

APELLIDOS, NOMBRE: _____ DNI: _____

Cuestiones Teóricas

- 1 (1 punto) Sabiendo que la matriz de varianzas covarianzas de los estimadores de un modelo econométrico en el que se cumplen las hipótesis básicas tiene la siguiente expresión $\text{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$, razone cómo se obtiene dicha expresión y explique el orden y los componentes de dicha matriz.
- 2 (0.75 puntos) Se quiere analizar mediante un modelo econométrico si es cierto que el precio de los combustibles baja determinados días de la semana. Razona la especificación del modelo y define la variable cualitativa a introducir.
- 3 (0.75 puntos) Contraste de Goldfeld-Quandt. Utilidad, hipótesis a contrastar, estadístico experimental y teórico.
- 4 (1 punto) Dado un modelo econométrico $y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ en el que se verifica que $u_t = \rho u_{t-1} + \epsilon_t$, explique para qué se utiliza y en qué consiste la transformación de Prains Winsten y en qué casos se recomienda su aplicación.

Cuestiones Prácticas

Se ha analizado un modelo econométrico para estimar las ventas de cierto producto (en miles de euros) de una empresa en base a la información del los últimos 365 días, en función del *Gasto en Publicidad* (en miles de euros), los *Ingresos Medios* (en miles de euros), el *Precio de la Competencia* (en euros), la evaluación de la *Calidad* del producto (evaluada de 1 a 10 estrellas), y si el día indicado era *Fin de Semana* (1: fin de semana, 0: no fin de semana). Se presentan las siguientes cuestiones sobre distintos resultados obtenidos por statsmodels en python para poder estudiar este modelo.

| | | | |
|-------------------|------------------|---------------------|-----------|
| Dep. Variable: | Log_Ventas | R-squared: | 0.824 |
| Model: | OLS | Adj. R-squared: | 0.822 |
| Method: | Least Squares | F-statistic: | 336.4 |
| Date: | Wed, 11 Dec 2024 | Prob (F-statistic): | 4.47e-133 |
| Time: | 18:42:27 | Log-Likelihood: | 297.36 |
| No. Observations: | 365 | AIC: | -582.7 |
| Df Residuals: | 359 | BIC: | -559.3 |
| Df Model: | 5 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|-------------------|---------|---------|--------|-------|--------|--------|
| const | 4.8529 | 0.073 | 66.094 | 0.000 | 4.708 | 4.997 |
| Publicidad | 0.0094 | 0.001 | 16.592 | 0.000 | 0.008 | 0.011 |
| Log_Ingresos | 0.0298 | 0.014 | 2.201 | 0.028 | 0.003 | 0.056 |
| Precio_Competidor | -0.0057 | 0.001 | -4.911 | 0.000 | -0.008 | -0.003 |
| Calidad | 0.0153 | 0.004 | 4.068 | 0.000 | 0.008 | 0.023 |
| Fin_de_Semana | 0.3953 | 0.028 | 14.296 | 0.000 | 0.341 | 0.450 |

| | | | |
|----------------|---------|-------------------|-----------|
| Omnibus: | 113.090 | Durbin-Watson: | 2.086 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 523.587 |
| Skew: | -1.248 | Prob(JB): | 2.02e-114 |
| Kurtosis: | 8.310 | Cond. No. | 747. |

donde a las variables que vienen precedidas de Log_ se les ha aplicado el logaritmo sobre la variable.

- 1 (0.5 puntos) Evaluar si el modelo es globalmente significativo, e indicar, de forma detallada, el contraste que se utiliza para determinarlo, y qué significa el valor que has utilizado para asegurarlo.
- 2 (0.5 puntos) Escribir la ecuación estimada del modelo en base en los coeficientes obtenidos. Interpretar los coeficientes asociados al factor de si es fin de semana y de los ingresos.
- 3 (0.5 puntos) ¿Como predicirías, utilizando el modelo estimado, las ventas del producto un miércoles en el que el gasto medio en publicidad fuera de 50 mil euros, los ingresos de 85 mil euros, el precio del competidor de 20 euros y la valoración de la calidad de 8 estrellas con el modelo obtenido?
- 4 (0.5 puntos) En los siguientes resultados se presentan pruebas diagnósticas sobre el modelo. Indica si, en base a estas y a otra posible información que conozcas sobre el modelo, crees que existe algún problema y detalla como podría resolverse.

