

Исследование влияния дискриминационного текста в информации о квартире на её цену.

Петраков Сергей, Семенев Данат э-301.

К анализу данной проблемы нас подтолкнула работа Ветеринарова (научный руководитель – В.В.Иванов) (2018) по оценке этнической дискриминации на рынке недвижимости в Москве. Мы решили реплицировать полученные достижения. Данные были взяты с открытой базы данных Циан.

Анализ данных

Мы начали с того, что провели графический анализ данных, построили диаграмму рассеяния, где по оси абсцисс расположили площадь квартиры, а по оси ординат – цену.

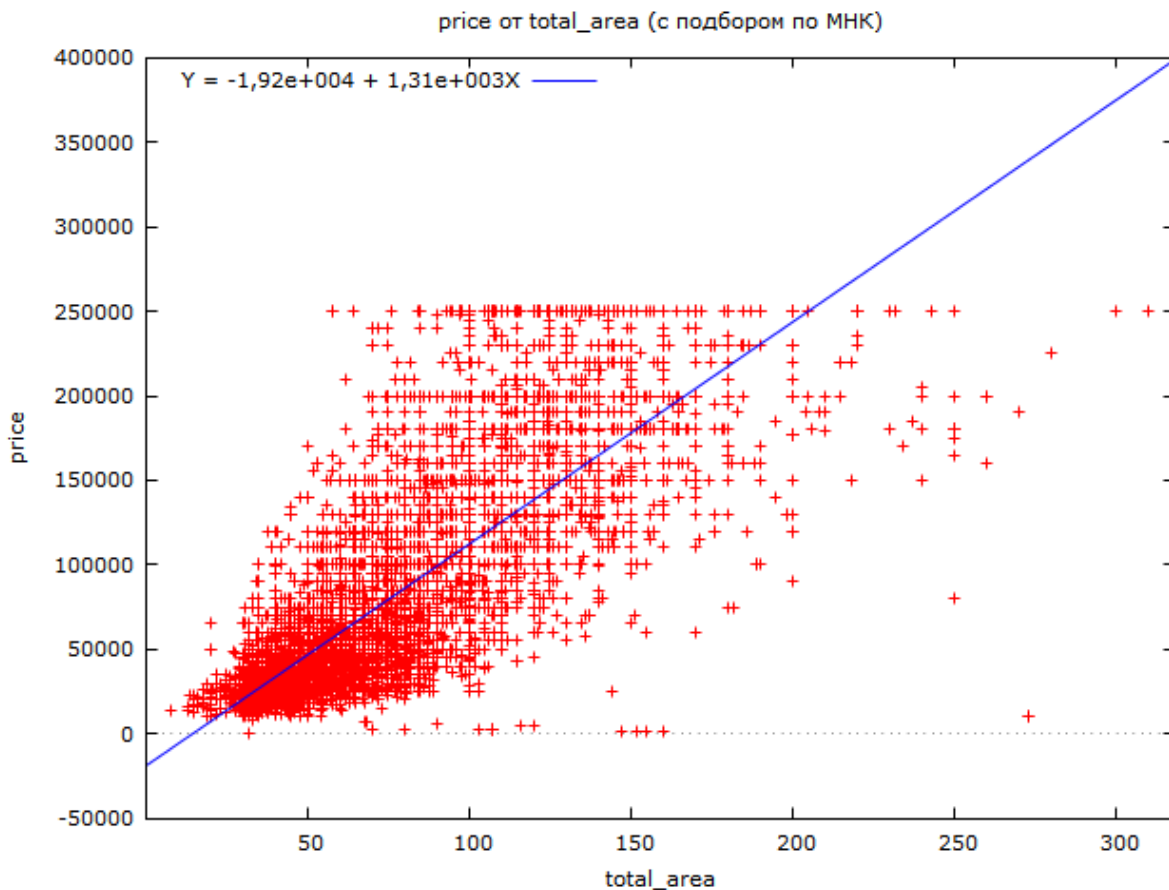


График рассеяния, где по оси абсцисс расположена площадь квартир, а по оси ординат – цена аренды. Исходная выборка (10000 наблюдений).

