Смоляк С.А. (ЦЭМИ РАН)

Робастные методы построения регрессионных зависимостей в задачах стоимостной оценки

Вестник ЦЭМИ №1 2019

**Аннотация**

При сравнительном подходе к оценке активов нередко формируется выборка аналогичных активов с известными ценами и на этой основе строится параметрическая регрессионная зависимость цены актива от его технико-экономических характеристик. Обычно калибровочные параметры функции регрессии оценивают методом наименьших квадратов, а отклонения цен от построенной зависимости предполагают нормальное распределенными. При этом оценщик должен выявить и исключить из выборки активы с «резко выделяющимися» ценами, что сильно усложняет процесс оценки. Однако этого можно и не делать, применяя робастные методы, уменьшающие влияние «резко выделяющихся» данных. Известные робастные методы ориентированы на ситуации, когда выборка может включать «посторонние» (не являющиеся аналогами) объекты, для которых отклонения от регрессионной зависимости могут иметь произвольное распределение. При использовании таких методов выбранным активам как бы приписывается «вес» тем меньший, чем больше их цены отклоняются от регрессионной зависимости. К недостаткам этих методов относится невозможность сравнивать различные спецификации регрессионных зависимостей с целью выбора «лучшей» из них. Мы рассматриваем «промежуточные» ситуации, типичные для задач стоимостной оценки, когда распределение отклонений от регрессионной зависимости близко к нормальному, но имеет «более тяжелые хвосты», убывающие экспоненциально. Для таких ситуаций предлагается ряд методов оценки калибровочных параметров регрессионных зависимостей, опирающихся на принцип максимального правдоподобия. Приведены примеры применения этих методов к стоимостной оценке активов.

**Ключевые слова:** параметрическая регрессия, устойчивая оценка, максимальное правдоподобие, распределения вероятностей, «выбросы», «тяжелые хвосты», стоимостная оценка, машины, квартиры

Using a comparative approach to asset valuation, the appraiser often forms a sample of similar assets with known prices and builds on this basis a parametric regression relationship between technical and economic characteristics of asset and its price. Usually, the calibration parameters of the regression function are estimated by the least squares method, and the price deviations from the constructed dependence assume normal distributed. In this case, the appraiser must identify and exclude from the sample assets with “sharply allocated” prices (outliers), which greatly complicates the valuation process. There is no need for this if we use robust methods that reduce the impact of outliers. Known robust methods focus on situations where the sample may include assets that are not similar to the asset being valued, so that the corresponding deviations from the regression may have an arbitrary distribution. When using such methods, the selected assets are essentially taken into account in calculating with the “weight” the smaller, the more their prices deviate from the regression. However, these methods do not allow comparing different specifications of regression in order to select the “best” of them. We consider “intermediate” situations typical for valuation problems, when the distribution of deviations from regression is close to normal, but has “heavier tails” that exponentially decrease. For such situations, we propose a number of methods for estimating the calibration parameters of regression based on the maximum likelihood principle, and give examples of their application to the valuation of assets.

**Key words**: parametric regression, robust regression, maximum likelihood estimation, probability distributions, “outliers”, “heavy tails”, valuation, equipment, apartments

JEL: C13, D46

# Регрессионные зависимости в стоимостной оценке

В задачах стоимостной оценки требуется оценить стоимость определенного объекта, используя информацию об этом объекте, его аналогах и о ситуации на соответствующем рынке. Имеется много видов стоимости, однако основным является рыночная стоимость, и далее мы будем говорить только о ней, опуская определение «рыночная». Согласно стандартам оценки [1,2,3], стоимость актива на определенную дату (дату оценки) отражает цену сделки с ним на эту дату, совершаемой типичными участниками рынка при определенных (указанных в стандартах оценки) условиях сделки. Оценкой стоимости объектов занимаются оценщики, оформляя результаты своей работы в виде отчета об оценке. Оценщики несут ответственность (в том числе, имущественную и уголовную) за результаты своей оценки, поэтому отчеты об оценке внимательно рассматриваются заказчиками оценки, банками и оценщиками-экспертами.

Мы рассмотрим типичную задачу стоимостной оценки с использованием так называемого сравнительного подхода. При этом подходе стоимость оцениваемого объекта оценивается по данным о ценах некоторых аналогичных или идентичных объектов (далее – аналогов). К числу аналогов обычно относят объекты, идентичные оцениваемому или сходные с ним по назначению и/или конструкции (а при оценке недвижимого имущества – близкие по местоположению). Процедура стоимостной оценки включает следующие этапы.