| Variable | VIF |
|-------------------|-----------|
| Publicidad | 16.302959 |
| Ingresos | 10.536765 |
| Precio_Competidor | 14.114326 |
| Calidad | 18.405485 |
| Fin_de_Semana | 3.420166 |

| | | | | | |
|-------------------|-------|------|-------|-------|------|
| Publicidad | 1.00 | 0.01 | -0.06 | -0.01 | 0.06 |
| Log_Ingresos | 0.01 | 1.00 | 0.05 | 0.04 | 0.89 |
| Precio_Competidor | -0.06 | 0.05 | 1.00 | 0.03 | 0.04 |
| Calidad | -0.01 | 0.04 | 0.03 | 1.00 | 0.02 |
| Fin_de_Semana | 0.06 | 0.89 | 0.04 | 0.02 | 1.00 |

- 5 (0.5 puntos) Utilizando los residuos obtenidos en la estimación del modelo, se ajustó el siguiente modelo auxiliar
- $$\text{residuo}^2 = \delta_0 + \delta_1 \text{Gasto en Publicidad} + \delta_2 \text{Log_Ingresos} + \delta_3 \text{Precio Competidor} + \delta_4 \text{Calidad} + \delta_5 \text{Fin de Semana} + v$$

| | | | |
|-------------------|------------------|---------------------|--------|
| Dep. Variable: | e^2 | R-squared: | 0.032 |
| Model: | OLS | Adj. R-squared: | 0.019 |
| Method: | Least Squares | F-statistic: | 2.378 |
| Date: | Wed, 11 Dec 2024 | Prob (F-statistic): | 0.0384 |
| Time: | 19:12:52 | Log-Likelihood: | 755.54 |
| No. Observations: | 365 | AIC: | -1499. |
| Df Residuals: | 359 | BIC: | -1476. |
| Df Model: | 5 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|-------------------|---------|---------|--------|-------|--------|-----------|
| const | 0.0373 | 0.021 | 1.781 | 0.076 | -0.004 | 0.078 |
| Publicidad | -0.0003 | 0.000 | -2.039 | 0.042 | -0.001 | -1.17e-05 |
| Log_Ingresos | -0.0044 | 0.004 | -1.151 | 0.251 | -0.012 | 0.003 |
| Precio_Competidor | 0.0004 | 0.000 | 1.102 | 0.271 | -0.000 | 0.001 |
| Calidad | 0.0004 | 0.001 | 0.416 | 0.678 | -0.002 | 0.003 |
| Fin_de_Semana | 0.0002 | 0.008 | 0.026 | 0.980 | -0.015 | 0.016 |

| | | | |
|----------------|---------|-------------------|------------|
| Omnibus: | 643.027 | Durbin-Watson: | 1.865 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 302352.667 |
| Skew: | 10.335 | Prob(JB): | 0.00 |
| Kurtosis: | 142.476 | Cond. No. | 747. |

¿Qué puedes concluir de los resultados obtenidos? En caso de existir algún problema en el modelo, ¿Cómo lo solventarías?

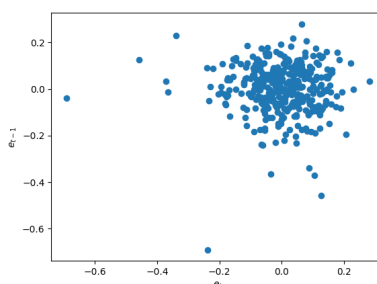
- 6 (0.5 puntos) Se ha aplicado el test de Glejser, obteniendo los siguientes resultados:

| Variable | h | pval | R ² |
|--------------|-----|--------|----------------|
| Publicidad | 2 | 0.0071 | 0.0198 |
| | -1 | 0.0018 | 0.0264 |
| | -2 | 0.0018 | 0.0264 |
| | 0.5 | 0.0046 | 0.0219 |
| Log_Ingresos | 2 | 0.0024 | 0.0250 |
| | -1 | 0.0009 | 0.0302 |
| | -2 | 0.0009 | 0.0302 |
| | 0.5 | 0.0019 | 0.0264 |

| Variable | h | pval | R ² |
|-------------------|-----|--------|----------------|
| Precio_Competidor | 2 | 0.9913 | 3.27e-07 |
| | -1 | 0.7670 | 0.0002 |
| | -2 | 0.7670 | 0.0002 |
| | 0.5 | 0.9410 | 1.51e-05 |
| Calidad | 2 | 0.9228 | 2.59e-05 |
| | -1 | 0.8401 | 0.0001 |
| | -2 | 0.8401 | 0.0001 |
| | 0.5 | 0.9851 | 9.62e-07 |