После графического анализа данных мы пришли к выводу целесообразности удаления выбросов: во-первых, данные, где цена менее 10000 руб. кажутся нецелесообразными в рассмотрении, так как они нерепрезентативны, и такие наблюдения, где стоимость

аренды менее 10 тыс., являются выбросами.

Также мы удалили все наблюдения, где площадь превышает 240 м², поскольку, их мало, они сильно сдвигают кривую регрессии, уменьшая R², и далеко не с каждой площадью >240м² в выборке существует соответствующая квартира (вследствие чего возможные оценки для отдельных значений total_area становятся непроверяемыми).

Итого мы удалили 37 наблюдений, и у нас осталось 9963 наблюдения.

Мы построили следующие графики рассеяния, и такие графики нас больше устроили.

Описательная статистика, используем наблюдения 1 - 9963					
Variable	Mean	Median	S.D.	Min	Max
price	70400	45000	58500	10000	25000
total_area	67.8	56.0	35.1	8	240
l_price	10.9	10.7	0.723	9.21	12.4

Графики рассеяния

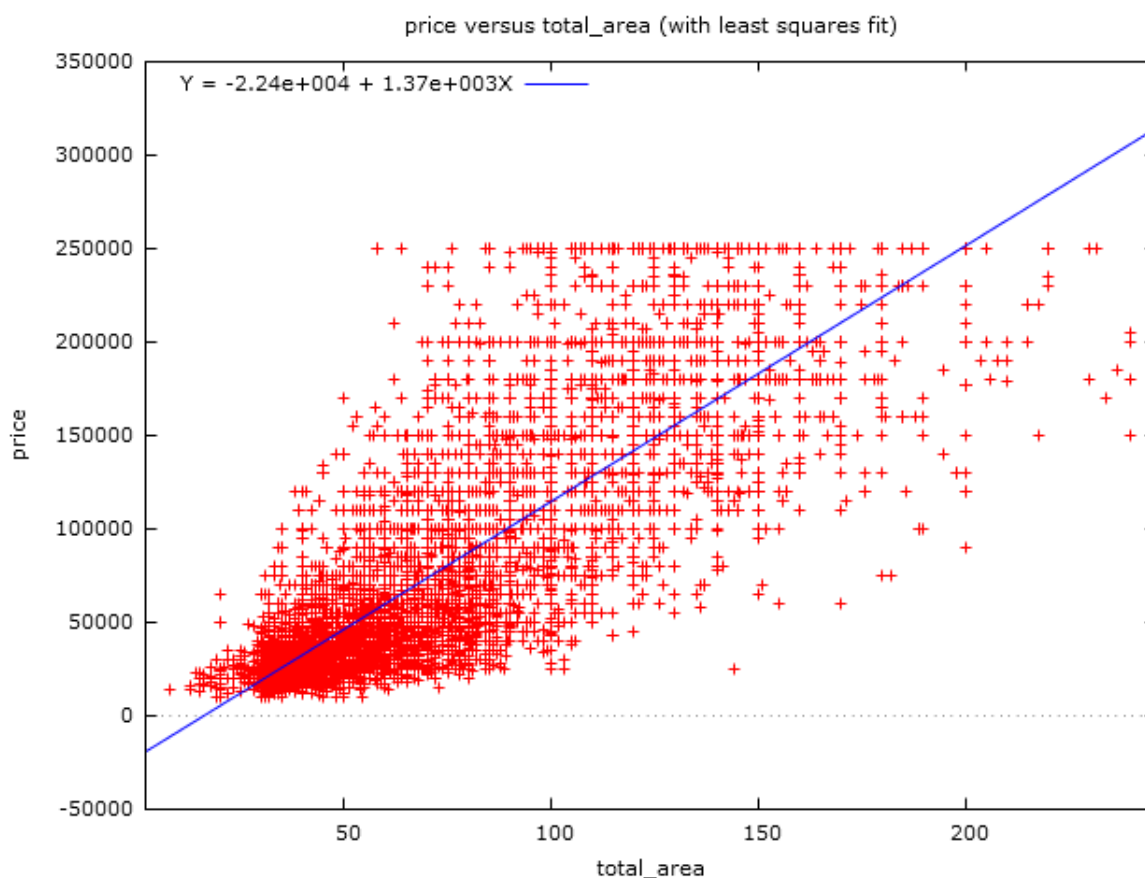


График рассеяния, где по оси абсцисс расположена площадь квартир, а по оси ординат – цена аренды. Урезанная выборка (9963 наблюдения).