1. Оценщик анализирует рынок, где обращаются аналоги, и поведение его участников. Анализируются совершенные на дату оценки или близкие к ней даты с аналогами, а также предложения о таких сделках. Разумеется, во внимание принимаются только такие сделки, условия которых близки к указанным в стандартах оценки. Проведенный анализ позволяет оценщику выбрать некоторый набор основных характеристик (их называют также ценообразующими факторами, предикторами, объясняющими или экзогенными переменными), принимаемых во внимание участниками рынка при совершении сделок с аналогами. Вектор, образованный основными характеристиками, мы будем рассматривать как один (векторный) **предиктор**. Другими словами, на этом этапе принимается, что стоимость аналога определяется только значением предиктора.
2. Оценщик формирует выборку аналогов, включая в нее объекты, проданные или выставленные на продажу в период, близкий к дате оценки, о каждом из которых известны значения предиктора (***y***) и цена (*z*). Для дальнейшего важно отметить, что правильность отбора аналогов и значения их характеристик подвергаются экспертизе. Поэтому далее мы будем исходить из того, что на этом этапе ошибок не допускается.
3. Принимается, что под влиянием случайных факторов цены аналогов отклоняются от их стоимостей. Это позволяет рассматривать (наблюдаемую) цену каждого аналога как реализацию случайной величины, распределение которой имеет центром стоимость этого аналога. Обычно в качестве такого центра принимается математическое ожидание и тогда стоимость аналога можно определить, построив по имеющейся выборке регрессионную зависимость цены аналога от значений его предиктора и подставив в нее значения предиктора у оцениваемого объекта.

Изложенная процедура оправдана, если (фактические или намечаемые) условия совершения сделки с аналогом близки к указанным в стандартах оценки. Однако иногда это предположение не выполняется (например, сделка предусматривала продажу сразу большой партии железнодорожных вагонов или ускоренную продажу квартиры). В таких случаях в состав компонент предиктора включаются и соответствующие характеристики условий сделки.

Основное внимание в данной статье мы уделим последнему этапу описанной процедуры – установлению (построению) регрессионной зависимости цены объекта (*z*) от значения его предиктора (***y***). Обычно считается известной *спецификация* такой зависимости. Иными словами, принимается, что стоимость объекта (*v*) является *известной* функцией от значения его предиктора ***y*** и некоторого (вообще говоря, векторного) *калибровочного* параметра ****: *v*=*F*(***y***,****). В таком случае возникает задача оценить неизвестный калибровочный параметр **** по данным о ценах и значениях предиктора объектов-аналогов. Другими словами, надо подобрать такое ****, чтобы равенство *z*=*F*(***y***,****) выполнялось возможно более точно для всех объектов выборки.

Для корректной постановки этой задачи необходимо также принять определенные допущения о характере случайных отклонений цен от стоимости. Нередко оценщики считают, что влияние случайных факторов на цену выражается аддитивной случайной добавкой к стоимости, имеющей нормальное распределение с нулевым средним значением и неизвестным среднеквадратичным отклонением (СКО). Тогда наблюдаемую цену (*zk*) *k*-го аналога можно рассматривать как реализацию нормально распределенной случайной величины с центром *F*(***y***,****) и неизвестным СКО . Для оценки объекта теперь надо будет установить не только калибровочные параметры зависимости, но и .

Однако допущение о нормальности распределения цен от стоимостей можно убедительно подтвердить статистическими методами только для достаточно больших по объему выборок. При этом, чем больше калибровочных параметров включает регрессионная зависимость, тем больше должен быть объем необходимой для этого выборки. Мало того, даже при отсутствии ошибок в исходной информации нередко возникают «чрезмерно большие» отклонения цен от (рассчитанных по модели) стоимостей ‑ так называемые «выбросы», а распределение наблюдаемых отклонений отклоняется от нормального. В случае однофакторной зависимости (единственная основная характеристика) выявить «выбросы» можно, анализируя соответствующие точки (***y***,*z*) на графике, однако для многофакторных зависимостей сделать это становится значительно сложнее. Более того, выбирать «подходящую» спецификацию регрессионной зависимости и «подходящие» для нее значения калибровочного параметра **** можно по-разному, и при этом «выбросами» могут оказаться каждый раз разные аналоги. Наконец, просто обнаружить «выброс» недостаточно, надо еще решить, что с ним делать. Казалось бы, проще всего исключить соответствующий объект из выборки. Сделать это необходимо, если такой объект оказался включенным в выборку по ошибке. В противном же случае такая операция может рассматриваться как «подгонка» результата оценки в «нужном» оценщику или его заказчику направлении. Более корректным было бы считать, что выборка сформирована правильно, а распределение цен отличается от нормального «более тяжелыми хвостами». Далее мы увидим, что такая ситуация на практике возникает достаточно часто.

Рассмотрим типичный случай, когда вероятностное распределение отклонений () цен от стоимостей известно с точностью до параметра масштаба (), т.е. когда плотность этого распределения имеет вид -1*h*(/). Будем считать, что плотность *h*(*x*) принимает максимальное значение при *x*=0. В таком случае стоимости объектов – центры распределений случайных цен ‑ будут отвечать модальным («наиболее вероятным») значениям их цен, что, по сути, предусматривается стандартами оценки [1,2,3]. Такую модель правомерно записывать в более привычном виде: *z*=*F*(***y***,****)+. Под *спецификацией* подобных моделей мы будем понимать набор учитываемых характеристик объектов (компонент вектора ***y***), функцию *F*(***y***,****) и вероятностное распределение случайных отклонений , т.е. функцию *h*(*x*). При правильно выбранном калибровочном параметре **** величина *v*=*F*(***y***,****) при этом будет трактоваться как (рассчитанная по модели) стоимость объектов с предиктором ***y***. Отметим, кстати, что если *h*(*x*) – четная функция (что обычно и принимается), то стоимости объектов *v*=*F*(***y***,****) будут одновременно и математическими ожиданиями их цен.