¿Qué conclusiones tomarías de los resultados obtenidos? ¿Qué metodología aplicarías si los resultados indicaran que existe algún problema en el cumplimiento de las hipótesis del modelo?

- 7 (0.5 puntos) Para analizar la existencia de autocorrelación en el modelo, se ha realizado el gráfico de los residuos con respecto a los residuos retardados un día, obteniendo el siguiente gráfico:



¿Qué puedes decir acerca de la correlación en base a esta gráfico? ¿Hay alguna medida obtenida en el modelo que te permita tomar más conclusiones acerca de la autocorrelación de éste? Justifica cuál es esta medida y razona la existencia o no de autocorrelación en base a esta.

RESOLUCIÓN EXAMEN ADECUADO

CUESTIONES TEÓRICAS

- 1 (1 punto) Sabiendo que la matriz de varianzas covarianzas de los estimadores de un modelo econométrico en el que se cumplen las hipótesis básicas tiene la siguiente expresión $\text{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$, razone cómo se obtiene dicha expresión y explique el orden y los componentes de dicha matriz.

Varianza de los ERCO

$$\text{var}(\vec{\hat{\beta}}) = E[(\vec{\hat{\beta}} - E[\vec{\hat{\beta}}]) \cdot (\vec{\hat{\beta}} - E[\vec{\hat{\beta}}])^t] \rightarrow \text{Sabiendo que la varianza es esto}$$
$$= E[(\vec{\hat{\beta}} - \vec{\beta}) \cdot (\vec{\hat{\beta}} - \vec{\beta})] = \rightarrow E[\vec{\hat{\beta}}] = \vec{\beta} \text{ al ser insesgado}$$

Sabemos que $\vec{\hat{\beta}} - \vec{\beta} = (X'X)^{-1} X' \vec{u}$ \rightarrow Sabemos esto

$$= E[(X'X)^{-1} X' \vec{u} \cdot (\vec{u}' X) (X'X)^{-1}] \rightarrow \text{calculamos}$$
$$= (X'X)^{-1} \underbrace{X' E(\vec{u} \cdot \vec{u}') X}_{\sigma^2} (X'X)^{-1} \rightarrow \text{Sabemos que } E[\vec{u} \cdot \vec{u}^t] = \sigma^2$$
$$= \sigma^2 (X'X)^{-1} = \text{var}(\vec{\hat{\beta}}) \rightarrow \text{Nos queda esto}$$

La matriz $X'X$ es una matriz simétrica definida como:

$$X'X = \begin{bmatrix} n & \sum_{i=0}^n X_{1i} & \dots & \sum_{i=0}^n X_{ki} \\ \sum_{i=0}^n X_{1i} & \sum_{i=0}^n X_{1i}^2 & \dots & \sum_{i=0}^n X_{1i} \cdot X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^n X_{ki} & \sum_{i=0}^n X_{ki} X_{1i} & \dots & \sum_{i=0}^n X_{ki}^2 \end{bmatrix}$$

Queda realizar la inversa que para ello sabemos que cumple con el supuesto de rango por columnas y por consiguiente tiene inversa y hay que medir el producto de σ^2 :

Como vemos hay K filas y K columnas por lo que el orden es $K \times K$.

Sabemos que es resultado de realizar operaciones algebraicas con X , siendo X :

$$X = \begin{pmatrix} 1 & x_{21} & \dots & x_{K1} \\ 1 & x_{22} & \dots & x_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2n} & \dots & x_{Kn} \end{pmatrix}$$

Siendo K el n.º de variables independientes del modelo lineal y n el número de observaciones de la muestra.

