

Метод остаточных ограничений (RCA): концепция и протокол

Русский перевод, подготовленный с использованием искусственного интеллекта и прошедший редакционную проверку; английская версия является основной.

William Bert Craytor

Том 1, Выпуск 1 · Лето 2026

Опубликовано: 1 июля 2026

DOI: [10.67330/kgebx46](https://doi.org/10.67330/kgebx46)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Эта статья открытого доступа распространяется на условиях [лицензии Creative Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

Аннотация. Существенное ограничение традиционных методов оценки — опора на устаревшие ручные приёмы, такие как анализ парных продаж, которые обходят стороной вклады характеристик объекта в стоимость: в традиционной таблице продаж нет столбца для вкладов в стоимость, есть лишь корректировки. Между тем именно вклады в стоимость дают необходимую основу для математических ограничений, не позволяющих оценщику завысить или занижить стоимость. Метод остаточных ограничений (RCA) развивает традиционный сравнительный подход, объединяя многоэтапный процесс оценки на основе многомерных адаптивных регрессионных сплайнов (MARS) со строгими математическими ограничениями на вклады характеристик — прежде всего неизмеряемых характеристик, то есть латентных переменных вроде эстетики, планировки, состояния и качества, стоимость которых традиционно определяется субъективным суждением оценщика.

Ключевые слова: оценка недвижимости, оценка жилья, метод остаточных ограничений, MARS, сравнительный подход, анализ остатков, латентные переменные, автоматизированная оценка

Содержание

1	Введение	3
2	Сведения об авторе	6
3	Инженер по оценке	9
4	Допущения	10
5	Рабочий процесс RCA	13
5.1	Виды анализа MARS	13
5.2	Рабочий процесс обработки	13
5.3	Рабочий процесс обработки данных	15
5.3.1	Требования к данным для MARS	15
5.3.2	Протоколы: Список закрытых сделок	17

5.4	Структура папок проекта	19
5.4.1	Протоколы: Несколько прогонов MARS (для рынка недвижимости)	20
5.5	Основная книга Excel «	22
5.6	Резервное копирование	22
6	Обозначения книг	23
6.1	Описательные измерения	23
6.1.1	Размер проекта (“a”)	24
6.1.2	Параметр версии (“r”)	25
6.1.3	Этапы рабочего процесса (“w”)	25
6.1.4	Листы книги рабочей Excel (“s”)	26
6.1.5	Список объектов недвижимости (“p”) — строки MLSData	26
6.1.6	Переменные MLS (“v”) Столбцы MLSData	27
6.1.7	Пример окончательной нотации	27
6.1.8	Обоснование	27
6.1.9	Матрицы?	28
6.1.10	Первое упрощение	28
6.1.11	Второе упрощение	28
7	Таблица продаж RCA	31
8	Переменные	35
8.1	Онтология «Variable Origin»	35
8.2	Онтология переменных приложений	36
8.3	Члены второй и третьей степени	38
8.3.1	Переменные корректировки	38
8.3.2	Агрегирующие переменные	38
8.3.3	Специальные переменные& Обозначения	39
8.3.4	Разбивка остатков	39
9	Основные доказательства метода RCA	41
10	Модель MARS «	43
10.1	Функции «g» для моделирования переменных и взаимодействий	43
10.2	Индексирование функций g	43
10.3	Размеры графика	44
10.4	Окончательное уравнение модели	45
10.5	Количество диаграмм, необходимых для построения графика модели	45

11	Характеристики данных	48
11.1	Цель программы «Residual»	48
11.2	Оценки состояния и качества в качестве приблизительных регрессоров	49
11.3	Ошибки в данных	49
11.4	Что такое точность?	51
11.5	Рассматриваемый объект оценки как аномалия	53
11.6	Важность CVR2 при работе с непредвиденными данными	53
12	Остатки	55
12.1	Рейтинг & Балл	55
12.2	Кривая CQA-остаточная величина	58
12.3	Характеристики кривой CQA	58
12.3.1	Анализ компонентов кривой CQA	59
12.3.2	Динамика рынка и социальные последствия	59
12.4	Объект оценки «Остаточные величины»	61
12.5	Выводы по оценке RCA	61
12.5.1	Общие соображения	61
12.5.2	Обзор, аудит, жалобы	62
13	Заключение	63

1 Введение

В данной статье представлены логические и математические основы современного подхода к точной оценке рыночной стоимости жилых и иных объектов недвижимости. Этот метод, разработанный на основе 20-летнего опыта применения программного обеспечения «Многомерные адаптивные регрессионные сплайны» (MARS) к жилым домам Северной Калифорнии — особенно в районах залива Сан-Франциско — обеспечивает значительное повышение точности оценки. MARS повышает качество оценки за счет умелого учета нелинейных взаимосвязей между характеристиками объекта недвижимости и его стоимостью, одновременно эффективно выявляя и приоритезируя наиболее значимые переменные. Эта возможность позволяет оценщикам выявлять динамику стоимости в новых рыночных зонах гораздо быстрее, чем это позволяют традиционные методы. MARS составляет основу предлагаемого Метода остаточных ограничений (RCA). Метод RCA получил своё название благодаря строгой методологической основе:

- Он налагает строгие ограничения на вклад в стоимость как измеримых, так и неизмеримых характеристик объекта недвижимости:
 1. Сумма всех вкладов в стоимость, полученных с помощью MARS, включая остаток, должна равняться чистой цене продажи каждого аналога.
 2. Остаток можно разложить на подкомпоненты с заданными значениями при условии, что их сумма равна остатку аналога.

- Метод RCA минимизирует систематическую ошибку при оценке субъективных характеристик — таких как эстетическая привлекательность, состояние и качество — путем привязки их совокупной стоимости к остатку, который точно количественно определяется при построении модели MARS компетентным инженером-оценщиком.

Хотя метод RCA в первую очередь разработан для жилой недвижимости, где субъективные характеристики существенно влияют на рыночную стоимость, он обладает потенциалом для более широкого применения в области оценки активов. Точные оценки имеют решающее значение, учитывая, что рынок жилой недвижимости США оценивается в 45 триллионов. Fannie Mae отмечает, что типичные оценки отклоняются примерно на 5% от фактических цен продажи, хотя эта цифра скрывает систематическую ошибку, поскольку оценщики часто сталкиваются с давлением, вынуждающим их привести оценку в соответствие с ценами продажи. Кроме того, поскольку 60% — 70% оценок проводятся для целей рефинансирования — при отсутствии аналогов по ценам продажи — необходимость в точности, которую обеспечивает RCA, становится очевидной. Согласно анализу Федеральной резервной системы 2012 года, 15% — 20% оценок, проведенных до кризиса 2008 года, отклонялись от цен продажи более чем на $\pm 10\%$; этот диапазон, вероятно, по-прежнему актуален для оценок при рефинансировании сегодня, причем он варьируется в зависимости от типа недвижимости и рынка. Почему важна точность? В случае сделок с низким отношением кредита к стоимости недвижимости (LTV) — например, когда покупатель с хорошей кредитной историей вносит первоначальный взнос в размере \$ 500 000 на дом стоимостью \$ 1 миллион — кредиторы могут отказаться от оценки, исходя из минимального риска обращения взыскания. Однако покупатели, которые обходятся без оценки, рискуют переплатить, обнаружив дефекты только после покупки. Это подчеркивает ценность надежных методов оценки, таких как RCA. Количественная оценка точности оценки по-прежнему представляет собой сложную задачу. Исследователи не располагают прямыми показателями отклонений от «рыночной стоимости» и вынуждены полагаться на косвенные показатели. Традиционные методы оценки, основанные на субъективном суждении и различном уровне компетентности оценщиков, затрудняют анализ. Стратегии включают:

1. Tzioumis (2016) В ходе исследования, охватившего 4 782 оценщиков и 259 436 оценок в 11 штатах, было установлено, что одна треть оценок находилась в пределах 1% от договорной цены, половина — в диапазоне от 1% до 3% выше, а небольшая часть в среднем на 5% выше — что свидетельствует о тенденции к завышению. В ходе исследования были выделены два типа оценщиков: «по рыночной стоимости» (близко к договорной цене) и «выше рыночной стоимости» (постоянно выше).
2. Krivorotov и LaCour-Little (2020) Было установлено, что 90% оценок при покупке соответствовали договорным ценам или превышали их, что подчеркивает «анкерный уклон» — тенденцию к привязке к желаемой стоимости — особенно в контексте рефинансирования.

Эти исследования выявляют предвзятость в практике оценки, усугубляемую ограничениями традиционных инструментов. Если бы оценщики были одинаково компетентны и беспристрастны, их оценки совпадали бы; постоянная завышенная оценка со стороны некоторых групп свидетельствует о системных проблемах.

Приведенные выше исследования, посвященные систематическим отклонениям, а также системная закономерность, на которую они указывают, отражают более глубокую особенность традиционной практики сравнения продаж, для устранения которой и была разработана методика RCA. Традиционная методология опирается на два канала с совершенно разным эпистемическим статусом. В канале измеримых характеристик — общая арендуемая площадь (GLA), количество комнат, площадь участка, возраст, местоположение — от оценщика

ожидается, что он будет выводить значения корректировок объективно, посредством анализа аналог пар или локальной линейной регрессии на основе аналогов продаж. Компетентные оценщики выполняют эту работу, и полученные в результате корректировки, по крайней мере, в общих чертах обоснованы с точки зрения данных. Однако даже в этом случае традиционная таблица фиксирует только корректировки; в ней даже нет столбца для вклада в стоимость. Традиционные корректировки также не рассчитываются как разности между вкладами в стоимость — они оцениваются непосредственно на основе сопоставленных пар или аналогичных процедур, без какого-либо расчета базовых вкладов. В RCA, напротив, каждая корректировка представляет собой разницу между вкладами в стоимость объекта оценки и аналога по данной характеристике — той структуре, последствия которой уточняет Доказательство II. Что касается канала субъективных характеристик — состояние, качество строительства, дизайн, функциональная полезность, эстетика — доступные вспомогательные процедуры варьируются от скудных до совершенно недостаточных. Корректировки по субъективным характеристикам могут подкрепляться анализом сопоставленных пар объектов недвижимости, выходящих за пределы непосредственного набора аналогов, ссылкой на признанную практику коллег или ссылкой на документально подтвержденный опыт оценщика на рынке. Все три метода формально соответствуют профессиональным стандартам. Ни один из них не достигает той статистической строгости, которую измеримый канал обеспечивает в ходе процедуры.

В результате возникает процедурная асимметрия между тем, что требуют стандарты, и тем, что может обеспечить данная методология. Стандарты требуют обоснования каждой корректировки, будь то измеримой или субъективной. Методология обеспечивает надежное обоснование в одном случае и целый спектр менее убедительных процедур в другом, начиная от единичного примера сопоставленной пары и заканчивая ссылкой на то, что обычно используют другие оценщики. Честный и компетентный оценщик, тщательно проводящий работу с сопоставленными парами при корректировке с видом, получает значительно более весомое обоснование, чем тот, кто без обоснования ссылается на практику коллег, однако эксперт, сравнивающий две оценки по стандартной форме, не может легко различить их, поскольку в форме фиксируется значение корректировки, а не строгость её вывода. Данная методология допускает такую асимметрию, поскольку более строгих процедур для оценки субъективных характеристик у практикующих оценщиков в целом нет.

Метод RCA улучшает оба канала, но совершенно разными способами. В измеримом канале регрессия MARS учитывает нелинейные зависимости, взаимодействия переменных и структуру базисных функций, которые анализ сопоставленных пар и обычная линейная регрессия не могут отразить. Благодаря построению модели на основе полного набора квалифицированных продаж в рыночной зоне — как правило, нескольких сотен — а не на трех – девяти аналогах традиционной таблицы, вклады в стоимость, полученные с помощью MARS, являются объективными и существенно более точными, чем методология, которую они заменяют. В субъективном канале RCA действует совершенно иначе, чем любые традиционные вспомогательные процедуры. Она вообще не пытается оценивать коэффициенты корректировки для отдельных субъективных характеристик. Вклады всех субъективных характеристик — латентных переменных — в совокупности проявляются в виде остатка: разницы между фактической чистой ценой продажи каждого аналога и его ценой, прогнозируемой по методу MARS, которая рассчитывается на основе данных без участия оценщика. Разложение этого остатка на названные компоненты — дизайн, функциональная полезность, вид, состояние, качество и т. д. — ограничено тем, что их сумма должна равняться общему остатку, и согласно доказательству II (раздел 9) любое разложение, удовлетворяющее этому ограничению, приводит к одному и тому же выводу о стоимости. Единственное суждение, которое остается в рабочем процессе, — это определение места объекта оценки в ранжированном распределении остаточных значений аналогов,

осуществляемое на основе поддающегося проверке ранжирования и явно задокументированное в отчёте. Все еще возможна слабая оценка по методу RCA — неопытный оценщик может создать переобученную модель MARS, разместить объект оценки неправдоподобно или пропустить аномалии в данных — но эта слабость проявляется в аспектах, которые рецензент может выявить путем проверки модели и ранжирования, а не в аспектах, которые невозможно выявить при стандартной рецензии.

Помимо рефинансирования, оценщики обязаны предоставлять владельцам точные оценки, особенно в случае дорогостоящих покупок, где ошибки могут обойтись в десятки или сотни тысяч. Тем не менее многие не хотят тратить \$ 600–\$ 1,000 на оценку, сомневаясь в её полезности. Передовые методы, такие как RCA, хотя и требуют значительных ресурсов на начальном этапе, обещают более высокую точность. С совершенствованием методов и накоплением опыта их эффективность — и распространение — должны возрасти.

2 Сведения об авторе

Автор имеет образование в области математики, информатики и статистики. По профессии он является инженером-программистом, имеющим дополнительный опыт в области оценки и бухгалтерского учета. Он занимается оценкой с перерывами с 2002 года и с 2004 года применяет систему MARS для определения корректировок стоимости и ценовых тенденций на рынке жилой недвижимости.

Оценка с использованием MARS — это «глубокий опыт» в области оценки. MARS генерирует модели, которые более детально, чем любой другой статистический метод, отражают как переменные, влияющие на рыночную стоимость, так и то, как изменения этих значений сказываются на цене продажи. Таким образом, при оценке с использованием MARS оценщик постоянно исследует причинно-следственные связи между значениями переменных и ценами продажи, генерируемыми моделями MARS.

В целях обеспечения прозрачности методология RCA наиболее эффективно применяется к жилым районам, обладающим следующими характеристиками:

1. Сложные районы, в которых объекты недвижимости значительно отличаются друг от друга
 - Старые дома
 - Различный уровень модернизации
 - Многообразие архитектурных стилей и дизайнов
 - Различия в характеристиках объектов недвижимости, таких как общая арендуемая площадь (GLA), площадь участка, высота над уровнем моря и вид из окна.
2. Сложные рыночные условия
 - Быстро меняющиеся рыночные условия
 - Множество различных типов покупателей
3. Надежные источники данных: протокол RCA основан на методах интеллектуального анализа данных и требует большого объема качественных данных. Хотя его можно использовать в разреженных регионах, затраты времени и средств могут оказаться неоправданными.

4. Желательно иметь точные оценки рыночной стоимости. В большинстве случаев RCA должен обеспечивать точность на уровне $\pm 1-2\%$. В отчёте инженера по оценке должны быть указаны условия, необходимые для обеспечения точности итогового заключения о стоимости. Если данных или информации не хватает, либо если в ходе интеллектуального анализа данных не удалось выявить устойчивых закономерностей на рынке, точность оценки снизится, и об этом следует проинформировать клиента. В частности, высокие значения R2 и CVR2 (перекрёстная проверка R2, рассчитанные путем многократного разделения набора данных MLS на случайные обучающие и тестовые подмножества), значимая модель MARS и выраженная закономерность в характеристиках CQA (состояние-качество-привлекательность) объектов недвижимости в ранжированных остатках являются основными средствами оценки точности и надежности вывода о стоимости.

Разработка этих протоколов RCA на протяжении последних более чем 20 лет основана на доступности данных MLSListings Inc. (Саннивейл, Калифорния) и других систем MLS, которые в настоящее время доступны через MLSListings и составляют более 95% опыта автора в области анализа данных. Опыт автора основан на обширной работе по оценке недвижимости в 15 основных округах Северной Калифорнии, а также на дополнительном опыте работы в 10 второстепенных округах.

Методология RCA наиболее эффективна на рынках, отвечающих двум ключевым критериям:

1. Широкая доступность высококачественных данных для оценки
2. Сложные, неоднородные жилые районы, характеризующиеся:
 - Разнообразием архитектурных стилей
 - Различной степенью обновления недвижимости
 - Многолетними тенденциями развития

Опыт округа Калифорнии

Утверждения, приведенные в данной статье, основаны на опыте оценки объектов недвижимости в следующих округах Северной Калифорнии:

Сан-Матео	Санта-Клара	Марин	Напа	Монтерей	Санта-Роза
Аламеда	Сан-Франциско	Контра-Коста	Сакраменто	Эль-Дорадо	Сан-Хоакин
Станислаус	Плейсер	Солано	Мендосино	Лейк	Колуса
Мерсед	Невада	Сан-Бенито	Санта-Круз	Саттер	Юба

Указанные выше районы, по сути, охватывают Кремниевую долину и прилегающие к ней округа, простираясь далее в более сельские районы, включая Центральную долину и предгорья Сьерры. Следует отметить, что автор всегда имел доступ к обширным и относительно точным данным по сравнению с тем, что имеется в распоряжении многих оценщиков в других, менее развитых регионах США.

Используемые данные MLS имеют сертификат RESO и соответствуют стандарту RCF (RESO Common Format). Несмотря на сертификацию RCF, данные, имеющие, как правило, второстепенное значение, по некоторым объектам недвижимости иногда отсутствуют или содержат неточности, что может быть вызвано ошибками при вводе данных в систему MLS либо первоначальным налоговым оценщиком, либо последующим агентом по продажам.

Часто эти ошибки следуют определённой схеме, например, просто не производится ввод данных, если данного параметра не существует. Например, вместо ввода 0 для «мест под навесом», когда навеса нет, данные не вводятся вовсе. В районах, где навесы для автомобилей встречаются редко, аналитик может заменить отсутствующие данные значением 0, чтобы получить полезную корректировку на наличие навесов. Хотя R/earth и другие реализации MARS могут обрабатывать отсутствующие данные, решение о том, как с ними поступать, остается за инженером по оценке (VE).

Данные MLSListings также доступны из систем MLS некоторых других штатов — от Флориды до Гавайев. Это может быть полезно для проверки эффективности работы протоколов и кода в разных штатах.

