

Año, kilometraje y motorización en el precio del BMW 320 usado

Un estudio hedónico de anuncios y una auditoría de comparabilidad internacional

Marcos Camacho

Documento de trabajo · versión 2 · 5 de octubre de 2026

Resumen

Este estudio analiza qué características se asocian con el precio anunciado de vehículos BMW 320 de segunda mano. A partir de 1.355 registros alemanes, se construye una muestra auditada de **1.289 anuncios**, con años declarados entre 2019 y 2025. Se comparan seis especificaciones por mínimos cuadrados ordinarios que incorporan kilometraje, año, motorización y menciones de xDrive y M Sport. La inferencia utiliza errores estándar HC3; la evaluación predictiva separa anuncios potencialmente repetidos por grupos.

El modelo semilogarítmico con kilometraje cuadrático relaciona el paso de 80.000 a 90.000 km con un precio geométrico ajustado **2,64 % menor**. Frente al 320e, las diferencias condicionadas son 7,4 % para el 320d y 5,5 % para el 320i. El modelo predictivo seleccionado por parsimonia alcanza un MAE de **1.814 €** en 259 anuncios reservados, un 62,4 % inferior al predictor de mediana.

La auditoría española identifica 42 precios ausentes, 75 valores superiores a 100.000 € y discrepancias de año con las URL. Estas incidencias impiden sostener una estimación fiable de rentabilidad de importación. Los resultados alemanes describen asociaciones de precios anunciados; el rechazo de RESET, la heteroscedasticidad y la selección de la muestra limitan su interpretación estructural.

Palabras clave: precios hedónicos; vehículos usados; MCO; HC3; variables categóricas; validación fuera de muestra.

1. Motivación

Dos coches con el mismo nombre comercial pueden diferir en uso, cohorte, tecnología y equipamiento. Comparar sus precios sin controlar esas diferencias mezcla composición con valoración. Este trabajo transforma una búsqueda de oportunidades de importación en una investigación reproducible sobre formación de precios, manteniendo una auditoría explícita del análisis original.

Las tablas y figuras proceden de los archivos conservados del proyecto, no de datos simulados. El análisis estadístico y la redacción de esta ampliación se realizaron con asistencia de IA. No constituye una publicación revisada por pares.

2. Pregunta de investigación y marco analítico

La pregunta principal es cuánto cambia el precio anunciado asociado a diferencias de kilometraje, año declarado y motorización, manteniendo constantes los atributos observados. Una pregunta secundaria es si la base española permite comparar vehículos equivalentes entre países.

2.1. Un vehículo como conjunto de características

La tradición de precios hedónicos estudia bienes diferenciados mediante sus características [4]. Aquí se estima una relación de precios observados con atributos del anuncio. La ecuación resultante no identifica por sí sola preferencias de los compradores ni costes marginales de los fabricantes.

Se espera una asociación negativa del kilometraje y una positiva de cohortes más recientes. La comparación de motorizaciones es abierta: el valor del híbrido enchufable, la demanda del diésel, la potencia y el equipamiento pueden influir de forma conjunta. Los signos se contrastan en los datos; no se imponen.

2.2. De los apuntes a una aplicación empírica

Se revisaron los seis temas de los apuntes de Econometría guardados en el Drive del proyecto [2]. El diseño utiliza la interpretación semilogarítmica del Tema 2, el modelo lineal general y las variables omitidas del Tema 3, las ficticias, polinomios y RESET del Tema 4 y los métodos ante heteroscedasticidad e influencia del Tema 5. El Tema 6 corresponde a series temporales; no se aplica aquí porque los años son atributos de vehículos distintos.

2.3. Diccionario de variables

Variable	Construcción	Interpretación
Precio	Precio numérico alemán, en euros	Precio solicitado; no precio de cierre
Año	Categorías 2019-2025	Año declarado en el archivo; definición registral sin verificar
Kilometraje	$k = \text{km} / 10.000 - 8$	Unidad de 10.000 km; centrado en 80.000 km
Motorización	320e, 320i o 320d, desde el título	Híbrido enchufable, gasolina, diésel
xDrive	1 si figura xDrive, 320xd o xD	Mención de tracción integral en el título
M Sport	1 si figura M Sport, M-Sport o Msport	Mención de acabado o paquete deportivo

Los indicadores extraídos del título miden una mención comercial. Un cero significa que no se detecta la mención; no certifica ausencia del equipamiento. El 320d puede incluir variantes de hibridación ligera y permanece dentro de la categoría diésel.

3. Datos, procedencia y reglas de depuración

Se preservaron tres hojas Excel y el programa del proyecto original [1]. Los archivos alemán y español constan en Drive desde el 3 y 4 de febrero de 2026; esa metainformación no acredita la fecha exacta de captura de cada anuncio. La ampliación se ejecutó el 5 de octubre de 2026 y no representa una cotización actualizada del mercado.