2 (0.75 puntos) Se quiere analizar mediante un modelo econométrico si es cierto que el precio de los combustibles baja determinados días de la semana. Razona la especificación del modelo y define la variable cualitativa a introducir.

La pregunta se limita a la especificación del modelo y a definir la variable cualitativa a introducir.

Especificación del modelo

La especificación del modelo trata de expresar matemáticamente la relación entre variables de interés. El Modelo Base será:

- Explicar P_t (precio del combustible) en función de las variables diarias D_t y otros factores ajenos que pueden influir, representados por el término aleatorio u_t :

$$P_t = \beta_0 + \beta_1 D_{1t} + \beta_2 D_{2t} + \dots + \beta_6 D_{6t} + u_t$$

-) $P_t \rightarrow$ precio combustible: variable dependiente
-) $D_{it} \rightarrow$ día de la semana: variable independiente

\rightarrow estas son las **variables cualitativas** ya que toman el valor 0 o 1, dependiendo si es ese día de la semana o no.

\rightarrow Usamos solo 6 días, ya que el 7 corresponde con β_0 que nos ayude a eliminar la multicolinealidad

perfecta, dado que la suma de todas las varianzas sería igual a 1.

Para centrarme a la pregunta del examen
davía por acabade la pregunta, pero estaría
bien añadir la interpretación de los coeficientes,
estimación con MCO, uso del modelo e
Inferencia.

3 (0.75 puntos) Contraste de Goldfeld-Quandt. Utilidad, hipótesis a contrastar, estadístico experimental y teórico.

Goldfeld - Quant. (Explicamos tal y como se
nos indica).

- Utilidad: Es para muestras pequeñas y si sospechamos
que σ^2_t está relacionada positivamente con la
variable X_i . Pasos:

1. Ordenamos los datos muestrales de menor
a mayor.
2. Quitamos los n centrales.
3. Ajustamos mediante MCO. obteniendo
 SCM y SCM_2

4. Contrastamos la hipótesis (de continuidad).

- Hipótesis a contrastar:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \text{ (homocedasticidad)}$$

$$H_1: \exists \sigma_i^2 \neq \sigma_1^2 \text{ (heterocedasticidad)}$$

- Estadísticas

$$*) \text{ Experimental: } F_{\text{exp}} = \frac{SCN_2}{SCN_1}$$

$$*) \text{ Teórico: } F_{\text{teo}} = F_{\frac{n-m-k}{2}, \frac{n-m}{2}, 1-\alpha}$$

4 (1 punto) Dado un modelo econométrico $y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ en el que se verifica que $u_t = \rho u_{t-1} + \varepsilon_t$, explique para qué se utiliza y en qué consiste la transformación de Prais Winsten y en qué casos se recomienda su aplicación.

$$\begin{cases} y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t \end{cases}$$

1. Explicación del modelo econométrico:

$y_t \rightarrow$ variable dependiente

$X_{2t}, X_{3t} \rightarrow$ variables explicativas

$u_t \rightarrow$ término de error, que sigue un proceso de autocorrelación de primer orden.

$$u_t = p_{u_{t-1}} + \varepsilon_t$$

$\rightarrow p \rightarrow$ mide la relación entre el error actual y el del periodo anterior

$\rightarrow \varepsilon_t \rightarrow$ es el ruido blanco que es no correlacionado y con media 0.

2. Problema del Modelo

Cuando podemos ver el problema presenta uno de los problemas de autocorrelación ya que uno de los supuestos es que los errores no pueden ser dependientes entre sí, el cual en este caso no se cumple, por lo que:

- El estimador sigue siendo insesgado, pero ineficiente.
- Problemas en cuanto a las conclusiones inferenciales

3. Prais Winsten

3.1. Para que se utiliza

Se usa para eliminar la autocorrelación de 1º orden en el término de error (u_t)

3.2. En que consiste la transformaci3n

Pasos de la transformaci3n:

1. Primera observaci3n ($t=1$)

Ajuste la 1era observaci3n para no producir, a diferencia de otras t3cnicas como Cochrane-Orcutt

Para ello modificamos las variables y_t (dependiente) e (independientes) X_{jt} :

$$y_1^* = \sqrt{1-p^2} \cdot y_1$$

$$X_{j1}^* = \sqrt{1-p^2} \cdot X_{j1} \quad j=2,3$$

2. Observaciones restantes

transformamos el modelo usando p :

$$y_t^* = y_t - p y_{t-1}$$

$$X_{jt}^* = X_{jt} - p X_{j(t-1)} \quad j=2,3$$

3. Modelo Transformado

obtenemos un nuevo m3todo sin autocorrelaci3n:

$$y_t^* = \beta_1(1-p) + \beta_2 X_{2t}^* + \beta_3 X_{3t}^* + u_t^*$$

u_t^* es independiente \rightarrow no presenta autocorrelaci3n.