3 Инженер по оценке

Definition 1: Valuation Engineer

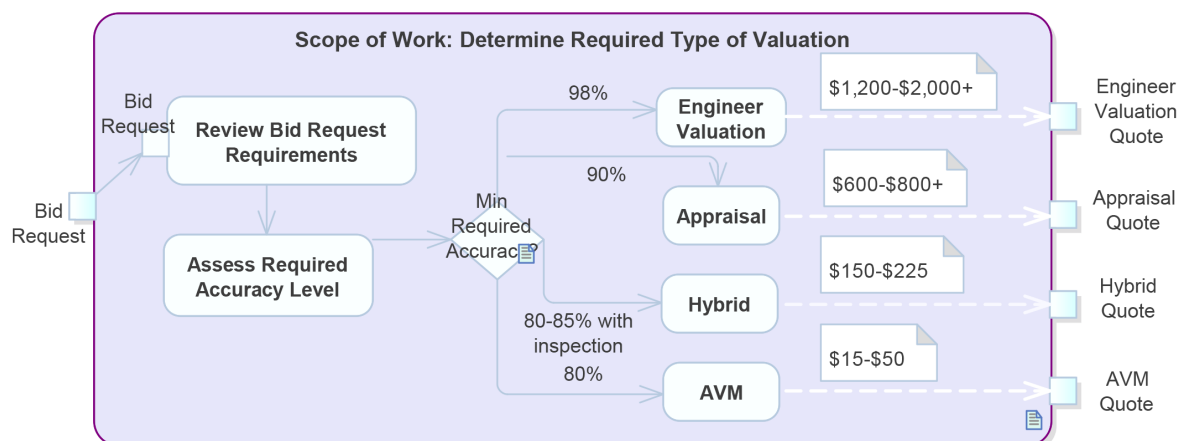
Во избежание двусмысленности, термин «инженер по оценке» (или сокращённо «ВЕ») используется для обозначения специалиста, обладающего квалификацией лицензированного оценщика, а также передовыми знаниями, навыками и опытом в области использования методов интеллектуального анализа данных, статистики, программирования, детальных протоколов и искусственного интеллекта для определения рыночной стоимости сложных объектов недвижимости. Также важно понимать, что оценка — это вид деятельности, требующий наличия той или иной формы лицензии в связи с существующим риском мошенничества.

Кредиторы, как правило, предлагают несколько вариантов оценки:

1. Автоматизированные модели оценки (AVM) — самый экономичный вариант
2. Гибридные оценки — с привлечением нелицензированных инспекторов под контролем лицензированных оценщиков
3. Стандартные оценки — выполняются лицензированными оценщиками

При проведении оценочных работ обычно не учитываются оценки, выполняемые специалистами инженерного уровня. Просто на рынке не так много оценщиков, обладающих необходимыми навыками для их проведения. В них, безусловно, должна быть потребность. При использовании методологии RCA (MARS и ограничительные условия) они могут обеспечить значительно более высокую точность и практически гарантировать достаточно высокую степень объективности.

Для объектов недвижимости, соответствующих нормам, расположенных в новых жилых районах, как правило, достаточно стандартной оценки. Однако для старых домов в районах, где в течение более 30 лет проводились значительные ремонтные работы, рекомендуется проводить инженерную оценку. Выбор между этими методами должен осуществляться в рамках структурированного процесса принятия решений с учетом характеристик объекта недвижимости и рыночных условий.



4 Допущения

Assumption 1: Geographical Area & Property Type

В данной работе основное внимание уделяется односемейным домам и квартирам в кондоминиумах, расположенным в мегаполисах и городских районах Калифорнии. Хотя сравнительный подход применим к другим типам недвижимости и географическим регионам, доходный и затратный подходы, как правило, играют более важную роль при оценке сельскохозяйственных земель, многоквартирных домов и коммерческой недвижимости. В таких случаях метод сравнительных продаж может потребовать модификации или иметь ограниченную применимость.

Assumption 2: Data Storage

Форматы CSV и Excel широко используются в системах MLS и считаются основным способом передачи данных в программы RCA. Хотя для непрерывной передачи данных в базы данных SQL можно использовать брокерские каналы. Для резервного копирования и промежуточного хранения часто используются базы данных SQL, и две из них, играющие важную роль, — это бесплатные системы с открытым исходным кодом SQLite и PostgreSQL. Текстовые или RTF-файлы также используются для фрагментов отчётов, а файлы PNG, JPG и TIFF — для графики. Обратите внимание, что многие также используют SQL Server от Microsoft, однако бесплатная версия «SQL Server Express» имеет значительные ограничения (например, можно использовать только 1 ядро процессора), а другие версии требуют уплаты значительных лицензионных сборов при коммерческом использовании. Базы данных по недвижимости в Калифорнии могут быть очень большими, а операции соединения (join) — весьма трудоемкими, поэтому для выполнения крупных соединений может пригодиться компьютер с 8 и более ядрами и 128 ГБ оперативной памяти. Базы данных SQL также важны, когда необходимо найти все объекты недвижимости в произвольно ограниченной области. Таким образом, если, например, вы использовали кластерный анализ для определения районов, вам, скорее всего, понадобится какая-либо база данных SQL. Однако разные базы данных имеют разные встроенные функции для работы с географической информацией и, следовательно, несовместимы друг с другом. В этой серии статей, посвящённых RCA, для сложных задач SQL будет использоваться PostgreSQL, а для простых — SQLite.

Assumption 3: “R/earth” Is Used For MARS processing.

До 2017 года программное обеспечение MARS от Salford Systems было доступным решением по цене \$ 270 в год. Однако после приобретения Salford Systems компанией Minitab MARS был включён в их набор SPM по цене \$ 16 000 в год — цена, недоступная для большинства малых предприятий. К счастью, пакет «earth» для языка программирования R ((Milborrow 2024a)), поддерживаемый Стивеном Милбороу, развился настолько, что теперь соответствует или даже превосходит возможности SPM MARS от Minitab. Пакет earth предлагает дополнительные преимущества: он доступен бесплатно, органично интегрируется с обширной статистической экосистемой R и доступен через Python с помощью библиотеки Pandas, что обеспечивает отличную гибкость для различных рабочих процессов. В обсуждениях, касающихся MARS, предполагается конкретное использование R/earth.

Assumption 4: Notation

Исходные данные из внешних источников, таких как файлы CSV и таблицы Excel, представляются с использованием стандартных латинских символов. Сюда входят данные о недвижимости из службы Multiple Listing Service (MLS), настройки MARS (Multivariate Adaptive Regression Splines) с использованием пакета R/earth, а также данные, специфичные для конкретных проектов. Эта нотация распространяется на соответствующие представления в виде фреймов данных в R или Python.

При обсуждении характеристик объектов недвижимости, а также их различных типов и подмножеств, для повышения точности и аналитической ясности используется нотация на основе греческих букв. Хотя такая математическая нотация может показаться более строгой, она выполняет важную функцию: позволяет отличать абстрактные характеристики объектов от их представлений в виде необработанных данных. По соглашению греческие буквы последовательно обозначают характеристики объектов недвижимости, а латинские — структуры данных в электронных таблицах. Например, символ M представляет матрицу данных MLS, в которой объекты недвижимости расположены в строках, а соответствующие им характеристики — в столбцах.

Assumption 5: Programming Languages

В этой области доминирующими языками являются R и Python с библиотекой Pandas, а языки C или C++ используются в тех случаях, когда важна скорость обработки. Сам R/earth написан на R и C.

Assumption 6: Indexing

Индексация может различаться в разных языках программирования и математике, что может сбивать с толку:

- В математике и R индексация начинается с 1.
- В C, C#, C++ и Python — с 0.

В данной статье, в целях единообразия и простоты, индексация везде начинается с 0 (как в Python или C). Вот как это работает:

- Для списков, матриц или дата-фреймов:
 - Индекс 0 зарезервирован для «объекта оценки» (основного свойства, которое изучается).
 - Пример: в таблице, где строки представляют объекты недвижимости, строка 0 — это объект оценки, а столбец 0 содержит уникальные идентификаторы всех объектов недвижимости.
 - Практический совет: располагайте объекты недвижимости так, чтобы сначала шло объект оценки (строка 0), а затем перечисляйте остальные от самой новой к самой старой дате продажи, пронумеровав их последовательно (0, 1, 2 и т. д.).
- Если «объекта оценки» нет:
 - Если структура данных не содержит объектов недвижимости (например, только переменные или функции), индекс 0 пропускается, и индексирование начинается с 1.
 - Особое примечание для пользователей R: поскольку в R используется индексация, начинающаяся с 1, при написании кода для математических формул из данной статьи в R необходимо прибавить 1 ко всем индексам, чтобы перейти от индексации, начинающейся с 0, к индексации, начинающейся с 1.

5 Рабочий процесс RCA

Понимание рабочего процесса RCA имеет решающее значение для эффективного анализа. Этот процесс состоит из структурированной последовательности аналитических задач и этапов принятия решений

5.1 Виды анализа MARS

Рабочий процесс анализа MARS зависит от того, требуется ли только рыночный анализ или же необходима оценка стоимости конкретного объекта недвижимости. Следует отметить, что для проведения оценки конкретного объекта недвижимости, как правило, требуется анализ его рыночной зоны.

- Рыночный анализ: фокусируется на обширных территориях, таких как округа, города или определенные регионы. Данный подход не затрагивает конкретный объект оценки и, следовательно, не требует вывода о его стоимости. Рабочий процесс сосредоточен на разработке модели, её валидации и формировании вспомогательной аналитики. Процесс завершается после определения и документирования оптимальной модели.
- Анализ отдельного объекта недвижимости включает в себя все элементы рыночного анализа, а также определение стоимости конкретного объекта. Это принципиально отличается от традиционных методов оценки. В традиционных подходах стоимость обычно определяется путем корректировок 3–9 аналогов сделок купли-продажи и расчета средневзвешенного значения. В отличие от этого, RCA определяет стоимость в основном на основе двух компонентов:

1. Прогнозируемая стоимость по модели MARS
2. Расчетная остаточная стоимость конкретного объекта недвижимости (с соответствующим баллом CQA)

Таблица продаж в RCA служит в первую очередь для объяснения различий в стоимости между объектом оценки и аналогами путем иллюстрации вклада отдельных переменных. Ключевым преимуществом RCA является ее стабильность: хотя клиенты могут спорить по поводу выбора аналогов, вывод о стоимости остается неизменным независимо от того, какие аналоги выбраны для представления, поскольку все сделки (будь то 15 или 600) подвергаются корректировкам с использованием одной и той же модели. Хотя более старые сделки автоматически подвергаются корректировкам с учетом текущих рыночных условий, в целом предпочтительны недавние сделки. Эффективность модели в учете ценовых факторов снижается с увеличением временного интервала, поскольку не измеренные рыночные характеристики приобретают все большее значение при более длительных периодах.

5.2 Рабочий процесс обработки

Общий рабочий процесс обработки можно описать следующими этапами:

1. Объем работ.
2. Настройка исходных папок и файлов проекта.
3. Загрузка данных MLS. Информацию о количестве необходимых записей об объектах недвижимости и установке ограничений по датам см. в протоколах 1.1–1.3.

4. Заполните листы настроек: Регрессионные переменные, Взаимодействия, Агрегации, Расчёты.
5. Запустите регрессию MARS (R/earth) для построения ценовой модели.
6. Сгенерируйте вклады в стоимость объекта недвижимости и остатки (чистый остаток и остаток/SF) на основе модели MARS.
7. Проведите ранжирование объектов недвижимости по показателю «остатки/SF».
8. Создайте балы CQA для каждого аналога сделки купли-продажи.
9. Предполагая, что был составлен рейтинг качества объектов недвижимости, который достоверно отражает их общую привлекательность, определите наилучшую позицию рассматриваемого объекта в рейтинге и присвойте ему балл CQA, изучив общие тенденции в рейтинге.
10. Присвойте рассматриваемому объекту соответствующий остаток/SF в расчёте на общую арендуемую площадь (GLA)· (объект оценки).
11. Сложите оценочную цену MARS и присвоенный остаток, чтобы получить итоговую оценку стоимости объекта .
12. Разбейте остаток объекта на описательные компоненты и вклады в стоимость.
13. Рассчитайте вклады в стоимость, полученные на основе модели аналогов & , корректировки, а также все промежуточные итоги и скорректированные цены продажи.
14. Скорректированные цены продажи будут равны окончательному выводу о стоимости объекта оценки.
15. Выберите 6–12 наиболее схожих объектов недвижимости для таблицы продаж.
16. Разбейте общие остаточные суммы по аналогам на остатки стоимости по компонентам, установив общую остаточную стоимость равной 0 после завершения разбивки.
17. Пересчитайте все промежуточные итоги и скорректированные цены продажи аналогов в таблице продаж. Проверьте значения.
18. Загрузите таблицу продаж в отчёт.

Обратите внимание, что при использовании метода RCA окончательный вывод о стоимости НЕ вытекает напрямую из аналогичных значений, приведенных в таблице продаж. При использовании метода RCA таблица продаж становится лишь инструментом для объяснения того, почему объекты недвижимости, внешне похожие на объект оценки, продаются дороже или дешевле, чем окончательное заключение о стоимости или оценочная стоимость объекта оценки. На это следует смотреть иначе: заключение о стоимости объекта оценки зависит от оценки цены по модели MARS и от остаточной величины, рассчитанной инженером-оценщиком. Однако эти две величины в значительной степени зависят от всех объектов недвижимости, включенных в регрессионный анализ R/earth, который сформировал ранжирование остатков объектов, определившее балл CQA для объекта оценки и, таким образом, определившее его остаточную стоимость посредством присвоения ему места в рейтинге инженером по оценке. Корректировки в таблице продаж на самом деле определяются наложенными математическими ограничениями RCA, хотя разбивка остаточной стоимости в некоторой степени зависит от оценщика, суммы вкладов по разбивке всегда должны в сумме равняться соответствующему остатку. Из доказательства 2 известно, что, хотя фактические значения компонентов остаточной стоимости могут варьироваться в зависимости от

суждения инженера по оценке, их суммы для данного объекта полностью ограничены и, следовательно, не изменят скорректированную цену продажи соответствующей недвижимости, независимо от того, как инженер по оценке изменяет разбивку остаточной стоимости.

5.3 Рабочий процесс обработки данных

Обратите внимание, что в данной статье представлен лишь общий рабочий процесс метода RCA. В других статьях будут подробно рассмотрены различные этапы и подэтапы анализа RCA.

5.3.1 Требования к данным для MARS

Нам необходимо указать дату начала продажных сделок и дату окончания, исходя из даты оценки или оценки, проведенной оценщиками. При этом обычно учитывается предполагаемое количество переменных регрессии и взаимодействий, которые потребуются для каждой степени взаимодействия, превышающей 1.

5.3.2 Протоколы: Список закрытых сделок

Protocol 1.1: Determining the Number of MLS Closed Sales Listings Required

В данном протоколе описывается, как рассчитать рекомендуемое минимальное количество завершённых сделок, которые необходимо провести ввод в регрессионный анализ MARS. Для начала вам потребуется провести приблизительную оценку, запустить MARS для анализа первой, второй и, возможно, третьей степени, чтобы сформировать исходную модель, а затем проверить количество строк в соответствии со следующими правилами, чтобы определить, требуется ли добавить дополнительные строки. В данном протоколе предполагается, что независимые переменные являются независимыми на 100% , что в сфере недвижимости встречается редко; например, общая арендуемая площадь (GLA) и количество комнат обычно сильно коррелируют, а GLA, вероятно, также коррелирует с площадью участка. Это связано с расчётом «степеней свободы (DOF)». Каждая действительно независимая переменная получает одну степень свободы. Однако при наличии зависимостей между так называемыми независимыми переменными точный расчёт DOF требует анализа, выходящего за рамки данной статьи. Просто имейте в виду, что из-за вероятной коллинеарности между переменными приведённые ниже оценки, скорее всего, завышены.

1. Для членов первого порядка (без взаимодействий) хорошо работает правило « ~ 15 » строк на переменную
2. Для каждого допустимого двустороннего взаимодействия следует добавить примерно 15–30 объектов недвижимости
3. Для взаимодействий более высокого порядка (3 и более переменных) каждое должно иметь 30 и более объектов недвижимости.

Однако это не жёсткое правило. Фактическое необходимое количество может зависеть от:

- Отношения сигнал/шум в ваших данных
- Сложности взаимосвязей, которые вы пытаетесь смоделировать
- Изменчивости вашей переменной отклика
- Качество ваших данных

Пример:

Итак, если у вас есть:

- 10 основных переменных
- 5 допустимых двусторонних взаимодействий
- 2 допустимых трёхсторонних взаимодействиях

Вы можете рассчитать минимальное количество продаж следующим образом:

$$\begin{aligned} N &= (10 \text{ переменных} * 15 \text{ строк/переменная}) + \\ &\quad (5 \text{ двусторонних взаимодействий} * 22 \text{ строк/двустороннее взаимодействие}) + \\ &\quad (2 \text{ трёхсторонних взаимодействия} * 30 \text{ строк/трёхстороннее взаимодействие}) \\ &= 320 \text{ строк} \end{aligned}$$

В качестве альтернативы можно поступить следующим образом: после запуска программы R/earth на наборе данных можно использовать количество базисных функций в модели, увеличенное на единицу, в качестве приблизительной оценки степени свободы, а затем умножить это число на 20–30, чтобы получить необходимое количество записей для ввода.

Protocol 1.2: Sales Periods Should Encompass Minimal Change in Buyer Tastes

При сборе исторических данных о недвижимости строгих временных ограничений не существует, если изменения рыночных предпочтений можно отразить с помощью соответствующих переменных. При анализе архитектурной эволюции на вашем рынке обратите внимание на следующие ключевые моменты: изменения доминирующих архитектурных стилей (например, переход от стиля «калифорнийское ранчо» к средиземноморскому) можно смоделировать с помощью:

- Идентификаторы микрорайонов
- Коды районов MLS
- Диапазоны дат строительства

Однако если эти изменения стилей невозможно адекватно учесть с помощью конкретных переменных, вы сталкиваетесь с важным ограничением: хотя «Дата продажи» обычно используется для корректировки с учётом рыночных условий, одновременное использование этой переменной для учета архитектурных предпочтений приведёт к смешению этих двух различных эффектов. В таких случаях предпочтительным подходом является либо:

1. Ограничить сбор данных периодом после появления более нового архитектурного стиля
2. Ограничить выбор аналогов микрорайонами, имеющими одинаковый архитектурный стиль

Это гарантирует, что ваш анализ сохранит согласованность характеристик недвижимости, при этом позволяя должным образом проводить корректировки рыночных условий с течением времени. В качестве альтернативы, по крайней мере в данном случае, если вы можете извлечь информацию об архитектурном стиле из данных, вы можете рассмотреть взаимосвязь между этими двумя переменными. Однако такое извлечение не всегда возможно.

