

Regresión hedónica

Traducción al español de la entrada publicada originalmente en inglés. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 1 de agosto de 2026

Versión 1.0.0

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

Definición formal. La regresión hedónica es la estimación de la función de precios hedónicos a partir de transacciones observadas en un mercado de bienes heterogéneos, que produce estimaciones empíricas de los precios implícitos de las características que entran en el modelo.

Planteamiento intuitivo. Las entradas anteriores describieron la función de precios hedónicos y sus precios implícitos como objetos teóricos. La función existe en principio, pero en la práctica observamos únicamente un conjunto finito de transacciones: conjuntos de características que se comerciaron y los precios a los que se comerciaron. La regresión hedónica es el puente empírico entre ambos.

El planteamiento es sencillo. Cada transacción aporta un par $(\mathbf{z}_i, p_i)$: el vector de características del conjunto transado y su precio de venta. La colección de estos pares es una muestra de la función de precios hedónicos (no observada). La regresión ajusta una forma funcional a esa muestra, estimando los parámetros que indexan la función. Una vez estimados los parámetros, las derivadas parciales de la función ajustada producen estimaciones de los precios implícitos.

Hay que tomar tres decisiones de modelado antes de que la regresión pueda ejecutarse:

1. *¿Qué características entran en el modelo?* La elección determina las dimensiones de la función estimada. Omitir una característica relevante hace que su efecto se filtre en los coeficientes de las características incluidas correlacionadas (sesgo por variables omitidas).
2. *¿Qué forma funcional?* Lineal, log-lineal, log-log, polinómica, con expansión de bases, semi-paramétrica, de aprendizaje automático. La elección restringe qué forma puede adoptar la función estimada y qué patrones de precios implícitos puede recuperar.
3. *¿Qué estimador?* MCO, mínimos cuadrados generalizados, M-estimadores robustos, regresión cuantílica, estimadores de econometría espacial, regresión penalizada. La elección determina qué optimiza el estimador y qué supuestos sustentan su inferencia.

Cada elección es metodológicamente sustantiva y cada una acarrea consecuencias de defendibilidad.

Dónde lo encuentran los tasadores. La regresión hedónica es el fundamento de la valuación masiva y el motor de la mayoría de los modelos de valuación automatizada (AVM). Cuando un asesor fiscal de condado revalúa 100,000 parcelas cada año, una regresión hedónica sobre ventas recientes está haciendo el trabajo. Cuando Zillow o Redfin producen una estimación automatizada, la está produciendo un modelo de estilo hedónico (a menudo un ensamble de árboles en lugar de MCO, pero conceptualmente un estimador hedónico de todos modos). Cuando el AVM de mercado secundario de un prestamista aprueba o rechaza un inmueble para un refinanciamiento, está involucrada la misma maquinaria.

La tasación de inmuebles individuales también recurre cada vez más a la regresión hedónica, incluso cuando la indicación final de valor proviene de la comparación de ventas. Los precios implícitos

estimados a partir de una regresión sobre una muestra del vecindario pueden sustentar las magnitudes de los ajustes de manera mucho más defendible que la intuición del tasador o los porcentajes de regla empírica. Los coeficientes de la regresión se convierten en evidencia, no en opinión, para los ajustes aplicados en el cuadro.

Por qué importa para la defendibilidad. Las estimaciones basadas en regresión tienen un perfil de defendibilidad distinto del análisis de pares comparables. Las ventajas: una regresión usa todas las ventas disponibles en lugar de dos a la vez; puede aislar el efecto de una característica mientras controla simultáneamente muchas otras; y produce errores estándar que cuantifican la incertidumbre de cada precio implícito. Las contrapartidas: una regresión impone una forma funcional que puede no coincidir con la verdadera función de precios; requiere más datos que un análisis de pares comparables; y sus resultados pueden ser sensibles a la elección de la muestra y de la especificación.

Varias cuestiones de defendibilidad se repiten en la tasación basada en regresión:

¿Es apropiada la muestra? La regresión estima la función hedónica tal como opera en el segmento de mercado y el periodo de la muestra. Una regresión sobre ventas del interior aplicada a un inmueble objeto frente a la playa es un fallo de defendibilidad por más limpias que luzcan las estadísticas.