3.3 Casos donde se recomienda su aplicación:

3.3.1 Autocorrelación de 1º orden detectada.

3.3.2. Dependencia de los errores en datos temporales

3.3.3 Procesos donde se quiere conservar datos.
ya que conserva + datos que Cochrane
Orcutt

EXTAD:

Pasos:

1. Estimamos por MCO y se obtienen los residuos e_t

2. Se estima $u_t = \rho u_{t-1} + \varepsilon_t$, siendo el valor de ρ empleando los e_t :

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_t^2}$$

3. Se estima por MCO $y^* = X^* \beta + u^*$: e_t^*

4. Estimamos $\hat{\rho}$ de la regresión $e_t^* = \hat{\rho} e_{t-1}^* + \varepsilon_t^*$:

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t^* e_{t-1}^*}{\sum_{t=2}^n e_t^{*2}}$$

5. Repetimos... la estimación de ρ es el valor que estabiliza la sucesión $\hat{\rho}, \hat{\hat{\rho}}, \hat{\hat{\hat{\rho}}}, \dots$ hasta que la diferencia sea $< 10^{-3}$.

Debemos de dar la primera observación en las transformaciones \rightarrow Proceso iterativo de Cochrane-Orcutt.

CUESTIONES PRÁCTICAS

Cuestiones Prácticas

Se ha analizado un modelo econométrico para estimar las ventas de cierto producto (en miles de euros) de una empresa en base a la información de los últimos 365 días, en función del *Gasto en Publicidad* (en miles de euros), los *Ingresos Medios* (en miles de euros), el *Precio de la Competencia* (en euros), la evaluación de la *Calidad* del producto (evaluada de 1 a 10 estrellas), y si el día indicado era *Fin de Semana* (1: fin de semana, 0: no fin de semana). Se presentan las siguientes cuestiones sobre distintos resultados obtenidos por statsmodels en python para poder estudiar este modelo.

| | | | |
|-------------------|------------------|---------------------|-----------|
| Dep. Variable: | Log_Ventas | R-squared: | 0.824 |
| Model: | OLS | Adj. R-squared: | 0.822 |
| Method: | Least Squares | F-statistic: | 336.4 |
| Date: | Wed, 11 Dec 2024 | Prob (F-statistic): | 4.47e-133 |
| Time: | 18:42:27 | Log-Likelihood: | 297.36 |
| No. Observations: | 365 | AIC: | -582.7 |
| Df Residuals: | 359 | BIC: | -559.3 |
| Df Model: | 5 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|-------------------|---------|---------|--------|-------|--------|--------|
| const | 4.8529 | 0.073 | 66.094 | 0.000 | 4.708 | 4.997 |
| Publicidad | 0.0094 | 0.001 | 16.592 | 0.000 | 0.008 | 0.011 |
| Log_Ingresos | 0.0298 | 0.014 | 2.201 | 0.028 | 0.003 | 0.056 |
| Precio_Competidor | -0.0057 | 0.001 | -4.911 | 0.000 | -0.008 | -0.003 |
| Calidad | 0.0153 | 0.004 | 4.068 | 0.000 | 0.008 | 0.023 |
| Fin_de_Semana | 0.3953 | 0.028 | 14.296 | 0.000 | 0.341 | 0.450 |

| | | | |
|----------------|---------|-------------------|-----------|
| Omnibus: | 113.090 | Durbin-Watson: | 2.086 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 523.587 |
| Skew: | -1.248 | Prob(JB): | 2.02e-114 |
| Kurtosis: | 8.310 | Cond. No. | 747. |

donde a las variables que vienen precedidas de Log_ se les ha aplicado el logaritmo sobre la variable.

