

Гедоническая регрессия

Foundations • The Valuation Engineer

Берт Крейтор • 8 августа 2026 г.

Версия 1.0.0

DOI: [10.67330/ams1t139](https://doi.org/10.67330/ams1t139)

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

Формальное определение. Hedonic regression is the estimation of the hedonic price function from observed transactions in a heterogeneous-goods market, yielding empirical estimates of the implicit prices of the characteristics that enter the model.

Интуитивное представление. В предыдущих разделах гедоническая функция цены и её неявные цены рассматривались как теоретические объекты. Эта функция существует в принципе, но на практике мы наблюдаем лишь конечное множество сделок: торгуемые наборы товаров и цены, по которым они торгуются. Гедоническая регрессия служит эмпирическим мостом между этими двумя аспектами.

Схема проста. Каждая сделка предоставляет пару « $(\mathbf{z}_i, p_i)$ »: вектор характеристик торгуемого набора и его цену продажи. Совокупность этих пар представляет собой выборку из (ненаблюдаемой) гедонической функции цены. В ходе регрессионного анализа к этой выборке подбирается функциональная форма, при этом оцениваются параметры, характеризующие функцию. После оценки параметров частные производные подгоняемой функции позволяют получить оценки неявных цен.

Прежде чем запустить регрессионный анализ, необходимо выбрать один из трёх вариантов моделирования:

1. *Какие характеристики включаются в модель?* Этот выбор определяет размерность оцениваемой функции. Исключение значимой характеристики приводит к тому, что её влияние переносится на коэффициенты коррелирующих включённых характеристик (систематическая ошибка, связанная с пропущенной переменной).
2. *Какую функциональную форму выбрать?* Линейную, логарифмически-линейную, логарифмически-логарифмическую, полиномиальную, с расширенной базой, полупараметрическую, основанную на машинном обучении. Этот выбор ограничивает форму, которую может принимать оцениваемая функция, и какие закономерности неявных цен она может воспроизвести.
3. *Какой оценщик?* OLS, обобщённые наименьшие квадраты, робастные М-оценщики, квантильная регрессия, пространственные эконометрические оценщики, регрессия с штрафными терминами. Выбор определяет, что именно оптимизирует оценщик и какие допущения лежат в основе его выводов.

Каждый выбор имеет методологическое значение и влечет за собой определенные последствия с точки зрения обоснованности.

Где оценщик встречается с этим понятием. Гедоническая регрессия лежит в основе массовой оценки и является движущей силой большинства автоматизированных моделей оценки (AVM). Когда оценщик округа ежегодно переоценивает 100,000 земельных участков, эту работу выполняет гедоническая регрессия на основе недавних сделок купли-продажи. Когда Zillow

или Redfin выдают автоматизированную оценку, за этим стоит модель гедонического типа (часто представляющая собой ансамбль деревьев, а не OLS, но концептуально все равно являющаяся гедоническим оценщиком). Когда АVM кредитора на вторичном рынке одобряет или отклоняет объект недвижимости для рефинансирования, задействуется тот же механизм.

При оценке отдельных объектов недвижимости всё чаще используется и гедоническая регрессия, даже если окончательная оценочная стоимость определяется на основе сравнения с аналогичными сделками. Неявные цены, рассчитанные на основе регрессии по выборке объектов в данном районе, позволяют обосновать величину корректировок гораздо более обоснованно, чем интуиция оценщика или процентные значения, основанные на эмпирических правилах. Коэффициенты регрессии становятся доказательством, а не мнением, при применении корректировок в таблице оценок.

Почему это важно для обоснованности. Оценки, полученные с помощью регрессионного анализа, отличаются по степени обоснованности от анализа парных продаж. Преимущества: регрессионный анализ использует все имеющиеся данные о продажах, а не по две единицы за раз; он позволяет выделить влияние одной характеристики, одновременно контролируя множество других; кроме того, он даёт стандартные ошибки, которые количественно оценивают неопределённость каждой неявной цены. Недостатки: регрессия навязывает функциональную форму, которая может не соответствовать истинной ценовой функции; она требует большего объема данных, чем анализ парных продаж; и её результаты могут быть чувствительны к выбору выборки и спецификации.