Protocol 1.3: If In An Area Of Few Sales Then Do The Best You Can

При работе с моделированием MARS для данных о продажах для проведения значимого анализа требуется примерно 15 сделок. Если у вас меньше продаж (например, 5 объектов недвижимости), вы можете продублировать каждую запись дважды, чтобы достичь порога в 15 объектов недвижимости. Важные моменты, которые следует учитывать при использовании этого подхода:

1. Используйте только регрессию первой степени, чтобы уменьшить переобучение
2. Игнорируйте или пропускайте перекрёстную проверку, поскольку набор данных слишком ограничен
3. Хотя в данном сценарии метод MARS, вероятно, покажет лучшие результаты, чем методы сопоставленных пар или стандартной линейной регрессии, имейте в виду, что дублирование данных влияет на достоверность оценок неопределённости и статистической значимости

Лучшим долгосрочным решением является расширение набора данных за счёт дополнительных фактических данных о продажах. Однако, если вам приходится работать с ограниченным объёмом данных, дублирование данных представляет собой приемлемое временное решение при условии учёта его статистических ограничений.

5.4 Структура папок проекта

Было бы неплохо обсудить структуру папок проекта. Рекомендуется создать отдельные папки для «Анализа рынка» и «Оценки отдельных объектов недвижимости». Обе эти папки имеют схожую структуру вложенных папок:

```

RCA_Projects/
├── MarketAnalysis/
│   ├── Burlingame_20240501/
│   ├── Pacifica_20240310/
│   ├── RedwoodCity_20240615/
│   └── ...
├── SingleFamily/
│   ├── Pacific_Rosewood_123_20240510/
│   │   ├── Code/
│   │   │   ├── R/
│   │   │   │   └── ...
│   │   │   └── Python/
│   │   │       └── ...
│   │   └── Data/
│   │       ├── MLS_Original.xlsx
│   │       ├── MLS/
│   │       │   └── VER202402510153248/
│   │       │       ├── MLS.xlsx
│   │       │       ├── MLS_Stage2.xlsx
│   │       │       ├── MLS_Stage3.xlsx
│   │       │       ├── MLS_Stage4.xlsx
│   │       │       ├── MLS_Stage5.xlsx
│   │       │       └── Documents/
│   │       │           ├── LinearModels.pdf
│   │       │           ├── Cum_Dist.pdf
│   │       │           ├── Model.txt
│   │       │           ├── PartContrib.pdf
│   │       │           ├── Pairs.pdf
│   │       │           ├── Residual1.png
│   │       │           ├── Residual2.png
│   │       │           ├── ResidualSF.png
│   │       │           ├── Rpart.pdf
│   │       │           ├── VariablesImp.pdf
│   │       │           ├── VariableVsSP.pdf
│   │       │           └── ...

```

На этапе 1 обработки будет создана папка «version» с подпапками. В неё будет скопирован файл «MLS_Original.xlsx», хотя эта таблица не должна изменяться, а должна точно отражать данные, загруженные из MLS на определенную дату и время; однако в случае длительного проекта может возникнуть необходимость в новой загрузке данных из MLS. Далее на этом этапе в папку «version» будет скопирована расширенная книга MLS.xlsx с параметрами конфигурации. Может существовать множество папок «version», каждая из которых создается для полного цикла работы MARS.

Затем на следующих этапах будут выполняться добавление столбцов, сортировка и другие

изменения в книге, а по завершении оценки обновленная версия будет сохранена в папке с данными проекта под именем `MLS_Stage(N)`. Повторим: корневая папка `MLS.xlsx` постоянно изменяется, поэтому копия версии представляет собой непрерывную модификацию исходной книги `MLS` с переменными удалениями и добавлениями, а также другими изменениями. Код для обработки на языке `R` или `Python`, скорее всего, будет запускаться с использованием ввода из рабочей книги «`Data MLS.xlsx`». Таким образом, у нас всегда имеется копия всех данных, лежащих в основе каждой версии запуска `RCA`.

Создается столько версий, сколько необходимо, пока не станет невозможно внести дальнейшие улучшения; после чего лучшая версия выбирается в качестве окончательной. Можно изменить название папки, добавив в конец «`_Final`», чтобы это было видно при просмотре каталога. Некоторые старые версии с интересными или конкурирующими моделями, как правило, следует сохранять. Каждая папка версии содержит ряд текстовых файлов, диаграмм и графиков, которые можно встроить в отчёт. Окончательная книга `MLS_Stage5.xlsx` содержит столько листов, сколько необходимо для отчёта. В частности, в ней может быть таблица для переноса данных в таблицу продаж `Fannie Mae`. Она также обязательно будет содержать расчёты или разбивку, лежащие в основе агрегированных данных. Это, возможно, самая трудоёмкая часть подхода `RCA` — поиск идеальной модели в сложной рыночной зоне.

Все необходимое для воссоздания таблиц продаж следует скопировать в папку версии, включая используемый код на языке `R` или `Python` — если в код вносятся частые изменения.

5.4.1 Протоколы: Несколько прогонов `MARS` (для рынка недвижимости)

Protocol 2.1: Overfitting Runs To Surgically Study Market Area Pricing Factors

В ходе первоначальных прогонов для рынка недвижимости переобучение следует использовать стратегически с целью выявления основных факторов, влияющих на цену продажи. Например, рыночная зона, представленная 800 сделками, может содержать лишь 3 объекта недвижимости с мастерскими, и в этих единичных случаях наличие мастерской может оказывать существенное влияние на цену продажи. Прогон только в прямом направлении даст модель, подвергнутую переобучению, показывающую приблизительный вклад этих мастерских в стоимость. Однако, поскольку из 800 сделок таких всего 3, термины, связанные с мастерскими, будут удалены при обратном прогоне и, безусловно, при перекрёстной проверке. Значение, которое проходит только в прямом направлении присваивает мастерским, можно записать и перенести в анализ латентных переменных, чтобы использовать его для любого объекта недвижимости с мастерской, фигурирующего в таблице продаж. Такие переобученные модели также дают подробные оценки для многих других характеристик с небольшим объемом данных, которые были бы исключены в результате обрезки при обратном проходе или в ходе прогонов с перекрёстной проверкой.

Сохраните каждый из этих пробных прогонов в отдельной папке версии: зафиксированные значения вкладов являются доказательствами, на которые `VE` будет ссылаться при последующем разборе остатков. Разумеется, следует понимать, что в конечном итоге `VE` должен опираться на процедуры обрезки и перекрёстной проверки, чтобы сделать окончательную модель устойчивой к объектам недвижимости, с которыми регрессия никогда не сталкивалась — например, к объекту оценки, который появляется на более поздних этапах. Прогон с переобучением служат инструментами для исследования; они никогда не являются окончательной моделью.

Protocol 2.2: Multiple Cross-Validated Runs Under Different earth() Parameters

Рыночные зоны различаются по структуре и размеру, и специалисту по оптимизации (VE) обычно приходится проводить серию экспериментальных прогонов: чтобы понять, какие переменные оказывают наибольшее влияние на цену; чтобы оценить, позволяет ли имеющийся набор измеренных переменных адекватно оценить стоимость, или же необходимо добавить новые данные (расстояние до водоема, высота над уровнем моря, расстояние до торговых центров и т. д.); а также для выявления новых видов взаимосвязей и исключения других как неэффективных или вводящих в заблуждение. В конечном итоге VE стремится получить хороший коэффициент R^2 на обучающих данных и наилучший возможный коэффициент CVR^2 на тестовых данных. Для достижения этой цели может потребоваться сбор новых данных, повышение точности существующих данных или настройка параметров earth() с учётом размера и характера рыночной зоны (например, с учётом доступных степеней свободы — см. Протокол 1.1).

Именно поэтому приведенная выше структура проекта поддерживает множество папок с версиями прогонов: каждая потенциальная модель сохраняется вместе со своими параметрами и показателями R^2/CVR^2 , что позволяет сравнивать динамику результатов по различным прогонам. Версия, у которой показатель R^2 растёт, в то время как CVR^2 остается на прежнем уровне или снижается, отражает скорее шум, чем рыночный сигнал, и именно история версий позволяет это увидеть.

Protocol 2.3: Prolog Expansion of Remarks and Descriptive Feature Columns

Хотя эта функция пока находится на несколько экспериментальной стадии, earthUI теперь поставляется со встроенным движком Prolog (vProlog) и использует грамматику определённых предложений (DCG) для разбора столбцов свободного текста в формате MLS — прежде всего «Public Remarks», но также и замечания агентов, а также разделенные запятыми списки характеристик, такие как характеристики помещений или существующие постройки. Благодаря этому расширению такие фразы, как «собственная солнечная батарея», «гараж на 3 машины» или «мастерская 40x50», превращаются в обычные столбцы характеристик, которые можно использовать в регрессии, а восстановленные таким образом характеристики с небольшим объёмом данных напрямую подаются в анализ латентных переменных (см. последующий подраздел, посвящённый оценкам состояния и качества в качестве приближённых регрессоров; сама характеристика описана в сопутствующей статье earthUI в этом выпуске).

Расширение текста увеличивает набор потенциальных предикторов, что требует проведения нескольких циклов анализа earth/RCA для тестирования: расширенные столбцы необходимо проверять как в составе модели, так и вне её, объединять в производные подсчёты или булевы значения в случаях коллинеарности, а также сверять с CVR^2 , чтобы подтвердить, что они несут рыночный сигнал, а не шум. Каждый такой эксперимент следует рассматривать как отдельную папку версии прогона, точно так же, как в протоколах 2.1 и 2.2. А поскольку сама грамматика разбора находится в стадии разработки, VE должен сверять извлеченные столбцы с исходным текстом, прежде чем полагаться на них в модели.

5.5 Основная книга Excel «

» Поскольку большинство сервисов MLS загружают данные об объектах недвижимости в таблицы Excel в формате CSV, Excel, вероятно, является наиболее удобным способом хранения данных и различных параметров настройки. Для каждого проекта имеется одна основная книга Excel с именем MLS.xlsx, которая содержит один лист под названием «MLSDData» для хранения списков завершённых сделок по продаже недвижимости, загруженных из MLS, а также несколько листов с настройками и параметрами:

```
Workbook: MLS.xlsx/  
├─ Sheet: Project Data  
├─ Sheet: MLSDData  
├─ Sheet: Regression Variables  
├─ Sheet: Calculations  
├─ Sheet: Allowed Interactions  
├─ Sheet: Aggregations  
└─ Sheet: (MARS run parameters)
```

Исходные данные хранятся в книге с именем типа «MLS_Original.xlsx». Необходимо создать её копию с другим именем, например «MLS.xlsx», которая будет загружаться в программу RCA на языке R или Python. Обе эти книги хранятся в папке «Data». Теперь рабочая книга «Data/MLS.xlsx», вероятно, пройдет множество итераций, в ходе которых будут пробоваться различные комбинации переменных, параметров MARS и изменений других настроек. Каждая итерация создает новую папку для ввода/вывода данных, названную по дате и времени в формате VER _ YYYYMMDDHHMMSS. В каждой папке будет находиться копия вводной книги MLS.xlsx, а также сгенерированные книги Excel для каждого из четырёх этапов обработки (этапы 2–5), а также документы вывода, диаграммы и графики. Обратите внимание, что этап 5, заключительный этап, может содержать ряд других листов для отчёта в зависимости от типа клиента и технического задания.

5.6 Резервное копирование

При записи новых или измененных таблиц Excel в папки версий проекта также можно создавать резервные копии в базах данных SQLite или PostgreSQL, расположенных в папке проекта. Вероятность непреднамеренного повреждения файлов баз данных впоследствии, в процессе проверки или использования в качестве справочных материалов, значительно ниже.

6 Обозначения книг

6.1 Описательные измерения

Для книги «М» существует примерно шесть описательных измерений:

1. Оценки «а»
2. Версии запуска «г»
3. Этапы рабочего процесса «w»
4. Листы «s»
5. Объявление о продаже недвижимости «р»
6. Набор переменных «v»

Итак, отдельную ячейку в конкретной книге М можно описать следующим образом:

$$\begin{matrix} r_* \\ w_* \end{matrix} \begin{matrix} a_* \\ W \\ s_* \end{matrix} \begin{matrix} p_* \\ v_* \end{matrix} \quad (1)$$

(Примечание: “*” означает любой допустимый индекс)

или:

$$\begin{matrix} r_k \\ w_m \end{matrix} \begin{matrix} a_g \\ W \\ s_h \end{matrix} \begin{matrix} p_i \\ v_j \end{matrix} \quad \begin{matrix} i = 1, \dots, n_p \\ j = 1, \dots, n_v \\ k = 1, \dots, n_r \\ m = 1, \dots, n_m \\ g = 1, \dots, n_g \\ h = 1, \dots, n_h \end{matrix} \quad (2)$$

where:

- n_p = # of rows (depends on sheet)
- n_v = # of columns (depends on sheet)
- n_r = # of run versions
- n_m = # of workflow stages
- n_g = # of analysis projects
- n_h = # of workbook sheets

Следует понимать, что «W» обозначает всю книгу Excel, состоящую из шести основных листов данных, каждый из которых содержит собственную таблицу, которая при запуске соответствующей программы на R или Python загружается в отдельные внутренние датафреймы. И в R, и в Python есть встроенные функции, которые по имени могут определить, к какой строке или столбцу датафрейма относится данное имя. Например, если в $index_v_j$, $v = ("GLA", "LotSize", "BathRms", "BedRms", \dots)$ и $j=0$, $tov_j = "GLA"$. Как R, так и Python внутренне смогут сопоставить «GLA» с столбцом, в котором оно хранится, которым, например, может быть 15-й столбец соответствующего фрейма данных. Поэтому в данной статье мы используем индексацию в наших массивах имен, которые передают имя запущенной программе, которая, в свою очередь, преобразует это имя в фактический индекс столбца или строки для внутреннего фрейма данных. Это немного сложно. Но, если быть точным, это необходимо и наиболее удобно. Речь идет о «разделении обязанностей» между ролью аналитика и ролью разработчика; хотя на самом деле инженер по оценке (VE) может выполнять обе роли. Обратите внимание, что «data frame» в R в Python называется «dataframe».

Также обратите внимание, что лист параметров выполнения MARS не является измерением W , поскольку W относится к версии данных, осуществляемого ввода в MARS, и не связана с параметрами, определяющими, как MARS/earth обрабатывает эти данные.

Для большей ясности представлена сложная многомерная нотация для книги, которая более непосредственно связана с внутренними массивами данных R или Python, в которые загружаются листы книги с данными. Это «тяжелая» нотация. Однако когда мы переходим к доказательствам и т. п., мы избавляемся от значительной части этого «багажа», делая допущения о том, с чем мы работаем. Тем не менее, эта нотация помогает с самого начала получить более полное представление о рабочем процессе и данных.

Начнём с обозначения множества книг, которые мы будем обрабатывать.

6.1.1 Размер проекта (“a”)

После утверждения объема работ (SOW) необходимо создать папку проекта для хранения данных ввода и вывода, включая параметры конфигурации, выходные графики, диаграммы, таблицы и фрагменты отчетов. VE должен идентифицировать все проекты с помощью простого последовательного идентификатора, такого как целое число, чтобы в случае пропажи отчёта это было очевидно. Если проекты отменяются, то в соответствующей папке проекта должен быть единообразный способ указания на то, что проект отменен, но сама папка никогда не должна удаляться. Мы предполагаем, что идентификаторы проектов начинаются с 0 и увеличиваются на 1, хотя возможны и другие варианты. Кроме того, мы предполагаем, что все действительные идентификаторы проектов хранятся в упорядоченном наборе «a», созданном с помощью программы обработки.

Данные проекта по оценке находятся в рабочем листе «Проект» этой книги, но также могут храниться в какой-либо центральной базе данных SQL,

Вот предлагаемые поля проекта:

1. Локальный идентификатор: Короткий целочисленный идентификатор, который генерируется последовательно для компании. Он отлично подойдет в качестве первичного ключа для хранения данных — и может автоматически генерироваться базой данных при создании каждого последующего проекта.
2. Глобальный GUID: При желании вы можете создать GUID (глобальный универсальный идентификатор) с помощью

<https://www.uuidgenerator.net/>

3. Местоположение объекта недвижимости (анализ рынка) или адрес
4. Дата оценки
5. Дата выдачи заказа
6. Дата осмотра
7. Дата составления первоначального отчёта
8. Дата составления первого обновленного отчёта
9. Дата составления второго обновленного отчёта
10. Дата составления третьего обновленного отчёта
11. Имя заказчика
12. Адрес заказчика
13. Название кредитной организации
14. Адрес кредитной организации
15. Имя владельца
16. Адрес владельца

17.

Пример списка идентификаторов проектов:

```
a = ("1",
      "2",
      "...",
      "141",
      "142")
```

6.1.2 Параметр версии ("r")

На втором этапе рабочего процесса мы выполняем MARS-регрессию над вводом. Вероятно, это потребует от нескольких до множества запусков, поскольку мы будем изменять различные параметры для улучшения модели. При каждом запуске для вывода создается новая папка версии, имя которой формируется на основе даты и времени выполнения. Например,

```
r = ("ver20241210_083124",
      "ver20241209_151130",
      "ver20241208_160944")
```

latest_version : r_1 = "ver20241210_083124"

Многообещающие версии, аналог по характеристикам и выводу с конкурентными моделями, следует сохранять до тех пор, пока не будет выбрана окончательная версия, или, возможно, дольше, если они представляют какую-либо полезность. Возможно, вам захочется изменить название папки с версией, чтобы обозначить, что это окончательная версия, добавив слово «Final» в конец названия.

6.1.3 Этапы рабочего процесса ("w")

Рабочий процесс RCA обычно делится на пять отдельных этапов, перечисленных ниже. Каждый этап должен быть успешно завершен перед началом следующего. Специалист по верификации (VE) должен иметь возможность остановить обработку по окончании этапа, чтобы проанализировать результаты перед продолжением работы. Это необходимо, поскольку рыночные зоны являются достаточно сложными и зачастую весьма уникальными. Могут возникнуть нестандартные проблемы, требующие нового подхода. Эксперт по оценке (VE) должен иметь возможность напрямую работать с кодом на языках R или Python, чтобы вносить необходимые изменения и даже пробовать новые подходы. Кроме того, в таких случаях ему может потребоваться экспериментировать и пробовать различные подходы, что потребует многократного запуска того же этапа, над которым он работает. Вместе с тем у VE должна быть возможность запускать все этапы одной командой — с некоторыми оговорками (например, агрегация 3-й степени должна быть уже определена в результате предыдущего прогона этапа 2). Это означает, что по завершении каждого этапа он должен сохранять все свои внутренние данные в файлах или базах данных, чтобы следующий этап можно было запустить так, как будто предыдущий этап все еще находится в памяти, путем повторной загрузки данных предыдущего этапа в память.