- 1 (0.5 puntos) Evaluar si el modelo es globalmente significativo, e indicar, de forma detallada, el contraste que se utiliza para determinarlo, y qué significa el valor que has utilizado para asegurarlo.
- 2 (0.5 puntos) Escribir la ecuación estimada del modelo en base en los coeficientes obtenidos. Interpretar los coeficientes asociados al factor de si es fin de semana y de los ingresos.
- 3 (0.5 puntos) ¿Como predicirías, utilizando el modelo estimado, las ventas del producto un miércoles en el que el gasto medio en publicidad fuera de 50 mil euros, los ingresos de 85 mil euros, el precio del competidor de 20 euros y la valoración de la calidad de 8 estrellas con el modelo obtenido?

1. Utilizamos el contraste:

$$\left\{ \begin{array}{l} H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0 \\ H_1: \exists \beta_i \neq 0 \text{ con } i = 1, \dots, 5. \end{array} \right.$$

$$\left\{ \begin{array}{l} \text{Valor F-statistic: } 336,4 \\ \text{P valor asociado: } 4,47 \cdot 10^{-133} \end{array} \right.$$

Como el P valor es muy pequeño podemos rechazar la H_0 , solo lo aceptamos cuando nuestro Pval es $> 0,05, > 0,01, \dots$ (niveles de significancia comunes vistos al darte).

Si, es globalmente significativo, las variables explicativas tienen un efecto conjunto significativo sobre los-ventas.

2. Ecuación estimada del modelo:

$$\text{log-ventas} = 4,8529 + 0,0094 \cdot \text{Publicidad} + 0,0298 \cdot \text{log-Ingresos} - 0,0057 \cdot \text{Precio-Competidor} + 0,0153 \cdot \text{Calidad} + 0,3953 \cdot \text{Fin-de-semana}$$

Interpretación:

- Fin-de-semana (0,3953)

*) Es dicotómica, puede valer 1 o 0, por lo que podemos ver que aumenta las

el promedio \leftarrow ventas en un 39,53% en promedio cuando su valor es = 1. la cantidad que cuando \leftarrow finde aumenta es $\leftarrow \left[e^{0,3953} - 1 \right] \approx 48,4\%$.

$$- \log - \text{Ingresos}(0,0298)$$

* Un aumento el 1% en los ingresos medios diarios está asociado con un aumento del 0,0298 %.
(se queda igual debido al logaritmo).

3. Sustituir valores:

$$\log - \text{ventas} = 7,8529 + 0,0094 \cdot 50 + 0,0298$$

$$+ \log(85) - 0,0257 \cdot 20 + 0,0153 \cdot 8 + 0 \approx$$

↓
debido a

que 0,0298 es el logaritmo.

↓
no es
lin de
seguir.

$\approx 4,99 \rightarrow$ DEBEROS de pasar a ventas

$$\boxed{\text{Ventas} = e^{7,9882} \approx 146,5 \text{ mil euros}}$$

- 4 (0.5 puntos) En los siguientes resultados se presentan pruebas diagnósticas sobre el modelo. Indica si, en base a estas y a otra posible información que conozcas sobre el modelo, crees que existe algún problema y detalla como podría resolverse.

Examen adelantado - 16/12/2024

| Variable | VIF |
|-------------------|-----------|
| Publicidad | 16.302959 |
| Ingresos | 10.536765 |
| Precio_Competidor | 14.114326 |
| Calidad | 18.405485 |
| Fin_de_Semana | 3.420166 |

| | | | | | |
|-------------------|-------|------|-------|-------|------|
| Publicidad | 1.00 | 0.01 | -0.06 | -0.01 | 0.06 |
| Log_Ingresos | 0.01 | 1.00 | 0.05 | 0.04 | 0.89 |
| Precio_Competidor | -0.06 | 0.05 | 1.00 | 0.03 | 0.04 |
| Calidad | -0.01 | 0.04 | 0.03 | 1.00 | 0.02 |
| Fin_de_Semana | 0.06 | 0.89 | 0.04 | 0.02 | 1.00 |

Matriz de correlación
(valores oscilan entre
1 y -1)

1 → correlación positiva
(variables ↑ juntas)

-1 → " negativa
(↑ aumento y otra ↓)

-0 → no relación lineal

VIF → >10 → problemas de colinealidad (graves)
→ entre 5 y 10 → problemas moderados
→ <5 → NO problemas

Vemos que según la matriz Log-Ingresos y Fin de Semana presenta problemas de colinealidad

Luego, en función del VIF:

* Publicidad, Ingresos, Precio-Competidor y Calidad tienen $VIF > 10$ = problemas graves multicolinealidad

* Fin de semana VIF bajo

¿Cómo resolverlo?