При оценке на основе регрессионного анализа неоднократно возникают следующие вопросы, касающиеся обоснованности:

Является ли выборка адекватной? Регрессия оценивает гедоническую функцию в том виде, в каком она действует в рыночном сегменте и периоде времени, охватываемых выборкой. Применение регрессии, построенной на данных о продажах недвижимости, расположенной в глубине суши, к объекту оценки, расположенному на берегу моря, является методологической ошибкой, независимо от того, насколько обоснованы статистические результаты.

Является ли данная спецификация уместной? Линейная спецификация — это серьёзное допущение. Если реальная функция демонстрирует убывающую доходность от GLA выше некоторого порогового значения, линейная подгонка приведёт к неверной оценке неявных цен в этом диапазоне. Диагностика, графики остатков и альтернативные спецификации являются частью обоснованного рабочего процесса.

Что означает стандартная ошибка? Стандартная ошибка отражает вариабельность выборки при принятой спецификации модели. Она не учитывает неверную спецификацию модели, систематическую ошибку, связанную с пропущенными переменными, или отбор выборки. Небольшая стандартная ошибка при неверной спецификации модели более опасна, чем большая стандартная ошибка при правильной спецификации.

Совпадает ли полученный результат с результатами других подходов? Обоснованная обоснованная неявная цена, рассчитанная на основе регрессионного анализа, должна согласовываться с выводами анализа парных сделок, с данными о стоимости компонентов, полученными по затратному подходу, а также с мнением оценщиков, хорошо знакомых с рынком. Регрессионный коэффициент, противоречащий всем трем факторам, является поводом для дополнительного изучения, а не для публикации.

Разобраный пример оценки. Теперь мы можем применить полный набор методов гедонической регрессии к восьми аналогам Pacifica. Набор данных доступен по адресу `shared/data/pacifica_comps`.

Код на R:

```
comps <- read.csv("shared/data/pacifica_comps.csv")
fit <- lm(price ~ gla + lot + view + cond, data = comps)
summary(fit)
```

Вывод:

```
Call:
lm(formula = price ~ gla + lot + view + cond, data = comps)

Residuals:
    Min       1Q   Median       3Q      Max
-17995   -8243   -5433   -1409    46455

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  80777.07   164509.35    0.491   0.6571
gla           574.93     109.76    5.238   0.0135 *
lot           22.70      10.38    2.187   0.1165
view        86179.03   31342.97    2.750   0.0708 .
cond        49824.33   27761.27    1.795   0.1706
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 30481 on 3 degrees of freedom
Multiple R-squared:  0.9707, Adjusted R-squared:  0.9317
F-statistic: 24.86 on 4 and 3 DF, p-value: 0.01231
```

Интерпретация. Модель, подгнанная для описания гедонической функции цены, описывает ее следующим образом:

$$\hat{p}(\mathbf{z}) = 80,777 + 575 z_{\text{GLA}} + 22.70 z_{\text{lot}} + 86,179 z_{\text{view}} + 49,824 z_{\text{cond}}.$$

Каждый коэффициент представляет собой расчетную неявную цену соответствующей характеристики. Показатель GLA вносит вклад в размере \$ 575 за квадратный фут (с наименьшей стандартной ошибкой и наибольшей статистической значимостью); вид вносит вклад в размере около \$ 86,000 в стоимость объекта недвижимости в Пасифике; дополнительная единица оценки состояния (например, переход от «хорошего» к «отличному») стоит примерно \$ 50,000.

Сразу бросаются в глаза несколько моментов, на которые следует обратить внимание:

Выборка очень мала. Восемь наблюдений и четыре предикторы оставляют три степени свободы остатков. Стандартные ошибки соответственно велики, и значения p для «лота», «вида» и «состояния» не достигают принятого порога в 0,05, несмотря на их экономическую обоснованность. Это следствие небольшого объема выборки, а не доказательство того, что эти характеристики не имеют значения. При наличии 80 аналогов, а не 8, значения p значительно уменьшились бы.