Вот этапы:

w = ("stage1", -Настройка
 "stage2", -Обработка MARS
 "stage3", -Итоговая таблица продаж
 "stage4", -Фрагменты отчётов
 "stage5") -Проверка/очистка

stage2 : w₂ = "stage2"

6.1.4 Листы книги рабочей Excel ("s")

В рабочей книге RCA Core Excel имеются следующие листы, каждый из которых сопоставлен с внутренним фреймом данных:

s = ("MLSData",
 "Регрессионные переменные",
 "Расчёты",
 "Взаимодействия",
 "Агрегации 1-й& 3-й степени",
 "Агрегации 2-й степени")

Причина, по которой агрегации 2-й степени выносятся на отдельный лист, заключается в том, что их проще всего определить с помощью симметричной таблицы, в которой показаны все возможные пары взаимодействий, что позволяет указать агрегационную переменную для каждого вклада значений, связанных с данной парой, а также корректировок, которые необходимо агрегировать (сложить). С агрегациями третьей степени мы не смогли бы этого сделать, поскольку для этого потребовалась бы трехмерная симметричная таблица, что невозможно в Excel. Кроме того, таблица размером 20×20×20 содержала бы 8 000 ячеек, с которыми пришлось бы работать. Проще просто перечислить названия всех членов третьей степени в таблице с агрегациями первой степени, например:

Переменная	Агрегировать в
Площадь отделанных помещений над уровнем земли	GLA
Площадь отделанных помещений под уровнем земли	BGLA
Площадь неотределанных помещений	UBA
Площадь участка	LotSize
...	...
GLA_ _ LotSize	GLA
GLA_ _ DateOfSale	GLA
BathRms_ _ DateOfSale	GLA
Lat_ _ Long	Location
Lat_ _ Long_ _ GLA	Location
...	...

6.1.5 Список объектов недвижимости ("p") — строки MLSData

Списки объектов недвижимости составляют строки листа MLSData. В других листах их нет до этапа 4, на котором формируются фрагменты итогового отчёта с окончательным перечнем аналогов. Ниже приведен примерный список идентификаторов объектов:

$$p = (\text{"MLS823211"}, \text{"MLS823456"}, \text{"MLS823542"}, \text{"MLS823721"}, \text{"..."})$$

6.1.6 Переменные MLS ("v") Столбцы MLSDData

О работе с переменными будет сказано подробнее позже. На данном этапе достаточно отметить, что лист MLSDData обычно содержит 30–40 переменных различных типов, однако для регрессии в MARS в качестве независимых переменных обычно выбирается лишь 10–20 из них. Эти переменные также встречаются на других листах, хотя, как правило, в виде списка всех возможных переменных, которые можно использовать для передачи в MARS. Пример списка переменных:

$$v = (\text{"GLA"}, \text{"SalePrice"}, \text{"BedRms"}, \text{"BathRms"}, \text{"GarageSF"}, \text{"..."})$$

6.1.7 Пример окончательной нотации

Давайте применим приведенную выше нотацию в заключительном примере. Используя приведенные выше списки, пусть

$$r_1 \quad w_1 \quad \overset{a_{141}}{W}_{s_0} \quad p_2 \quad v_0 \quad (3)$$

Обратите внимание, что здесь мы используем нулевую индексацию, то есть считается, что первый элемент списка имеет индекс 0. В Python также используется нулевая индексация, а в R — индексация, начинающаяся с 1.

Итак, из вышесказанного следует, что речь идет о значении ячейки для:

- Переменная v_0 или столбец «GLA»
- Для свойства r_2 или ID «MLS823542»
- Из листа s_0 или «MLSDData»
- Для этапа рабочего процесса w_1 или «stage2»
- Из версии r_1 или «ver20241209_151130»
- Из проекта a_{141} или «142»

Приведённую выше запись можно также записать следующим образом:

$$\begin{matrix} \text{"ver20241209_151130"} \\ \text{"stage2"} \end{matrix} \quad \overset{\text{"142"}}{W}_{\begin{matrix} \text{"MLSDData"} \\ \text{"GLA"} \end{matrix}} \quad \text{"MLS823542"} \quad (4)$$

6.1.8 Обоснование

«Почему бы нам просто не использовать индексацию или цифры в обозначениях?» К нашим обозначениям предъявляются два требования:

1. Используются подмножества: например, аналитик часто выбирает подмножества из списка, такие как подмножества переменных. Например, на листе `MLSDData` может содержаться 35–40 переменных, но для конкретного прогона используется только 10–20 из них. От одного прогона к другому `VE` часто меняет перечень переменных, передаваемых в `MARS/earth`.
2. Итерация, например, $i = 1, 2, \dots, n_1$
 Наши алгоритмы, протоколы и доказательства обычно должны иметь возможность проходить по всем элементам упорядоченного списка, от первого до последнего, или выбрать какой-либо конкретный элемент.

Итак, нам нужна буква для обозначения подмножества или списка элементов, и у этой буквы должен быть еще один индекс, с помощью которого мы сможем перебирать элементы списка, чтобы получить строки, используемые для индексации строк или столбцов датафрейма.

6.1.9 Матрицы?

А как насчёт матриц из линейной алгебры? Матрицы работают только с числами — целыми или действительными. Матрицы как структура данных поддерживаются как в `R`, так и в `Python`. Однако многие переменные, с которыми мы работаем в данных `MLS`, называются «строками» или символьными данными. Даже если некоторые переменные представляют собой только числа, зачастую это не действительные числа, а просто имена. Например, коды `MLS` для города могут быть следующими: («660», «661», «662», «663», ..., «672»). Регрессия `Earth` позволяет задавать переменные как «факторные», и в этом случае числа рассматриваются просто как имена, без какой-либо последовательности. В таких случаях построенная модель будет просто предоставлять вклад каждого числа в значение, а не функцию, пробегающую по последовательности. Другие примеры включают «C1», ..., «C6» для параметра «Condition» и «Q1», ..., «Q6» для параметра «quality».

6.1.10 Первое упрощение

Приведенная выше нотация для основной книги проекта `Excel` удобна для связывания всех элементов воедино, но слишком громоздка для использования при объяснении алгоритмов или математических доказательств. При выполнении рутинной работы мы можем исходить из того, что она относится к данному проекту, версии, этапу и листу. Мы просто указываем, с чем работаем, если это не следует из контекста. На самом деле, после того как мы указали, с чем работаем, из приведённого выше нам нужно только следующее:

$$M_{s_h}^{p_i, v_j} \quad \begin{array}{l} i = 1, 2, \dots, n_p \\ j = 1, 2, \dots, n_v \\ h = 1, 2, \dots, n_s \end{array} \quad (5)$$

where: n_p = # properties
 n_v = # input variables
 n_s = # input sheets

6.1.11 Второе упрощение

Удобнее переименовать листы в соответствии с конкретными типами данных, которые они поддерживают, вместо использования индексов. В каждом листе строки и столбцы

выполняют разные функции. Таким образом, набор данных W разбивается на массивы данных M , V , A , I и C следующим образом:

$$\diamond \text{ MLS Data} \quad M_{v_j}^{p_i} = W_0^{p_i} \quad \begin{matrix} i = 0, 1, \dots, n_p \\ j = 0, 1, \dots, n_v \end{matrix} \quad (6)$$

where: $n_p = \# \text{ properties}$
 $n_v = \# \text{ input variables}$

Примечание: По умолчанию объект оценки всегда находится первым в списках свойств В таблице Excel «MLSDData» и соответствующем фрейме данных первая строка всегда зарезервирована для объекта оценки. Оно никогда не сортируется вместе с другими объектами недвижимости MLS, расположенными ниже. Оно никогда не вводится в регрессию MARS в качестве независимых данных, а вводится в качестве целевых данных или в качестве зависимой переменной. То же самое относится и к спискам объектов недвижимости. Во всех списках объектов недвижимости первый элемент с индексом 0 всегда считается идентификатором объекта оценки, если не указано иное.

Даже если задача состоит в создании модели для рыночной зоны без каких-либо объектов оценки, первая строка всё равно заполняется фиктивными данными о свойствах. Это значительно упрощает написание программ, определение алгоритмов или проведение доказательств. Итак, если мы используем c для представления списка объектов недвижимости, то c_0 является объектом оценки, а c_i и $i = 1, \dots, n_p$ — аналогами, которые будут введены в регрессионный анализ MARS/earth.

◇ Спецификации переменных регрессии

$$V_{s_j}^{v_i} = W_1^{v_i} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_s \end{matrix} \quad (7)$$

where: $n_v = \# \text{ available variables}$
 $n_s = \# \text{ of specifications}$

Обратите внимание, $n_s \leq n_v$, что обычно мы выбираем подмножество доступных переменных для ввода в регрессию.

◇ Расчеты

$$C_{c_j}^{v_i} = W_2^{v_i} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_c \end{matrix} \quad (8)$$

where: $n_v = \# \text{ of variables}$
 $n_c = \# \text{ of calc specs}$

◇ Interactions

$$I_{t_j}^{f_i} = W_3^{f_i} \quad \begin{matrix} i = 1, 2, \dots, n_f \\ j = 1, 2, \dots, n_t \end{matrix} \quad (9)$$

where: $n_f = \# \text{ from variables}$
 $n_t = \# \text{ to variables}$

Обратите внимание, что таблица данных I является симметричной, с переменными, перечисленными в алфавитном порядке в строках «from» и столбцах «to». Ячейки помечены цифрой «1», чтобы указать, что взаимодействие между соответствующей парой переменных допускается. В противном случае цифра «0» означает, что взаимодействие не допускается. Для взаимодействия второй степени это относительно просто, так как возможно только одно взаимодействие. В случае регрессии третьей степени у нас есть три парных взаимодействия между всеми тремя переменными, причём каждая пара

взаимодействий должна быть разрешена, чтобы сформировать терм (или переменную) взаимодействия трёх переменных.

◇ Агрегации

$$A_{v_j}^{v_i} = W_4^{v_i}_{v_j} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_v \end{matrix} \quad (10)$$

where: $n_v = \#$ of variables

7 Таблица продаж RCA

Имейте в виду, что протокол RCA требует использования программы на языке R или Python, способной обращаться к R/earth и выполнять все необходимые шаги. Этот процесс требует обработки больших объемов данных, и выполнить эту работу вручную просто невозможно. В то же время, даже если предположить, что программа не содержит ошибок, вам понадобится программное обеспечение, чтобы гарантировать, что необходимая работа будет выполнена точно и без человеческих ошибок.

Кроме того, специалистам по оценке (VE), выполняющим эту работу, необходимо иметь возможность модифицировать такие программы по мере необходимости, особенно когда они сталкиваются с новыми непредвиденными рыночными зонами, районами, объектами недвижимости или другими новыми требованиями и проблемами.

В рамках обработки RCA мы используем три типа таблиц продаж:

1. Таблица продаж RCA Processing, обозначенная в данной статье как «MLSDData», представляет собой очень большой фрейм данных и связанную с ним электронную таблицу, которые могут содержать множество промежуточных сведений, подвергающихся впоследствии процессу агрегирования для получения таблицы продаж RCA, упомянутой ниже. Привести пример здесь было бы невозможно из-за его большого объёма. Скорее всего, инженеры по оценке будут адаптировать эту электронную таблицу под свои предпочтения. В будущем я планирую написать отдельную статью о разработке такой таблицы.
2. Окончательная таблица продаж RCA, используемая в отчетах, не относящихся к GSE, выглядит примерно так: [Table 2](#). Она действительно имитирует, например, форму Fannie Mae, что позволяет читателю почувствовать некоторую знакомость. Однако обратите внимание, что в ней больше столбцов на каждый аналог продажи, чем в таблице Fannie Mae: у каждого аналога есть столбец «Вклад в стоимость» наряду со столбцами «Показатель» и «Корректировка», и у объекта оценки он также присутствует — этих столбцов нет в форме Fannie Mae. Еще более простая форма — это таблица продаж [«Table 1»](#).
3. Это таблица продаж GSE Upload, которая используется для загрузки таблицы продаж в программное обеспечение, такое как Alamode. Однако обратите внимание, что не все программные пакеты для оценки недвижимости предоставляют возможность загрузки данных таблицы продаж из электронной таблицы. По необходимости такие таблицы не могут содержать отдельный столбец для вкладов в оценку, понятия о которых у государственных ипотечных агентств (GSE), таких как Fannie Mae, отсутствует. Стратегия заключается в том, чтобы предоставить им требуемую форму, но подкрепить её таблицей продаж RCA, а также другой сопроводительной документацией, такой как подробная информация о любых агрегированных переменных.

Таблицы продаж RCA Processing не похожи на таблицы продаж Fannie Mae, но, грубо говоря, представляют собой транспонированную версию таблицы продаж Fannie Mae с переставленными строками и столбцами: каждая строка содержит данные по какому-либо объекту недвижимости, а каждый столбец — данные по какой-либо характеристике объекта. Количество строк с данными об объектах может легко достигать сотен или тысяч, а количество столбцов — 200 и более. Первая строка зарезервирована для имен переменных, а вторая — для объекта оценки, даже если мы проводим лишь рыночный анализ. Когда лист импортируется в дата-фрейм, первая строка листа становится характеристикой «names», а вторая строка, относящаяся к объекту оценки, становится первой строкой дата-фрейма.

Столбцы в таблице продаж RCA включают идентификационные столбцы, а затем первичные, внешние и синтетические переменные, которые объединяются и передаются в MARS в качестве независимых переменных. На основе этих переменных MARS делает вывод о том, какие из них оказывают существенное влияние на цену продажи, также известную как зависимая или целевая переменная. При регрессии второй степени мы можем получить пары этих переменных, которые выступают в качестве новых переменных. При регрессии третьей степени мы также можем получить тройки переменных, которые выступают в качестве так называемых независимых переменных, хотя они, конечно же, сильно коррелируют с исходными переменными.

Таблица 1: Пример таблицы RCA: вертикальная компоновка

Subject			Comp 1			Comp 2		
	Measure	Value Contribution	Measure	Value Contribution	Adjustment	Measure	Value Contribution	Adjustment
Sale Price			\$1,150,000			\$1,500,000		
Base Value		\$500,000		\$500,000	\$0.00		\$500,000	\$0
Measured Features			Measured Features			Measured Features		
Living Area	1,530	\$463,500	1,350	\$382,500	\$81,000	1,700	\$540,000	-\$76,500
Lot Size	6,000	\$200,000	6,500	\$220,000	-\$20,000	7,000	\$240,000	-\$40,000
Age	53	\$63,600	57	\$68,400	-\$4,800	51	\$61,200	\$2,400
Bathrooms	2	\$15,000	3	\$30,000	-\$15,000	3	\$30,000	-\$15,000
Bedrooms	3	\$6,000	4	\$9,000	-\$3,000	4	\$9,000	-\$3,000
Total Measured		\$1,248,100.00		\$1,209,900.00	\$38,200		\$1,380,200.00	-\$132,100
Residual Features			Residual Features			Residual Features		
Residual	CQA 5.8	\$30,000	CQA 2.5	-\$59,900	\$89,900	CQA 8.2	\$119,800	-\$89,800
Residual Breakdown			Residual Breakdown			Residual Breakdown		
Design		7,000		-\$15,000	\$22,000		\$35,000	-\$28,000
Condition		5,000		-\$8,000	\$13,000		\$20,000	-\$15,000
Quality		8,000		-\$10,000	\$18,000		\$30,000	-\$22,000
Func Utility		5,000		-\$12,000	\$17,000		\$15,000	-\$10,000
Amenities		5,000		-\$14,900	\$19,900		\$19,800	-\$14,800
Sub-Total		30,000		-\$59,900	\$89,900		119,800	-\$89,800
Total			Total			Total		
Tot Value Contribution		\$1,278,100.00		\$1,150,000.00			\$1,500,000.00	
Adjusted Sale Price					\$1,278,100.00			\$1,278,100.00

8 Переменные

Что касается ввода при запуске MARS, у нас есть только простые переменные, а именно:

1. Характеристики объектов недвижимости, загруженные из таблицы MLS.
2. Производные переменные, созданные оценщиком для уточнения переменных MLS или преобразования существующих переменных. Например, оценщик может преобразовать «Дату продажи» в «DaysOffMarket», равную дате оценки за вычетом «Даты продажи».
3. Оценщик может импортировать данные из сторонних источников, таких как ставки по ипотечным кредитам. Или, если в данных MLS отсутствуют данные о широте, долготе или высоте над уровнем моря, он может импортировать их из других источников. Эти переменные можно назвать «внешними переменными».

На выводе системы MARS мы работаем с ценовой моделью, представляющей собой комбинацию слагаемых, которые либо:

1. Содержат только одну из вводных переменных и предоставляют уравнение, определяющее вклад этой переменной в стоимость. Их называют слагаемыми первой степени или (выходными) переменными. Само слагаемое рассматривается как величина вклада в стоимость и получает название.
2. Содержат две вводные переменные и представляют собой уравнение, определяющее вклад в значение при взаимодействии данных двух переменных. Эти члены можно называть членами второй степени или (выводными) переменными.
3. Содержат три вводные переменные и представляют собой уравнение, определяющее вклад в значение при взаимодействии этих трех переменных. В данной работе эти члены могут называться членами третьей степени или (выводными) переменными.

Таким образом, выводными переменными модели MARS являются вклады в значение, включая вклад остаточной величины. В свою очередь, они рассматриваются как переменные при расчетах в таблице Sales Grid, например при вычислении корректировок или построении графиков.

Существуют различные типы переменных, которые подразделяются на две основные категории: (1) переменные происхождения и (2) переменные применения, т. е. неостаточные и остаточные.

8.1 Онтология «Variable Origin»

1. Переменные первой степени
 - 1.1 Переменные регрессии MLS
 - 1.2 Внешние переменные регрессии
 - 1.3 Производные переменные регрессии

2. Факторные переменные:

Для работы с факторными переменными требуется функция, способная их идентифицировать. Поскольку они задаются в рабочей книге конфигурации проекта Excel, создается функция, которая возвращает значение «true», если переданная ей переменная была классифицирована как факторная.