Archivo	Registros	Función en el estudio
BMW 320 alemanes.xlsx	1.355	Fuente de la muestra econométrica
BMW SERIE 3 ESPAÑA.xlsx	119	Auditoría de comparabilidad internacional
mejores_opciones_importacion_bmw320.xlsx	1.352	Resultado original reproducido
analyze_margins.py	Programa	Reglas originales de precios y emparejamiento

3.1. Selección de la muestra alemana

Regla aplicada, en este orden	Excluidos	Restantes
Registros originales, todos con ID distinto	0	1.355
Títulos con otro modelo explícito (318d o 330d)	6	1.349
Versión 320 no identificable con la regla conservadora	32	1.317
Mención explícita de precio neto (NETTO)	2	1.315
Valores numéricos inválidos o 320e anterior a 2021	0	1.315
Título normalizado, precio, año, km y combustible idénticos	26	1.289

La extracción del modelo reconoce 320d, 320i, 320e y la abreviatura 320xd. Se excluyen títulos con códigos de otros modelos aunque comiencen por un prefijo 320. No se infiere una versión solo por el combustible. Se mantiene una fila por duplicado exacto; un ID distinto no garantiza un vehículo distinto.

Los filtros numéricos exigen $5.000 < \text{precio} < 80.000$ €, año entre 2019 y 2025 y $0 \leq \text{km} \leq 150.000$. Ningún registro adicional se elimina por esos filtros. El rango observado final es 15.780-48.890 € y 34-150.000 km. El límite de kilometraje observado puede reflejar filtros anteriores de captura, cuya documentación no se conserva.

3.2. Corrección de motorización y base de selección

En la base original, 253 registros 320e se etiquetan como gasolina y uno como eléctrico. BMW documenta el 320e como híbrido enchufable, comercializado desde marzo de 2021 [3]. La muestra final contiene 247 anuncios 320e. Se modela la motorización sin exigir que el anuncio tenga un comparable español, evitando seleccionar observaciones por disponibilidad de la referencia internacional.

4. Composición de la muestra y soporte común

Versión	n	Mediana €	Mediana km	xDrive ¹	M Sport ¹
320e	247	27.990	64.693	19,8 %	25,9 %
320i	199	32.500	52.000	1,5 %	36,7 %
320d	843	28.850	85.155	43,7 %	28,5 %

¹ Proporción con mención en el título. La mediana del año es 2022 en las tres versiones. Muestra total: 1.289 anuncios. Fuente de todas las tablas y figuras empíricas: cálculo propio sobre los archivos [1].

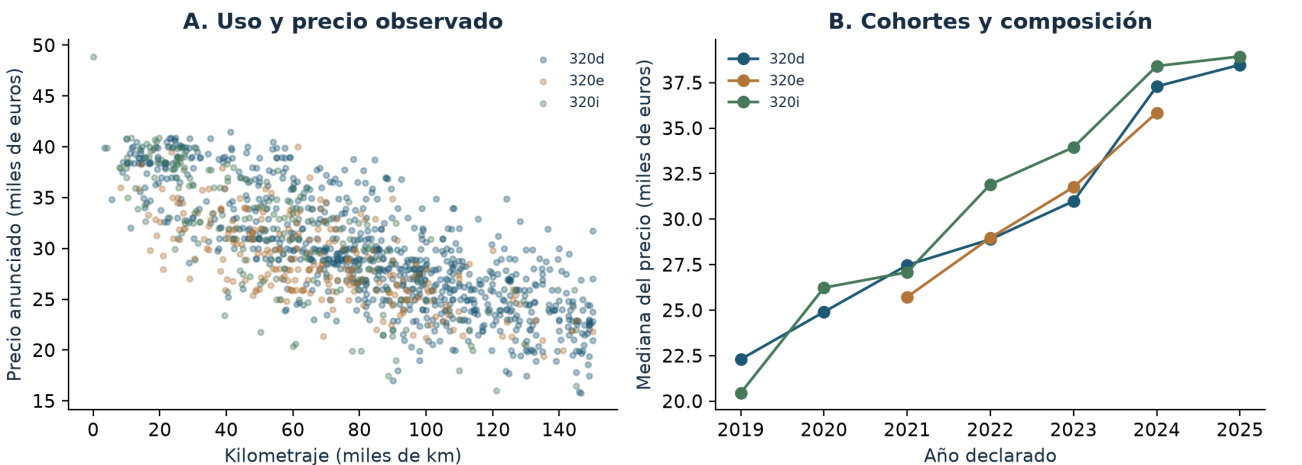


Figura 1. Precios y kilometraje observados, y medianas por año y versión. Las líneas unen cohortes de vehículos distintos; no representan la trayectoria del mismo coche.

