Цены квартир в Москве*

Я. Р. Магнус
Department of Econometrics & OR, Tilburg University,
Тилбург, Нидерланды.

А. А. Пересецкий
ЦЭМИ РАН, РЭШ,
Москва, Россия.

10 марта, 2010

JEL: C21, C51, C53, R21, R31.

Ключевые слова: гедонистические цены, Москва, предварительное тестирование.

Аннотация. Мы строим простую гедонистическую модель цен квартир в Москве в начале 2003 г. Используя около 15 000 наблюдений мы оцениваем параметры модели и строим прогнозные значения цен. Отдельно рассматриваются проблемы, связанные с предварительным тестированием (pretesting).

Для связи с авторами: Jan R. Magnus, Department of Econometrics & OR, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: +31-13-466-3092, fax: +31-13-466-3066, e-mail: magnus@uvt.nl.

*Мы благодарны Dmitry Danilov, Arthur van Soest, Marina Velikova и Marno Verbeek за ценные обсуждения, Chris Muris и Hongying Sun за помощь в работе с данными, а также Г. Мальгинову и Г. Стернику за данные к рис. 1.

1. Введение

В данной статье мы делаем попытку объяснить структуру цен квартир в Москве в 2003 г. и прогнозировать эти цены. В работе предполагается, что цена квартиры зависит от ее характеристик, таких как размер и расположение. Таким образом, мы следуем концепции «гедонистических» (hedonic) цен, которая впервые была сформулирована в работе (Haas, 1922) и получила развитие в работах (Griliches, 1961) по ценам автомашин и (Chow, 1967) по ценам на компьютеры. В работе (Lancaster, 1966) содержится изложение экономической теории, на которой основан метод гедонистических цен.

Динамика цен спроса квартир в Москве за 16-летний период представлена на рис. 1.[1] После падения Берлинской стены 9 ноября 1989 г.

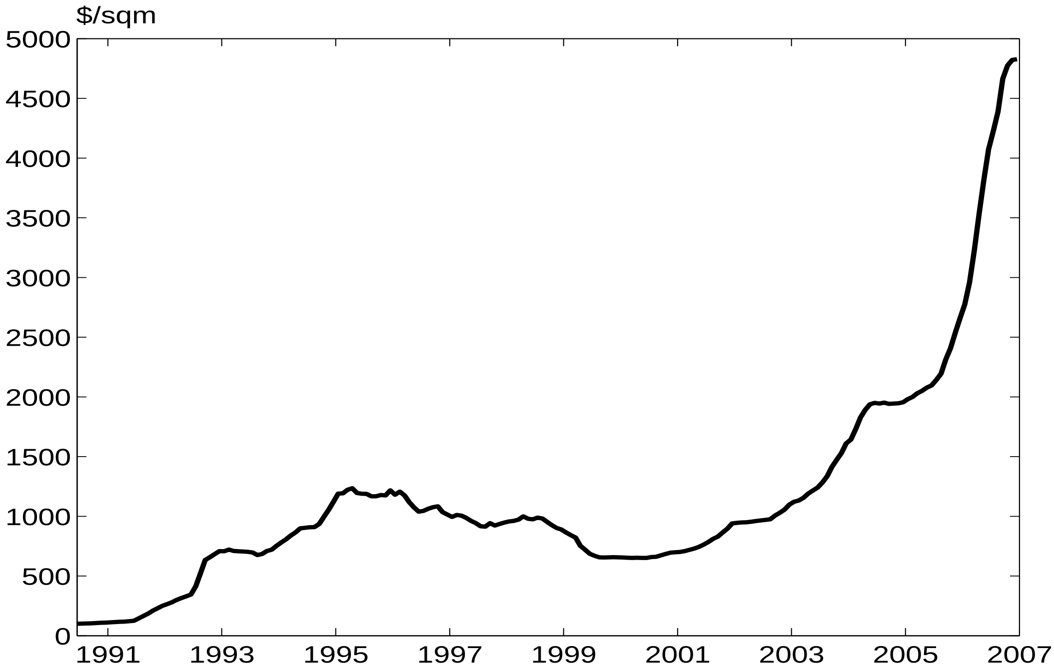

Рис. 1. Цены спроса на квартиры в Москве, 1990–2006 гг.

и путча в августе 1991 г. Советский Союз распался и Б. Ельцин стал первым президентом России. Рубль претерпел два серьезных кризиса в 1990-е годы: первый — 11 октября 1994 г. («черный вторник») и второй

[1]График построен по ежемесячным наблюдениям за период с июня 1990 г. по декабрь 2006 г. (см. Мальгинов, Стерник (2006), данные до декабря 2005 г.). Данные за 2006 г. любезно предоставлены Г. Стерником.

— в понедельник 17 августа 1998 г. Эффект этих кризисов виден на графике. После 2000 г. цены на недвижимость стремительно росли. Средняя цена спроса в июне 2000 г. была 652 $/кв.м., а через 6 лет, в июне 2006 года она уже достигла 4072 $/кв.м. В настоящее время Москва является одним из самых дорогих городов мира и по ценам на недвижимость занимает третье место после Монако и Лондона. По количеству долларовых миллиардеров Москва уступает только Нью-Йорку и Лондону. В процессе строительства находятся амбициозные проекты, как, например, 64-этажное здание башни Федерации на Краснопресненской набережной.

При таком нестабильном макроэкономическом окружении оценка и прогноз цен на недвижимость являются рискованным предприятием. Тем не менее обладание выборкой большого размера (около 15 000 наблюдений) дает нам основание надеяться получить результаты, проливающие свет на этот рынок.

В том случае, если бы нас интересовала динамика цен на недвижимость, было необходимо включить в рассмотрение состояние экономики, финансового рынка, ожидания. Однако нас интересуют относительные цены в один момент времени, что позволяет использовать гедонистический подход.

Работа преследует три основных цели. Во-первых, моделировать стоимость квартиры в зависимости от ее характеристик по данным веб-сайтов. Такие модели оцениваются на половине выборки. Во-вторых, другая половина выборки используется для тестирования точности прогноза по построенной модели. Поскольку цены на недвижимость зависят от множества индивидуальных характеристик объектов недвижимости, что особенно важно для такого нестабильного рынка как московский, было бы наивно предполагать, что цены могут определяться простой моделью, в которой пропущены многие важные факторы. Однако полученные по моделям прогнозы представляются разумными и полезными. В-третьих, мы рассматриваем эффект предварительного тестирования (pretesting). Опасность игнорирования этого эффекта не учитывается в большинстве прикладных работ.

В экономике недвижимости метод гедонистических цен используется для того, чтобы принять во внимание неоднородность зданий, результатом которой является трудность оценивания спроса на здания в целом. Гедонистический подход предполагает, что дом или квартира могут быть представлены как набор их характеристик, например, число и площадь комнат, расстояние до центра города, качество окружающей среды (загрязнение воздуха, воды, шум), красивый вид из окна, близость к парку или реке и т. п.