3. Переменные взаимодействия: они создаются программой MARS при построении модели, если степень взаимодействия установлена на 2 или выше. Однако рекомендуется использовать не более 3 взаимодействующих переменных из-за сложности интерпретации, вероятности переобучения, а также влияния на степень свободы и связанного с этим увеличения количества записей о недвижимости, необходимого для обеспечения надежности модели.
 - 3.1 Переменные второй степени: Они встречаются часто; хорошими примерами могут служить GLA:BathRms, GLA:LotSize, DateOfSale:BathRms, Latitude_ Longitude.
 - 3.2 Переменные третьей степени: менее распространённые взаимодействия, требующие осторожности и указывающие на возможную подгонку.
4. Переменные разбивки остатков

8.2 Онтология переменных приложений

1. Вклад MARS в создание стоимости
 - 1.1 Переменные первого порядка
 - 1.1.1. Переменные регрессии MLS
 - 1.1.2. Внешние переменные регрессии
 - 1.1.3. Производные переменные регрессии
 - 1.2 Переменные взаимодействия
 - 1.2.1. Переменные второй степени
 - 1.2.2. Переменные третьей степени
 - 1.3 Переменные разбивки остатков
 - 1.4 Агрегированные переменные
2. Переменные, сгенерированные инженерами& Значения
 - 2.1 Переменные разбивки остатков
 - 2.1.1. Дизайн/эстетика
 - 2.1.2. Ландшафт
 - 2.1.3. Функциональная полезность
 - 2.1.4. Вид
 - 2.1.5. Состояние
 - 2.1.6. Качество
 - 2.1.7. ...

Эти группы переменных будут обозначены соответствующими символами и более подробно описаны в следующем разделе

Protocol 3.1: Factor Variables

Факторные переменные, или просто факторы, могут встречаться во всех слагаемых, включая взаимодействия. Они важны, поскольку при проведении регрессии MARS и построении графиков с ними необходимо обращаться особым образом.

В то время как переменные с символьными значениями по определению являются факторами, числовые переменные необходимо пометить как факторные, чтобы MARS распознавал их как таковые. Например, районы MLS часто обозначаются номерами. В частности, районы в районе залива Сан-Франциско часто обозначаются уникальными целыми числами. Если эти значения передаются в MARS в качестве факторов, то программа попытается определить, имеют ли определенные районы средний вклад в объем продаж недвижимости в пределах своих границ, не учитывая при этом никаких взаимосвязей с другими районами. Напротив, в случае обычных переменных программа будет предполагать наличие некоей непрерывности в вкладе значений по мере изменения значения переменной. Это важное различие. Кроме того, эти факторные переменные могут встречаться в членах модели вместе с нефакторными переменными, что затрудняет построение графиков. В частности, при построении графиков членов с одной переменной и факторами мы объединяем все такие факторы в одну группу и просто строим гистограмму, показывающую их вклад — положительный или отрицательный.

Если факторы фигурируют в выражениях вместе с другими переменными, они всегда будут иметь только одно возможное значение, поскольку факторные переменные могут принимать только одно значение для данного объекта недвижимости: если значение площади равно 662, то по определению она не может иметь никакого другого значения для данной площади. Если вам это кажется запутанным, то, что ещё хуже, в модели, сгенерированной MARS, нет единого имени переменной для фактора; MARS возьмёт корневое имя факторной переменной, введённой в качестве вводных данных, а затем создаст новые столбцы с именами переменных, образованными путём добавления возможных значений. Таким образом, один столбец в таблице Excel, например MlsArea, будет соответствовать нескольким столбцам в массиве данных, созданным для каждого возможного значения, таким как MlsArea660, MlsArea661 и т. д. Переменная, такая как MlsArea661, будет иметь значение 1 для объектов недвижимости в зоне 661; в остальных случаях — 0. Затем MARS выполнит регрессионный анализ по этим столбцам, как если бы они представляли булевы переменные. Эти измененные имена будут использоваться в модели, предоставленной MARS.

Вероятно, вам захочется преобразовать эти имена моделей обратно в их исходные названия, чтобы сделать графики более понятными для клиентов, которые предпочитают, чтобы всё было красиво и просто (по крайней мере, настолько просто, насколько это возможно). Вместо того чтобы видеть MlsArea661=1, мы будем видеть MlsArea=661. Затем мы сможем построить графики членов, просто указав в заголовке графика что-то вроде «Для MLS Area = 662». Таким образом, члены степени N с M факторными переменными отображаются как члены степени (N-M), а факторное измерение исчезает в заголовке.

Обратите внимание, что в одном члене могут присутствовать две или более факторных переменных, поэтому, скорее всего, вам понадобится график для каждой возможной пары значений. Если переменная x с тремя различными значениями и переменная y с двумя различными значениями появляются в одном члене вместе с нефакторной переменной z, то вам понадобится $2 \cdot 3 = 6$ графиков. Примером может служить сочетание «MlsArea» + «Architectural Style»: например, стоимость викторианских домов в Сан-Франциско может варьироваться в зависимости от района города, в котором они расположены. К счастью, в большинстве районов обычно имеется лишь 1–3 архитектурных стиля, достаточно значимых с точки зрения влияния на стоимость, чтобы их можно было использовать в модели (хотя может быть 6 и более различных архитектурных стилей).

8.3 Члены второй и третьей степени

При регрессии второй или третьей степени некоторые члены выводной модели могут включать две или, соответственно, три переменные. Например:

$$\text{BathRms_GLA} = -52 * \max(0, 1200 - \text{GLA}) * \max(0, \text{BathRms} - 1)$$

Обратите внимание, что мы присваиваем им имена, сортируя имена переменных в алфавитном порядке и объединяя их, возможно, с использованием подчеркивания. То же самое мы делаем с членами третьей степени в тройках имен переменных, сгенерированных моделью. Например, «BathRms_GLA_LotSz».

Эти переменные добавляются к основным вводным переменным на этапе 3, когда программа RCA рассчитывает вклад всех переменных для каждого объекта недвижимости, включая переменные взаимодействия.

8.3.1 Переменные корректировки

Программа RCA также рассчитывает корректировки для всех переменных и объектов недвижимости на этапе 3, после расчета вклада каждого элемента в стоимость, путем вычитания вклада соответствующего аналога объекта недвижимости по данному элементу из вклада объекта оценки по этому же элементу.

8.3.2 Агрегирующие переменные

Агрегирующие переменные определяются для замены подмножеств переменных, объединяя все их значения и корректировки в одну общую переменную, которая будет использоваться в отчёте. Например, мы могли бы разбить как готовую, так и незавершённую жилую площадь на более детализированные переменные:

1. Законно обустроенная площадь над уровнем земли
2. Законно обустроенная площадь ниже уровня земли
3. Незаконно обустроенная площадь над уровнем земли
4. Незаконно обустроенная площадь ниже уровня земли
5. Необустроенная площадь над уровнем земли
6. Необустроенная площадь ниже уровня земли

Да, есть сферы, в которых это действительно имеет смысл.

Регрессия MARS будет проводиться по всем переменным и позволит построить модель, которая определит вклады и корректировки для каждой из них. Однако для отчёта в установленной форме эти различные вклады, вероятно, придется объединить или агрегировать в меньшее количество переменных. Подробности агрегирования следует привести в приложении. Более распространённым примером может служить агрегирование корректировок на взаимодействие в одну из корректировок на основные переменные. Например, мы можем агрегировать корректировку на GLA_LotSize в корректировку на GLA или на LotSize, либо даже распределить их поровну между обеими переменными, как указано в листе «Агрегирование».

8.3.3 Специальные переменные & Обозначения

π_i	Фактическая цена продажи недвижимости $i = 1, 2, \dots, n_p$; для объекта оценки π_0 обозначает вывод о стоимости (см. Доказательство I)
α_i	Скорректированная цена продажи недвижимости $i = 1, 2, \dots, n_p$
ρ_i	Расчетная цена продажи по MARS для недвижимости $i = 0, 1, \dots, n_p$
β	Базовая стоимость модели
$\varphi_{i,j}$	Вклад переменной регрессии $j = 1, 2, \dots, n_r$ в значение для объекта $i = 0, 1, \dots, n_p$
$\delta_{i,j}$	Корректировка для переменной регрессии $j = 1, 2, \dots, n_r$ для объекта $i = 1, 2, \dots, n_p$
ϵ_i	Общая величина остатка для объекта $i = 0, 1, \dots, n_p$
$\epsilon_{i,c}$	Вклад компонента остаточной величины $c, c = 1, 2, \dots, n_c$ для объекта $i = 0, 1, \dots, n_p$
$\xi_{i,c}$	Корректировка для компонента остаточной величины $c, c = 1, 2, \dots, n_c$ для объекта $i = 0, 1, \dots, n_p$

где

$n_p = \#$ of comparable properties

$n_r = \#$ of regression variables

$n_c = \#$ of residual breakdown variables

8.3.4 Разбивка остатков

После построения модели с помощью регрессии MARS система генерирует расчетные цены продажи на основе входных переменных, выбранных системой MARS. Затем это значение вычитается из соответствующей чистой цены продажи, чтобы получить остаток для каждого объекта недвижимости. Теперь, согласно протоколу RCA, после того как оценщик

(VE) определил окончательный набор аналогов (обычно 6–12 объявлений о продаже), ему необходимо проанализировать каждое объявление и разбить рассчитанный остаток на наиболее вероятные характеристики (латентные переменные), которые, по его мнению, влияют на стоимость, игнорируя возможность случайных ошибок в данных. Например, для данного объекта недвижимости у него могут быть следующие переменные разбивки:

1. Дизайн
2. Функциональность
3. Продажа по цене выше/ниже рыночной
4. Состояние
5. Качество

Различные объекты недвижимости, вероятно, будут иметь несколько разных переменных, характеризующих остаточную стоимость, которые необходимо собрать в один упорядоченный набор.

Оценка вклада редких признаков. Некоторые значимые признаки встречаются слишком редко для регрессионного канала: автомастерская, присутствующая в 3 из 780 сделок, не может обеспечить обоснованный коэффициент, и при прямом проходе такая переменная обычно полностью игнорируется. Это остаточные признаки, и их вклад можно оценить в остаточном пространстве: выполните подгонку модели без данного признака, а затем проведите регрессию остатков модели на исключенный столбец (столбцы) методом обычных наименьших квадратов. Для бинарного флага коэффициент представляет собой единовременный вклад (что эквивалентно среднему смещению остатков по сделкам с этим флагом); для числового столбца — вклад на единицу. Когда оцениваются совместно несколько коррелирующих удобств (например, площадь пристройки вместе с флагом «мастерская»), регрессия распределяет кредит между ними, а не учитывает одни и те же доллары дважды в двух строках разбивки. Оценки, основанные на менее чем десяти наблюдениях, носят лишь ориентировочный характер и должны быть подтверждены амортизированной стоимостью или расширенной выборкой (например, флагами, извлеченными из публичных комментариев); само ограничение по остаточной величине ограничивает любое значение, которое в конечном итоге будет присвоено. earthUI (v0.11+) реализует этот метод непосредственно в виде вкладки «Вклад остаточной величины», включая совместную оценку, данные об остаточной величине по каждой сделке и примечание к рабочему файлу.

Обозначения: Используя обозначения из приведенной выше таблицы, получаем:

$$\epsilon_i = \epsilon_{i,1} + \epsilon_{i,2} + \dots + \epsilon_{i,n_c} \quad (11)$$

где n_c — количество компонентов разбивки остатка.

Эксперт по оценке (VE) должен самостоятельно определить, как разбить остаток на указанные компоненты. Можно было бы подумать, что это может привести к появлению предвзятости в окончательном заключении о стоимости. Однако это не так, поскольку у нас есть математическое доказательство того, что если сумма корректировок по разбивке равна остаточной стоимости, они не повлияют на заключение о стоимости. Цель разбивки заключается не в установлении стоимости, а в объяснении того, почему остаточная стоимость объекта такова, какова она есть, по сравнению с остаточными стоимостями других объектов недвижимости, проданных по рыночной цене. Единственный момент, в котором субъективный подход инженера может повлиять на окончательный вывод о стоимости, — это оценка остатка стоимости объекта оценки, поскольку, разумеется, для него отсутствует фактическая цена продажи, необходимая для расчета остатка стоимости.

Каждому объекту недвижимости соответствует свой набор вкладов в остаточную стоимость, однако их необходимо объединить в один набор C для всех объектов недвижимости:

$$\text{Letting: } A_i = \{\{\epsilon_{i,c}\}_{c=1}^{n_c}\}$$

$$C = \{x \mid x \in \bigcup_{i=0}^{n_p} A_i\}$$

где $[\epsilon_{i,c}]$ — имя переменной, соответствующей значению $\epsilon_{i,c}$, а A_i — множество таких переменных, и n_c — количество значений в каждом ϵ_i .

Таким образом, если данные по объекту оценки и аналогичные данные по характеристикам расположены в отдельных столбцах, то под всеми вкладами значений переменных регрессионной модели для каждой характеристики мы должны увидеть множество C из n_c соответствующих вкладов значений остатков.

9 Основные доказательства метода RCA

Пусть:

$$n_p = \# \text{ of properties}$$

$$n_r = \# \text{ of regression variables}$$

$$n_c = \# \text{ of residual component variables}$$

Доказательство I: Каждая аналог скорректированная цена продажи равна расчетной цене продажи рассматриваемого объекта.

Для любого аналога i скорректированная цена продажи α_i представляет собой сумму чистой цены продажи π_i и следующих корректировок: разницы между вкладом аналога и объекта оценки в базовую стоимость, регрессионных переменных $j = 1, 2, \dots, n_r$ и остатка. Для объекта оценки, по которому отсутствуют данные о фактической продаже, чистая цена продажи π_0 определяется как итоговая оценка стоимости $\beta + \varphi_{0,1} + \varphi_{0,2} + \dots + \varphi_{0,n_r} + \epsilon_0$. Тогда:

$$\begin{aligned} \alpha_i &= \pi_i + (\beta - \beta) + (\varphi_{0,1} - \varphi_{i,1}) \\ &\quad + (\varphi_{0,2} - \varphi_{i,2}) + \dots + \\ &\quad (\varphi_{0,n_r} - \varphi_{i,n_r}) \\ &\quad + (\epsilon_0 - \epsilon_i) \\ &= \pi_i + (\beta + \varphi_{0,1} + \varphi_{0,2} + \dots \\ &\quad + \varphi_{0,n_r} + \epsilon_0) \\ &\quad - (\beta + \varphi_{i,1} + \varphi_{i,2} + \dots + \\ &\quad \varphi_{i,n_r} + \epsilon_i) \\ &= \pi_i + \pi_0 - \pi_i \\ &= \pi_0 \end{aligned} \tag{12}$$

Доказательство II: Изменение структуры остаточной ошибки при условии, что сумма вкладов всех компонентов остаточной ошибки равна аналогу остаточной ошибки, не влияет на окончательный вывод протокола RCA относительно значения.

Это почти тривиально. Однако, не задумываясь о математике, об этом легко упустить из виду, поскольку традиционные оценщики сосредотачиваются на корректировках, а не на вкладе в стоимость. Именно вклад в стоимость в данном случае компенсирует друг друга — это упущение, существовавшее на протяжении десятилетий. Для некоторого аналога объекта k , $k > 0$, общая корректировка остатков равна сумме остаточных компонентов разбивки n_c , которая, в свою очередь, равна сумме разниц в вкладах в стоимость, равной общей корректировке остатков объекта оценки минус общая корректировка остатков аналога объекта. И это не зависит от количества созданных остаточных компонентов или от того, что они представляют. До тех пор, пока значение остаточных компонентов равно общему остатку как для объекта оценки, так и для выбранного аналога, ξ_k не изменится.

$$\begin{aligned}
 \xi_k &= \xi_{k,1} + \xi_{k,2} + \dots + \xi_{k,n_c} \\
 &= (\epsilon_{0,1} - \epsilon_{k,1}) + (\epsilon_{0,2} - \epsilon_{k,2}) + \dots + \\
 &\quad (\epsilon_{0,n_c} - \epsilon_{k,n_c}) \\
 &= (\epsilon_{0,1} + \epsilon_{0,2} + \dots + \epsilon_{0,n_c}) - \\
 &\quad (\epsilon_{k,1} + \epsilon_{k,2} + \dots + \epsilon_{k,n_c}) \\
 &= \epsilon_0 - \epsilon_k
 \end{aligned} \tag{13}$$

10 Модель MARS «

»

10.1 Функции «g» для моделирования переменных и взаимодействий

Одним из основных принципов метода RCA является то, что «объяснительное рассуждение» имеет первостепенное значение. К сожалению, построение содержательных графиков для различных уравнений вклада значений, особенно при использовании взаимодействий и факторных переменных, может представлять собой сложную задачу. Поэтому считается крайне важным разработать четкий и строгий протокол для создания этих графиков. Данный протокол приводится ниже, однако следует понимать, что он предназначен для разработчиков, пишущих код на языках R или Python для генерации графиков, и может представлять лишь поверхностный интерес для тех, кто не занимается программированием.

Ценовая модель, сгенерированная системой MARS, представляет собой сумму базовой константы и дополнительных слагаемых, каждое из которых включает от 1 до 3 переменных. Для её анализа мы сначала разбиваем модель на отдельные слагаемые. Затем несколько слагаемых, имеющих общий набор переменных и представляющих одно взаимодействие, объединяются в отдельные выражения взаимодействия. Предположим, что максимальная степень регрессии равна 3. Рассмотрим модель регрессии третьей степени, состоящую из базовой константы, n_1 выражений первой степени, n_2 выражений взаимодействия второй степени и n_3 выражений взаимодействия третьей степени. В результате мы получаем четыре отдельные группы выражений, представляющие собой либо однопеременные, либо многопеременные взаимодействия, каждое из которых отображается на графике по-разному в зависимости от двух факторов: (1) включают ли они факторную переменную (категориальную переменную), и (2) количество уникальных переменных, которые они содержат. В частности:

- Выражения первой степени (1 переменная) отображаются в виде двумерных графиков.
- Выражения второй степени (2 переменные) отображаются в виде трехмерных графиков.
- Выражения третьей степени (3 переменные) отображаются в виде набора трехмерных графиков.
- Если выражение содержит факторную переменную, каждая уникальная факторная переменная уменьшает размерность графика на единицу (например, трехмерный график становится двумерным, если одна из переменных является факторной).

Индексация каждой функции g предоставляет информацию, необходимую для определения способа построения её графика, отражая количество уникальных переменных и наличие факторных переменных.

10.2 Индексирование функций g

Каждая функция g индексируется как $f_{j,k}^j g_k^j$, где:

- $f_{j,k}$ — первый индекс (верхний индекс перед g):
 - j принимает значения от 0 до 3 и определяет группу:
 - * $j = 0$: базовая константа (исключение из обычной системы индексации).