El diésel reúne 843 anuncios, el 65,4 % de la muestra. Su kilometraje mediano supera al de las otras motorizaciones. La correlación año-kilometraje es -0,606: ambos atributos están relacionados y deben introducirse conjuntamente. El precio bruto mediano del 320i es mayor, pero también presenta menos uso y más menciones de M Sport.

Año	320e	320i	320d	Total
2019	0	19	61	80
2020	0	12	133	145
2021	46	26	160	232
2022	186	60	239	485
2023	8	27	137	172
2024	7	40	78	125
2025	0	15	35	50

La comparación de tecnologías exige cautela: 186 de los 247 híbridos enchufables pertenecen a 2022, y no hay 320e en 2019, 2020 ni 2025. El modelo impone una diferencia de versión común entre años. Por ello se añade una sensibilidad limitada a 2021-2022, donde hay observaciones de las tres versiones.

5. Especificaciones y estrategia de estimación

5.1. Ecuación de interpretación: modelo M3

$$\ln(P_i) = \alpha + \beta_1 k_i + \beta_2 k_i^2 + \sum_{t=2020..2025} \gamma_t D_{ti} + \delta_g I_{320i,i} + \delta_d I_{320d,i} + \theta X_i + \varphi M_i + u_i$$

P es el precio en euros; k es el kilometraje centrado; D representa el año; X y M son las menciones de xDrive y M Sport. Las categorías de referencia son 2019 y 320e. La constante implica una combinación sin observaciones (320e de 2019) y no se interpreta como valoración de un vehículo real.

Modelo	Variable dependiente	Regresores adicionales al año
M0	Precio en euros	k + combustible original (dos grupos)
M1	Log-precio	k + combustible original (dos grupos)
M2	Log-precio	k + versión + xDrive + M Sport
M3	Log-precio	k + k ² + versión + xDrive + M Sport
M4	Log-precio	M3 + interacciones k × versión
M5	Precio en euros	k + k ² + versión + xDrive + M Sport

M0 y M1 conservan como contraste la división original diésel/gasolina, que agrupa el 320e con el 320i. El registro originalmente eléctrico se incorpora a ese grupo solo para reproducir una clasificación binaria. En M2-M5 se utiliza la versión corregida. Las versiones y sus tres categorías tecnológicas coinciden en esta muestra: incluir ambos juegos de ficticias produciría colinealidad perfecta. No se puede aislar un efecto puro del combustible de las demás diferencias entre versiones.

5.2. Inferencia, comparación y predicción

Todas las especificaciones utilizan la misma muestra. M3 sirve como descripción semilogarítmica flexible del kilometraje y M4 contrasta pendientes distintas. Los intervalos son del 95 %, con covarianza HC3 y referencia t; los contrastes conjuntos emplean Wald F robusto. Se ajustan mediante Holm los cuatro contrastes de año, versión, curvatura e interacciones. Son análisis exploratorios, sin registro previo de hipótesis.

Para evaluar la predicción se reserva el 20 % de grupos con semilla 2102. En el resto se realiza validación cruzada de cinco particiones. Se escoge el modelo con menos parámetros dentro de un error estándar del mejor MAE medio. El test no interviene en esa selección. Los R² en niveles y logaritmos no son directamente comparables; la comparación entre escalas se realiza en euros fuera de muestra.

Los modelos logarítmicos se convierten a euros mediante $\exp(\text{predicción}) \times \text{media de } \exp(\text{residuo})$, calculada solo con cada entrenamiento. Esta corrección global de retransformación aproxima la media en euros; con heteroscedasticidad puede introducir sesgos por subgrupo.

6. Estimaciones y contraste de hipótesis

Tabla 1. MCO semilogarítmico con kilometraje cuadrático (M3). Variable dependiente: logaritmo natural del precio alemán. n = 1.289; 13 parámetros.

Variable	Coefficiente	EE HC3	IC 95 %	p
Constante	9,944407	0,014199	[9,916551; 9,972264]	< 0,001
Año 2020 vs. 2019	0,159339	0,014496	[0,130900; 0,187779]	< 0,001
Año 2021 vs. 2019	0,182103	0,013936	[0,154763; 0,209443]	< 0,001
Año 2022 vs. 2019	0,256231	0,013471	[0,229803; 0,282660]	< 0,001
Año 2023 vs. 2019	0,283580	0,014474	[0,255184; 0,311976]	< 0,001
Año 2024 vs. 2019	0,336043	0,015196	[0,306231; 0,365855]	< 0,001
Año 2025 vs. 2019	0,312723	0,017617	[0,278162; 0,347285]	< 0,001
320i frente a 320e	0,053464	0,007516	[0,038719; 0,068210]	< 0,001
320d frente a 320e	0,071169	0,006070	[0,059261; 0,083078]	< 0,001
k (10.000 km, centrado)	-0,026307	0,000827	[-0,027930; -0,024684]	< 0,001
k ²	-0,000428	0,000180	[-0,000782; -0,000075]	0,018
Mención xDrive	0,025461	0,005412	[0,014844; 0,036077]	< 0,001
Mención M Sport	0,123863	0,004766	[0,114512; 0,133214]	< 0,001

R² = 0,8241; R² ajustado = 0,8224. Los coeficientes se expresan en unidades de log-precio, no directamente en euros ni porcentajes. Los p-valores individuales son nominales; la corrección Holm se aplica a los cuatro contrastes de la tabla siguiente.