Общая подгонка модели высокая. Скорректированный коэффициент $R^2 = 0.93$ и значение p в 0,012 указывают на то, что включенные характеристики в совокупности объясняют большую часть колебаний цен в выборке. Сочетание сильной совокупной подгонки со слабой индивидуальной значимостью является классическим признаком мультиколлинеарности, которая в данном случае носит структурный характер: GLA, участок и состояние объекта являются ковариатами в выборке.

Один остаток выделяется. Максимальный остаток составляет +\$46,455, что соответствует компании Н. Остальные семь остатков находятся в диапазоне от $\pm \$18,000$. Что-то в компании

Comp H не учитывается функцией подгонки — модель систематически занижает её цену на величину, почти в три раза превышающую следующий самый большой остаток.

Это отклонение от среднего значения не является шумом. Оно свидетельствует о том, что ненаблюдаемая характеристика Comp H влияет на его цену таким образом, который не может быть учтён с помощью включённых в модель переменных. В разделе, посвящённом латентным переменным, рассматривается именно этот вопрос.

Проверка на соответствие логике парных продаж. Анализ парных продаж, приведенный в записи № 004, предсказывал разницу в \$ 127,039 между аналогами A и B, исходя исключительно из данных о просмотрах и лотах. Неявные цены в регрессии идентичны (поскольку они являются источником данного расчёта). Регрессия обобщает логику парных продаж на все восемь аналогов одновременно, но при данном размере выборки не даёт ничего принципиально нового; она подтверждает результаты парного анализа, одновременно предоставляя отчёт о стандартных ошибках, чего не может сделать метод парных продаж.

Список литературы.

Корт, А. Т. (1939). «Гедонические ценовые индексы на примере автомобилей». В сборнике «Динамика спроса на автомобили». General Motors Corporation.

Гриличес, З. (1961). Гедонические ценовые индексы для автомобилей: эконометрический анализ изменения качества. В сборнике «Статистика цен федерального правительства». Национальное бюро экономических исследований.

Розен, С. (1974). Гедонические цены и неявные рынки: дифференциация продукции в условиях чистой конкуренции. «*Journal of Political Economy*», 82(1), 34–55.

Триплетт, Дж. Э. (2006). *Справочник по гедоническим индексам и корректировкам качества в ценовых индексах*. Издательство ОЭСР.

Перекрёстные ссылки: heterogeneous good; characteristics space; hedonic price function; implicit price; latent variable.

Приложение: формализация на Prolog

Следующий машиночитаемый вспомогательный файл (005-hedonic-regression.pl) формализует эту статью в базе знаний журнала на Prolog. Он также публикуется рядом с этой статьёй как отдельный просматриваемый файл (galley).

```
%% 005-hedonic-regression.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 005
%% Title: Hedonic Regression
%% Author: Bert Craytor
%% Concept: hedonic_regression
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('.././../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(hedonic_regression, "Hedonic Regression", econometric_method,
```

```

%      "Hedonic regression is the estimation of the hedonic price
%      function from observed transactions in a heterogeneous-goods
%      market, yielding empirical estimates of the implicit prices of
%      the characteristics that enter the model.",
%      published, 'dictionary-2026.1').
%
% term(mass_appraisal, ...).
% term(automated_valuation_model, ...). % synonym: avm
% term(omitted_variable_bias, ...).
% term(multicollinearity, ...).
% term(residual, ...).
% term(ordinary_least_squares, ...).
%
% introduced_by(hedonic_regression, court_1939).
% introduced_by(hedonic_regression, griliches_1961).
% formalized_by(hedonic_regression, rosen_1974).
% depends_on(hedonic_regression, hedonic_price_function).
% depends_on(hedonic_regression, implicit_price).
% -----

:- use_terms([ hedonic_regression, hedonic_price_function, implicit_price,
               characteristics_space, heterogeneous_good, latent_variable,
               mass_appraisal, automated_valuation_model,
               omitted_variable_bias, multicollinearity, residual,
               ordinary_least_squares,
               sales_comparison_approach, paired_sales_analysis, adjustment ]).
:- use_relations([related/2, depends_on/2,
                  introduced_by/2, formalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(hedonic_regression).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.005').

% =====
% LOCAL VOCABULARY
% =====

local_term(specification).
local_term(coefficient).
local_term(standard_error).
local_term(p_value).
local_term(r_squared).
local_term(adjusted_r_squared).
local_term(residual_standard_error).
local_term(degrees_of_freedom).
local_term(f_statistic).
local_term(significance_level).
local_term(diagnostic).

% =====
% THE EMPIRICAL BRIDGE
% =====
%
% Hedonic regression is the bridge from theoretical  $p(z)$  to a sample of
% observed  $(z_i, p_i)$  pairs.