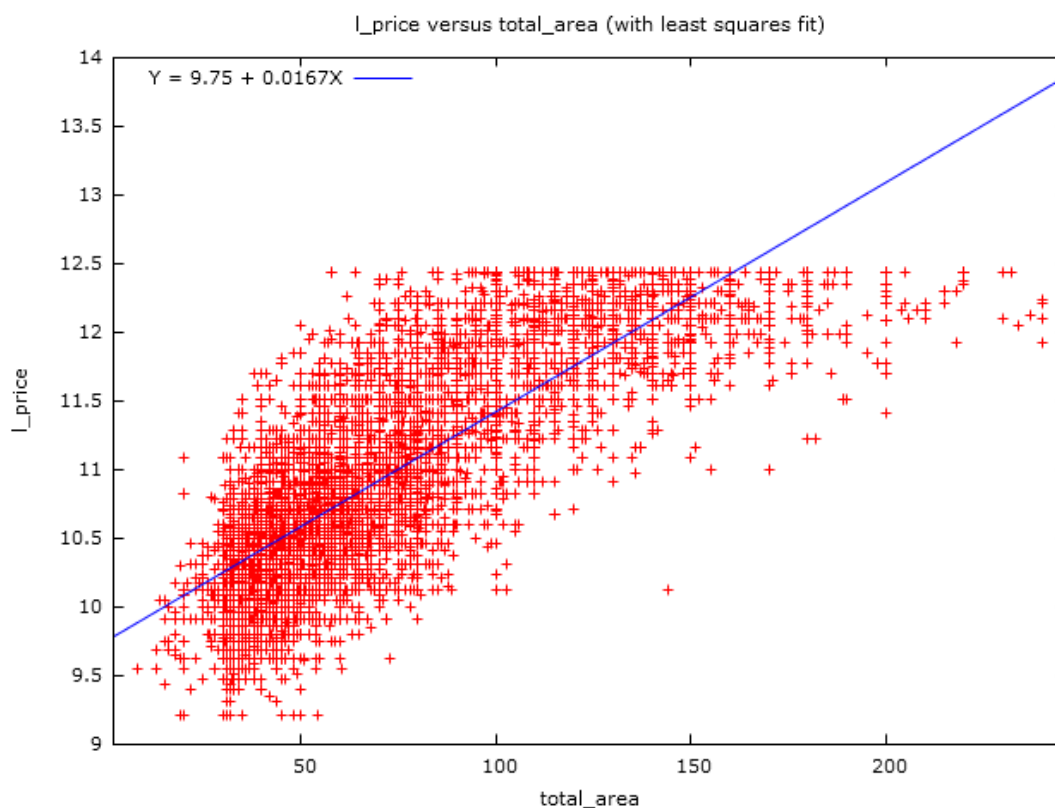
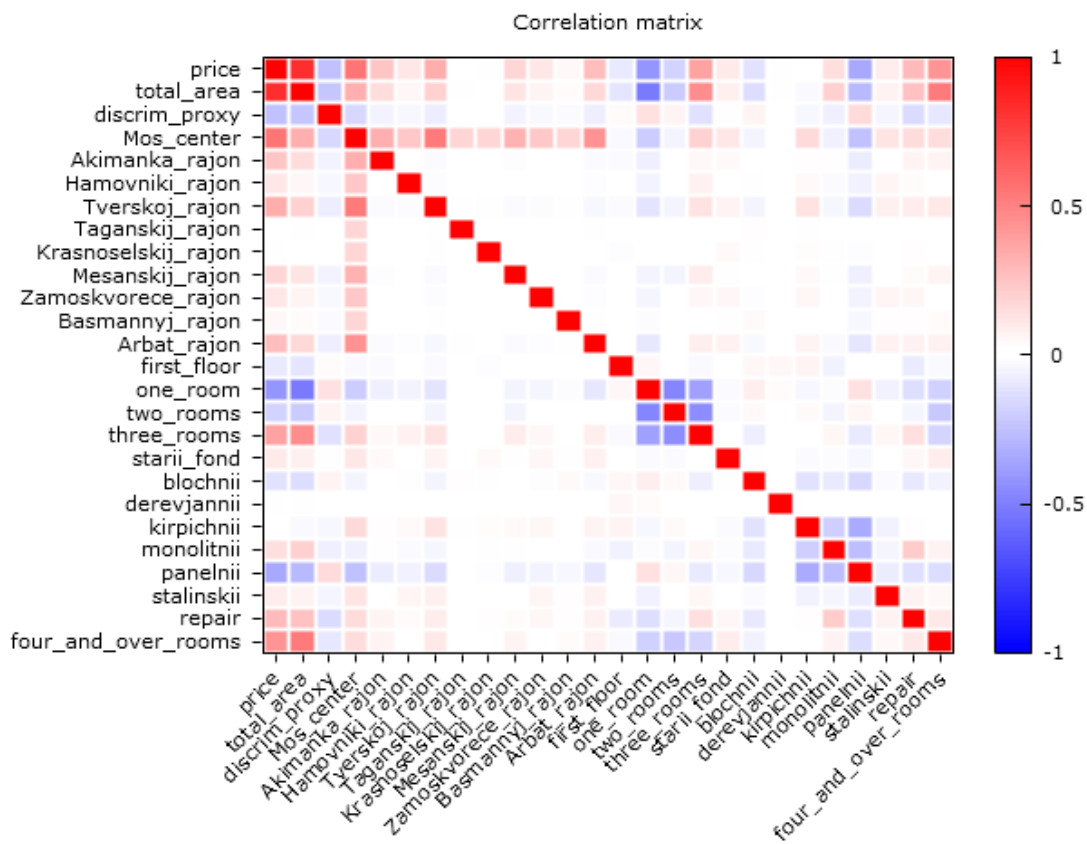