6.1. Contrastes conjuntos robustos

Hipótesis	Modelo	F robusto	p	p Holm
Año	M3	112,79	< 0,001	< 0,001
Motorización	M3	69,38	< 0,001	< 0,001
Curvatura	M3	5,64	0,018	0,018
Pendientes por versión	M4	4,77	0,009	0,017

Los cuatro contrastes se rechazan al 5 % tras el ajuste Holm. El término cuadrático y las interacciones aportan evidencia de asociaciones no constantes, aunque su mejora predictiva es pequeña. La significación estadística no demuestra causalidad ni garantiza una especificación completa.

7. Magnitudes económicas e interpretación

7.1. Kilometraje: cambios finitos, no una pendiente universal

Comparación	Cambio ajustado	IC 95 %
40.000 a 50.000 km	-2,30 %	[-2,59; -2,02] %
80.000 a 90.000 km	-2,64 %	[-2,80; -2,48] %
120.000 a 130.000 km	-2,97 %	[-3,32; -2,62] %

En M3, el cambio porcentual exacto es $100 \times \{\exp[\beta_1(k_2 - k_1) + \beta_2(k_2^2 - k_1^2)] - 1\}$. Los intervalos utilizan la covarianza conjunta de ambos coeficientes. El descuento asociado a 10.000 km crece en valor relativo dentro del rango estudiado. No describe la depreciación causal que sufriría un mismo vehículo al circular esos kilómetros.

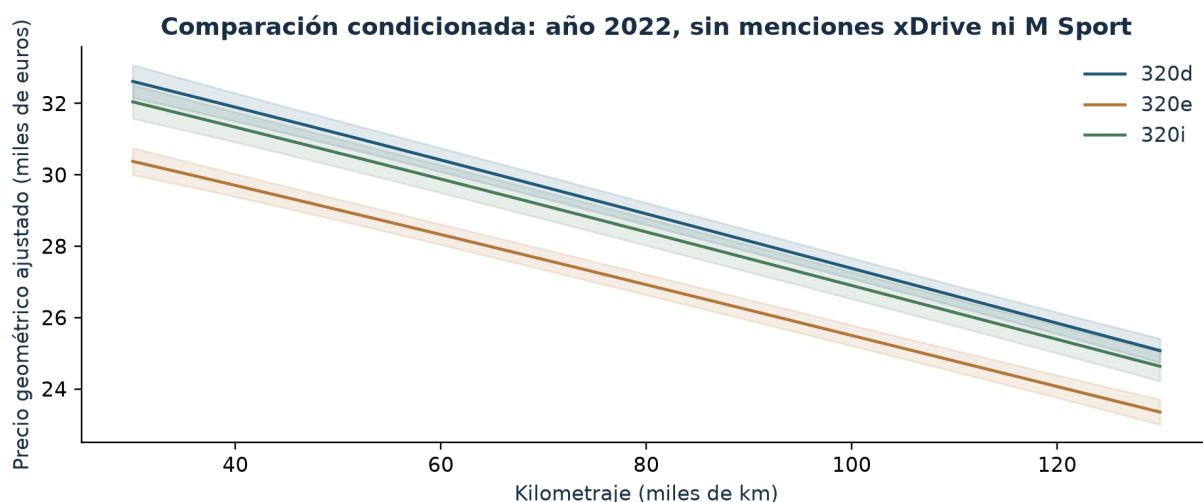


Figura 2. Perfil de precio geométrico ajustado, $\exp(E \text{ estimada de log-precio})$, en la cohorte 2022. Bandas: IC 95 % de la función ajustada HC3, no intervalos de predicción de un anuncio individual. Se mantiene un mismo perfil de menciones en el título.

7.2. Año, versión y equipamiento

La cohorte 2024 presenta un precio ajustado 16,6 % superior a 2021, con IC [14,4; 18,9] %. Esto combina edad, posibles cambios de producto y composición. El coeficiente de 2025 es inferior al de 2024; no se fuerza una tendencia monótona ni se deduce una tasa anual de depreciación.

Frente al 320e, la diferencia del 320d es 7,4 % [6,1; 8,7] % y la del 320i es 5,5 % [3,9; 7,1] %. Para menciones xDrive y M Sport, las asociaciones son 2,6 % y 13,2 %. Se calculan como $100 \times [\exp(\text{coeficiente}) - 1]$.