```

```

setup_of_regression(
    each_transaction_supplies_pair("(z_i, p_i): characteristics vector and sale price"),
    sample_from("the unobserved hedonic price function"),
    fits("a functional form to that sample, estimating parameters")).

% =====
% THREE MODELING CHOICES (per the article)
% =====

modeling_choice(which_characteristics,
    "Determines the dimensions of the estimated function. Omission causes omitted_variable_bias.").
modeling_choice(functional_form,
    "Constrains what shape the estimated function can take. Linear, log-linear, polynomial, basis-
    ↪ expanded, semiparametric, ML.").
modeling_choice(estimator,
    "OLS, GLS, robust, quantile, spatial, penalized. Determines optimization criterion and
    ↪ inference assumptions.").

% =====
% PRACTICE CONTEXTS WHERE REGRESSION IS THE ENGINE
% =====

regression_engine_context(mass_appraisal,
    "Assessor revaluation of 100,000 parcels annually.").
regression_engine_context(automated_valuation_model,
    "Zillow/Redfin/lender AVMs; tree ensembles are conceptually hedonic estimators.").
regression_engine_context(single_property_appraisal,
    "Regression coefficients support adjustments more defensibly than appraiser intuition.").

% =====
% DEFENSIBILITY PROFILE: REGRESSION vs PAIRED-SALES
% =====

advantage_over_paired_sales(uses_all_data,
    "Regression uses all available sales rather than two at a time.").
advantage_over_paired_sales(simultaneous_control,
    "Isolates one characteristic while controlling for many others.").
advantage_over_paired_sales(quantified_uncertainty,
    "Produces standard errors quantifying uncertainty in each implicit price.").

tradeoff(functional_form_constraint,
    "Regression imposes a form that may not match the true price function.").
tradeoff(data_requirement,
    "Requires more observations than paired-sales analysis.").
tradeoff(sample_specification_sensitivity,
    "Results can be sensitive to the choice of sample and specification.").

% =====
% RECURRING DEFENSIBILITY QUESTIONS
% =====

defensibility_question(sample_appropriateness,
    "The regression estimates the function for the sample's market segment and time period;
    ↪ mismatch is a defensibility failure regardless of statistical fit.").
defensibility_question(specification_appropriateness,
    "Linear is a strong assumption; diagnostics, residual plots, and alternative specifications
    ↪ belong in the defensible workflow.").
defensibility_question(standard_error_interpretation,