Описанная выше регрессионная модель содержит два неизвестных параметра ‑ **** и . Они должны оцениваться по данным о ценах и характеристиках выбранных аналогов. Некоторые методы такой оценки будут обсуждаться в следующем разделе статьи.

# Робастная оценка параметров регрессионной зависимости

Обычно оценщики оценивают значение калибровочного параметра **** методом наименьших квадратов (МНК), решая задачу:

 (1)

где нижний индекс означает номер аналога в выборке.

МНК дает разумные оценки, если отклонения  цен от стоимостей имеют нормальное распределение. Однако, как уже отмечалось, на практике такое требование не всегда выполняется, и тогда возможные «выбросы» могут сильно исказить «оптимальные» значения ****. Подобные ситуации моделируют следующим способом. Принимается, что для «основной массы» объектов выборки отклонения  имеют нормальное распределение с нулевым средним и неизвестной дисперсией 2, тогда как для остальных объектов эти отклонения имеют какое-то иное, неизвестное заранее «засоряющее» распределение. Для оценки параметров уравнения регрессии в такой ситуации был предложен ряд процедур, более устойчивых к подобному «засорению». Подобные процедуры, начиная с работы Хьюбера [4], принято называть робастными.

Приведем несколько примеров подобных процедур для простейшего случая оценки параметров центра *m* и СКО  нормального распределения цен в условиях «засорения» выборки.

**Пример 1**. Хорошие результаты дает метод наименьших модулей, при котором центр распределения оценивается из условия:

 (2)

где sign(*x*) – функция, равная 1 при положительных *x*, -1 – при отрицательных *x*, и нулю при *x*=0.

В данной конкретной ситуации центр распределения оценивается медианой выборочного распределения цен. При этом робастную оценку  находят, решая следующую задачу [5, 6]:



где  ≈ 0.676 – 75%-й квантиль стандартного нормального распределения.

**Пример 2**. В книге [4] предлагается следующий общий метод нахождения робастных оценок параметров сдвига *m* и масштаба  вероятностного распределения. Для этого определенным образом (в зависимости от вида распределения) подбираются функции  и  и решается система уравнений:



В частности, для оценки центра и дисперсии нормального распределения предлагается принять:  где  а  – функция стандартного нормального распределения.

Для той же ситуации в [6] рекомендуется использовать, например, следующие функции  и : 

Отметим, что в [4,6] и ряде последующих работ предлагались и основанные на тех же идеях робастные методы оценки параметров регрессионных зависимостей. Не входя в подробное обсуждение достоинств и недостатков этих методов, отметим лишь три их характерные особенности.

1. Методы, предложенные в [4] ориентированы на выборки любого объема, в которых вероятность «засорения» заранее ограничена сверху (пусть и малым числом), тогда как в [6] принимается, что вероятность «засорения» убывает с увеличением объема выборки. Соответственно в литературе уделяется большое внимание асимптотическому поведению «оптимальных» оценок параметров при неограниченном увеличении объема выборки. К сожалению, в задачах оценки реальных активов объемы выборки невелики. Исключение составляют легковые автомобили, квартиры и офисные помещения (для них имеются большие базы данных), однако их стоимости определяются десятками ценообразующих факторов, измеряемых не только в числовой, но и в номинальной шкале. А тогда становится неясным, применимы ли «асимптотически оптимальные» методы к оценке, скажем двух десятков параметров зависимости при выборке в несколько тысяч единиц. Поэтому «асимптотическая оптимальность» робастного метода оценки параметров регрессионной зависимости не может быть веским доводом для его применения в условиях малой выборки или большого числа калибровочных параметров.
2. Известные робастные методы ориентированы на то, чтобы по возможности смягчить влияние *сколь угодно больших* по величине «выбросов». Другими словами, они учитывают возможность появления «сколь угодно грубых» ошибок в исходной выборке. Мы видим две основные причины, по которым цена какого-либо объекта выборки будет сильно отличаться от рассчитанной по уравнению регрессии. Одна из них ‑ существенно «нестандартные» условия сделки с этим объектом (скажем, вынужденная сделка, малый срок экспозиции на рынке или продажа в составе крупной партии идентичных объектов). В условиях, когда правильность формирования выборки контролируется заказчиком оценки, оценщиками-экспертами и банками, подобная ситуация обычно невозможна. Другая причина, по которой цена объекта может сильно отличаться от рассчитанной по уравнению регрессии ‑ отличие этого объекта от других аналогов по каким-то характеристикам, не отнесенным к числу основных (предикторов). Например, к аналогам оцениваемой машины может быть отнесена машина той же марки, оснащенная какими-то дополнительными приспособлениями или модернизированная силами владельца для более качественного выполнения некоторых операций. Однако обычно в состав основных характеристик объектов выборки включаются те, которые влияют на цену наиболее сильно (не случайно оценщики называют эти характеристики «ценообразующими»). А тогда влияние характеристик, не отнесенных к числу основных, не может быть слишком большим.
3. В известных робастных методах предполагается, что вид «истинного» (не «засоренного») распределения наблюдаемых цен известен точно, неизвестны лишь его параметры. Между тем, при оценках идентичных объектов разные оценщики нередко выбирают разные виды зависимости стоимости объектов от их основных характеристик, рассматривая несколько альтернативных вариантов функций *F*(***y***,****). При этом разные варианты могут различаться по составу основных характеристик и по количеству параметров, образующих вектор . Для каждого из таких вариантов можно с помощью указанных и аналогичных методов оценить значения параметров, однако остается неясным, как следует выбирать «наиболее подходящий» вариант функции регрессии.