- Eliminar variables altamente correlacionadas.
- Transformar las variables, usando por ej. PC.
- Disminuir el modelo \rightarrow evitar redundancia.

5 (0.5 puntos) Utilizando los residuos obtenidos en la estimación del modelo, se ajustó el siguiente modelo auxiliar $\text{residuo}^2 = \delta_0 + \delta_1 \text{Gasto en Publicidad} + \delta_2 \text{Log-Ingresos} + \delta_3 \text{Precio Competidor} + \delta_4 \text{Calidad} + \delta_5 \text{Fin de Semana} + v$

| | | | |
|-------------------|------------------|---------------------|--------|
| Dep. Variable: | e ² | R-squared: | 0.032 |
| Model: | OLS | Adj. R-squared: | 0.019 |
| Method: | Least Squares | F-statistic: | 2.378 |
| Date: | Wed, 11 Dec 2024 | Prob (F-statistic): | 0.0384 |
| Time: | 19:12:52 | Log-Likelihood: | 755.54 |
| No. Observations: | 365 | AIC: | -1499. |
| Df Residuals: | 359 | BIC: | -1476. |
| Df Model: | 5 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|-------------------|---------|---------|--------|-------|--------|-----------|
| const | 0.0373 | 0.021 | 1.781 | 0.076 | -0.004 | 0.078 |
| Publicidad | -0.0003 | 0.000 | -2.039 | 0.042 | -0.001 | -1.17e-05 |
| Log_Ingresos | -0.0044 | 0.004 | -1.151 | 0.251 | -0.012 | 0.003 |
| Precio_Competidor | 0.0004 | 0.000 | 1.102 | 0.271 | -0.000 | 0.001 |
| Calidad | 0.0004 | 0.001 | 0.416 | 0.678 | -0.002 | 0.003 |
| Fin_de_Semana | 0.0002 | 0.008 | 0.026 | 0.980 | -0.015 | 0.016 |

| | | | |
|----------------|---------|-------------------|------------|
| Omnibus: | 643.027 | Durbin-Watson: | 1.865 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 302352.667 |
| Skew: | 10.335 | Prob(JB): | 0.00 |
| Kurtosis: | 142.476 | Cond. No. | 747. |

¿Qué puedes concluir de los resultados obtenidos? En caso de existir algún problema en el modelo, ¿Cómo lo solventarías?

5. Debemos de analizar los p-valores de las variables ya que un p valor pequeño, menor a 0,05 sugiere que la variable asociada tiene un efecto significativo en los residuos al cuadrado, si no es significativo quiere decir que el modelo presenta heterocedasticidad.

Esto se debe a que el ajuste expone que los residuos² se pueden expresar en función de las variables explicativas por lo que si es significativo, quiere decir que hay heterocedasticidad.

- Aunque el R^2 es pequeño no debemos de descartar la posible presencia de heterocedasticidad.

| | coef | std err | t | P> t |
|-------------------|---------|---------|--------|-------|
| const | 0.0373 | 0.021 | 1.781 | 0.076 |
| Publicidad | -0.0003 | 0.000 | -2.039 | 0.042 |
| Log_Ingresos | -0.0044 | 0.004 | -1.151 | 0.251 |
| Precio_Competidor | 0.0004 | 0.000 | 1.102 | 0.271 |
| Calidad | 0.0004 | 0.001 | 0.416 | 0.678 |
| Fin_de_Semana | 0.0002 | 0.008 | 0.026 | 0.980 |

Vemos que Publicidad influye, por lo que podemos afirmar que efectivamente hay heteroskedasticidad.