M Sport puede recoger otros extras, selección del vendedor y estilo de anuncio. La mención xDrive es poco frecuente en el 320i (3 anuncios); la estimación agrupada proviene sobre todo de las otras versiones. Las diferencias tecnológicas no son experimentos de sustitución de combustible.

8. Validación predictiva fuera de muestra

Modelo	Parámetros	MAE CV €	EE del MAE €	RMSE CV €
M0	9	2.291	32	2.824
M1	9	2.302	35	2.841
M2	12	1.810	44	2.283
M3	13	1.805	43	2.274
M4	15	1.810	44	2.273
M5	13	1.796	32	2.258

Tabla 2. Media de cinco particiones agrupadas en 1.030 anuncios de entrenamiento. El error estándar se obtiene de la variación entre particiones y se emplea como regla heurística de parsimonia; no es un contraste de igualdad entre modelos.

M5 obtiene el menor MAE CV (1.796 €). M2 tiene un parámetro menos y su MAE (1.810 €) queda dentro del umbral de 1.828 €: mejor MAE más su error estándar (32 €). Se selecciona M2 antes de evaluar el test. M3 se mantiene para interpretar la curvatura; no se presenta como ganador predictivo.

Indicador del test reservado	M2	Referencia: mediana
Anuncios / grupos	259 / 258	Los mismos
MAE en euros	1.814	4.826
RMSE en euros	2.259	5.880
R ² del test	0,852	--
Mediana de error porcentual absoluto	5,4 %	--

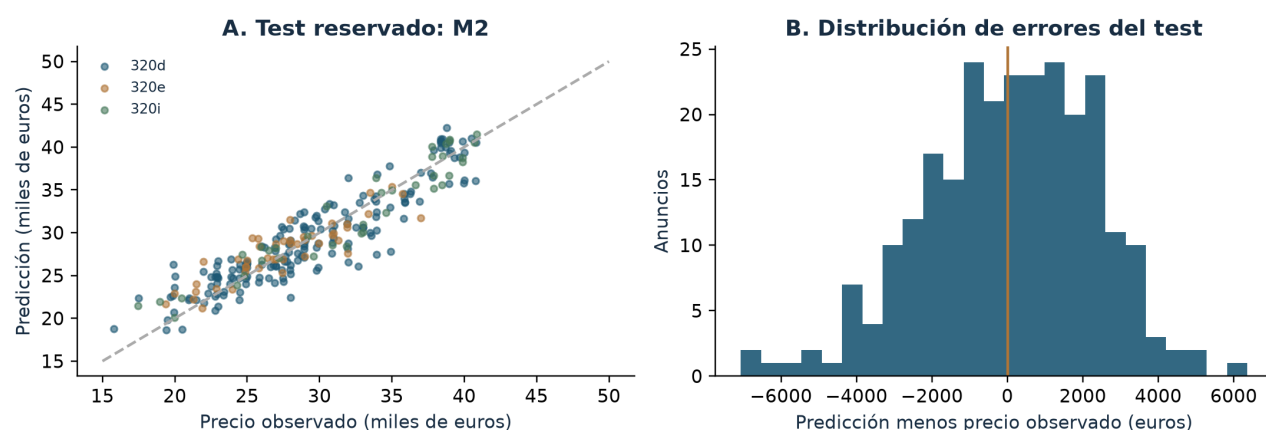


Figura 3. Predicciones en los anuncios reservados. La diagonal señala predicción perfecta. Las mayores desviaciones recuerdan que un buen ajuste agregado no elimina errores de valoración individual.

El MAE tiene un intervalo bootstrap por grupos de 1.655-1.978 € (2.000 remuestras, 95 %). Ese intervalo refleja variación del test con el modelo entrenado fijo. La reducción de MAE frente a la mediana es 62,4 %. Los MAE por versión son 1.544 € (320e, n=46), 1.672 € (320i, n=39) y 1.917 € (320d, n=174).

9. Diagnósticos: qué resuelve y qué queda abierto

Modelo	Koenker-BP, p	RESET HC3, p	VIF máximo	Cook > 4/n
M0	< 0,001	< 0,001	4,63	47
M2	< 0,001	< 0,001	4,83	70
M3	< 0,001	< 0,001	4,83	64
M4	< 0,001	< 0,001	8,75	68
M5	0,032	< 0,001	4,83	61

Koenker-Breusch-Pagan contrasta varianza residual constante frente a asociación con los regresores. RESET añade el cuadrado y el cubo de los valores ajustados estandarizados y contrasta ambos términos conjuntamente con HC3. Centrar y escalar evita problemas numéricos al elevar precios en euros al cubo. Los umbrales Cook son señales de revisión, no pruebas de error de un registro.