Среди работ, применяющих гедонистический подход к оценке недвижимости в США,наиболее известными являются (Haas, 1922) в Миннесоте, 1916–1919; (Bailey et al., 1963) в Сент-Луисе, 1937–1959; (Witte et al., 1979), основанная на теории, развитой в (Rosen, 1974), в Северной Каролине, 1972; (Milton et al., 1984) во Флориде; и (Mills, Simenauer, 1996) в регионах США, 1986–1992.

Среди работ по европейской недвижимости отметим работы (Bender et al., 1994) в Женеве, 1978–1992; (Lansink, Thijssen, 1998) в Нидерландах, 1970–1988; (Maurer et al., 2004) в Пеариже, 1990–1999; и (van Soest, Verbeek, 2010) в Москве, 1994 и 1996.

В большинство работ, за исключением работы (Maurer et al., 2004) используются данные лишь по нескольким сотням сделок. Наш набор данных имеет значительно больший объем и состоит примерно из 15 000 наблюдений.

План статьи следующий. Раздел 2 содержит описание данных. В разделе 3 строится модель и обсуждаются некоторые результаты. Затем в разделе 4 рассматривается качество прогноза по построенным моделям. Раздел 5 содержит обсуждение эффекта предварительного тестирования и, наконец, в разделе 6 содержатся выводы.

2. Данные

Используемые в работе данные были собраны студентами Российской экономической школы (РЭШ) в январе-феврале 2003 г. в рамках курсовых проектов по курсу «эконометрика-1». Исходные данные состояли из 16 115 наблюдений по одно-, двух-, трех- и четырехкомнатным квартирам. Большую часть выборки (43 %) составляют двухкомнатные квартиры, затем идут трехкомнатные (28 %) и однокомнатные (25 %). Четырехкомнатные квартиры составляют лишь 4 % выборки и — согласно мнению московских риелтеров — цена каждой квартиры индивидуальна и не следует каким-то общим закономерностям. Поскольку целью статьи является моделирование стандартных квартир, составляющих большую часть рынка, мы исключили четырехкомнатные квартиры из выборки, после чего в ней осталось 15 476 наблюдений.

Данные по одно-, двух- и трехкомнатным квартирам были собраны 66 студентами, организованными в 19 групп. Каждой группе было дано задание собрать данные по одной из категорий квартир (одно-, двух- или трехкомнатных). Группы должны были координировать свою работу (например, по линиям метро или административным округам) так, чтобы все районы Москвы были представлены, и ни одна из квартир не встречалась бы одновременно в данных двух групп.

Источником данных являлись веб-сайты агентств недвижимости, наиболее популярных в Москве в 2003 г.:

www.appartment.ru, www.astet.ru, www.babilon.ru, www.kdo.ru.
www.estate.msk.ru, www.kont.ru, www.mian.ru, www.miel.ru.
www.novostroy.ru, www.orsn.ru, www.realty.ru, www.trigon.ru,

а так же данные агентства недвижимости «Фили». Три типичных примера информации по предложению квартир приведены в табл. 1, в которой

Таблица 1. Три объявления о продаже квартир

Количество комнат	3	3	1
Метро	Авиамоторная	Автозаводская	Академическая
Адрес	ул. Синичкина, 4 кор. 2	ул. Трофимова, 1/17	ул. Шверника, 7
Расстояние до метро	5 мин на транспорте	10 мин пешком	10 мин пешком
Цена	75 000 $	86 000 $	39 000 $
Этаж	5/5	4/8	4/12
Тип дома	кирпичный	кирпичный	панельный
Площадь: общая/жилая/кухня	73/49/9	83/54/9	35/14/10

представлены данные по двум трехкомнатным и одной однокомнатной квартирам. Указана ближайшая станция метро, адрес. Номер этажа 4/8 означает, что квартира находится на 4-м этаже 8-этажного дома. В графе общая/жилая/кухня указаны общая, жилая площади и площадь кухни (в кв.м.).

Все цены являются ценами предложения, мы их используем, поскольку информация по ценам реальных сделок недоступна. В Москве (и в Санкт-Петербурге) в 2003 г. цены указывались в долларах США, в то время как в других городах большинство цен указывалось в рублях. Данные содержат предложения как первичного, так и вторичного рынков квартир. В наборе данных присутствуют следующие переменные.

price: цена предложения квартиры (в 1000 $), указанная в объявлении, используемая как прокси для неизвестной цены реальной сделки;

площади:

- *totsp:* общая площадь квартиры, в кв.м.;
- *livsp:* жилая площадь квартиры, в кв.м.;
- *kitsp:* площадь кухни, в кв.м. Рассматриваются только стандартные квартиры с отдельной кухней.

Кроме *livsp* и *kitsp* квартира имеет дополнительную площадь (*addsp*), включающую ванну, туалет, коридор и др. По определению,

$$totsp = livsp + kitsp + addsp.$$

rooms: количество комнат в квартире (1, 2 или 3).

переменные, характеризующие расстояния:

- *dist:* расстояние до центра Москвы, в км. Москва имеет радиально-кольцевую структуру с центром в районе Красной площади. Расстояние измеряется от ближайшей станции метро до центра города;
- *metrdist:* время дороги до ближайшей станции метро, в минутах пешком или на общественном транспорте (автобус, троллейбус, трамвай, маршрутное такси); см. фиктивная переменная *walk*;

фиктивные переменные:

- *walk:* 1 — если расстояние *metrdist* указано в минутах пешком; 0 — иначе. Пара *metrdist*, *walk* характеризует расстояние от квартиры до метро. Например, если в объявлении указано «5 минут до метро пешком», то $metrdist = 5$ и $walk = 1$, а если указано «10 минут до метро на транспорте», то $metrdist = 10$ и $walk = 0$.
- *brick:* 1 — если квартира в доме построенном из кирпича или монолитного железобетона; 0 — иначе (панельные или блочные дома). Эта переменная отражает качество здания;
- *tel:* 1 — если в квартире есть городская телефонная линия; 0 — иначе. Как правило, $tel = 1$, кроме новых зданий в районах с неразвитой инфраструктурой. Таким образом, *tel* служит прокси для развитой инфраструктуры. Если телефонная линия отсутствует, то это создает определенные неудобства, телефонизация дома может занять 1-2 года. Конечно, с развитием мобильной связи фактор наличия городской телефонной линии становится менее значимым;

- *balc:* 1 — если в квартире есть хотя бы один балкон или лоджия; 0 — иначе. Как правило, в квартирах на первом этаже нет балкона по причине безопасности;

- *floor:* 1 – если квартира не на первом и не на последнем этаже; 0 — иначе. В Москве квартиры на первом и последнем (кроме пентхаусов) этажах менее популярны, поскольку, как правило, в них нет балкона. Кроме того в квартире на последнем этаже может быть шумно из-за механики лифтов, могут быть протечки при нарушении гидроизоляции крыши, а летом может быть слишком жарко при плохой теплоизоляции. Квартира на первом этаже считается менее безопасной, может страдать от шума лифта или входной двери, плохого запаха при нарушении изоляции подвала, в ней может быть холодный пол, а также ей свойственна недостаточная приватность — в квартиру легко заглянуть с улицы.