График рассеяния, где по оси абсцисс расположена площадь квартир, а по оси ординат – логарифм цены аренды. Урезанная выборка (9963 наблюдения).

Описательная статистика, using the observations 1 - 9963

Variable	Mean	Median	S.D.	Min	Max
discrim_proxy	0,19	0	0,39	0	1
Mos_center	0,16	0	0,37	0	1
Akimanka_rajon	0,02	0	0,14	0	1
Hamovniki_rajon	0,01	0	0,10	0	1
Tverskoj_rajon	0,05	0	0,22	0	1
Taganskij_rajon	0,01	0	0,08	0	1
Krasnoselskij_rajon	0,01	0	0,08	0	1
Mesanskij_rajon	0,02	0	0,14	0	1
Zamoskvorece_rajon	0,01	0	0,10	0	1
Basmannyj_rajon	0,01	0	0,08	0	1
Arbat_rajon	0,04	0	0,19	0	1

first_floor	0,06	0	0,24	0	1
one_room	0,29	0	0,45	0	1
two_rooms	0,37	0	0,48	0	1
three_rooms	0,26	0	0,44	0	1
starii_fond	0,00	0	0,06	0	1
blochnii	0,06	0	0,23	0	1
derevjannii	0,00	0	0,02	0	1
kirpichnii	0,21	0	0,40	0	1
monolitnii	0,14	0	0,34	0	1
panelnii	0,31	0	0,46	0	1
stalinskii	0,02	0	0,13	0	1
repair	0,50	0	0,50	0	1
four_and_over_rooms	0,08	0	0,27	0	1



Схематичное Изображение корреляционной матрицы для всех присутствующих в исходных данных переменных.