- * $j = 1$: выражения первой степени.
- * $j = 2$: выражения второй степени.
- * $j = 3$: выражения третьей степени.
- k — индекс конкретной функции внутри её группы (например, $k = 1$ для первой функции в группе j).
- Значение $f_{j,k}$ указывает на количество уникальных факторных переменных в функции:
 - * $f_{j,k} = 0$: факторных переменных нет.
 - * $f_{j,k} > 0$: количество факторных переменных (например, 1, 2 или 3).
- Индексы в конце g :
 - Верхний индекс j после g (например, g^j) повторяет номер группы для ясности.
 - Индекс k после g (например, g_k) соответствует k в $f_{j,k}$, идентифицируя функцию внутри её группы.

Хотя индексы f и g совпадают по значению, их роли различаются: $f_{j,k}$ подсчитывает факторные переменные, что имеет решающее значение для принятия решений при построении графиков, в то время как j и k в g определяют её группу и положение.

10.3 Размеры графика

Размер d графика для функции $f_{j,k} g_k^j$ вычисляется следующим образом:

$$d = j - f_{j,k} \quad (14)$$

где:

- j — количество уникальных переменных (или номер группы, без учета случая с константой).
- $f_{j,k}$ — количество факторных переменных.
- d — количество непрерывных (нефакторных) переменных, которые будут отображаться на графике, при этом максимальное количество составляет 3 (поскольку максимальная степень равна 3). Функция sd непрерывными переменными строится в $d + 1$ измерениях: $d = 1$ даёт двумерный график, $d = 2$ — трёхмерный график (или двумерную тепловую карту), $d = 3$ — набор трёхмерных графиков.

Например:

- Если $j = 3$ (3 переменные) и $f_{j,k} = 0$ (без факторных переменных), то $d = 3 - 0 = 3$, что требует набора 3D-графиков, каждый из которых фиксирует одну из переменных.
- Если $j = 3$ и $f_{j,k} = 1$ (одна факторная переменная), то $d = 3 - 1 = 2$, что требует 3D-графика (или 2D-тепловой карты) для каждого значения факторной переменной, фигурирующей в модели.
- Если $j = 0$ (константа), построение графика, как правило, не требуется, так как это одно значение.

Когда $d = 3$ (например, выражение третьей степени без факторных переменных), для визуализации всей функции может потребоваться несколько 3D-графиков, каждый из которых фиксирует одну переменную на разных значениях, чтобы показать подмножество поведения функции.

10.4 Окончательное уравнение модели

Модель MARS разбита на четыре группы выражений (постоянные, выражения первой, второй и третьей степени). Индексация $f_{j,k}^j g_k^j$ указывает:

- j : группу и количество переменных.
- $f_{j,k}$: количество факторных переменных.
- $d = j - f_{j,k}$: размерность графика.

Эта структура определяет, как визуализируется каждая функция g , обеспечивая соответствие метода построения графика сложности выражения и типам переменных.

Оценка цены продажи $\hat{P}$, полученная с помощью модели MARS, может быть выражена следующим образом:

$$\begin{aligned} \hat{P} = & f_{0,1} g_1^0 + \sum_{q=1}^{n_1} f_{1,q} g_q^1(x_*) \\ & + \sum_{r=1}^{n_2} f_{2,r} g_r^2(x_*, y_*) \\ & + \sum_{s=1}^{n_3} f_{3,s} g_s^3(x_*, y_*, z_*) \end{aligned} \quad (15)$$

$n_1 = \#$ of first degree terms

$n_2 = \#$ of second degree terms

$n_3 = \#$ of third degree terms

10.5 Количество диаграмм, необходимых для построения графика модели

Сколько графиков потребуется для визуализации всех функций в модели? Ответ зависит от переменных, их взаимодействий, а также от того, являются ли они факторными или нефакторными переменными.

VE может ограничивать переменные и взаимодействия в конфигурационном файле Excel для R/earth, влияя на то, какие взаимодействия допускаются. Практичный подход заключается в том, чтобы позволить R/earth оценить все возможные взаимодействия и выбрать те из них, которые вносят значительный вклад в оценку цены продажи, одновременно проверяя наличие таких проблем, как переобучение или коллинеарность. Некоторые взаимодействия могут быть запрещены, что повлияет на взаимодействия более высокой степени (например, если пара исключена из взаимодействия третьей степени, исключается вся комбинация).

В случае факторных переменных (например, «Зона MLS» с 12 значениями, взаимодействующей с «GLA») построение графиков становится более сложным. Для этого может потребоваться 12 графиков — по одному на каждую зону. Если другой фактор, такой как «Архитектурный стиль» (6 значений), взаимодействует с «Зоной MLS» и «GLA», может потребоваться до $12 \cdot 6 = 60$ графиков. Однако R/earth редко признает все комбинации значимыми; как правило, построение графиков оправдано только для нескольких ключевых взаимодействий (например, определённых зон и стилей), выявленных моделью.

Вот упрощённое руководство по количеству необходимых графиков:

1. Функции одной переменной (без факторов): 1 график.
2. Функции одной переменной (с факторами): 1 гистограмма, объединяющая все значения факторов.

3. Базовая функция: график не требуется.
4. Взаимодействия второй степени (без факторов): 1 график (например, двумерная тепловая карта с цветовой кодировкой цены продажи).
5. Взаимодействия второй степени (1 фактор): по 1 графику на каждое значение фактора в модели (например, 2 значения = 2 графика).
6. Взаимодействия второй степени (2 фактора): 1 график с матрицей комбинаций факторов (например, 3 области $\times$ 2 стиля = 6 ячеек).
7. Взаимодействия 3-й степени (без факторов): несколько графиков (например, 4 графика для количества ванн = 1, 2, 3, 4).
8. Взаимодействия 3-й степени (1 фактор): по одному графику на каждое значение фактора в модели.
9. Взаимодействия третьей степени (2 фактора): по 1 графику на каждую комбинацию значений факторов в модели.
10. Взаимодействия третьей степени (3 фактора): по 1 графику на каждое значение фактора с наименьшим количеством различных значений, с использованием цветовой матрицы для двух остальных.

Для более сложных сценариев могут потребоваться передовые методы построения графиков. Существуют и другие варианты с, вероятно, более сложными графиками.

Пример полной модели, сгенерированной с помощью MARS, можно найти по адресу [Figure 1](#).

11 Характеристики данных

В следующем разделе рассматривается система MARS. Система MARS полностью зависит от получаемых данных, и большая часть этих данных поступает из какой-либо службы MLS (Multiple Listing Service) или от налоговых оценщиков. Однако данные могут проходить через других поставщиков, которые занимаются их обработкой и улучшением, а затем продают с определенной наценкой. В некоторых регионах система достаточно надежна, в других — не столь надежна. В критических случаях инспектор по недвижимости вынужден полагаться на собственные наблюдения и измерения, точность которых может быть поставлена под сомнение.

Инженер по оценке должен понимать, что отражают данные и насколько точно они отражают то, что должны отражать. Это влечет за собой целый ряд вопросов:

- Как определить, что переменная является измерением?
- Если переменная не является измерением, то чем же она может быть?
- Какую пользу может принести регрессионный анализ, если получаемые им данные неточны?

Большая часть данных, которые мы получаем из списков MLS и от налоговых органов для проведения регрессионного анализа, представляет собой числовые данные, отражающие измерения или подсчеты. Часть этих данных имеет логический характер — «Истина» или «Ложь».

11.1 Цель программы «Residual»

Остаток служит косвенным показателем добавленной стоимости, создаваемой переменными, не включенными в регрессионную модель. На языке гедонической литературы эти неизмеренные характеристики называются латентными переменными: переменными, которые влияют на цену продажи, но не измеряются напрямую, в большинстве случаев из-за отсутствия практической возможности объективно их измерить. Их значения приходится выводить косвенно, и на практике значение латентной переменной часто оценивается на основе субъективного суждения — в методе RCA это суждение опирается на ранжирование остаточных значений. Ранжируя объекты недвижимости по их остаточным значениям или остаточным значениям на квадратный фут, мы получаем объективный рейтинг от объектов с наименьшей привлекательностью до объектов с наибольшей. Этот метод ранжирования эффективен при условии, что модель MARS (многомерные адаптивные регрессионные сплайны) разрабатывается квалифицированным аналитиком, глубоко разбирающимся в местном рынке. Ниже приведены некоторые дополнительные требования:

- Качество данных и модели: Правильно построенная регрессионная модель и высококачественные данные являются необходимыми условиями для получения достоверных остатков. В районе залива Сан-Франциско моя цель состоит в разработке модели с коэффициентом R^2 примерно 75–80% и коэффициентом R^2 перекрёстной проверки (CV R^2) 55–70%.
- Избегайте перегонки, ограничивайте R^2 : Инженер по оценке должен стремиться к максимальному значению R^2 около 80 % на большинстве зрелых рынков в районах с высокими ценами, таких как район залива Сан-Франциско, поскольку большая часть конкурентной стоимости жилой недвижимости заключается в ее субъективной привлекательности,

которая не поддается количественному измерению и, следовательно, не может быть напрямую включена в регрессионные модели. Значение R^2 , значительно превышающее 80 %, может указывать на переобученную модель, что требует дополнительного тщательного анализа. Однако следует учесть: существуют рыночные зоны со значениями R^2 выше 90 % без переобучения, а также районы, которые сложно оценить, где максимальное достижимое значение может составлять 50–60 % или ниже.

- Высокие значения R^2 в определенных районах: Однако существуют районы, такие как жилые микрорайоны или другие однородные жилые зоны, где более высокие значения R^2 встречаются на удивление часто из-за однородности объектов недвижимости.

11.2 Оценки состояния и качества в качестве приблизительных регрессоров

Шкалы оценки состояния и качества Fannie Mae (C1–C6 и Q1–Q6) заслуживают особого внимания. Эти шкалы слишком грубы, чтобы рассматривать их как подлинные измеряемые характеристики, а поскольку они присваиваются оценщиком, на практике они представляют собой субъективные оценки — латентные переменные сами по себе. Целесообразно явно различать два уровня: грубую оценку оценщика, используемую в качестве регрессора — «состояние-грубое» (condition-1) и «качество-грубое» (quality-1), с одной стороны, и мелкозернистые латентные переменные, восстановленные из остатка: «состояние-мелкозернистое» (состояние-2) и «качество-мелкозернистое» (качество-2), — с другой. VE может принимать грубозернистые оценки оценщика в качестве приближений и проводить регрессию по ним, получая приблизительные вклады в стоимость, которые полезны в качестве начальных значений для тонкозернистых латентных переменных. В этом случае протокол заключается в том, чтобы вычесть оцененные с помощью регрессии вклады в стоимость этих двух крупнозернистых переменных из оценки модели — удалив их из измеренных характеристик в разделе регрессии таблицы продаж — и добавить их обратно в остаток. Затем они удаляются из остатка и присваиваются соответствующим латентным переменным с тонкой детализацией: `res_condition1` и `res_quality1`. Дополнительная информация, извлеченная из публичных комментариев MLS (`public_remarks`) и связанных списков строк, затем используется для микроуточнения значений этих латентных переменных. В результате вклады значений с высокой степенью детализации заменяют вклады значений с низкой степенью детализации: грубые оценки служат лишь основой для приближенной модели, а значения, в конечном итоге представленные для параметров «состояние» и «качество», являются значениями с высокой степенью детализации, полученными на основе остатков.

11.3 Ошибки в данных

Судя по комментариям в социальных сетях, данные Калифорнийской системы совместного листинга (MLS) в целом отличаются более высоким качеством, чем данные по многим другим регионам США. Как правило, можно ожидать, что данные MLS будут хорошего качества в большинстве развитых мегаполисов, но более низкого качества — в сельских районах.

Хотя в данных по разным причинам иногда встречаются ошибки, при наличии достаточного объема данных эти ошибки, как правило, должны взаимно компенсироваться и оказывать минимальное влияние на средние значения, используемые в ценовых моделях.

Однако в отдельных районах могут возникать систематические искажения. Например, рассмотрим район, где многие дома изначально строились с полноценным вторым этажом, но некоторым покупателям предоставлялась возможность убрать комнаты на втором этаже, чтобы создать сводчатые потолки в одной или нескольких комнатах первого этажа, что

значительно уменьшало общую жилую площадь (GLA). По тем или иным причинам эти изменения впоследствии не были отражены в документации оценщика. В результате в системе MLS может указываться площадь в 2 400 квадратных футов, тогда как фактическая измеренная площадь составляет всего 2 000 квадратных футов. В таких случаях данные MLS следует исправить перед выполнением регрессии MARS (многомерные адаптивные регрессионные сплайны). Важно отметить, что получение этой информации по аналогам может потребовать определенных исследовательских усилий. Несоответствия следует отметить для использования в будущем.

Учитывая вышесказанное, можно обоснованно предположить, что в районах с высококачественной системой MLS погрешности, проникающие в остаток, минимальны по сравнению с влиянием субъективных и неизмеримых характеристик, таких как планировка, состояние и качество. Это является важным фактором при оценке точности метода RCA. Например, если значение коэффициента CQA (R^2) для данной рыночной зоны составляет 80%, то примерно 20% цены приходится на остаток CQA. Погрешность $\pm 10\%$ при отнесении объекта оценки к ранжированию по остатку, следовательно, приводит к отклонению в размере $\pm 10\% \cdot 20\% = \pm 2\%$ error in the final value conclusion. This is how you improve accuracy! In many cases, accurate value estimates well below $\pm 1\%$.

Поэтому рассмотрение остатков в регрессии MARS в качестве косвенных показателей совокупной стоимости немоделируемых переменных является обоснованным при условии, что специалист по оценке обладает достаточным опытом работы на местном рынке, чтобы определить, где и почему возникают системные погрешности измерения. Такой подход гарантирует, что анализ точно отражает фактическую стоимость объектов недвижимости, учитывая ограничения и нюансы данных по недвижимости.

Пример из практики: измерение неизмеренных факторов в Си-Ранч

Когда достижимый уровень « R^2 » в рыночной зоне оказывается ниже того, чего можно было бы ожидать, исходя из кажущейся однородности рынка, данная концепция предполагает выявление и измерение недостающей переменной, а не принятие этого ограничения. Сообщество «Sea Ranch» в округе Сонома наглядно иллюстрирует этот подход.

Конкретный пример реакции «оценивай, но не соглашайся». Сообщество «Sea Ranch» в округе Сонома — это прибрежный жилой комплекс, где расстояние до океана является основным фактором, определяющим стоимость: дом, расположенный в нескольких сот футов от утеса, стоит значительно дороже, чем дом, расположенный на четверть мили вглубь суши, при прочих равных условиях. В данных MLS по этому жилому комплексу расстояние до океана не учитывается в качестве переменной. Первоначальная модель MARS, проведенная на доступных переменных MLS, продемонстрировала более низкую объясняющую способность, чем следовало бы ожидать, учитывая кажущуюся однородность рынка, при этом структура остатков не позволяла достаточно четко упорядочить объекты недвижимости по каким-либо видимым в данных характеристикам — это было сигналом того, что в модели отсутствовал важный фактор, а не того, что рынок действительно был «шумным».

Решением стала прямая съемка. Автор получил карту района, на которой было показано расположение объектов недвижимости относительно береговой линии, и измерил расстояние до океана для нескольких сотен домов. Полученные значения расстояний были добавлены в качестве переменной в набор данных MLS, после чего регрессионный анализ MARS был проведен заново. Новая переменная была признана моделью значимой, объясняющая способность существенно возросла, а ранжирование остатков впоследствии упорядочило объекты недвижимости по признакам, соответствующим остальным субъективным характеристикам — виду, состоянию, качеству — а не по скрытому сигналу расстояния до океана, который

ранее искажал результаты. Вопрос высоты над уровнем моря был решен аналогичным образом в районах, расположенных на склонах холмов, где переменная MLS отсутствует или является недостоверной.

Методологический принцип носит общий характер. Когда достижимый показатель « R^2 » опускается ниже уровня, который должен обеспечиваться видимой однородностью рынка, первая гипотеза заключается в том, что в модели отсутствует важная переменная, и надлежащим ответом является выявление этой переменной и сбор данных непосредственно на местах. Источниками информации служат карты населенных пунктов, данные ГИС, отчеты налоговых инспекторов, отчеты строительных ведомств, спутниковые снимки и результаты выездных осмотров. Затраты на такие измерительные работы реальны, но ограничены; как только переменная собрана для данной рыночной зоны, её можно использовать в будущих заданиях в этой зоне. RCA легко интегрирует новые переменные, поскольку данная методология не зависит от того, какие именно переменные входят в регрессию MARS — структура остатков указывает VE, отражает ли новая переменная реальный сигнал, и вклады переменных подвергаются соответствующим корректировкам.

Альтернативный подход — принятие низкого коэффициента « R^2 » и подготовка отчёта с более широкими диапазонами погрешности — имеет обоснованность в тех случаях, когда отсутствующие факторы действительно невозможно измерить (индивидуальные предпочтения покупателей, кратковременные колебания настроений на рынке, специфические обстоятельства сделок), но не в тех случаях, когда они просто не были учтены. Первым шагом эксперта по оценке (VE) при работе со слабой моделью должно быть установление того, относится ли отсутствующий сигнал к последнему случаю.

11.4 Что такое точность?

Для обеспечения абсолютной точности оценки необходимо:

1. Соответствие рыночной стоимости: гипотетическая сделка купли-продажи объекта оценки на 100% соответствует приведенному определению рыночной стоимости
2. Разработка точных и надежных моделей: Разработана высококачественная модель MARS с коэффициентом R^2 около 75–80 % и коэффициентом CVR^2 примерно 55–70 % в зависимости от местоположения.
3. Качественный анализ ошибок данных: все вероятные ошибки данных по измеряемым переменным проанализированы и устранены, при этом их влияние на остатки минимально.
4. Учет системных отклонений: Любые системные отклонения в данных должны быть учтены и скорректированы.
5. Анализ рангов остатков: Убедитесь, что свойства рангов остатков следуют разумной последовательности характеристик от самого низкого до самого высокого ранга. Аномалии, такие как продажи из наследственной массы, продажи с недостаточной площадью и аукционные продажи, должны быть изучены, объяснены и, возможно, исключены.
6. Определение места объекта оценки: на основе глубокого понимания ранжирования остаточных значений объект недвижимости должен быть точно позиционирован. Необходимо выявить недвижимость с значительно более высокой привлекательностью и другую — со значительно более низкой, а рассматриваемая недвижимость

разместить между ними и присвоить ей балл. Затем следует постараться максимально сузить разрыв в рейтинге между объектами недвижимости с более высокой и более низкой привлекательностью.