Перед тем как использовать данные в расчетах, они должны быть очищены от опечаток и проверены. Очевидные опечатки были удалены, например такие, как $totsp = 100\,000$, что маловероятно, $floor < 0$, что невозможно, также невозможна комбинация ($walk = 0$ и $metrdist = 0$), которая означала бы путешествие 0 минут на транспорте от ближайшей станции метро. Кроме того, мы наложили следующие ограничения на данные:

$20 \le price \le 250$. Если цена менее 20 000 $ это может означать особую ситуацию. Например, цена в объявлении 18 000 $ может означать, что истинная цена равна 36 000 $, но предполагается, что покупатель платит 50 % сейчас и 50 % через год. Возможно, также что агент скажет что-то вроде «да, цена 18 000 $. Но владелец квартиры — пожилая женщина и нуждается в деньгах сейчас. Мы подписываем контракт, по которому Вы платите деньги сейчас, но становитесь владельцем квартиры, когда женщина умрет». Если цена превышает 250 000 $, то она, скорее всего, относится к «элитной квартире», которая принадлежит совершенно другому сегменту рынка недвижимости.

$25 \le totsp \le 150$. Если общая площадь однокомнатной квартиры менее 25 кв.м., это, скорее всего, означает, что квартира «гостиничного» типа, т. е. категории, которую мы не рассматриваем.

$10 \le livsp \le 100$. Трудно вообразить однокомнатную квартиру с площадью комнаты менее 10 кв.м.

$5 \leq kitsp \leq 25$. Площадь кухни менее 5 кв.м. может означать квартиру «гостиничного» типа, или опечатку.

$6 \leq addsp \leq 45$. Маловероятно, что коридор и совмещенный санузел имеют площадь менее 6 кв.м.

$0.4 \leq psqm \leq 3.0$, где $psqm = price/totsp$ означает цену квадратного метра квартиры. В январе-феврале 2003 г. цена менее 400 $ за кв.м. подозрительно мала и может означать предоплату, престарелого владельца, общежитие или какую-то другую особую ситуацию. Цены выше 3000 $ за кв.м. относятся к сегменту «элитных» квартир.

$9 \leq livsp/rooms \leq 30$ означает, что средняя площадь комнаты в квартире не должна быть слишком большой или слишком маленькой. Если средняя площадь превышает 30 кв.м., это может означать квартиру со свободной планировкой.

Наложив эти ограничения, мы удаляем из выборки 754 наблюдения, и наш набор данных сокращается с 15 476 до 14 722 наблюдений. С этим «очищенным» набором данных мы и работаем в дальнейшем.

Студенты, участвовавшие в сборе данных — серьезные и ответственные студенты мастер-программы по экономике РЭШ. В процессе сбора данных они консультировались с преподавателями. Тем не менее можно предположить недобросовестность в работе одной-двух групп, например, студенты могли взять данные одного из прошлогодних проектов и умножить цены на некоторую константу. Ниже мы покажем, как можно контролировать наличие возможной недобросовестности студентов.

В качестве предварительного анализа данных рассмотрим описательные статистики переменных, приведенные в табл. 2, рассчитанные по 3533 однокомнатным квартирам, 6785 двухкомнатным квартирам и 4404 трехкомнатным квартирам — всего 14 722 квартир.

Таблица 2. Описательные статистики

Комнат		*price*	*totsp*	*livsp*	*kitsp*	*addsp*	*psqm*	*dist*	*metrdist*
1	Среднее	38.3	37.3	19.3	8.2	9.8	1.03	11.9	9.4
	Медиана	36.0	37.0	19.0	8.0	9.0	1.00	12.5	10.0
2	Среднее	58.6	52.1	31.3	8.0	12.8	1.11	9.9	8.4
	Медиана	50.0	52.0	31.0	8.0	12.0	1.00	10.5	10.0
3	Среднее	76.4	72.1	46.6	8.7	16.9	1.04	10.1	8.7
	Медиана	68.0	72.6	45.0	8.5	16.0	0.97	10.5	10.0

Из табл. 2 видно, что квартиры с различным количеством комнат практически не различаются по расстоянию от метро, расстоянию до центра Москвы, цене квадратного метра и качеству дома. Это не удивительно, поскольку в типовых жилых домах в Москве обычно есть квартиры различных категорий. Средняя площадь кухни также практически одинаковая. Однако другие площади *totsp*, *livsp*, *addsp* (и, конечно, цена *price*) существенно различаются. Среднее значение переменной *floor*, не включенной в таблицу, несколько меньше для однокомнатных квартир, что может означать, что в старых малоэтажных домах выше пропорция однокомнатных квартир или что однокомнатные квартиры на первом этаже чаще других выставляются на продажу.

3. Модели

Для построения модели стоимости квартиры мы выбрали в качестве зависимой переменной $\log(price)$, поскольку она гораздо лучше, чем *price*, аппроксимируется нормальным распределением. Кроме того, мы проводим различие между переменными, представляющими особый интерес для исследования (focus variables), и вспомогательными переменными (auxiliary variables).

Основные переменные — те, которые мы обязательно хотим включить в модель, поскольку оценка их влияния является целью исследования или поскольку экономическая теория или здравый смысл указывают, что они должны быть включены в модель. Вспомогательные переменные не представляют прямого интереса и включаются в модель лишь потому, что у нас есть априорное представление, что это поможет точнее оценить влияние основных переменных. Такое априорное разделение переменных на основные и вспомогательные означает, что переменные могут быть оставлены в модели даже в том случае, если соответствующие им t-статистики малы. Как будет показано ниже, это также позволяет точнее учесть смещение в pretest-оценках (оценках, полученных после процедуры предварительного тестирования) и их дисперсиях.

В качестве основных выделены 8 переменных и 3 выбраны в качестве вспомогательных.

Основные переменные: константа (*constant*) рассматривается как основная переменная. Площадь квартиры представлена переменной $\log(totsp)$ и отношением *kitsp/totsp*. Расположение квартиры представлено ее расстоянием от центра $\log(dist)$ и тремя переменными, совместно характеризующими расстояние от квартиры до ближайшей станции метро: *metrdist* $\times$ *walk*, *metrdist* $\times$ $(1 - walk)$ и *walk*. Переменная *floor* также является основной, поскольку нас интересует, оказывает ли она влияние на цену.

Вспомогательные переменные: наличие балкона (*balc*) рассматривается как вспомогательная переменная. Поскольку на первом этаже балконов, как правило, не бывает, в уравнение также включен перекрестный член *balc* $\times$ *floor*. Кроме того, включается фиктивная переменная *group 18* — индикатор группы студентов, которая подозревается в недобросовестном сборе данных.

Итак, модель без ограничения включает все 11 переменных, а модель с ограничениями — только 8 основных переменных. Таким образом, процедура отбора касается только трех вспомогательных переменных.

Перед выбором модели и оцениванием ее параметров мы делим выборку на две части, поскольку хотим оценивать модель на одной части выборки, а тестировать точность прогноза по полученной модели на независимой выборке. Случайным образом выборка разбивается на 2 части: первую («выборка 1», 7395 наблюдений), на которой оцениваются коэффициенты модели, и вторую («выборка 2», 7327 наблюдений), которая используется для анализа точности прогноза.