Soluciones:

1. Transformaciones de datos : garantizamos que los datos después de las transformaciones son homocedásticos:

$$y^* = \frac{y}{\sqrt{w_i}} \quad x^* = \frac{x}{\sqrt{w_i}}$$

2. Mínimos Cuadrados Ponderados

3. Mínimos Cuadrados Generalizados

4. Uso de Estimadores Robustos

Nota: No pide explicarlos.

6 (0.5 puntos) Se ha aplicado el test de Glejser, obteniendo los siguientes resultados:

| Variable | h | pval | R ² |
|--------------|-----|--------|----------------|
| Publicidad | 2 | 0.0071 | 0.0198 |
| | -1 | 0.0018 | 0.0264 |
| | -2 | 0.0018 | 0.0264 |
| | 0.5 | 0.0046 | 0.0219 |
| Log_Ingresos | 2 | 0.0024 | 0.0250 |
| | -1 | 0.0009 | 0.0302 |
| | -2 | 0.0009 | 0.0302 |
| | 0.5 | 0.0019 | 0.0264 |

| Variable | h | pval | R ² |
|-------------------|-----|--------|----------------|
| Precio_Competidor | 2 | 0.9913 | 3.27e-07 |
| | -1 | 0.7670 | 0.0002 |
| | -2 | 0.7670 | 0.0002 |
| | 0.5 | 0.9410 | 1.51e-05 |
| Calidad | 2 | 0.9228 | 2.59e-05 |
| | -1 | 0.8401 | 0.0001 |
| | -2 | 0.8401 | 0.0001 |
| | 0.5 | 0.9851 | 9.62e-07 |

¿Qué conclusiones tomarías de los resultados obtenidos? ¿Qué metodología aplicarías si los resultados indicaran que existe algún problema en el cumplimiento de las hipótesis del modelo?

6. Test de Glejser

$$\begin{cases} H_0: \text{homocedasticidad} \\ H_1: \text{Heterocedasticidad} \end{cases}$$

Si $p\text{-valores} < 0,05$ rechazamos $H_0 = \text{homocedasticidad}$

Debemos quedarnos con el que tenga un $R^2 >$ porque es el + fiable

→ Publicidad $pval = 0,018$ (+ hay 4 R^2 que es = 0,0264)

$0,018 < 0,05 \rightarrow \text{Significativo} = \text{Heterocedasticidad}$

→ Log-Ingresos $pval = 0,0009 < 0,05 \rightarrow //$

→ Precio-competidor $pval = 0,07670 > 0,05$ NO H_1

→ Calidad $pval = 0,8401 \rightarrow \text{NO } H_1$

Soluciones.

- Eliminar las variables
- Reformular el modelo
- Las de la pregunta anterior.

Soluciones:

1. Transformaciones de datos : garantizamos que los datos después de las transformaciones son homocedásticos:

$$y^* = \frac{y}{\sqrt{w_i}} \quad x^* = \frac{x}{\sqrt{w_i}}$$

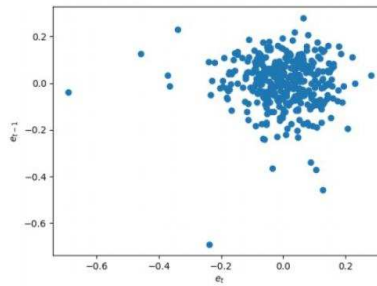
2. Mínimos Cuadrados Ponderados

3. Mínimos Cuadrados Generalizados

4. Uso de Estimadores Robustos

Nota: No pide explicaciones.

7 (0.5 puntos) Para analizar la existencia de autocorrelación en el modelo, se ha realizado el gráfico de los residuos con respecto a los residuos retardados un día, obteniendo el siguiente gráfico:



¿Qué puedes decir acerca de la correlación en base a este gráfico? ¿Hay alguna medida obtenida en el modelo que te permita tomar más conclusiones acerca de la autocorrelación de éste? Justifica cuál es esta medida y razona la existencia o no de autocorrelación en base a esta.

- Se distribuyen de manera aleatoria alrededor del centro \rightarrow no correlacionados.

- Podemos apoyarnos en $DW =$

| | |
|----------------|-------|
| Durbin-Watson: | 1.865 |
|----------------|-------|

Al ser cercano a 2 podemos afirmar que no hay correlación.