¿Es apropiada la especificación? Una especificación lineal es un supuesto fuerte. Si la función verdadera exhibe rendimientos decrecientes de la GLA por encima de algún umbral, el ajuste lineal estimará mal los precios implícitos en ese rango. Los diagnósticos, los gráficos de residuos y las especificaciones alternativas forman parte del flujo de trabajo defendible.

¿Qué significa el error estándar? Un error estándar captura la variabilidad muestral bajo la especificación supuesta. No captura la especificación errónea del modelo, el sesgo por variables omitidas ni la selección de la muestra. Un error estándar pequeño sobre una especificación equivocada es más peligroso que un error estándar grande sobre una correcta.

¿Triangula el resultado con los demás enfoques? Un precio implícito basado en regresión y defendible debería ser coherente con lo que sugiere el análisis de pares comparables, con lo que implican los costos de los componentes del enfoque del costo y con lo que creen los tasadores que conocen el mercado. Un coeficiente de regresión en desacuerdo con los tres es una invitación a investigar, no a publicar.

Ejemplo de tasación resuelto. Podemos ahora ejercitar toda la maquinaria de la regresión hedónica sobre los ocho comparables de Pacifica. El conjunto de datos está disponible en `shared/data/pacifica.co`

Código en R:

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

Salida:

```
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

```

Interpretación. El modelo ajustado estima la función de precios hedónicos como:

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

Cada coeficiente es el precio implícito estimado de la característica correspondiente. La GLA contribuye \$575 por pie cuadrado (con el error estándar más estrecho y la significancia más fuerte); la vista contribuye alrededor de \$86,000 a un inmueble de Pacífica; una unidad adicional de estado de conservación (p. ej., pasar de “bueno” a “excelente”) vale aproximadamente \$50,000.

Varias salvedades son inmediatamente visibles:

La muestra es diminuta. Ocho observaciones y cuatro predictores dejan tres grados de libertad residuales. Los errores estándar son correspondientemente amplios, y los valores p del lote, la vista y el estado no alcanzan el umbral convencional de 0.05 a pesar de su plausibilidad económica. Se trata de un artefacto de muestra pequeña, no de evidencia de que estas características no importen. Con 80 comparables en lugar de 8, los valores p se reducirían drásticamente.

El ajuste global es fuerte. Un R^2 ajustado = 0.93 y un valor p del modelo de 0.012 indican que las características incluidas explican conjuntamente la mayor parte de la variación de precios en la muestra. La combinación de un ajuste conjunto fuerte con una significancia individual débil es una señal de manual de multicolinealidad, que aquí es estructural: la GLA, el lote y el estado covarían en la muestra.

Un residuo sobresale. El residuo máximo es de +\$46,455 y corresponde al comparable H. Los otros siete residuos caen todos dentro de una banda de $\pm \$18,000$. Algo del comparable H no queda capturado por la función ajustada — el modelo está subprediciendo sistemáticamente su precio por una cantidad casi tres veces mayor que el siguiente residuo más grande.

Este residuo atípico no es ruido. Es la señal de que una característica no observada del comparable H está contribuyendo a su precio de una manera que las variables incluidas no pueden recuperar. La entrada sobre la variable latente aborda exactamente esta cuestión.

Validación frente al razonamiento de pares comparables. El análisis de pares comparables de la entrada 004 predijo una diferencia de \$127,039 entre los comparables A y B a partir de la vista y el lote únicamente. Los precios implícitos de la regresión son idénticos (puesto que son la fuente de ese cálculo). La regresión generaliza la lógica de los pares comparables a los ocho comparables simultáneamente, pero no produce nada fundamentalmente nuevo con este tamaño de muestra; confirma el análisis por pares a la vez que informa errores estándar que el razonamiento por pares no puede ofrecer.

Referencias.

- Court, A. T. (1939). Hedonic Price Indexes with Automotive Examples. In *The Dynamics of Automobile Demand*. General Motors Corporation.
- Griliches, Z. (1961). Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change. In *The Price Statistics of the Federal Government*. National Bureau of Economic Research.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.
- Tripllett, J. E. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*. OECD Publishing.

Referencias cruzadas: bien heterogéneo; espacio de características; función de precios hedónicos; precio implícito; variable latente.

Adéndum: formalización en Prolog

El siguiente material complementario legible por máquina (005-hedonic-regression.pl) formaliza esta entrada en la base de conocimiento Prolog de la revista. También se publica como galerada navegable junto con este artículo. El código no se traduce.

```
%% 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).

```