```

```

    "Standard errors capture sampling variability under the assumed specification, NOT
    ↪ misspecification, omitted-variable bias, or sample selection.>").
defensibility_question(triangulation,
    "Regression coefficients should be consistent with paired-sales, cost components, and local
    ↪ appraiser opinion; isolated coefficients are an invitation to investigate.>").

% =====
% WORKED EXAMPLE: lm(price ~ gla + lot + view + cond)
% =====

:- consult('.../..kb/pacifica_comps').

% The fitted model.
fit_call("lm(formula = price ~ gla + lot + view + cond, data = comps)").
fit_estimator(ordinary_least_squares).
fit_software(r_lm).

% Fitted coefficients, with standard errors and significance from
% the R output reproduced in the article.
fitted_coefficient(intercept, 80777.07, 164509.35, 0.491, 0.6571, ' ').
fitted_coefficient(gla, 574.93, 109.76, 5.238, 0.0135, '*').
fitted_coefficient(lot, 22.70, 10.38, 2.187, 0.1165, ' ').
fitted_coefficient(view, 86179.03, 31342.97, 2.750, 0.0708, '.').
fitted_coefficient(cond, 49824.33, 27761.27, 1.795, 0.1706, ' ').

% Diagnostics from the R output.
residual_summary(min, -17995).
residual_summary(q1, -8243).
residual_summary(median, -5433).
residual_summary(q3, -1409).
residual_summary(max, 46455).

model_summary(residual_standard_error, 30481).
model_summary(degrees_of_freedom, 3).
model_summary(r_squared, 0.9707).
model_summary(adjusted_r_squared, 0.9317).
model_summary(f_statistic, 24.86).
model_summary(model_p_value, 0.01231).

% =====
% INTERPRETIVE CAVEATS FROM THE ARTICLE
% =====

caveat(tiny_sample,
    "n=8, four predictors, three residual df. Standard errors are correspondingly wide; lot/view/
    ↪ cond fail to reach the 0.05 threshold despite economic plausibility. Small-sample artifact.>").
    ↪

caveat(structural_multicollinearity,
    "Adj R^2 = 0.93 with model p = 0.012 indicates strong joint fit; weak individual significance
    ↪ is a textbook multicollinearity signature. Here GLA, lot, and condition covary in the sample
    ↪ .").

caveat(comp_h_residual,
    "Comp H's residual is +$46,455; the other seven fall within +/- $18,000. The model
    ↪ systematically under-predicts Comp H by ~3x the next-largest residual. NOT noise. Signal of
    ↪ an unobserved characteristic.>").

% =====

```

```

% RELATING TO PAIRED-SALES (entry 004)
% =====
%
% The regression's implicit prices ARE the values used in entry 004's
% paired-sales decomposition of A vs B. The regression generalizes that
% logic to all 8 comps simultaneously while reporting standard errors
% that paired-sales reasoning cannot produce.

confirms_paired_sales_from(entry_004,
  "A vs B differential of $127,039 from view and lot alone - coefficients are the same;
  ↪ regression confirms the paired analysis while adding standard errors.>").

% =====
% RESIDUAL OF COMP H POINTS FORWARD TO ENTRY 006
% =====

points_forward_to(latent_variable,
  "Comp H's anomalous +$46,455 residual is the signal that an unobserved characteristic is
  ↪ contributing to price. Entry 006 takes up the question.>").

% =====
% PREDICTED PRICES FROM THE FITTED MODEL
% =====
%
% Computes predicted price from the fitted coefficients. The residual
% for each comp is sale_price - predicted_price.

predicted_price(P, Predicted) :-
  coord(P, gla_sqft, GLA),
  coord(P, lot_sqft, Lot),
  coord(P, view, View),
  coord(P, cond_numeric, Cond),
  fitted_coefficient(intercept, B0, _, _, _, _),
  fitted_coefficient(gla, Bg, _, _, _, _),
  fitted_coefficient(lot, B1, _, _, _, _),
  fitted_coefficient(view, Bv, _, _, _, _),
  fitted_coefficient(cond, Bc, _, _, _, _),
  Predicted is B0 + Bg * GLA + B1 * Lot + Bv * View + Bc * Cond.

residual_of(P, R) :-
  sale_price(P, Observed),
  predicted_price(P, Predicted),
  R is Observed - Predicted.

coord(P, A, N) :-
  attribute_value(P, A, V),
  ( number(V) -> N = V
  ; A == view, V == yes -> N = 1
  ; A == view, V == no -> N = 0
  ; V = N
  ).

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(characteristics_space).
cross_reference(hedonic_price_function).

```

```
cross_reference(implicit_price).  
cross_reference(latent_variable).
```