Из корреляционной матрицы нам особенно интересны интенсивности взаимосвязей между различными районами и содержательными характеристиками квартир, с одной стороны, а также связи характеристик квартир между собой, с другой.

price	area	discr	1st floor	1 room	2 rms	3 rms	>4 rms	
-0,252	-0,233	1,000	0,027	0,128	0,057	-0,128	-0,107	discrim proxy
0,550	0,323	-0,165	-0,036	-0,210	-0,055	0,188	0,143	Mos center
0,236	0,145	-0,061	-0,031	-0,076	0,000	0,042	0,059	Akimanka
0,106	0,043	-0,044	-0,010	-0,061	-0,006	0,068	0,002	Hamovniki
0,329	0,192	-0,084	-0,032	-0,118	-0,057	0,120	0,103	Tverskoj
-0,002	-0,018	-0,007	-0,010	0,005	0,006	-0,002	-0,014	Taganskij
0,017	-0,001	-0,012	-0,020	-0,010	0,014	-0,006	0,002	Krasnoselskij
0,170	0,116	-0,058	0,007	-0,057	-0,054	0,085	0,054	Mesanskij
0,105	0,057	-0,040	-0,006	-0,053	0,007	0,045	0,003	Zamoskvorece
0,041	0,030	-0,026	0,013	-0,023	-0,006	0,012	0,030	Basmannyj
0,267	0,163	-0,077	0,001	-0,107	-0,009	0,077	0,069	Arbat

Таблица корреляций между инструментальными переменными, соответствующим районам ЦАО (по строкам) и переменными, отражающими цены, размеры, структуру **самих** квартир и исследуемый фактор дискриминации.

star	bloch	derev	kirpich	monolit	panel	stalinsk	repair	
-0,018	0,061	-0,008	-0,043	-0,076	0,151	-0,052	-0,150	discrim proxy
0,107	-0,057	-0,008	0,156	-0,070	-0,254	0,116	0,147	Mos center
0,037	-0,011	-0,003	0,021	-0,017	-0,085	0,011	0,058	Akimanka
-0,006	-0,017	-0,002	0,037	-0,029	-0,069	0,051	0,029	Hamovniki
0,058	-0,055	-0,004	0,125	-0,048	-0,151	0,069	0,085	Tverskoj
-0,005	0,020	-0,001	0,015	0,015	-0,002	0,001	0,008	Taganskij
0,038	0,016	-0,001	0,024	-0,019	-0,022	0,001	0,020	Krasnoselskij
0,015	-0,012	-0,002	0,036	-0,018	-0,077	0,012	0,027	Mesanskij
0,042	-0,020	-0,002	0,049	0,001	-0,061	0,052	0,047	Zamoskvorece
0,016	0,032	-0,001	0,007	-0,007	-0,046	0,022	0,022	Basmannyj
0,066	-0,040	-0,003	0,056	-0,036	-0,109	0,067	0,066	Arbat

Таблица корреляций между инструментальными переменными, соответствующим районам ЦАО (по строкам) и переменными, отражающими и исследуемый фактор дискриминации.

Результаты регрессий

Метод оценки - МНК
Зависимая переменная: price

	(1)	(2)	(3)
const	-12485,54*** (2669,44)	-12567,91*** (2669,84)	-8901,06*** (2736,22)
total_area	1144,68*** (19,93)	1144,31*** (19,93)	1122,61*** (20,07)
discrim_proxy	-3862,54*** (455,67)	-3396,54*** (453,69)	-12351,91*** (1349,28)
Mos_center	51822,62*** (2069,51)	52233,59*** (2083,11)	51185,67*** (2114,44)
Akimanka_rajon	7790,04** (3249,86)	7679,19** (3242,98)	8456,84** (3298,42)
Hamovniki_rajon	-3180,20 (4458,43)	-3296,01 (4448,66)	-2972,62 (4517,91)
Tverskoj_rajon	2999,10 (2603,89)	3148,29 (2604,88)	3361,35 (2660,69)
Taganskij_rajon	-35688,52*** (3222,30)	-34442,62*** (3246,91)	-34148,95*** (3524,55)
Krasnoselskij_rajon	-31829,51*** (3682,36)	-30880,25*** (3673,52)	-30228,21*** (3963,32)
Mesanskij_rajon	-8288,65** (3397,79)	-8375,60** (3392,96)	-7757,28** (3432,41)
Zamoskvorece_rajon	-8680,97** (4115,32)	-8700,70** (4103,96)	-8085,40* (4183,95)
Basmannyj_rajon	-31766,26*** (3475,82)	-31555,48*** (3507,96)	-31934,94*** (3498,13)
first_floor	-499,76 (1040,93)	-475,78 (1040,97)	-246,31 (1310,91)
one_room	4350,27** (2020,07)	4304,91** (2019,91)	1836,80 (2042,47)
two_rooms	855,91 (1887,48)	858,11 (1887,75)	-187,70 (1896,36)
three_rooms	-334,08 (1716,95)	-370,91 (1717,48)	-399,79 (1726,85)
starii_fond	4593,66 (5441,31)	4823,11 (5458,35)	4130,31 (5456,05)
blochnii	-6596,86*** (975,30)	-6540,67*** (973,73)	-8722,63*** (1202,19)
derevjannii	-18234,37***	-18120,13***	-18939,68***