Таким образом, известные робастные методы оценивания плохо приспособлены к решению задач стоимостной оценки, особенно в условиях малых выборок. Более подходящие для этой цели методы можно предложить, опираясь на известный в статистике принцип **максимального правдоподобия**.

Заметим для начала, что если бы вероятностное распределение случайных цен было известно с точностью до его калибровочного параметра , то найти подходящую оценку этого параметра можно было бы с помощью метода максимального правдоподобия (ММП). Так, если бы распределение цен было нормальным, а его дисперсия не зависела от характеристик объекта, ММП привел бы к формуле (1) для нахождения центра распределения, т.е. превратился бы в метод наименьших квадратов (МНК). А в случае, когда цены имеют распределение Лапласа с плотностью , ММП приводит к методу наименьших модулей (2), устойчивому по отношению к любым «выбросам», но недостаточно точному (не говоря уже о том, что фактически наблюдаемые распределения отклонений цен от регрессионных зависимостей довольно далеки от распределения Лапласа).

Заметим теперь, что оценки, полученные методом наименьших квадратов в его традиционной форме, неустойчивы, т.е. могут сильно измениться при появлении «выбросов». В то же время, как показано выше, подобные «выбросы» в практике стоимостной оценки не могут быть связаны с ошибками в формировании выборки или с влиянием неучтенных характеристик объектов-аналогов. Единственной причиной, по которой могут возникать большие отклонения цен от регрессионной зависимости, мы считаем *ненормальность* закона распределения таких отклонений.

Казалось бы, такой вывод не согласуется с оценочной практикой. Ведь в подавляющем большинстве случаев оценщики используют методы, ориентированные на нормальное распределение отклонений, и получают при этом разумные результаты. Более того, во многих случаях они проверяют нормальность распределения, используя соответствующие статистические критерии. Однако эти обстоятельства не противоречат нашему мнению о ненормальности закона распределения таких отклонений. Дело в том, что в подавляющем большинстве случаев характеристики оцениваемого объекта близки к средним по выборке. При этом наличие больших «выбросов может сильно изменить калибровочные параметры регрессионной модели, но мало скажется на рассчитанной по этой модели стоимости оцениваемого объекта. Далее, применяемые оценщиками критерии нормальности проверяют близость выборочного и распределения к нормальному по некоторым характеристикам (например, эксцессу). Если выборка мала, то «доверительный интервал», в котором должна лежать характеристика выборочного распределения, оказывается очень широким. Но тогда критерии нормальности всего лишь *не позволяют отвергнуть* гипотезу о нормальности распределения цен (или отклонений цен от стоимостей). А вот критериев для сравнения выборочного распределения с другими возможными распределениями оценщики не применяют. Более того, не применяют они и достаточно мощных критериев Колмогорова, Смирнова, 2, Андерсона-Дарлинга и др. [7], оценивающих близость функций выборочного распределения к функции нормального распределения во всем диапазоне изменения аргумента. Все дело в том, что такие критерии могут подтвердить нормальность лишь при достаточно большом объеме выборки. Приведем пример.

Проверяется гипотеза о нормальности распределения цен объектов определенного вида. Для этого по выборочным данным оценивается среднее значение и дисперсия распределения, затем строится функция нормального распределения с этими характеристиками и проверяется, что она отличается от выборочной функции распределения не более, чем на 0.02. Допустим, что цены объектов в генеральной совокупности на самом деле имеют нормальное распределение. Оказывается, чтобы подтвердить это в 95% случаев, необходимо, чтобы объем выборки был не менее двух тысяч единиц – такие большие выборки в оценочной деятельности почти не встречаются. Требования к объему выборки возрастают еще больше, если речь идет об отклонениях от многофакторной регрессионной зависимости. Все это позволяет сделать вывод, что надежно обосновать вид распределения цен в задачах стоимостной оценки обычно оказывается невозможным.

В то же время «успешность» методов, ориентированных на нормальные распределения, показывает, что отклонения от нормальности возникают редко, хотя порой оказываются сравнительно большими. Так, относительно большие (скажем, превышающие «две сигмы» или «три сигмы») отклонения цен от средних значений в оценочной практике встречаются чаще, чем это должно было бы быть, если бы цены имели нормальное распределение. Это свидетельствует о том, что распределения наблюдаемых цен отличается от нормального «более тяжелыми хвостами». А в этих случаях удовлетворительные, хотя и не очень точные результаты дает метод наименьших модулей, ориентированный, как мы видели, на двустороннее экспоненциальное распределение отклонений цен от стоимостей. Это дает основание считать, что хвосты распределения наблюдаемых цен должны быть не «тяжелее», чем у распределения Лапласа.

