. California: IHS Global Inc, 2013.
3. *Jesse Russell Econometric* / Jesse Russell. – М.: VSD, 2017. – 717 с.
4. IdeaHost – Хостинг серверов на территории РБ [Электронный ресурс]. – Режим доступа: <https://ideahost.by/>. – Дата доступа: 23.04 2022.