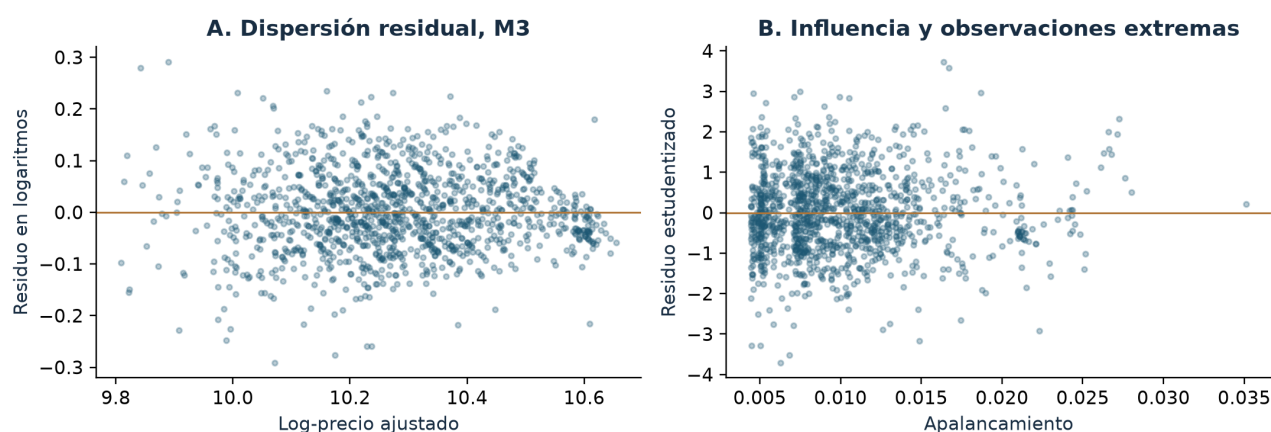


Figura 4. Residuos y apalancamiento en la especificación M3. Los gráficos complementan los contrastes y orientan la sensibilidad de la página siguiente.

9.1. Heteroscedasticidad

M3 rechaza homoscedasticidad ($LM = 76,09$; $p < 0,001$). Los errores estándar convencionales no son la referencia del estudio; se informa HC3 [5]. Esta corrección protege la inferencia ante varianza no constante bajo condiciones regulares, pero no corrige variables omitidas ni errores de medida. M5 también rechaza al 5 %, con $p = 0,032$.

9.2. Especificación y dependencia

RESET rechaza incluso con curvatura e interacciones: para M3, $F = 17,01$ y $p < 0,001$. Por tanto, la ecuación no se considera una descripción estructural completa del precio. Pueden faltar características, interacciones o formas funcionales. Se conservan el diagnóstico y las estimaciones; no se selecciona una variante solo para lograr un p-valor favorable.

M3 tiene rango completo (13 columnas) y VIF máximo 4,83. M4 llega a 8,75, consistente con mayor colinealidad de las interacciones. No se aplica Durbin-Watson al orden de las filas: son datos de sección cruzada y no hay una secuencia temporal documentada. Se desconoce el vendedor de cada anuncio, de modo que la dependencia por concesionario no puede estimarse.

10. Robustez, solapamiento y observaciones influyentes

Muestra	n	β de k	320d vs e ¹	320i vs e ¹	M Sport ¹
Principal	1.289	-0,0263	7,4 %	5,5 %	13,2 %
Sin alertas Cook	1.225	-0,0273	7,9 %	6,0 %	13,3 %
Un anuncio por grupo	1.286	-0,0262	7,3 %	5,5 %	13,2 %
Años 2021-2025	1.064	-0,0265	7,2 %	6,3 %	13,1 %
Años 2021-2022	717	-0,0281	7,9 %	6,5 %	13,5 %

¹ Transformación exacta de los coeficientes semilogarítmicos a porcentaje. Las sensibilidades reestiman M3 y mantienen las categorías de año observadas en cada submuestra. β de k es la pendiente local en log-precio a 80.000 km, no el cambio finito de la Tabla de la sección 7.

10.1. Influencia

64 observaciones superan $\text{Cook} > 4/n$. Se mantienen en la muestra principal porque la alerta no demuestra que sean erróneas. Su exclusión se informa como sensibilidad: el coeficiente local de kilómetros pasa de -0,0263 a -0,0273 y la diferencia diésel-híbrido de 7,4 % a 7,9 %. La dirección y el orden de magnitud permanecen estables; no se afirma inmunidad a todos los posibles errores de captura.

10.2. Duplicación y agrupación

Tras quitar los duplicados exactos quedan 1.286 grupos definidos por versión, año, kilómetros y precio. Tres registros comparten grupo aunque difieran en texto. Esta regla conservadora mantiene anuncios potencialmente repetidos en la misma partición de entrenamiento o test [7]. Puede agrupar vehículos distintos por coincidencia y no detecta repeticiones con precio o kilometraje cambiado.

La reestimación con una fila por grupo altera poco las magnitudes. Una sensibilidad adicional con covarianza agrupada por esa clave produce errores estándar entre 0,988 y 1,002 veces los HC3. El resultado solo evalúa esa repetición observable; no sustituye la agrupación por vendedor ni prueba independencia entre todos los vehículos.