7. Обоснование: Никогда не жертвуйте обоснованием ради точности. Оценщик должен быть способен объяснить различия в стоимости между объектом оценки и аналогами, по крайней мере, человеку с разумным уровнем интеллекта. Например, хорошую модель MARS с оптимальными показателями R^2 и CVR^2 можно использовать для объяснения вклада в стоимость измеряемых характеристик, в то время как разбивка остатков, выполненная инженером по оценке, может служить для объяснения вклада в стоимость других характеристик.
8. Переобучение: Следует избегать переобучения моделей, не используя характеристики объектов недвижимости, которые могут идентифицировать — как по отдельности, так и в сочетании — один или несколько конкретных объектов. В частности, следует избегать использования факторных переменных, значения которых могут указывать на небольшое количество конкретных объектов. Факторная переменная, идентифицирующая районы в значительных наборах данных, является уместной, но та, которая идентифицирует объекты недвижимости с бассейнами в районе, где таких бассейнов очень мало, должна использоваться с осторожностью. Использование характеристик, идентифицирующих уникальные объекты недвижимости, таких как точные ГИС-координаты, номера земельных участков или адреса, категорически запрещено.

Если коэффициент детерминации R^2 составляет 80 %, то можно ожидать, что примерно 20 % значения объекта оценки обусловлено остатком. Однако это лишь приблизительная оценка. Из кривой «CQA-остаток» известно, что большая часть распределения остатков приходится на нижние и верхние крайние значения CQA. Исходя из кривой «CQA-остаток» и крайних значений наклона для низких и высоких значений CQA, можно ожидать, что оценка остатка в этих областях будет менее точной. Однако это компенсируется относительной редкостью объектов недвижимости в этих зонах и более значительной разницей в привлекательности, которую мы наблюдаем между соседними объектами недвижимости. Таким образом, хотя мы и можем говорить о влиянии диапазона значений $\pm 5\%$ в средней зоне, в крайних зонах нас будет больше интересовать несколько более узкий диапазон значений $\pm 2 - \pm 3\%$

Расчет точности RCA, как правило, представляет собой некоторую сложность из-за нелинейности кривой «CQA-остаточная величина». Однако в каждом конкретном случае это возможно, при условии наличия адекватного ранжирования баллов по остаткам. Диапазон баллов CQA составляет от 0,0 до 10,0. При наличии качественной модели MARS ваш рейтинг остаточных значений должен демонстрировать чёткую последовательность от наименьшей до наибольшей привлекательности, и вам не должно составить труда оценить балл CQA с точностью до $\pm 5\%$ от общего диапазона или ± 0.5 . Если в регрессии участвуют 100 объектов недвижимости, это означает, что рассматриваемый объект оценки будет точно отнесён к одной из групп по 10 объектов недвижимости. Итак, обращаясь к данным о продажах в Берлингейме, приведенным в таблице 3, рассчитаем приблизительную точность (или погрешность) для CQA, оцененного VE, равного 7,5, 5,1 и 1,9, где его заключение о стоимости объекта оценки составляет \$ 3 000 000. Обратите внимание, что, поскольку в таблице 3 представлена выборка из нескольких сотен объектов недвижимости, вам потребуется интерполировать остатки там, где это необходимо (при условии линейности). Для ясности: мы предполагаем, что в каждом случае VE с высокой степенью уверенности утверждает, что его рейтинг объекта оценки находится в диапазоне от ± 5 объектов недвижимости в рейтинге остатков, включающем всего 100 объектов недвижимости.

CQA= 7,5±0.5				
CQA	Остаток	Абсолютная погрешность	Погрешность эксперимента	Гипотетическая п
8,00	\$ 257 112			
7,50	\$ 197 113	\$ 59 999		
7,00	\$ 154 615	\$ 42 498	±\$51249	\$ 3 000

CQA=5,1±0.5				
CQA	Остаточная ошибка	Абсолютная разница	Статистическая погрешность	Гипотетич
5,60	\$ 50 911			
5,10	\$ 19 559	\$ 31 352		
4,60	-\$ 28 207	\$ 47 766	±\$39559	\$

CQA= 1,9±0.5				
CQA	Остаток	Абсолютная разница	Ошибка эксперимента	Гипотетическая цена прод
2,40	-\$ 228 411			
1,90	-\$ 276 381	\$ 47 970		
1,40	-\$ 326 938	\$ 50 557	±\$49264	\$ 3 000 000

Наконец, единственный объективный способ проверки точности заключается в разработке строгого протокола (о котором в данной статье говорится лишь отчасти и в общих чертах) и контроле его соблюдения. Это необходимо, поскольку фактические цены продажи могут отклоняться от рыночной стоимости из-за несовершенств рынка. Сравнение оценок и последующих продаж может дать некоторое представление о точности. Однако сложность и нестабильность рынка недвижимости требуют привлечения экспертов-оценщиков или высококвалифицированных инженеров-оценщиков для проведения оценок стоимости на основе установленных стандартов и протоколов. Эти стандарты и протоколы должны быть достаточно строгими, чтобы гарантировать, что компетентные инженеры-оценщики независимо друг от друга придут к практически одинаковым выводам относительно стоимости. В большинстве случаев, когда имеются качественные данные, эта цель должна быть достижимой.

11.5 Рассматриваемый объект оценки как аномалия

Поскольку для оценки рыночной стоимости объекта недвижимости требуется анализ недавних сделок купли-продажи, вполне уместно задаться вопросом: насколько точно недавние сделки отражают различные особенности старого объекта оценки, не прошедшего модернизацию и ремонт. Это особенно актуально для района залива Сан-Франциско, где недавние сделки купли-продажи, как правило, касаются домов, которые были обновлены с целью повышения их продажной цены. Таким образом, среднее качество и состояние аналогов, как правило, превосходят более старые дома, которые довольно долгое время не выставлялись на продажу. Эта разница в стоимости должна быть учтена при оценке остатка объекта оценки, или, другими словами, его балла CQA. Это важный фактор, который необходимо учитывать с точки зрения качества и точности данных при оценке объекта оценки.

11.6 Важность CVR2 при работе с непредвиденными данными

Регрессия MARS основана на многократных итерациях алгоритма R/earth, применяемого к случайно выбранному подмножеству объектов недвижимости, которое, как правило, составляет от 80 % до 90 % набора данных MLS, называемого обучающим набором. Остальные 10

% – 20 % объектов недвижимости служат в качестве тестового набора. На этом этапе на основе обучающего набора данных создаётся модель MARS, которая затем используется для прогнозирования цен продажи объектов недвижимости из тестового набора данных; полученные результаты сравниваются с фактическими ценами продажи. Данная методология признана специалистами по анализу данных наиболее эффективной стратегией для разработки надёжной модели, способной прогнозировать цены продажи объектов недвижимости, данные о которых ещё не были собраны.

Однако основная ценность показателя CVR2 заключается не в том, что он отражает фактическую точность регрессионной модели в отношении реальных продаж в рыночной зоне, а в том, что он позволяет оптимизировать прогнозные возможности модели в отношении гипотетических продаж недвижимости на определенную дату оценки. Данные объекты недвижимости не включены в набор обучающих данных и могут демонстрировать новые или неожиданные комбинации характеристик, отсутствующих как в данных об объектах недвижимости из набора обучающих данных, так и в данных из набора тестовых данных.

Учитывая это, как мы можем оценить цену продажи объекта недвижимости, который не обновлялся в течение последних 15 лет, если предположить, что он был выставлен на продажу, например, за месяц до даты оценки? Модель, демонстрирующая высокий показатель CVR2, с большей вероятностью даст наиболее точную оценку, чем модель, не прошедшая такую многократную перекрёстную проверку. Также крайне важно, чтобы набор данных включал достаточное количество аналогичных объектов недвижимости с различной степенью обновления, ремонта и модернизации.

В конечном итоге, если у нас есть надёжная модель, которая точно прогнозирует вклад различных характеристик объектов недвижимости в нашем наборе данных по аналогичным объектам, мы сможем эффективно определить место нового объекта оценки в рейтинге этих объектов, составленном моделью на основе остатка. Такое определение места позволяет нам присвоить данному объекту соответствующий балл CQA и остаток.

12 Остатки

Как следует из названия «Метод остаточных ограничений», остатки занимают центральное место в этой методологии. Под остатком здесь понимается разница между чистой ценой продажи объекта недвижимости и ценой продажи, прогнозируемой с помощью регрессионного анализа.

Положительный остаток свидетельствует о том, что объект недвижимости был продан по цене, превышающей его оценочную стоимость, рассчитанную на основе типичных измеримых характеристик, таких как общая жилая площадь (GLA) и площадь участка, что позволяет предположить, что он, возможно, обладает более высокой привлекательностью, чем в среднем. Отрицательный остаток указывает на то, что объект недвижимости был продан по цене ниже оценочной, что, возможно, свидетельствует о его меньшей привлекательности или наличии других негативных характеристик, не учтенных в оценке.

Учитывая, что у более просторных домов, скорее всего, остатки выше, для более обоснованного сравнения мы могли бы использовать показатель остатков на квадратный фут, рассчитанный на основе общей арендуемой площади (GLA). Более подробное рассмотрение этого вопроса будет посвящено другой статье.

12.1 Рейтинг & Балл

После построения модели MARS (многомерные адаптивные регрессионные сплайны) для измеряемых переменных для каждого объекта недвижимости рассчитываются остатки. Эти остатки представляют собой разницу между наблюдаемыми чистыми ценами продажи и ценами продажи, прогнозируемыми моделью. Следует создать дополнительный столбец для показателя «Остаток/SF», то есть остатка, деленного на жилую площадь объекта недвижимости, исходя из того, что, как правило, остаток для конкретных объектов недвижимости тем значительнее, чем больше жилая площадь. В некоторых районах могут лучше работать другие статистические показатели, но это тема для другого разговора.

Assumption 7: “Ranking by Residual”

В данной статье под «ранжированием по остаточной стоимости» следует понимать ранжирование по остаточной стоимости, по остаточной стоимости, деленной на SF, либо ранжирование по какой-либо другой функции остатков, которую инженер по оценке сочтет наиболее подходящей для рыночной зоны. Опыт автора в районе залива Сан-Франциско показывает, что показатель «остаточная стоимость/SF» обычно является наиболее надежным, особенно при оценке объекта оценки с необычно малой или большой общей жилой площадью (GLA).

Далее объекты недвижимости сортируются по величине их остатков — от самых отрицательных к самым положительным. Затем для каждого объекта рассчитывается балл «Состояние-Качество-Привлекательность» (CQA) в диапазоне от 0,0 до 10,0. Этот балл рассчитывается следующим образом:

Балл CQA соответствует процентильному рангу остаточной стоимости объекта недвижимости среди всех объектов недвижимости. В частности, балл рассчитывается путём определения процента объектов недвижимости, у которых остатки ниже, чем у данного объекта, и деления этого процента на 10. Например, если остаток объекта недвижимости выше, чем у 50 % объектов, то его балл составит 5,0.

Из-за округления одно свойство может получить балл 10,0, что соответствует собственности с наибольшим остатком. Однако идеальный балл 10,0 теоретически невозможен, поскольку,

по логике вещей, ни одна собственность не может иметь остаток, превышающий остаток 100%-х собственностей, включая само эту собственность. С другой стороны, минимально возможный балл 0,0 возможен, поскольку по крайней мере у одного объекта недвижимости нет других объектов с более низкими остатками. Обратите внимание, что из-за округления балл 0,0 может быть присвоен более чем одному объекту недвижимости, если сравнивается большое количество объектов.

Таким образом, балл CQA служит интуитивно понятным критерием того, насколько характеристики объекта недвижимости соответствуют ожиданиям модели, причем более высокие баллы свидетельствуют о лучшем соответствии прогнозам модели.

Важно признать, что балл CQA можно рассматривать лишь как косвенный индикатор привлекательности аналога недвижимости в том случае, если условия продажи соответствуют его рыночной стоимости. Инженер по оценке несет ответственность за установление того, включает ли остаточная стоимость элементы, завышающие или занижающие рыночную стоимость, путем предоставления исчерпывающего объяснения и тщательного расследования сделки купли-продажи. Хотя значительные продажи по ценам, значительно превышающим или занижающим рыночную стоимость, происходят нечасто, они могут иметь место в таких обстоятельствах, как вынужденная продажа, наследственное дело, недееспособность, сговор и т. д.

В таблице 3 представлен полуреалистичный выдержка из данных MLS вместе с остатками и баллами CQA — как в общем, так и в расчете на квадратный фут. Общее количество объектов недвижимости, проанализированных в рамках данного исследования, составило примерно 600. Следует отметить, что некоторые переменные, включенные в регрессионный анализ, в таблице не отображены. Другими словами, представленные остатки получены после того, как в рамках регрессии MARS было учтено влияние других переменных, таких как дата продажи.

Таблица 3: Аналоги продажи, отсортированные по остаткам

Цена продажи	Расчетная цена продажи	Остаток	CQA	Остаток/SF	CQA SF	GLA	Дата продажи
\$ 3 450 000	\$ 2 318 865	\$ 1 131 135	10	\$ 420,50	9,98	2 690	2021-04
\$ 4 360 757	\$ 3 399 922	\$ 960 835	9,9	\$ 290,28	9,82	3 310	2021-09
\$ 4 000 000	\$ 3 134 444	\$ 865 556	9,8	\$ 234,57	9,60	3 690	2020-11
\$ 5 850 000	\$ 5 150 966	\$ 699 034	9,7	\$ 109,91	8,35	6 360	2021-12
\$ 3 600 000	\$ 3,023 366	\$ 576 634	9,5	\$ 269,46	9,79	2 140	2022-06
\$ 4 425 000	\$ 3 987 873	\$ 437,127	9,2	\$ 127,44	8,68	3 430	2021-08
\$ 2 530 000	\$ 2 164 564	\$ 365 436	8,9	\$ 180,02	9,28	2 030	2019-12
\$ 4 250 000	\$ 3 884 719	\$ 365 281	8,8	\$ 87,18	7,91	4 190	2021-08
\$ 3 350 000	\$ 3 034 613	\$ 315 387	8,5	\$ 130,87	8,75	2 410	2021-09
\$ 3 800 000	\$ 3 522 028	\$ 277 972	8,2	\$ 77,43	7,72	3 590	2020-09
\$ 4 200 000	\$ 3,963 748	\$ 236 252	7,8	\$ 56,12	6,89	4 210	2021-01
\$ 2 900 000	\$ 2 702 887	\$ 197,113	7,5	\$ 89,19	7,92	2 210	2021-07
\$ 4 250 000	\$ 4 103 885	\$ 146 115	6,9	\$ 39,38	6,29	3 710	2022-08
\$ 5 000 000	\$ 4 864 958	\$ 135 042	6,8	\$ 28,25	5,85	4 780	2021-07
\$ 4 320 000	\$ 4 199 672	\$ 120 328	6,5	\$ 33,90	6,04	3 550	2022-07
\$ 2 600 000	\$ 2,479 904	\$ 120 096	6,5	\$ 60,96	7,00	1 970	2021-10
\$ 3 400 000	\$ 3 301 146	\$ 98 854	6,2	\$ 30,32	5,90	3 260	2021-06
\$ 2 000 000	\$ 1 901 987	\$ 98 013	6,2	\$ 73,14	7,54	1 340	2019-11
\$ 3 000 000	\$ 2 919 718	\$ 80 282	5,9	\$ 26,50	5,79	3 030	2020-05
\$ 6 750 000	\$ 6 699 089	\$ 50 911	5,6	\$ 7,75	5,20	6 570	2020-08
\$ 5 652 000	\$ 5 601,570	\$ 50 430	5,5	\$ 12,83	5,37	3 930	2022-05
\$ 2 600 000	\$ 2 579 082	\$ 20 918	5,2	\$ 10,51	5,33	1 990	2022-03
\$ 1 925 000	\$ 1 905 441	\$ 19 559	5,1	\$ 11,51	5,36	1 700	2021-01
\$ 3 200 000	\$ 3 209 500	-\$ 9 500	4,8	-\$ 3,21	4,88	2 960	2020-11
\$ 4 150 000	\$ 4,187 560	-\$ 37 560	4,5	-\$ 10,49	4,68	3 580	2021-04
\$ 2 700 000	\$ 2 776 510	-\$ 76 510	3,9	-\$ 23,61	4,14	3 240	2021-07
\$ 3 450 000	\$ 3 567 668	-\$ 117 668	3,4	-\$ 31,05	3,90	3 790	2020-09
\$ 1 975 000	\$ 2 109 255	-\$ 134 255	3,3	-\$ 55,48	3,24	2 420	2020-02
\$ 1 950 000	\$ 2 085 431	-\$ 135 431	3,2	-\$ 68,40	2,86	1 980	2021-02
\$ 3 695 000	\$ 3 859 154	-\$ 164 154	3,1	-\$ 47,04	3,56	3 490	2021-05
\$ 2 288 000	\$ 2 455 747	-\$ 167 747	2,9	-\$ 56,86	3,19	2 950	2020-06
\$ 2 635 000	\$ 2 802 781	-\$ 167 781	2,9	-\$ 56,87	3,18	2 950	2020-09
\$ 2 050 000	\$ 2 217 794	-\$ 167 794	2,9	-\$ 82,66	2,46	2 030	2021-07
\$ 1 550 000	\$ 1 731 804	-\$ 181 804	2,7	-\$ 130,79	1,32	1 390	2021-05
\$ 2 800 000	\$ 3 028 411	-\$ 228 411	2,4	-\$ 81,58	2,51	2 800	2021-02
\$ 3 700 000	\$ 3 976 381	-\$ 276 381	1,9	-\$ 76,99	2,62	3 590	2021-08
\$ 2 517 500	\$ 2 834 974	-\$ 317 474	1,5	-\$ 120,71	1,48	2 630	2019-11
\$ 2 550 000	\$ 2 905 330	-\$ 355 330	1,1	-\$ 115,74	1,64	3 070	2019-12
\$ 1 815 000	\$ 2 263 944	-\$ 448 944	0,7	-\$ 273,75	0,15	1 640	2021-11
\$ 3 711 000	\$ 4 683 454	-\$ 972 454	0	-\$ 226,68	0,37	4 290	2021-12
\$ 2 225 000	\$ 3 332 960	-\$ 1 107 960	0	-\$ 370,56	0,03	2 990	2021-08
\$ 3 550 000	\$ 4 746 489	-\$ 1 196 489	0	-\$ 274,42	0,14	4 360	2022-06

12.2 Кривая CQA-остаточная величина

График зависимости CQA от остаточной величины/SF для того или иного района неизменно приведет к построению характеристики CQA-остаточная величина, представленной по адресу [Figure 2](#). Данный конкретный график отражает данные по городу Берлингем (штат Калифорния) примерно за 2022 год. Аналогичные кривые, как правило, наблюдаются при построении графиков, где по оси абсцисс отложены баллы CQA, а по оси ординат — значения остаточной стоимости; однако я предпочитаю использовать показатель «остаточная стоимость на квадратный фут» (Residual/SF) из-за его значимости для контекстуализации показателя CQA с учётом общей жилой площади (GLA) объекта оценки. Умножив показатель «Остаточная стоимость/SF» на GLA, можно преобразовать показатель CQA объекта недвижимости в соответствующую остаточную стоимость.