3.1. Выбор модели

Результаты оценивания параметров моделей основаны на выборке 1 и представлены в табл. 3.

Сначала мы оцениваем модель без ограничений, M1. В этой модели t-статистики трех вспомогательных параметров равны соответственно 7.5, -4.8 и -1.4. Поэтому мы решили удалить из модели переменную *group 18* и оставить переменные *balc* и *balc* $\times$ *floor*. Получаем модель M2 (отобранная модель). Заметим, что результат не зависит от процедуры выбора модели: от частного к общему или от общего к частному — результат один и тот же в обеих процедурах. Оценка отобранной модели приведена во второй колонке табл. 3 и мы видим, что оценки коэффициентов изменились незначительно.

Колонки 3 и 4 (модели M3 и M4) включены в табл. 3 для сравнения. Хотя модель с ограничением M3 и отвергается в пользу модели M2, наложенное ограничение не изменяет значительно оценки коэффициентов при основных переменных. Простейшая модель M4 хуже остальных, но и она дает разумные значения коэффициентов.

Таблица 3. Результаты регрессий

Переменная	Модель M1	Модель M2	Модель M3	Модель M4
constant	−0.053	−0.060	−0.032	—
	(0.044)	(0.043)	(0.043)	—
$\log(totsp)$	1.080	1.082	1.083	1.094
	(0.009)	(0.009)	(0.009)	(0.002)
$kitsp/totsp$	0.904	0.891	0.936	—
	(0.058)	(0.057)	(0.057)	—
$\log(dist)$	−0.203	−0.203	−0.199	−0.180
	(0.004)	(0.004)	(0.004)	(0.004)
$metrdist \times walk$	−0.004	−0.004	−0.004	—
	(0.001)	(0.001)	(0.001)	—
$metrdist \times (1 - walk)$	−0.007	−0.007	−0.007	—
	(0.001)	(0.001)	(0.001)	—
walk	0.053	0.053	0.052	0.093
	(0.011)	(0.011)	(0.011)	(0.005)
floor	0.093	0.093	0.058	—
	(0.010)	(0.010)	(0.005)	—
balc	0.069	0.069	—	—
	(0.009)	(0.009)	—	—
$balc \times floor$	−0.054	−0.054	—	—
	(0.011)	(0.011)	—	—
group 18	−0.014	—	—	—
	(0.010)	—	—	—
$\hat{\sigma}^2$	0.033	0.033	0.033	0.036
R^2	0.781	0.781	0.779	0.760

M1 — модель без ограничения;
M2 — отобранная модель;
M3 — модель с ограничением;
M4 — простейшая (наивная) модель.

Знаки всех коэффициентов соответствуют первоначальным ожиданиям. Согласно общему мнению, цена квартиры в основном зависит от размера и расположения. Расположение и в самом деле значимо: как расстояние до центра, так и расстояние до ближайшей станции метро являются значимыми регрессорами. Поскольку у нас нет данных по качеству воздуха, близости зеленых насаждений и т. п., мы не в состоянии более точно учесть влияние расположения квартиры на ее стоимость.

Площадь квартиры — еще более значимый фактор. Если удалить переменную $\log(totsp)$ из регрессии, то значение R^2 падает с 0.781 до 0.352. Оказывается, цена квадратного метра *не* является постоянной. Квартиры большой площади не только более дорогие, но и *относительно* более дорогие: коэффициент при $\log(totsp)$ статистически значимо превосходит 1. Это может объясняться тем, что большие квартиры встречаются достаточно редко, а также тем, что они часто являются более «элитными».

3.2. Проверка работы групп по сбору данных

Работа по сбору данных довольно трудоемкая, и, возможно, какая-то группа студентов работала недобросовестно. В предыдущем учебном году студенты выполняли аналогичное задание. Поэтому было бы легко взять данные прошлого года и умножить все цены на фиксированный множитель. Такую недобросовестность в сборе данных трудно обнаружить. Предположим, мы предполагаем возможность подобной проблемы с данными группы 18.

Эта группа была одной из групп, которые собирали данные по однокомнатным квартирам. Их данные состоят из 710 наблюдений (из них 364 в выборке 1). Из табл. 3 видно, что коэффициент при *group 18* незначим, (t-статистика равна -1.4). Поэтому нет достаточных доказательств того, что данные были собраны недобросовестно.

Конечно, можно сказать, что этот результат неубедителен, так как группа 18 собирала данные только по однокомнатным квартирам. Для того чтобы учесть это возражение, вставим в регрессию фиктивные переменные *R1* и *R2* — индикаторы одно- и двухкомнатных квартир. При этом коэффициент при *group 18* становится равен -0.006 (0.011), это означает, что он также незначимо отличается от 0.

Поэтому в дальнейшем мы предполагаем, что все группы студентов работали добросовестно.

3.3. Оптимальная планировка квартиры

Коэффициент при $kitsp/totsp$ равен 0.891, положителен, следовательно, в рассматриваемый момент кухни занимали слишком малую часть площади квартиры. В нашей выборке среднее значение доли кухни в общей площади равно 15.9 %. До какого значения увеличение этой доли ведет к удорожанию квартиры? Каково оптимальное значение площади кухни?

Для ответа на эти вопросы добавим переменную $(kitsp/totsp)^2$ в уравнение M2. В общем случае в уравнении

$$y = \alpha + \beta_1 x + \beta_2 x^2 + \dots,$$

где $\beta_1 > 0$ и $\beta_2 < 0$, значение x, при котором y достигает максимума, равно $x^* = -\beta_1/(2\beta_2)$. Взяв $x = kitsp/totsp$, мы получаем $\hat{\beta}_1 = 3.636$ и $\hat{\beta}_2 = -7.411$, откуда $x^* = 0.245$. Таким образом, для того чтобы стоимость квартиры была максимальной (т. е. ее привлекательность для покупателя была максимальной), кухня должна занимать 24.5 % общей площади квартиры. Заметим, что новые серии многоквартирных домов в Москве приближаются к этому оптимальному значению.

3.4. Расстояние до центра

Большинство московских квартир двухкомнатные, и наблюдается нехватка однокомнатных квартир. Повышенный спрос на однокомнатные квартиры может привести к тому, что их цена убывает с увеличением расстояния до центра медленнее, чем цена двухкомнатных квартир.

Для проверки этой гипотезы заменим в уравнении M2 константу на три фиктивные переменные *R1*, *R2* и *R3*, а также заменим $\log(dist)$ на $\log(dist) \times R1$, $\log(dist) \times R2$ и $\log(dist) \times R3$. В модели M2 увеличение расстояния от центра на 1 % приводит к уменьшению цены на 0.20 %. В новом уравнении это уменьшение цены равно −0.16 %, −0.21 % и −0.19 % для одно- двух- и трехкомнатных квартир соответственно. Причем полученное различие между оценками эластичности статистически значимо, что подтверждает нашу гипотезу.