10.3. Cohortes compartidas

La submuestra 2021-2022 contiene 717 anuncios y proporciona presencia de las tres versiones en ambos años. Allí las diferencias frente al 320e son 7,9 % (320d) y 6,5 % (320i), y M Sport se asocia con 13,5 %. Esta comprobación reduce la dependencia de comparaciones entre años sin híbridos enchufables. Sigue sin equiparar batería, carrocería, estado, extras o vendedores.

La estabilidad entre submuestras refuerza la descripción de estas asociaciones dentro de los archivos disponibles. No elimina el sesgo de selección de anuncios ni la incertidumbre por variables no observadas.

11. Auditoría de la comparación Alemania-España

El programa original utiliza combustible, año y kilometraje para localizar anuncios españoles. Busca primero año ± 1 y kilómetros ± 25.000 ; si encuentra menos de tres comparables amplía a ± 2 años y ± 40.000 km. La referencia es la mediana de los precios españoles seleccionados y se resta el precio alemán. No exige el mismo modelo ni la misma carrocería.

Incidencia en España	Recuento	Consecuencia
Registros originales	119	Muestra pequeña y selección sin documentar
Precio ausente	42	Sin referencia monetaria utilizable
Precio superior a 100.000 €	75	Valor sospechoso; reparación no verificada
Precio plausible sin modificación	2	Insuficiente para una regresión española
Kilometraje ausente	1	No permite emparejamiento por uso
Año en celda distinto del visible en URL	32	Inconsistencia, sin poder decidir qué dato es correcto

Las incidencias se solapan. Hay 77 precios no ausentes, de los cuales 75 presentan valores de seis dígitos. El programa los divide por diez y redondea hacia abajo. Por ejemplo, transforma 299.012 en 29.901 €, sin acreditar si el precio correcto era 29.900 €, 29.901 € u otro. Esa regla puede mezclar precio con elementos capturados del anuncio. No se adopta como dato confirmado.

11.1. Reproducción y crítica del resultado original

La reproducción devuelve exactamente las 1.352 referencias del Excel original, con diferencia máxima de 0 € en las medianas españolas estimadas. La diferencia bruta mediana es 3.047 €, y 1.049 diferencias son positivas. En 548 anuncios (40,5 %) hay cinco comparables o menos. Reproducir el cálculo valida su implementación; no valida sus supuestos.

Además del precio, los títulos españoles mezclan 318d, 320d, 320i, 320e y 330i. El 320e alemán se compara con gasolina mientras sus equivalentes españoles figuran como híbridos. Un diferencial alto puede proceder de esta incompatibilidad. Los comparables se reutilizan para numerosos anuncios alemanes, por lo que los 1.352 diferenciales no representan 1.352 referencias españolas independientes.

11.2. Conclusión que sí permite la evidencia

Solo dos registros españoles presentan precio plausible sin modificar, kilometraje disponible y año coherente con la URL. No se estima una prima de país ni rentabilidad neta. Harían falta datos españoles corregidos en origen, equivalencia de versión y carrocería, soporte común, condiciones fiscales homogéneas y costes completos de importación. La diferencia entre precios anunciados tampoco acredita precio de venta ni beneficio obtenido.

12. Discusión, conclusión y lectura para el portfolio

12.1. Aportación y alcance

El estudio aporta una secuencia verificable: auditoría de fuentes, depuración trazable, extracción de características, comparación de modelos, inferencia robusta y evaluación reservada. Año y uso explican una parte sustancial de la dispersión; distinguir la motorización y las menciones de equipamiento mejora la predicción frente a la clasificación binaria original.

El descuento asociado al kilometraje es consistente en las sensibilidades. Las diferencias entre 320d, 320i y 320e son relevantes dentro de la muestra, y M Sport muestra una asociación positiva amplia. La validación del modelo M2 reduce el MAE un 62,4 % frente a la mediana de entrenamiento. Este resultado describe anuncios similares a los observados, no ventas futuras ni todo el mercado europeo.

12.2. Límites que condicionan la interpretación

La muestra se concentra en anuncios alemanes con títulos mayoritariamente Touring y tiene filtros de captura no documentados. No se observan precio de cierre, vendedor, historial de daños, mantenimiento, batería, potencia exacta, garantía, equipamiento completo ni condición fiscal uniforme. Excluir dos menciones NETTO no certifica que el resto de precios sean brutos comparables. Esos atributos pueden estar relacionados con año, versión y precio.

La heteroscedasticidad y el rechazo de RESET señalan que la relación estimada es incompleta. Los intervalos robustos no incluyen incertidumbre por medición, selección o elección del conjunto de modelos. Las diferencias por combustible están unidas a la versión comercial. La generalización temporal exigiría capturas repetidas y una evaluación en fechas posteriores.