Отметим теперь, что, в отличие от методов наименьших квадратов и наименьших модулей, ММП позволяет выбирать и наиболее правдоподобную спецификацию регрессионной зависимости. Например, с его помощью можно сравнивать зависимость , в которой отклонение  имеет нормальное распределение, с другими ее вариантами:

* , где отклонение  имеет нормальное распределение;
* , где отклонение  имеет логистическое распределение.

Все это особенно необходимо в стоимостной оценке, где всегда имеется несколько гипотез о вероятностном распределении цен объектов, аналогичных оцениваемому.

В связи с этим при стоимостной оценке желательно применять такие методы оценки параметров регрессионных зависимостей, которые:

* слабо реагируют на достаточно большие «выбросы»;
* дают достаточно точные результаты в случае, когда распределения цен близки к нормальному;
* позволяют выбирать лучший из нескольких вариантов функции регрессии.

В следующем разделе мы рассмотрим возможность применения ММП к оценке параметров регрессионной зависимости в предположении, что отклонения цен от стоимостей имеют распределение, близкое к нормальному при небольших отклонениях от центра, и близкое к распределения Лапласа при больших таких отклонениях.

# Методы максимального правдоподобия, ориентированные на альтернативные распределения цен

Разумеется, существует много распределений, которые при малых отклонениях от центра близки к нормальному, но имеют экспоненциально убывающие хвосты. Как уже отмечалось, мы ограничиваемся распределениями с плотностями вида , где *m* и  ‑ параметры центра и масштаба. В табл. 1 приведены четыре таких распределения (с параметрами соответственно *m*=0 и =1), пригодных, по нашему мнению, для использования в задачах стоимостной оценки машин и оборудования. Для сравнения там же приведено и нормальное распределение.

Таблица 1

|  |  |  |  |
| --- | --- | --- | --- |
| Распределение | Функция распределения *F*(*x*) | Функция плотности распределения *p*(*x*) | СКО |
| Нормальное (N) |  |  | 1 |
| Логистическое (LS) |  |  |  |
| Гиперболического секанса (HS) |  |  | /2 |
| (EL) |  |  | 2 |

В последней графе таблицы даются среднеквадратичные отклонения (СКО) распределений, которые могут не совпадать с параметром масштаба. Последнему из приведенных в табл. 1 распределений, обозначенному EL, трудно дать какое-то приемлемое название. Отметим лишь, что такое распределение имеет разность двух независимых случайных величин, имеющих экспоненциальное распределение, или сумма двух независимых случайных величин, имеющих распределение Лапласа.

Как уже отмечалось, распределения LS, HS и EL близки к нормальным при малых отклонениях от центра, но имеют «более тяжелые», экспоненциально убывающие «хвосты». Это позволяет применять такие распределения для оценки параметров регрессионных зависимостей и тогда, когда имеют место большие отклонения цен от линии регрессии, не отбрасывая соответствующие объекты-аналоги, а просто учитывая их как бы «с меньшим весом». Общая схема применения таких распределений к оценке параметров регрессионных зависимостей при этом остается традиционной. Опишем ее.

Рассматривается группа отобранных оценщиком объектов-аналогов. Все объекты характеризуются одним и тем же набором характеристик. Значения этих характеристик для *k*-го объекта образуют вектор ***y****k*. Цена каждого *k*-го объекта (*zk*) рассматривается как реализация случайной величины, имеющей одно из указанных распределений с центром *F*(***y****k*, ****) и параметром масштаба . Векторный параметр **** и параметр масштаба  теперь могут быть оценены методом максимального правдоподобия, т.е. из условия:



или, что то же самое,

 (3)

Нередко более подходящими оказываются модели, где вместо цен и стоимостей объектов используются их логарифмы. Тогда соответствующая регрессионная модель имеет вид: , а наиболее правдоподобные значения параметров **** и  оцениваются. из условия:

 (4)

Удобно выполнять расчеты в одной электронной таблице сразу для всех четырех указанных в таблице распределений. Наименьшему значению *L* будет отвечать и наиболее правдоподобное распределение, а значит, и наиболее правдоподобное значение искомого параметра регрессионной зависимости ****. Отметим особо, что этому параметру может не отвечать минимальное значение параметра масштаба  (и пропорциональное ему СКО).

Обратим внимание, что при сравнении нескольких вариантов спецификации регрессионной зависимости можно **по тому же критерию** выбрать и наиболее правдоподобную спецификацию. Далее мы приведем несколько примеров применения изложенного подхода к оценке активов.

# Примеры робастной оценки параметров регрессионной зависимости

В приводимых ниже примерах основные характеристики объектов и калибровочные параметры модели обозначаются по-разному, однако обозначение *z* для цен сохранено. В ряде примеров зависимости цен объектов от их характеристик принимаются степенными, при переходе к логарифмам они становятся линейными. Таким образом, в этих примерах речь, по существу, идет о зависимостях логарифмов цен объектов от логарифмов их основных характеристик. Все примеры носят иллюстративный характер (на практике объекты характеризуются более широкими наборами основных характеристик), однако основаны на реальных ценах 2014, 2015 или 2019 года.