Стоимость трехкомнатных квартир также убывает с расстоянием медленнее, чем стоимость двухкомнатных квартир. Этот факт не означает повышенного спроса на трехкомнатные квартиры. Возможно, владельцы трехкомнатных квартир более состоятельны и имеют автомобили, и потому расстояние для них имеет меньшее значение. Кроме того, в трехкомнатных квартирах часто проживают семьи с детьми, которые, вероятно, предпочитают экологически более благоприятные районы на удалении от центра, в то время как одиночки — владельцы однокомнатных квартир предпочитают близость к центру.

3.5. Расстояние до метро

Напомним, что переменная *metrdist* означает время в пути от ближайшей станции метро до квартиры пешком или на общественном транспорте. Из уравнения M2 (табл. 3) мы видим, что

$$\log(price) = \begin{cases} \dots - 0.007 \times metrdist, & \text{если } walk = 0, \\ \dots + 0.053 - 0.004 \times metrdist, & \text{если } walk = 1. \end{cases}$$

Таким образом, в сравнении с квартирой, расположенной рядом с метро, такая же квартира на удалении 5, 10, 20, 30 минут пешком от метро дешевле соответственно на 2.0 %, 3.9 %, 7.7 %, 11.3 %. Если же до квартиры надо добираться на общественном транспорте, то она дешевле соответственно на 8.6 %, 11.6 %, 17.6 %, 23.1 %.

Более детальный анализ, аналогичный проведенному выше, показывает, что для покупателей трехкомнатных квартир близость к метро менее важна, чем для покупателей одно- и двухкомнатных квартир. Возможно, они достаточно состоятельны и пользуются автомобилями.

3.6. Влияние этажа

Оценка коэффициента при переменной *floor* равна 0.093, что означает предпочтение этажей, кроме первого и последнего. Влияние переменной *floor* на однокомнатные квартиры оказывается значительно меньшим, чем на многокомнатные. Вероятно, покупатели однокомнатных квартир обычно являются молодыми семьями, с относительно невысокими доходами, которые менее чувствительны к «плохим» этажам.

4. Прогноз стоимости

Трудно ожидать, что цена квартиры будет точно прогнозироваться нашей моделью. Цены на недвижимость зависят от множества характеристик, и многие из них, особенно касающиеся расположения дома, отсутствуют в наших данных. Прогноз, полученный по нашей модели, может рассматриваться лишь как начальная точка отсчета, которую покупатель, продавец или агент по продаже недвижимости могут принимать во внимание.

Мы рассматриваем отобранную модель M2 (табл. 3), оценки коэффициентов которой получены по набору данных выборки 1. Поскольку значения регрессоров, входящих в модель, известны на всем наборе данных, то прогнозные значения цен можно рассчитать для каждого наблюдения в выборке и сравнить с их истинными значениями. Пусть p обозначает истинную цену, а $\hat{p}$ прогноз цены. На рис. 2 представлен график прогнозной цены $\log(\hat{p})$ по сравнению с истинной ценой для каждого из 7327

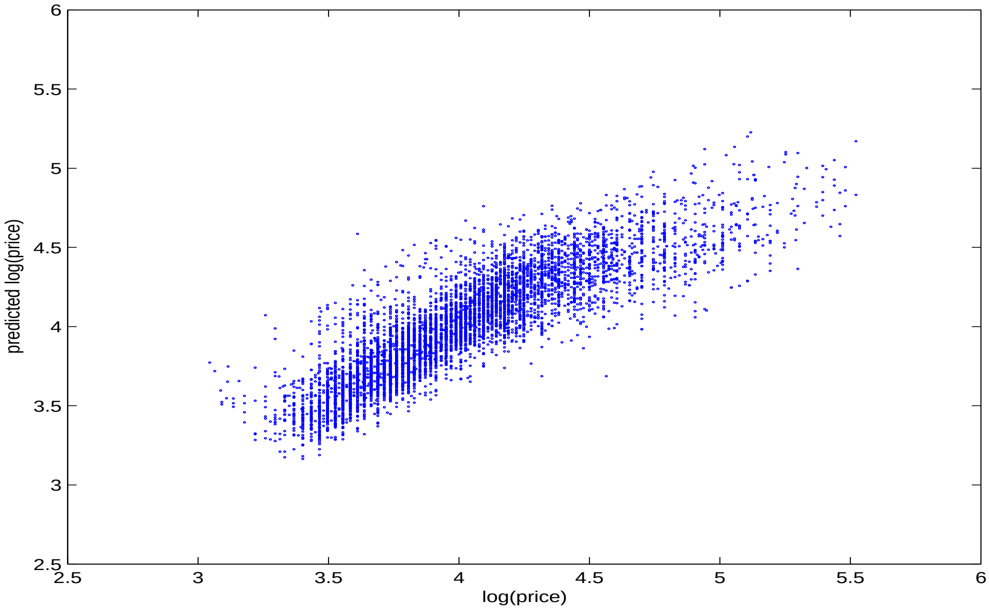

Рис. 2. Диаграмма рассеивания $(\log(\hat{p}), \log(p))$

наблюдений в выборке 2. Поскольку цены спроса дискретны, а прогнозы цен непрерывны, то график имеет структуру дискретных вертикальных линий. Облако точек несколько изогнуто, модель недооценивает стоимость наиболее дорогих квартир. Возможно, цена таких квартир зависит от факторов, не включенных в наш набор данных: вид на парк или реку, охраняемая территория, ванна с джакузи и т. п.

Гистограмма относительных ошибок прогноза цены $(\hat{p} - p)/p$ представлена на рис. 3. Как и ожидалось, гистограмма несколько скошена вправо. Лишь 28.6 % прогнозов отличаются менее чем на 5 % от действительных цен; 52.2 % — менее чем на 10 %; 85.6 % — менее, чем на 25 %. Кроме того, 6.5 % прогнозных значений цены на 25 % ниже истинных значений, а 7.8 % прогнозных значений цены на 25 % превышают истинную цену. В целом прогнозные значения цены дают полезную точку отсчета для оценки стоимости квартиры, но они также и подтверждают гипотезу, что цены на недвижимость определяются бо́льшим количеством факторов, чем те, что включены в наш набор данных.