12.3. Miniartículo para acompañar el proyecto web

Qué revela un anuncio sobre el precio de un coche usado

Una comparación de coches usados empieza con los precios, pero solo gana valor cuando examina cómo se han obtenido los datos. En este proyecto se revisó una base de 1.355 anuncios de BMW 320 y se detectaron repeticiones, títulos ambiguos y una clasificación del híbrido enchufable como gasolina. La depuración dejó 1.289 anuncios para estudiar sus características de forma conjunta.

El análisis relaciona 10.000 km adicionales, entre 80.000 y 90.000 km, con un precio ajustado un 2,64 % menor. También encuentra diferencias por cohorte, motorización y menciones de equipamiento. Para comprobar su utilidad se reservaron 259 anuncios: el modelo elegido predice sus precios con un error absoluto medio de 1.814 €, frente a 4.826 € al usar únicamente la mediana de entrenamiento.

La revisión más importante afecta a la comparación internacional. Los precios españoles contenían suficientes incidencias como para impedir una conclusión fiable sobre beneficio de importación. El valor del proyecto está en combinar econometría, validación y criterio sobre la calidad de los datos: una predicción útil exige explicar también dónde termina su evidencia.

Descripción breve para CV: Proyecto de econometría aplicada a precios de vehículos usados: auditoría de 1.355 anuncios, modelos hedónicos con inferencia HC3 y validación agrupada; MAE de 1.814 € en una muestra reservada de 259 anuncios.

Referencias y anexo de reproducibilidad

[1] **Datos y programa del proyecto original.** BMW 320 alemanes.xlsx; BMW SERIE 3 ESPAÑA.xlsx; mejores_opciones_importacion_bmw320.xlsx; analyze_margins.py. Archivos privados conservados en Drive, febrero de 2026. Consulta: 5-10-2026.

[2] **Apuntes de Econometría (curso 2025-2026).** Temas 1-6, Dpto. de Análisis Económico y Economía Cuantitativa. Material docente privado conservado como Tema_1_GADE_202526_alumnos.pdf a Tema_6_GADE_202526_alumnos.pdf. Pasajes aplicados: T2, pp. 79-82; T3, modelo lineal y omisión; T4, pp. 23-35; T5, heteroscedasticidad e influencia, pp. 24-50.

[3] **BMW Group (27-1-2021).** New entry-level models with plug-in hybrid drive for the BMW 3 Series and BMW 5 Series. Comunicado técnico, pp. 1-3. [Consultar fuente oficial](#).

[4] **Rosen, S. (1974).** Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy, 82(1), 34-55. Referencia fundacional; metadatos editoriales consultados. doi:10.1086/260169.

[5] **statsmodels (documentación oficial, versión 0.15.0).** OLSResults.HC3_se; het_breuschpagan; linear_reset. HC3 según MacKinnon-White; diagnóstico de varianza y forma funcional. HC3 · Koenker-BP · RESET.

[6] **Universidad Carlos III de Madrid (OpenCourseWare).** Econometrics: Lecture notes. Temas 4 y 7: variables cualitativas y heteroscedasticidad. Consulta de contraste metodológico. [Material docente abierto](#).

[7] **scikit-learn (documentación oficial, versión 1.9.1).** GroupShuffleSplit y GroupKFold. Separación y validación por grupos sin solapamiento entre entrenamiento y validación. [Separación](#) · [Validación cruzada](#).

A. Replicación y comprobaciones

El paquete de reproducción contiene los tres Excel originales, el programa original, analisis_reproducible.py, requirements.txt, el generador del PDF y los resultados en CSV y JSON. Los apuntes docentes y el entorno instalado se mantienen fuera del paquete. Ejecutar el análisis con Python y las versiones fijadas regenera muestras, estimaciones, pruebas, predicciones y figuras; ejecutar la maquetación construye este informe.

Parámetro o verificación	Valor
Python / semilla	3.12.14 / 2102
Partición y CV	80/20 por grupos; 5 particiones en entrenamiento
Inferencia	HC3, referencia t; Wald F; Holm en 4 contrastes
Grupos sin solapamiento	Verificado mediante intersección vacía
Reproducción del Excel original	1.352 referencias; máximo desvío = 0 €
Archivos de resultados	auditoría, muestra, coeficientes, predicciones, sensibilidad
Versiones de análisis	numpy 2.3.5; pandas 3.0.1; statsmodels 0.15.0; scikit-learn 1.9.1

Las huellas SHA-256 completas de los tres Excel se registran en resultados.json. No se imputan observaciones ni se crean precios sintéticos. Los precios españoles reparados solo se reproducen para auditar el método anterior; no entran en las regresiones ni fundamentan una conclusión de rentabilidad.

Siguiente investigación: recapturar y verificar la muestra española; estructurar carrocería, fiscalidad y equipamiento; incorporar vendedor y estado del vehículo; evaluar generalización temporal con una captura posterior.