**Пример 3**. Ищется степенная регрессионная зависимость цен (*z*) 13 гидравлических листогибочных прессов DURMA разных марок (моделей) от их главного параметра ‑ (номинального развиваемого) усилия (*X*, т). В логарифмической форме она имеет вид:

ln*z*=ln*A* + ln*X* +  (5)

Исходная информация о прессах представлена в табл. 2.

Таблица 2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| № | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Марка пресса | R1260 | R2060 | R25100 | R30100 | R30135 | R30175 | R30220 |
| Усилие, т | 60 | 60 | 100 | 100 | 135 | 175 | 220 |
| Цена, млн. руб. | 1.550 | 2.638 | 2.580 | 3.162 | 3.465 | 3.752 | 3.991 |
|  |  |  |  |  |  |  |  |
| № | 8 | 9 | 10 | 11 | 12 | 13 |  |
| Марка | R30320 | R37175 | R37220 | R40175 | R40220 | R40320 |  |
| Усилие, т | 320 | 175 | 220 | 175 | 220 | 320 |  |
| Цена, млн. руб. | 5.114 | 4.365 | 4.656 | 4.705 | 5.578 | 5.723 |  |

Расчеты проводились при четырех (указанных в табл. 1) распределениях случайной ошибки . На рис. 1 представлены исходные данные (обозначенные *fact*) и регрессионные зависимости, отвечающие нормальному и наиболее правдоподобному *HS* распределениям, в табл. 3 – значения параметров модели *a*,  и  и рассчитанного по формуле (4) логарифмического правдоподобия *L* для всех вариантов расчета.

Таблица 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Распределение  | *A* |  |  | *L* |
| *N* | 0.172 | 0.613 | 0.136 | -9.87 |
| *LS* | 0.223 | 0.564 | 0.074 | -9.70 |
| *HS* | 0.252 | 0.541 | 0.086 | **-9.57** |
| *EL* | 0.236 | 0.553 | 0.068 | -9.59 |



Рис. 1. Зависимости логарифмов цены прессов от логарифмов их номинального развиваемого усилия

**Пример 4**. По данным 20 автокранов «Галичанин» КС 55713 и КС 55731 грузоподъемностью 25 т разных марок (моделей, модификаций) строится степенная регрессионная зависимость их цен (*z*, млн. руб.) от длины стрелы (*y*, м). В логарифмической форме она имеет вид: . На рис. 2 представлены исходные данные и регрессионные зависимости, отвечающие нормальному и наиболее правдоподобному *HS* распределениям, в табл. 4 – значения параметров модели *a*,  и  и рассчитанного по формуле (4) логарифмического правдоподобия *L* для всех вариантов расчета.



Рис. 2. Зависимости логарифмов цены автокранов от логарифмов длины их стрелы

Таблица 4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Распределение  | *A* |  |  | *L* |
| *N* | 0.874 | 0.697 | 0.0510 | -11.70 |
| *LS* | 1.227 | 0.591 | 0.0277 | -11.88 |
| *HS* | 1.637 | 0.501 | 0.0294 | **-10.86** |
| *EL* | 1.460 | 0.537 | 0.0249 | -11.44 |

Для этого же примера целесообразно рассмотреть аддитивную модель цены: . Результаты расчетов здесь представлены в табл. 5.

Таблица 5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Распределение  | *a* | *b* |  | *L* |
| *N* | 2.152 | 0.242 | 0.453 | **-12.54** |
| *LS* | 2.642 | 0.223 | 0.266 | -13.27 |
| *HS* | 3.661 | 0.181 | 0.285 | -12.92 |
| *EL* | 3.147 | 0.204 | 0.237 | -13.21 |

Как видим, здесь наиболее правдоподобным оказалось нормальное распределение. Однако, поскольку -12.54<-10.86, аддитивная модель оказалась менее правдоподобной, чем мультипликативная.

**Пример 5**. Строится регрессионная зависимость цены дорожных катков ДУ85 (*z*, млн. руб.) от их возраста (*t*, годы). В соответствии с [8,9] принимается, что эта зависимость имеет вид: .

Выборка включала 49 подержанных машин, выставленных на продажу в 2014 г., а коэффициент *A*, отражающий стоимость новой машины (возраста 0 лет) был определен отдельно по данным первичного рынка и составил 2.45 млн. руб. Таким образом, необходимо было оценить только параметры *a*,  и  этой модели. Исходные данные, результаты расчетов и значения рассчитанного по формуле (4) логарифмического правдоподобия *L* для всех вариантов распределения случайной ошибки  представлены на рис.3 и в табл. 6.



Рис. 3. Зависимости цены дорожных катков от возраста

Таблица 6

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Распределение  |  |  |  | *L* |
| *N* | 0.723 | 0.1624 | 0.308 | 11.90 |
| *LS* | 0.713 | 0.1625 | 0.172 | 11.71 |
| *HS* | 0.703 | 0.1680 | 0.206 | **11.96** |
| *EL* | 0.698 | 0.1732 | 0.159 | 11.95 |

**Пример 6**. По данным о 72 квартирах в многоэтажных домах в г.о. Мытищи Московской области строится регрессионная зависимость их цен (*z*) от полезной площади (*s*) и количества комнат (*n*). Принимается, что влияние этих характеристик на стоимость квартиры – мультипликативное, что позволяет принять следующую спецификацию регрессионной модели: , или, в логарифмах, .