5. Влияние предварительного тестирования

В эконометрике обычно один и тот же набор данных используется для выбора и оценивания модели (или прогноза). Поэтому стандартная ста-

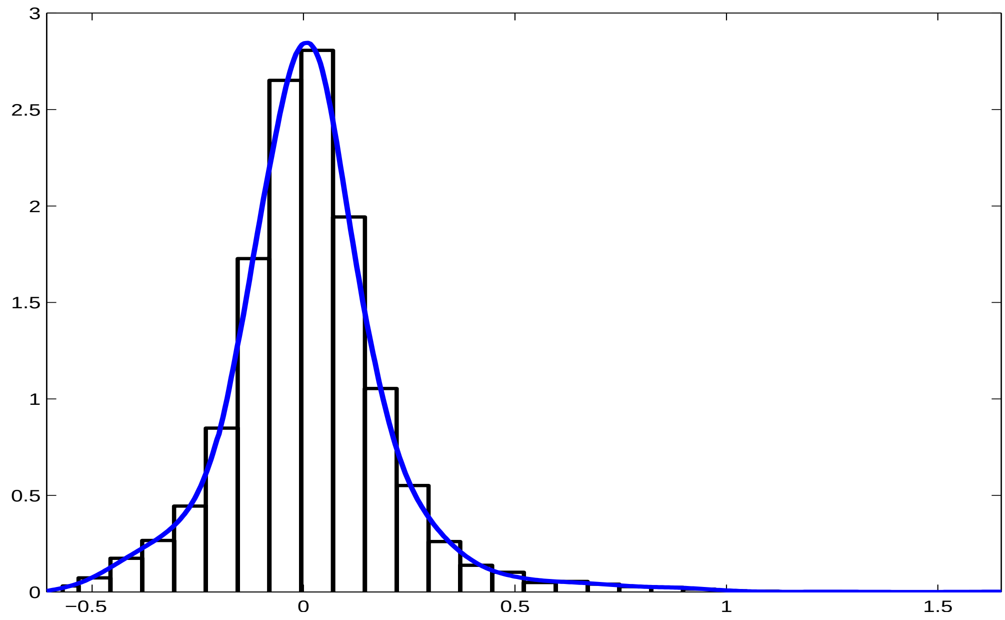

Рис. 3. Гистограмма относительных ошибок прогноза цены $(\hat{p} - p)/p$.

тистическая теория неприменима напрямую, поскольку свойства оценок зависят не только от статистической природы модели, но и от того способа, каким модель была выбрана. Соответствующая «теория предварительного тестирования» (pretest theory), которая рассматривает эту проблему, была (по крайней мере, частично) разработана в работах (Magnus, Durbin, 1999) и (Danilov, Magnus, 2004a,b). В данном разделе мы анализируем возможный эффект предварительного тестирования на представленные выше оценки.

Рассмотрим стандартную модель линейной регрессии:

$$y = X\beta + Z\gamma + \varepsilon,$$

где y — $(n \times 1)$ вектор наблюдений, X $(n \times k)$ и $Z\,(n \times m)$ — матрицы регрессоров (которые предполагаются неслучайными), ε — $(n \times 1)$ вектор ошибок (ненаблюдаемый, случайный), $\beta\,(k \times 1)$ и $\gamma\,(m \times 1)$ — неизвестные, неслучайные векторы параметров уравнения. Предполагается, что $k \geq 1$, $m \geq 1$, $n - k - m \geq 1$, также предполагается, что матрица $(X : Z)$ имеет полный ранг по столбцам $k + m$, а ошибки $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ есть независимые, одинаково распределенные нормальные случайные величины (i.i.d. $\mathrm{N}(0, \sigma^2)$).

Причина, по которой регрессоры разделены на две группы X и Z, состоит в том, что X содержит основные объясняющие переменные (‘focus’ regressors). Они должны быть включены в модель по теоретическим или каким-то другим основаниям, в то время как Z содержит вспомогательные объясняющие переменные (‘auxiliary’ regressors), в необходимости включения которых в уравнение нет твердой уверенности, и их роль состоит только в улучшении свойств оценок β. В нашем случае имеется $k = 8$ основных регрессоров и $m = 3$ вспомогательных.

Определим следующие матрицы:

$$M := I_n - X(X'X)^{-1}X' \quad \text{и} \quad Q := (X'X)^{-1}X'Z(Z'MZ)^{-1/2},$$

и нормализованный вектор параметров $\theta = (Z'MZ)^{1/2}\gamma$. МНК-оценки векторов β и γ равны соответственно $b_u = b_r - Q\hat{\theta}$ и $\hat{\gamma} = (Z'MZ)^{-1/2}\hat{\theta}$, где $b_r = (X'X)^{-1}X'y$ и $\hat{\theta} = (Z'MZ)^{-1/2}Z'My$. Индексы u и r обозначают оценки без ограничения (unrestricted) и оценки с ограничением (restricted, т.е. при $\gamma = 0$). Заметим, что $\hat{\theta} \sim \mathrm{N}(\theta, \sigma^2 I_m)$, а также то, что b_r и $\hat{\theta}$ являются независимыми случайными векторами.

Обозначим через S_i матрицу ограничений (selection matrix) размера $m \times r_i$ и ранга r_i $(0 \leq r_i \leq m)$, такую, что $S_i' = (I_{r_i} : 0)$ или получается из нее перестановкой столбцов. Уравнение $S_i'\gamma = 0$ выделяет подмножество компонент вектора γ, которые положены равными нулю. Согласно работе (Danilov, Magnus, 2004a), МНК-оценки коэффициентов β и γ при ограничении $S_i'\gamma = 0$ задаются формулами:

$$b_{(i)} = b_r - QW_i\hat{\theta}, \qquad c_{(i)} = (Z'MZ)^{-1/2}W_i\hat{\theta},$$

где

$$W_i = I_m - (Z'MZ)^{-1/2}S_i\big(S_i'(Z'MZ)^{-1}S_i\big)^{-1}S_i'(Z'MZ)^{-1/2}$$

является симметрической идемпотентной $m \times m$ – матрицей ранга $m - r_i$. (В том случае, если $r_i = 0$, получаем $W_i = I_m$). Распределение вектора $b_{(i)}$ равно:

$$b_{(i)} \sim \mathrm{N}\left(\beta + Q(I_m - W_i)\theta,\ \sigma^2\left((X'X)^{-1} + QW_iQ'\right)\right).$$

Выделяя в наборе $(\gamma_1, \ldots, \gamma_m)$ одно из 2^m (в нашем случае 8) подмножеств и приравнивая выделенные γ_i к нулю, получаем одну из 2^m возможных моделей с ограничением. Оценка, полученная предварительным тестированием (*pretest*-оценка) вектора β получается в результате следующей процедуры. Во-первых, выбирается одна из этих 2^m моделей на основе t- или F-тестов или других критериев отбора модели. Во-вторых, оценивается вектор β по этой выбранной на первом шаге модели. Предположим, что процедура выбора модели основана исключительно на анализе остатков модели с ограничением, т.е. на My. Это предположение выполняется во всех стандартных критериях отбора. Таким образом, pretest-оценка вектора β может быть записана в виде $b = \sum_i \lambda_i b_{(i)}$, где веса λ_i удовлетворяют условиям:

$$\lambda_i = \lambda_i(My), \qquad \lambda_i \geq 0, \qquad \sum_i \lambda_i = 1,$$

где суммирование идет по всем 2^m моделям, и все λ_i равны 0, за исключением одного, равного 1. Отсюда pretest-оценка вектора β равна $b = b_r - QW\hat{\theta}$, где $W = \sum_i \lambda_i W_i$. Заметим, что W является случайной матрицей, поскольку $\{\lambda_i = \lambda_i(My)\}$ есть случайные величины.