	(3130,50)	(3132,42)	(3272,36)
kirpichnii	-7957,00*** (880,13)	-7908,20*** (880,83)	-9881,33*** (1008,93)
monolitnii	-3490,68*** (1224,63)	-3463,59*** (1224,73)	-3727,73*** (1365,96)
panelnii	-12095,29*** (724,93)	-12110,72*** (724,87)	-15040,25*** (863,35)
stalinskii	-4324,75 (3009,91)	-4330,81 (3009,01)	-5769,86* (3075,17)
repair	5759,32*** (595,11)	5794,48*** (595,46)	6649,24*** (694,79)
Mos_center_discr		-10936,25*** (3086,06)	5620,94 (6667,86)
Akimanka_discr			-33439,63*** (8712,06)
Hamovniki_rajon_discr			-34039,89*** (9823,82)
Tverskoj_rajon_discr			-7305,27 (9870,58)
Taganskij_rajon_discr			-13946,49* (7530,36)
Krasnoselskij_rajon_discr			-15384,97** (7799,89)
Mesanskij_rajon_discr			-27992,47** (14278,37)
Zamoskvorece_rajon_discr			-29737,83*** (7889,53)
first_floor_discr			-710,90 (1629,07)
one_room_discr			5108,03*** (817,06)
three_rooms_discr			-5253,59*** (1664,85)
blochnii_discr			11023,90*** (1683,63)
kirpichnii_discr			11682,15*** (1640,59)
monolitnii_discr			-1418,80 (2215,47)
panelnii_discr			13968,38*** (1310,07)
repair_discr			-3502,58*** (1046,88)

stalinskii_discr			3242,93 (10367,69)
n	9963	9963	9963
Испр. R ²	0,78	0,78	0,78
lnL	-1,2e+005	-1,2e+005	-1,2e+005

В скобках указаны стандартные ошибки

* significant at the 10 percent level

** significant at the 5 percent level

*** significant at the 1 percent level

Тест Рамсея (RESET) для модели (1) -

Нулевая гипотеза: спецификация адекватна

Тестовая статистика: $F(2, 9937) = 417,224$

р-значение = $P(F(2, 9937) > 417,224) = 1,02134e-174$

Тест Рамсея (RESET) для модели (2) -

Нулевая гипотеза: спецификация адекватна

Тестовая статистика: $F(2, 9936) = 412,235$

р-значение = $P(F(2, 9936) > 412,235) = 1,02262e-172$

Тест Рамсея (RESET) для модели (3) -

Нулевая гипотеза: спецификация адекватна

Тестовая статистика: $F(2, 9920) = 391,34$

р-значение = $P(F(2, 9920) > 391,34) = 2,60065e-164$

Метод оценки - МНК

Зависимая переменная: l_price

	(1)	(2)	(3)
const	9,92*** (0,03)	9,92*** (0,03)	9,96*** (0,03)
total_area	0,01*** (0,00)	0,01*** (0,00)	0,01*** (0,00)
discrim_proxy	-0,07*** (0,01)	-0,07*** (0,01)	-0,19*** (0,02)
Mos_center	0,58*** (0,02)	0,58*** (0,02)	0,56*** (0,02)
Akimanka_rajon	0,03 (0,03)	0,03 (0,03)	0,03 (0,03)
Hamovniki_rajon	-0,01 (0,03)	-0,01 (0,03)	-0,01 (0,03)
Tverskoj_rajon	0,02 (0,02)	0,02 (0,02)	0,02 (0,02)
Taganskij_rajon	-0,26***	-0,25***	-0,24***