В отличие от предыдущих примеров одна из ценообразующих характеристик – количество комнат – выражена здесь в порядковой шкале. Поэтому соответствующая функция *f*(*n*), отражающая влияние количества комнат, должна задаваться таблицей (перечнем своих значений). Здесь важно учесть, что, если все *f*(*n*) умножить на некоторое число, а коэффициент *A* разделить на него, то результат от этого не изменится. Это позволяет принять *f*(1)=1. Исходная информация, случайно отобранная из большой базы данных, представлена в табл. 7.

Поскольку в выборке представлены только квартиры с числом комнат от 1 до 4, то в данной задаче необходимо оценить пять калибровочных параметров: *A*, , *f*(2), *f*(3), *f*(4), а также параметр масштаба распределения . Результаты расчетов наиболее правдоподобных значений этих параметров при разных спецификациях распределения случайной ошибки  представлены в табл. 8.

Как видим, нормальное распределение оказалось наименее правдоподобным, а наиболее правдоподобным оказалось *HS*-распределение – в  раза по сравнению с нормальным. В то же время оценки калибровочных параметров, ориентированные на нормальное распределение, отличаются от других оценок, хотя и не слишком сильно.

На рис. 4‑5 представлены (в логарифмах) зависимости цены квартир с разным количеством комнат от их площади и стоимости этих квартир, оцененные применительно к наиболее правдоподобному и нормальному распределениям.

Обратим внимание, что наибольшие отклонения от регрессионной зависимости наблюдаются здесь для двухкомнатных квартир.

Таблица 7.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № | *n* | *s* | *z* |  | № | *n* | *s* | *z* |  | № | *n* | *s* | *z* |
| 1 | 1 | 25.0 | 4.2 |  | 25 | 2 | 55.5 | 6.5 |  | 49 | 3 | 88.0 | 10.0 |
| 2 | 1 | 25.4 | 3.7 |  | 26 | 2 | 58.0 | 7.3 |  | 50 | 3 | 97.0 | 12.8 |
| 3 | 1 | 25.6 | 3.7 |  | 27 | 2 | 61.0 | 10.7 |  | 51 | 3 | 98.9 | 13.2 |
| 4 | 1 | 28.6 | 4.0 |  | 28 | 2 | 61.6 | 6.6 |  | 52 | 3 | 99.8 | 12.7 |
| 5 | 1 | 30.0 | 3.8 |  | 29 | 2 | 64.0 | 8.0 |  | 53 | 3 | 103.9 | 13.6 |
| 6 | 1 | 30.0 | 4.2 |  | 30 | 2 | 68.0 | 8.2 |  | 54 | 3 | 105.1 | 13.6 |
| 7 | 1 | 31.5 | 5.2 |  | 31 | 2 | 80.0 | 9.0 |  | 55 | 3 | 107.0 | 14.3 |
| 8 | 1 | 31.7 | 3.7 |  | 32 | 2 | 84.0 | 9.7 |  | 56 | 4 | 70.0 | 7.5 |
| 9 | 1 | 37.1 | 5.5 |  | 33 | 3 | 64.0 | 6.8 |  | 57 | 4 | 78.0 | 7.8 |
| 10 | 1 | 38.1 | 5.1 |  | 34 | 3 | 59.1 | 6.5 |  | 58 | 4 | 90.7 | 11.3 |
| 11 | 1 | 40.7 | 6.2 |  | 35 | 3 | 59.6 | 7.2 |  | 59 | 4 | 92.2 | 9.5 |
| 12 | 1 | 42.0 | 5.4 |  | 36 | 3 | 59.9 | 6.6 |  | 60 | 4 | 94.0 | 12.3 |
| 13 | 1 | 45.6 | 6.4 |  | 37 | 3 | 60.0 | 8.8 |  | 61 | 4 | 94.8 | 12.3 |
| 14 | 1 | 46.7 | 5.4 |  | 38 | 3 | 63.5 | 7.2 |  | 62 | 4 | 95.1 | 8.6 |
| 15 | 1 | 50.4 | 6.1 |  | 39 | 3 | 66.0 | 9.9 |  | 63 | 4 | 102.0 | 10.0 |
| 16 | 2 | 40.0 | 4.5 |  | 40 | 3 | 66.0 | 12.9 |  | 64 | 4 | 102.0 | 10.5 |
| 17 | 2 | 41.0 | 4.6 |  | 41 | 3 | 68.0 | 6.3 |  | 65 | 4 | 117.7 | 13.7 |
| 18 | 2 | 44.3 | 4.5 |  | 42 | 3 | 73.0 | 8.5 |  | 66 | 4 | 119.5 | 16.0 |
| 19 | 2 | 45.0 | 6.0 |  | 43 | 3 | 74.5 | 8.0 |  | 67 | 4 | 121.0 | 15.0 |
| 20 | 2 | 46.0 | 5.1 |  | 44 | 3 | 77.6 | 8.7 |  | 68 | 4 | 129.0 | 14.3 |
| 21 | 2 | 48.0 | 5.0 |  | 45 | 3 | 78.8 | 10.3 |  | 69 | 4 | 142.5 | 18.5 |
| 22 | 2 | 49.0 | 5.3 |  | 46 | 3 | 82.0 | 11.3 |  | 70 | 4 | 147.0 | 18.5 |
| 23 | 2 | 51.4 | 6.3 |  | 47 | 3 | 84.6 | 9.8 |  | 71 | 4 | 150.3 | 19.9 |
| 24 | 2 | 53.1 | 7.5 |  | 48 | 3 | 90.0 | 11.2 |  | 72 | 4 | 232.5 | 17.9 |