Введем обозначения $\eta = \theta/\sigma$ и $\hat{\eta} = \hat{\theta}/\sigma$. Обозначим также через $V = \sigma^2(X'X)^{-1}$ матрицу ковариаций в модели с ограничением, с диагональными элементами v_{jj} $(j = 1, \ldots, k)$. По теореме эквивалентности (Danilov, Magnus, 2004a, теорема 1) получаем, что смещение, дисперсия и среднеквадратичное отклонение компонент вектора b для $j = 1, \ldots, k$ задаются формулами:

$$\begin{aligned} \operatorname{bias}(b_j) &:= \operatorname{E}(b_j - \beta_j) = -\sqrt{v_{jj}\, q_{0j}^2} \cdot q_j' \operatorname{E}(W\hat{\eta} - \eta), \\ \operatorname{var}(b_j) &= v_{jj}\left(1 + q_{0j}^2 \cdot q_j' \operatorname{var}(W\hat{\eta}) q_j\right), \end{aligned}$$

и, следовательно,

$$\operatorname{MSE}(b_j) = v_{jj}\left(1 + q_{0j}^2 \cdot q_j' \operatorname{MSE}(W\hat{\eta}) q_j\right),$$

где

$$q_{0j}^2 = \frac{e_j' Q Q' e_j}{e_j'(X'X)^{-1} e_j}, \qquad q_j = \frac{Q' e_j}{\sqrt{e_j' Q Q' e_j}},$$

а e_j обозначает j-ый единичный вектор. Получаем, что свойства сложно устроенной pretest-оценки существенно зависят от свойств более простой оценки $W\hat{\eta}$ вектора η. Различные процедуры отбора модели приводят к различным эффектам предварительного тестирования, которые целиком зависят от свойств оценки $W\hat{\eta}$.

В дальнейшем предположим, что дисперсия σ^2 известна, а именно равна оценке $\hat{\sigma}^2$ в модели без ограничения. Это несколько упрощает последующий анализ, а как показано в (Danilov, 2005), при большом количестве наблюдений этой неточностью можно пренебречь. Удобно шкалировать вспомогательные регрессоры z_1, z_2 и z_3 следующим образом: $z_i^* = z_i/\sqrt{z_i' M z_i}$, $i = 1, 2, 3$. Это шкалирование не влияет на дальнейший анализ. Получаем:

$$Z^{*\prime} M Z^* = \begin{pmatrix} 1.000 & 0.809 & -0.012 \\ 0.809 & 1.000 & 0.001 \\ -0.012 & 0.001 & 1.000 \end{pmatrix}.$$

Видно, что первые два вспомогательных регрессора сильно коррелированы, а третий вспомогательный регрессор (*group 18*) почти некоррелирован с первыми двумя. При высокой корреляции обычно $\mathrm{MSE}(W\hat{\eta})$ велика. Таким образом, можно ожидать более сильное влияние предварительного тестирования для двух первых вспомогательных регрессоров. В нашем случае это влияние мало, поскольку значения t-статистик велики (также велико и количество наблюдений).

Таблица 4. Статистики вспомогательных переменных

Переменная	$\hat{\gamma}$	t_γ	$\hat{\gamma}^*$	$\hat{\theta}^*$	$\hat{\eta}^* = \hat{\theta}^*/\hat{\sigma}$
balc	0.069	7.541	2.338	1.403	7.707
balc $\times$ *floor*	-0.054	-4.839	-1.500	-0.275	-1.510
group 18	-0.014	-1.352	-0.246	-0.266	-1.463

В табл. 4 представлены значения t-статистик и значения сопутствующих статистик для трех вспомогательных регрессоров в модели без ограничения. Конечно, $\hat{\gamma}^*$ и $\hat{\gamma}$ различаются — результат шкалирования, но соответствующие значения t-статистик совпадают. Значения $\hat{\theta}$ и $\hat{\eta}$ (не приведены) близки к значениям $\hat{\theta}^*$ и $\hat{\eta}^*$ (но не совпадают с ними). Сравнение значений t_γ и $\hat{\eta}^*$ представляет интерес, поскольку η есть ненаблюдаемое «теоретическое» значение t-статистики. Если бы $Z^{*\prime}MZ^*$ было равно единичной матрице, то мы имели бы $t_\gamma = \hat{\eta}^*$. Поэтому неудивительно, что это равенство почти выполняется для третьей вспомогательной переменной *group 18*.

Эффекты предварительного тестирования приведены в табл. 5 для случая отбора модели от общего к частному. Первые две колонки таблицы повторяют оценки b и их стандартные ошибки $\sqrt{\widetilde{\mathrm{var}}(b)}$ основных параметров β из табл. 3. Сами оценки не изменяются от того, принимается ли во внимание эффект предварительного тестирования или нет — изменяются только их свойства. Следовательно, $\widetilde{\mathrm{var}}(b)$ не является правильной оценкой дисперсии, если принимать во внимание предварительное тестирование.

Значения величин q_0^2 важны, поскольку от них зависит величина эффекта. Малое значение q_0^2 соответствует небольшому различию между полученной и истинной оценками дисперсии и среднеквадратичного отклонения MSE, т. е. при этом эффектом предварительного тестирования можно пренебречь. В нашем случае значение q_0^2 весьма мало для всех регрессоров, за исключением *floor*. Таким образом, если эффект предварительного тестирования имеется, то только для этого регрессора.

Таблица 5. Влияние предварительного тестирования на основные переменные

Переменная	b	$\sqrt{\widehat{\mathrm{var}}(b)}$	q_0^2	bias	RMSE	RMSE*
constant	−0.053	0.044	0.027	−0.0037	0.044	0.045
$\log(totsp)$	1.080	0.009	0.029	0.0012	0.009	0.009
$kitsp/totsp$	0.904	0.058	0.037	−0.0075	0.057	0.058
$\log(dist)$	−0.203	0.004	0.023	−0.0002	0.004	0.004
$metrdist \times walk$	−0.004	0.001	0.001	0.0001	0.001	0.001
$metrdist \times (1 - walk)$	−0.007	0.001	0.001	−0.0000	0.001	0.001
walk	0.053	0.011	0.001	−0.0007	0.011	0.011
floor	0.093	0.010	2.770	0.0002	0.010	0.013

Смещение, дисперсия и среднеквадратичное отклонение оценок коэффициентов, полученные в результате процедуры предварительного тестирования, являются функциями вектора параметров η. Этот вектор неизвестен, но известна его оценка $\hat{\eta}^*$. Она несмещенная, но ее дисперсия постоянна, поскольку $\hat{\eta}^* \sim \mathrm{N}(\eta, I_m)$. Следовательно, дисперсия не стремится к нулю при увеличении размера выборки. В колонках 4 и 5 табл. 5 представлены смещение и корень из среднеквадратичной ошибки, рассчитанные при $\eta = \hat{\eta}^*$. Конечно, η не равно $\hat{\eta}^*$. Поэтому в последней колонке таблицы приведено значение RMSE* — 95%-ой верхней границы RMSE, рассчитанной как максимум по трехмерному параллелепипеду вероятностей вокруг $\hat{\eta}^*$.