	(0,03)	(0,03)	(0,03)
Krasnoselskij_rajon	-0,20*** (0,04)	-0,19*** (0,04)	-0,18*** (0,04)
Mesanskij_rajon	-0,07*** (0,03)	-0,07*** (0,03)	-0,07** (0,03)
Zamoskvorece_rajon	-0,08** (0,04)	-0,08** (0,04)	-0,07** (0,04)
Basmannyj_rajon	-0,21*** (0,04)	-0,21*** (0,04)	-0,21*** (0,04)
first_floor	-0,08*** (0,01)	-0,08*** (0,01)	-0,08*** (0,02)
one_room	-0,11*** (0,02)	-0,11*** (0,02)	-0,13*** (0,02)
two_rooms	0,04** (0,02)	0,04** (0,02)	0,04* (0,02)
three_rooms	0,10*** (0,02)	0,10*** (0,02)	0,09*** (0,02)
starii_fond	0,07 (0,05)	0,08 (0,05)	0,07 (0,05)
blochnii	-0,03** (0,01)	-0,03** (0,01)	-0,06*** (0,02)
derevjannii	-0,47*** (0,06)	-0,47*** (0,06)	-0,48*** (0,06)
kirpichnii	-0,04*** (0,01)	-0,04*** (0,01)	-0,06*** (0,01)
monolitnii	-0,01 (0,01)	-0,01 (0,01)	-0,02 (0,02)
panelnii	-0,11*** (0,01)	-0,11*** (0,01)	-0,15*** (0,01)
stalinskii	0,06** (0,03)	0,06** (0,03)	0,04 (0,03)
repair	0,18*** (0,01)	0,18*** (0,01)	0,20*** (0,01)
Mos_center_discr		-0,03 (0,03)	0,17*** (0,05)
Akimanka_discr			-0,35*** (0,05)
Hamovniki_rajon_discr			-0,27** (0,11)
Tverskoj_rajon_discr			-0,10 (0,08)
Taganskij_rajon_discr			-0,23*** (0,06)

Krasnoselskij_rajon_discr			-0,22*** (0,08)
Mesanskij_rajon_discr			-0,25 (0,18)
Zamoskvorece_rajon_discr			-0,26*** (0,07)
first_floor_discr			-0,01 (0,03)
one_room_discr			0,07*** (0,02)
three_rooms_discr			0,02 (0,02)
blochnii_discr			0,16*** (0,03)
kirpichnii_discr			0,13*** (0,03)
monolitnii_discr			-0,01 (0,03)
panelnii_discr			0,18*** (0,02)
repair_discr			-0,06*** (0,02)
stalinskii_discr			0,11 (0,16)
n	9963	9963	9963
Испр. R ²	0,78	0,78	0,78
lnL	-3,4e+003	-3,3e+003	-3,3e+003

В скобках указаны стандартные ошибки

* significant at the 10 percent level

** significant at the 5 percent level

*** significant at the 1 percent level

Тест Рамсея (RESET) для модели (1) -

Нулевая гипотеза: спецификация адекватна

Тестовая статистика: $F(2, 9937) = 475,062$

р-значение = $P(F(2, 9937) > 475,062) = 9,11988e-198$

Тест Рамсея (RESET) для модели (2) -

Нулевая гипотеза: спецификация адекватна

Тестовая статистика: $F(2, 9936) = 476,084$

р-значение = $P(F(2, 9936) > 476,084) = 3,59377e-198$

Тест Рамсея (RESET) для модели (3) -

Нулевая гипотеза: спецификация адекватна

Тестовая статистика: $F(2, 9920) = 457,823$

р-значение = $P(F(2, 9920) > 457,823) = 6,57793e-191$