Таблица 8

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Распределение  | *A* |  | *f*(2) | *f*(3) | *f*(4) |  | *L* |
| *N* | 0.147 | 0.982 | 0.869 | 0.913 | 0.833 | 0.137 | -40.71 |
| *LS* | 0.116 | 1.050 | 0.827 | 0.853 | 0.783 | 0.074 | -42.50 |
| *HS* | 0.104 | 1.082 | 0.807 | 0.825 | 0.757 | 0.088 | **-42.59** |
| *EL* | 1.077 | 1.077 | 0.808 | 0.830 | 0.756 | 0.069 | -42.17 |



Рис. 4. Зависимости цен одно- и двухкомнатных квартир от площади при разных распределениях случайного отклонения



Рис. 5. Зависимости цен трех- и четырехкомнатных квартир от площади при разных распределениях случайного отклонения

# Выводы

При построении регрессионных зависимостей методом наименьших квадратов в целях стоимостной оценки реальных активов нередко наблюдаются большие отклонения («выбросы») фактических цен от построенной зависимости. Исключать из выборки соответствующие объекты не всегда правильно, поскольку вероятностное распределение цен может иметь «тяжелые хвосты». Это подтверждается и приводимыми примерами, основанными на реальных данных. Предлагается строить регрессионные зависимости, ориентируясь на альтернативные виды вероятностных распределений и применяя метод максимального правдоподобия. Предлагается одновременно выбирать наиболее правдоподобное из нескольких альтернативных распределений и оценивать наиболее правдоподобные значения отвечающих этому распределению калибровочных параметров регрессионной зависимости, что позволяет более глубоко обосновывать принимаемые оценщиками спецификации регрессионных зависимостей. При этом активы с резко выделяющимися ценами остаются в выборке, но учитываются в расчетах как бы с меньшими «весами».

**ЛИТЕРАТУРА**

1. МСО 2017. Международные стандарты оценки 2017 / пер. с англ. М.: Российское общество оценщиков. 2017. 168 с.
2. ЕСО 2016. Европейские стандарты оценки 2016. Восьмое издание. / Пер. с англ. М.: Российское общество оценщиков. 2017. 428 с.
3. ФСО № 1. Общие понятия оценки, подходы и требования к проведению оценки. Приказ Минэкономразвития России от 20.05.2015 N 297
4. Хьюбер П. Робастность в статистике. М.: Мир. 1984. 304 с.
5. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. Robust statistics: the approach based on influence functions (2nd ed.). 1986. New York: Wiley.
6. Шурыгин А.М. Прикладная статистика: робастность, оценивание, прогноз. М.: Финансы и статистика. 2000. 224 с.
7. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: монография / Б.Ю. Лемешко [и др.]. Новосибирск: Изд-во НГТУ, 2011. 888 с. (серия «Монографии НГТУ»)
8. Смоляк С.А. Зависимость стоимости машин от возраста: проблемы и модели // Аудит и финансовый анализ. 2014. №5. С. 138-150.
9. Смоляк С.А. Стоимостная оценка машин и оборудования (секреты метода ДДП). М.: Издательский дом «Опцион». 2016. 377 с.

**REFERENCES**

1. International Valuation Standards 2017. International Valuation Standards Council: London. 2017. 119 p.
2. European Valuation Standards 2016. Eighth edition. TEGoVA. 2016. 376 p.
3. Federal Valuation Standard No 1. General concepts of valuation, approaches and requirements for valuation conducting. Order of the Ministry of Economic Development of Russia of 20.05.2015 No 297. (in Russian).
4. Huber, Peter J. Robust Statistics. New York, Chichester, Brisbane, Toronto, Singapore: Wiley. 1981.
5. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. Robust statistics: the approach based on influence functions (2nd ed.). 1986. New York: Wiley.
6. Shurygin, A.M. Applied stochastics: robustness, estimation, forecast. Мoscow: Financy i statistica. 2000. 224 p. (in Russian).
7. Statistical data analysis, modeling and research of probability laws. Computer approach: monograph / B.Yu. Lemeshko [et al.]. Novosibirsk: NSTU Publishing House, 2011. 888 p. (series “Monographs of NSTU”). (in Russian).
8. Smolyak S.A. The dependence of the equipment value on age: problems and models // Audit and financial analysis. 2014. No 5. Pp. 138-150. (in Russian).
9. Smolyak S.A. Machinery and equipment valuation (secrets of the DCF method). Мoscow: Option Publishing House. 2016. 377 p. (in Russian).