Из данных табл. 5 следует заключить, что в нашем случае эффект предварительного тестирования весьма мал благодаря большим значениям t-статистик и малым значениям q_0^2. Таким образом, оценки всех восьми основных переменных не смещены, и их стандартные ошибки оценены корректно. Единственным возможным исключением является переменная *floor*, для которой среднеквадратичное отклонение может быть на 37 % выше, чем рассчитанное.

Для метода отбора моделей от частного к общему получаются в основном аналогичные результаты. Среднеквадратичное отклонение в этом случае на 44 % выше, чем рассчитанное в соответствии с результатом, полученным с помощью симуляций в (Danilov, Magnus, 2004a).

6. Выводы

Ограничения гедонистического подхода хорошо известны. В частности, этот подход принимает во внимание только готовность покупателя платить за наблюдаемые различия в свойствах объекта и их непосредственные последствия. Таким образом, если покупатель не связывает свойства объекта и выгоды, связанные с ними, то эти свойства не будут отражены в цене недвижимости. Гедонистический подход также предполагает, что покупатели могут выбирать комбинации свойств недвижимости, которые они предпочитают, при ограничениях, накладываемых их доходами. Однако рынок недвижимости может быть подвержен внешним воздействиям, как налоги, процентные ставки и др. Несмотря на эти возможные проблемы, мы получили результаты, робастные, в том смысле, что различные спецификации приводят к одинаковым выводам.

Главным фактором, определяющим стоимость квартиры в Москве, является ее общая площадь (*totsp*), а не ее расположение. Эластичность цены по общей площади больше 1, так что большие квартиры не только более дорогие, но и относительно более дорогие. Возможным объяснением этого эффекта является то, что в рассматриваемом сегменте рынка недвижимости ($totsp < 150$) большие квартиры часто соответствуют элитным квартирам, и потому они относительно более дорогие.

Существует «оптимальная» планировка квартиры, при которой кухня занимает 24.5 % общей площади квартиры — гораздо больше, чем среднее значение в 2003 г. (15.9 %). Небезынтересно, что новые серии многоквартирных домов предлагают квартиры, почти достигающие этого оптимального значения.

Поскольку Москва имеет радиально-кольцевую структуру, то можно предположить, что расстояние до центра является важным фактором в определении цены квартиры. Мы получили, что увеличение расстояния до центра на 1 % снижает стоимость квартиры на 0.20 %. Стоимость однокомнатных квартир убывает с расстоянием в меньшей степени, чем стоимость других квартир, что объясняется повышенным спросом на однокомнатные квартиры.

Метро в Москве является эффективным и надежным общественным транспортом, оно также имеет радиальную структуру, это основной вид общественного транспорта по объему перевозок. Поэтому расстояние до метро также важно для стоимости квартиры. По сравнению с квартирой, находящейся рядом со станцией метро, квартира, расположенная в 10 минутах ходьбы пешком, дешевле на 3.9 %, а аналогичная квартира на расстоянии 10 минут на общественно транспорте, дешевле на 11.6 %. Расстояние до метро менее существенно для владельцев трехкомнатных квартир, возможно, они чаще пользуются личными автомобилями.

Москвичам не нравятся квартиры на первом и последнем этажах, такие квартиры, в среднем, на 9.3 % дешевле. Этот эффект менее значим для однокомнатных квартир, чем для многокомнатных квартир, возможно, потому, что обычно однокомнатные квартиры покупают молодые семьи с относительно невысоким уровнем доходов.

Трудно ожидать, что модель, принимающая во внимание только несколько основных свойств квартиры, будет давать точные оценки цены. Стоимость недвижимости зависит от множества других факторов, по которым у нас нет данных: местная экология, хороший вид из окна, охраняемая территория, качество строительства и отделки и многое другое.

Тем не менее относительная ошибка прогноза цены по нашей модели в 85.6 % случаев не превосходит 25 % и может служить полезным инструментом оценки цены для покупателей, продавцов и агентов недвижимости.

Неожиданно эффект процедуры предварительного тестирования оказался весьма мал в нашем случае. Это является следствием большого объема выборки, а также особых свойств набора данных (малые значения q_0^2). В этом случае процедура отбора модели оказывает лишь незначительное влияние на свойства оценок и прогнозов. Таким образом, обычные оценки и прогнозы могут рассматриваться как несмещенные, с корректно оцененными стандартными ошибками.

Литература

Bailey, M.J., R.F. Muth, and H.O. Nourse (1963). A regression method for real estate price index construction, *Journal of the American Statistical Association*, 58, 933–942.

Bender, A.R., B. Gacem, and M. Hoesli (1994). Construction d'indices immoboliers selon l'approche hédoniste, *Finanzmarkt und Portfolio Management*, 8, 522–534.

Chow, G.C. (1967). Technological change and the demand for computers, *The American Economic Review*, 57, 1117–1130.

Danilov, D. (2005). Estimation of the mean of a univariate normal distribution when the variance is not known, *The Econometrics Journal*, 8, 277–291.

Danilov, D. and J.R. Magnus (2004a). On the harm that ignoring pretesting can cause, *Journal of Econometrics*, 122, 27–46.

Danilov, D. and J.R. Magnus (2004b). Forecast accuracy after pretesting with an application to the stock market, *Journal of Forecasting*, 23, 251–274.

Griliches, Z. (1961). Hedonic price indices for automobiles: An econometric analysis of quality change, in: *The Price Statistics of the Federal Government*, National Bureau for Economic Research, General Series No. 73, New York, 137–196.

Haas, G.C. (1922). Sales prices as a basis for farm land appraisal, Technical Bulletin No. 9, The University of Minnesota Agricultural Experiment Station, St. Paul.

Lancaster, K.J. (1966). A new approach to consumer theory, *Journal of Political Economy*, 74, 132–157.

Lansink, A.O. and G. Thijssen (1998). Testing among functional forms: An extension of the generalized Box-Cox formulation, *Applied Economics*, 30, 1001–1010.

Magnus, J.R. and J. Durbin (1999). Estimation of regression coefficients of interest when other regression coefficients are of no interest, *Econometrica*, 67, 639–643.

Malginov, G. and G. Sternik (2006). The housing market of Moscow region, Section 4.9 in: *Russian Economy in 2005: Trends and Outlooks*, Institute for the Economy in Transition, Issue 27, Moscow, 429–448.

Maurer, R., M. Pitzer, and S. Sebastian (2004). Hedonic price indices for the Paris house market, *Allgemeines Statistisches Archiv*, 88, 303–326.

Mills, E.S. and R. Simenauer (1996). New hedonic estimates of regional constant quality housing prices, *Journal of Urban Economics*, 39, 209–215.

Milton, J.W., J. Gressel, and D. Mulkey (1984). Hedonic amenity and functional form specification, *Land Economics*, 60, 378–388.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of Political Economy*, 82, 34–55.

van Soest, A. and M. Verbeek (2010). *Empirical Applications I*, Econometric Exercises, Volume 4, Cambridge University Press, New York.

Witte, A.D., H.J. Sumka, and H. Erekson (1979). An estimate of a structural hedonic price model of the housing market: An application of Rosen's theory of implicit markets, *Econometrica*, 47, 1151–1174.