Инженер-оценщик должен тщательно проанализировать, оказывает ли размер объекта недвижимости, а именно общая жилая площадь (GLA), существенное влияние на величину остатков. Эта оценка имеет решающее значение, поскольку от нее зависит выбор наиболее точного метода расчета стоимости недвижимости. Например, высококлассная кухонная техника и роскошная сантехника в ванных комнатах могут существенно повысить остаточную стоимость дома. Однако в районах, где размеры и планировка кухонь и ванных комнат, как правило, однородны, эти усовершенствования могут не иметь сильной корреляции с GLA. Их влияние на остаточную стоимость может зависеть не столько от занимаемой ими площади, сколько от качества, которое они представляют.

С другой стороны, дорогостоящие элементы экстерьера и материалы, которые заметно повышают привлекательность недвижимости с улицы, могут иметь более прямую связь с ее размером. В более просторных домах можно реализовать более сложные элементы экстерьера, такие как обширное благоустройство территории, высококачественные кровельные материалы или просторные патио, что естественным образом повышает общую привлекательность недвижимости и, как следствие, ее остаточную стоимость.

Поэтому специалистам по оценке крайне важно учитывать эту динамику при анализе остатков. Они должны не только оценивать внутреннюю стоимость, создаваемую высококачественной отделкой и дополнительными элементами, но и понимать, как эти улучшения взаимодействуют с физическими характеристиками объекта недвижимости. Такой комплексный подход гарантирует, что результаты оценки точно отражают как качественные, так и количественные характеристики объекта недвижимости.

Следовательно, рекомендуется построить графики как кривых «CQA-Residual», так и «CQA-Residual/SF» и проанализировать различия между ними, которые могут указывать на закономерности в конкретной рыночной зоне. Инженер по оценке может поручить своей программе рассчитать оба вида остатков и взять их средневзвешенное значение для расчета остатка, который будет присвоен объекту оценки. Это может иметь значительное влияние, а может и нет.

12.3 Характеристики кривой CQA

Кривая «Состояние-Качество-Привлекательность» (CQA) представляет собой сложную модель, позволяющую понять, как стоимость недвижимости колеблется в зависимости от привлекательности объекта, его состояния и динамики рынка в конкретном районе. Эта модель учитывает взаимодействие экономических и физических факторов, влияющих на рынок недвижимости, отражая все нюансы оценки недвижимости в различных сегментах рынка.

12.3.1 Анализ компонентов кривой CQA

1. Экспоненциальный спад в левой части кривой:

По мере снижения балла CQA с уровня около 2 до 0 объекты недвижимости демонстрируют характеристики быстрого износа. Эта часть кривой хорошо описывается функцией экспоненциального спада, что подчеркивает, насколько быстро объект недвижимости может утратить свою привлекательность и стоимость без надлежащего обслуживания и обновлений. Физическое изнашивание, такое как гниение дерева, выцветание красок и ржавление металла, ускоряет этот процесс, особенно в домах, которые не подвергаются регулярному техническому обслуживанию.

2. Линейный рост в среднем диапазоне:

Основная часть рынка, как правило, относится к этой категории, в которой привлекательность растёт почти линейно при оценках от 2 до 8 баллов. В этот раздел входят объекты недвижимости, которые, как правило, поддерживаются в надлежащем состоянии, но не обязательно отличаются исключительными качествами или выгодным расположением. Он представляет широкий средний сегмент жилья, охватывающий большую часть рынка жилья, где дома предлагаются по конкурентоспособным ценам с учетом их стандартных характеристик и общего состояния.

3. Экспоненциальный рост в правой части кривой:

Здесь кривая растёт экспоненциально в диапазоне баллов от 8 до 10, что отражает сегмент рынка с наиболее высокой привлекательностью. Объекты недвижимости в этом сегменте часто принадлежат состоятельным людям, которые не только вкладывают средства в безупречное содержание объекта, но и в улучшения, значительно повышающие его привлекательность и стоимость. Экспоненциальный рост отражает дефицит таких высококачественных объектов недвижимости и высокий спрос со стороны состоятельных покупателей. При этом следует понимать, что дом по более низкой цене может получить очень высокую оценку с точки зрения привлекательности, в то время как большой и дорогой дом может получить низкую оценку. Эта динамика выявляется благодаря грамотному использованию системы MARS и обычно остается незамеченной большинством оценщиков.

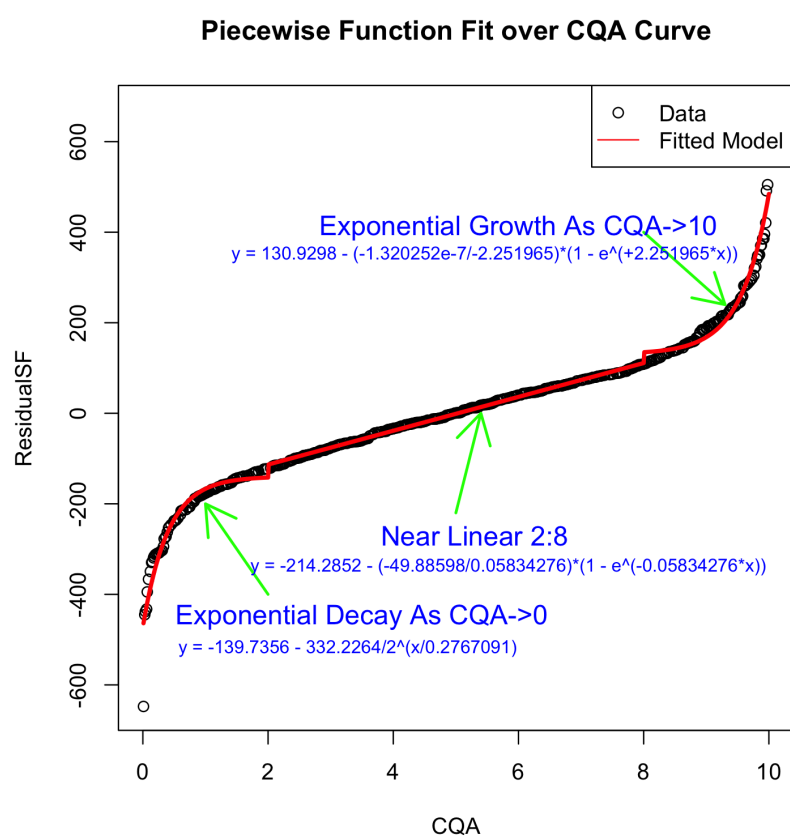
12.3.2 Динамика рынка и социальные последствия

Кривая CQA косвенно отражает ситуацию с относительным благосостоянием и экономическим неравенством в различных регионах. В таких местах, как округ Сан-Матео, разница в стоимости недвижимости между районами вроде Берлинггейма и более состоятельными районами, такими как Хиллсборо, Вудсайд или Атертон, свидетельствует о значительных социально-экономических различиях. Это неравенство сказывается на всех аспектах — от содержания недвижимости до рыночной стоимости — и, в конечном итоге, влияет на демографический состав покупателей и тенденции рынка недвижимости.

Кроме того, история об относительно состоятельном человеке, инвестирующем в высококачественный дом в районе среднего ценового сегмента, где ему комфортно и где он по разным причинам предпочел бы жить, подчеркивает еще один важный аспект экономики недвижимости — концепцию «избыточного строительства». Такие объекты недвижимости, несмотря на свою привлекательность, могут не найти местных покупателей, способных заплатить ту цену, которую они могли бы получить в более высокой рыночной зоне.

Кривая CQA-Остаток является характеристикой для каждой рыночной зоны, и её не следует упускать из виду при проведении анализа.

Рис. 2: CQA — остаточная характеристическая функция



12.4 Объект оценки «Остаточные величины»

В отличие от аналогов, рассматриваемый объект недвижимости не имеет объективно определенной остаточной стоимости; она должна быть оценена инженером-оценщиком (VE). Такая оценка вносит определенную долю субъективности в протокол применения Метода остаточных ограничений (RCA), что является критическим моментом, в котором личное суждение играет ключевую роль.

Для оценки бала «Состояние-Качество-Привлекательность» (CQA) эксперт по оценке (VE) тщательно анализирует объекты недвижимости, отсортированные по рейтингу, выявляя заметные закономерности в рейтинговой таблице. Этот анализ часто включает подробный просмотр фотографий из базы данных MLS, связанных с каждым объектом. Исчерпывающий анализ нецелесообразен при работе с большим набором данных, например, с 600 объектами. Вместо этого VE применяет практическое суждение, чтобы приблизительно определить вероятное положение объекта оценки в общем рейтинге. Объекты недвижимости, требующие капитального ремонта, или «объекты для ремонта», как правило, занимают места в нижней части рейтинга, тогда как объекты с превосходной отделкой и привлекательным внешним видом, как правило, занимают места в верхней части рейтинга.

Затем VE применяет более целенаправленный подход, «перемещаясь» в пределах рейтинга для выявления аналогов, которые максимально соответствуют привлекательности объекта оценки. По мере сужения круга поиска VE уделяет больше внимания сходствам между объектами недвижимости, особенно в таких аспектах, как качество, состояние, эстетика дизайна (включая такие элементы, как покраска, деревянные и каменные отделочные элементы, а также окна), функциональность и другие значимые характеристики.

Крайне важно понимать, что оценки VE в основном основываются на рейтингах, сформированных рынком. Это гарантирует, что субъективные элементы процесса опираются на рыночные реалии, что позволяет добиться справедливой и точной оценки объекта оценки.

12.5 Выводы по оценке RCA

12.5.1 Общие соображения

Определение показателя CQA имеет решающее значение для установления ориентировочной остаточной стоимости объекта недвижимости, поскольку эти факторы неразрывно связаны между собой. Хотя оценка объекта недвижимости возможна и без учета показателя CQA, этот показатель предоставляет ценную информацию, позволяя увидеть, как данный объект позиционируется по сравнению с другими объектами на рынке в рыночной зоне. Например, балл CQA, равный 3,2, указывает на то, что данный объект недвижимости обладает большей привлекательностью, чем 32 % аналогичных объектов в окрестностях. Если имеются возражения против присвоенного бала, стороны спора должны ознакомиться с рейтингами и обосновать любые предлагаемые корректировки на основе фотографических доказательств или других существенных данных. Если система ранжирования надежна, а инженер-оценщик (VE) точно определил место объекта недвижимости в этой системе, маловероятно, что возникнут значительные расхождения.

На этапе подготовки отчета, как правило, отбирается от 6 до 12 актуальных и релевантных аналогов для обоснования вывода об оценке. Проводится анализ различий между этими аналогами и объектом оценки, при этом каждая переменная корректируется соответствующим образом. Затем эти корректировки применяются к чистым ценам продажи аналогов для получения скорректированных цен продажи. Как показано математически, эти скорректированные цены продажи должны сходиться с расчетной ценой продажи объекта

оценки.

12.5.2 Обзор, аудит, жалобы

После определения расчетной остаточной стоимости объекта оценки можно рассчитать корректировки на остаточную стоимость для всех аналогов. Эти корректировки служат ограничительными условиями для дальнейшей ручной декомпозиции, проводимой оценщиком с учетом конкретных переменных CQA, которые считаются уместными. Важно подчеркнуть, что эти подробные разбивки носят пояснительный характер и не влияют на окончательную оценку, как подтверждается в Доказательстве II.

Что касается жалоб, поданных пользователями, не удовлетворёнными выводом об оценке стоимости, традиционному оценщику, как правило, приходится тратить время на объяснение того, почему тот или иной объект недвижимости с более высокой или более низкой ценой не оказывает ожидаемого пользователем влияния на стоимость. При использовании метода RCA все возможные аналогичные сделки (обычно от 100 до 600) оцениваются с точностью до той же стоимости, что и итоговая цена, поэтому не требуется больших усилий, чтобы продемонстрировать корректировки для какого-либо произвольного объекта недвижимости. Весь процесс RCA автоматизирован и объективен, за исключением оценки остаточного коэффициента объекта оценки. Таким образом, еще одним преимуществом метода RCA является то, что критика в отношении несовершенств субъективного суждения оценщика в основном ограничивается именно этой областью — и с ней можно эффективно и рационально справиться. Кроме того, если предположить, что R2 составляет около 80 %, влияние ошибки ± 10 % при оценке субъективного остатка приводит к погрешности ± 2 % в окончательной оценке стоимости.

Во многих случаях объект оценки настолько идеально вписывается в остаточный рейтинг объектов недвижимости, что можно ожидать, что точность будет находиться в пределах ± 1 %.

Сумма расчетного остаточного значения и расчетной цены продажи объекта оценки, полученной с помощью модели MARS, позволяет сформировать окончательный вывод о стоимости, известный как «вывод о стоимости по методу остаточных ограничений» (RCA). Данный процесс обеспечивает комплексный и прозрачный подход к оценке недвижимости, укрепляя доверие к точности и объективности оценки.

13 Заключение

Метод остаточных ограничений представляет собой концептуальную основу, включающую три существенных вклада: свойство структурной антианкорированности, вытекающее из подгонки модели к всему рынку, а не к небольшому аналогу набора данных (Доказательство I); результат инвариантности, позволяющий использовать разбивку остатков в качестве объясняющего уровня без нарушения вывода о значении (Доказательство II и следующее за ним замечание); и эмпирическая закономерность — характеристическая кривая CQA-остаточных значений — которая, как представляется, повторяется во всех рыночных зонах в характерной экспоненциально-линейно-экспоненциальной форме. Ограничения данной концептуальной рамки столь же существенны: декомпозиция остаточных величин является инвариантной, но не идентифицирующей; этап размещения CQA основан на субъективной оценке VE, которая ещё не была подтверждена в ходе формального исследования межэкспертной согласованности, а утверждения о точности требуют сравнительного анализа с альтернативными подходами.

В данной статье представлены концептуальная основа и протокол. То, чего в ней пока нет, но что входит в более обширную исследовательскую программу, реализуемую в настоящее время, — это исполняемый конвейер с возможностью отслеживания аудита, который обеспечивает реализацию протокола в программном обеспечении с обоснованными настройками по умолчанию. Эта конвейерная система — пакет `earthUI` на CRAN для этапа базисных функций MARS, пакет `glmnetUI` для регуляризованной структурной модели на основе этих базисных функций и пакет `mgcvUI` для пространственного сглаживания — является объектом оценки сопутствующих статей в этом выпуске и служит мостом от RCA как метода к RCA как конечному результату. Разрыв между теорией и практикой в регулировании оценки (Craytor, 2026, настоящий выпуск) объясняет, почему этот «мост» важен: выводы оценки используются в экономике в качестве прогнозных данных о рисках, и система, которая делает заключение о текущей стоимости поддающимся аудиту и воспроизводимым, является необходимым условием для любого добросовестного подхода к такому прогнозному использованию. RCA — это один из шагов в рамках этого более широкого проекта. Это не весь проект, и в данной статье не утверждается, что это так.

Список литературы

- Ambrose, Brent W. и др. (2025). «Do Appraiser and Borrower Race Affect Mortgage Collateral Valuation?» B: Review of Finance. DOI: [10.1093/rof/rfaf046](https://doi.org/10.1093/rof/rfaf046).
- Craytor, William Bert (2026). «The Doctrine–Practice Gap in Real Estate Appraisal: A Structured Account of Functions, Boundaries, and Tensions». B: Valuation Engineer Journal 1.1.
- Friedman, Jerome H. (1991). «Multivariate Adaptive Regression Splines». B: The Annals of Statistics 19.1, с. 1–141.
- Hastie, Trevor, Robert Tibshirani и Jerome H. Friedman (2017). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second. Springer Series in Statistics. New York, NY: Springer. ISBN: 978-0-387-84857-0.
- James, Gareth и др. (2013). An Introduction to Statistical Learning: With Applications in R. Springer Texts in Statistics 103. New York: Springer. ISBN: 978-1-4614-7137-0.
- Krivorotov, George и Michael LaCour-Little (2020). «AVM versus Appraisal-Based Underwriting in Refinance Mortgages: The Trade-off Between Noise and Bias». B: SSRN Electronic Journal. DOI: [10.2139/ssrn.3603327](https://doi.org/10.2139/ssrn.3603327).
- Kuhn, Max и Kjell Johnson (2016). Applied Predictive Modeling. Corrected at 5th printing. New York: Springer. ISBN: 978-1-4614-6848-6.

- Lundberg, Scott M. и Su-In Lee (2017). «A Unified Approach to Interpreting Model Predictions». B: Advances in Neural Information Processing Systems 30 (NeurIPS 2017), с. 4765—4774.
- Milborrow, Stephen (2024a). Earth: Multivariate Adaptive Regression Splines. R Package. URL: <https://CRAN.R-project.org/package=earth>.
- (2024b). Notes on the Earth Package. Tex. отч. URL: <http://www.milbo.org/doc/earth-notes.pdf>.
- (2024c). Variance Models in Earth. Tex. отч. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.
- Perry, Andre M., Jonathan Rothwell и David Harshbarger (2018). The Devaluation of Assets in Black Neighborhoods. Tex. отч. Brookings Institution.
- Perry, Andre M., Jonathan Rothwell, J. Williamson и др. (2024). How Racial Bias in Appraisals Affects the Devaluation of Homes in Majority-Black Neighborhoods. Tex. отч. Brookings Institution.
- Property Appraisal and Valuation Equity Task Force (2022). Action Plan to Advance Property Appraisal and Valuation Equity. Tex. отч. U.S. Department of Housing и Urban Development.
- Shapley, Lloyd S. (1953). «A Value for n -Person Games». B: Contributions to the Theory of Games, Volume II. Под ред. Harold W. Kuhn и Albert W. Tucker. Annals of Mathematics Studies 28. Princeton: Princeton University Press, с. 307—317.
- Tzioumis, Konstantinos (2016). «Appraisers and Valuation Bias: An Empirical Analysis». B: Real Estate Economics 45.3, с. 679—712. DOI: [10.1111/1540-6229.12133](https://doi.org/10.1111/1540-6229.12133).
- Williamson, Jake и Mark Palim (2022). Appraising the Appraisal: A Closer Look at Divergent Appraisal Values for Black and White Borrowers Refinancing Their Home. Tex. отч. Fannie Mae.
- Zhang, Wengang (2020). MARS Applications in Geotechnical Engineering Systems: Multi-Dimension with Big Data. Singapore: Springer. ISBN: 978-981-13-7421-0.