Econometría I

Heterocedasticidad

Carlos Yanes | Departamento de Economía | 2023-04-10

😅

Preguntas de la sesion anterior?

😅

Preguntas de la sesion anterior?

Heterocedasticidad

Debemos recordar los **supuestos** de M.C.O

1. Nuestra ***muestra*** es aleatoria las variables (x_k) y (y_i) son **representativas** de una población.
2. La variable (y) es una **función lineal** de los (β_k)'s del modelo y del residuo (u_i).
3. No hay **multicolinealidad perfecta** (relación) de las variables explicativas.
4. Las variables explicativas son **exogenas**: $\boldsymbol{E}[u|X] = 0\,(\implies \boldsymbol{E}[u] = 0)$.
5. Tenemos **varianza constante** de los residuos del modelo, es decir, σ^2 se mantiene siempre o es estable, *p.e.*,
   - $\boldsymbol{E}\left[u_i^2|X\right] = \mathrm{Var}(u_i|X) = \sigma^2 \implies \mathrm{Var}(u_i) = \sigma^2$
   - $\mathrm{Cov}\left(u_i,\, u_j|X\right) = \boldsymbol{E}\left[u_iu_j|X\right] = 0$ para $i \neq j$
6. Los residuos tienen distribución normal, *p.e.*, $u_i \overset{\text{iid}}{\sim} \mathrm{N}\left(0, \sigma^2\right)$.

Heterocedasticidad

Nos enfocamos hoy en el supuesto # 5:

Tenemos **varianza constante** de los residuos del modelo, es decir, σ^2 se mantiene siempre o es estable, *p.e.*,

- $\boldsymbol{E}\left[u_i^2|X\right] = \text{Var}(u_i|X) = \sigma^2 \implies \text{Var}(u_i) = \sigma^2$
- $\text{Cov}\left(u_i,\, u_j|X\right) = \boldsymbol{E}\left[u_i u_j|X\right] = 0$ para $i \neq j$
- Nos enfocamos en la violación de este **supuesto** porque nos va a generar problemas en el modelo, *p.e*:.

Heterocedasticidad $\text{Var}(u_i) = \sigma_i^2$ y $\sigma_i^2 \neq \sigma_j^2$ para algunos $i \neq j$.

En otras palabras: nuestros residuos o (*perturbaciones*) tienen **varianzas** distintas o diferentes.

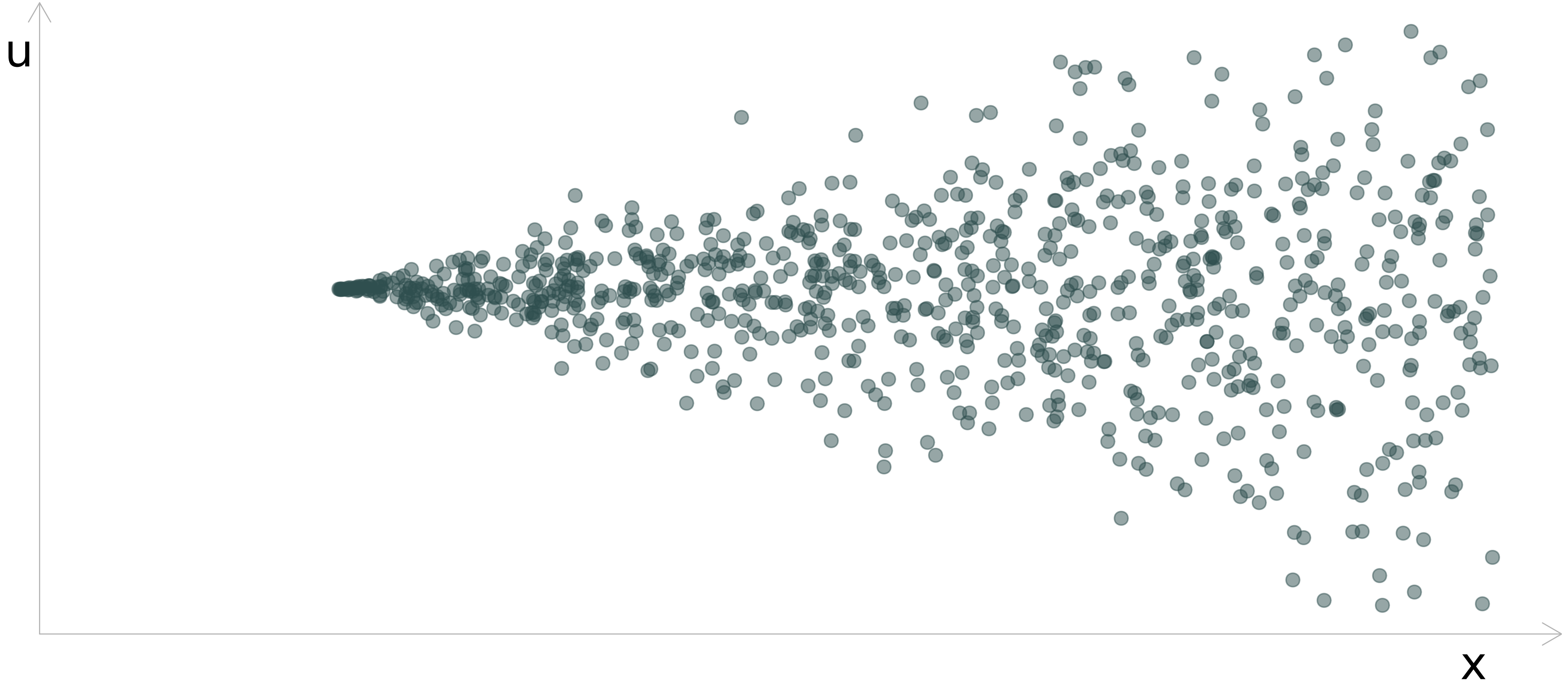
Heterocedasticidad

Aunque se hace necesario que un modelo de regresión cumpla con este supuesto, es bueno saber que en otros momentos es bueno *relajarlo*. Piense que esta midiendo la relación existente entre `educación` y la `habilidad` de una persona (a veces no observable) esta se mantenga constante es muy estricto.

El problema de **heterocedasticidad** va mas que todo enfocada en el sesgo pero de los errores estandar de los estimadores de la regresión.

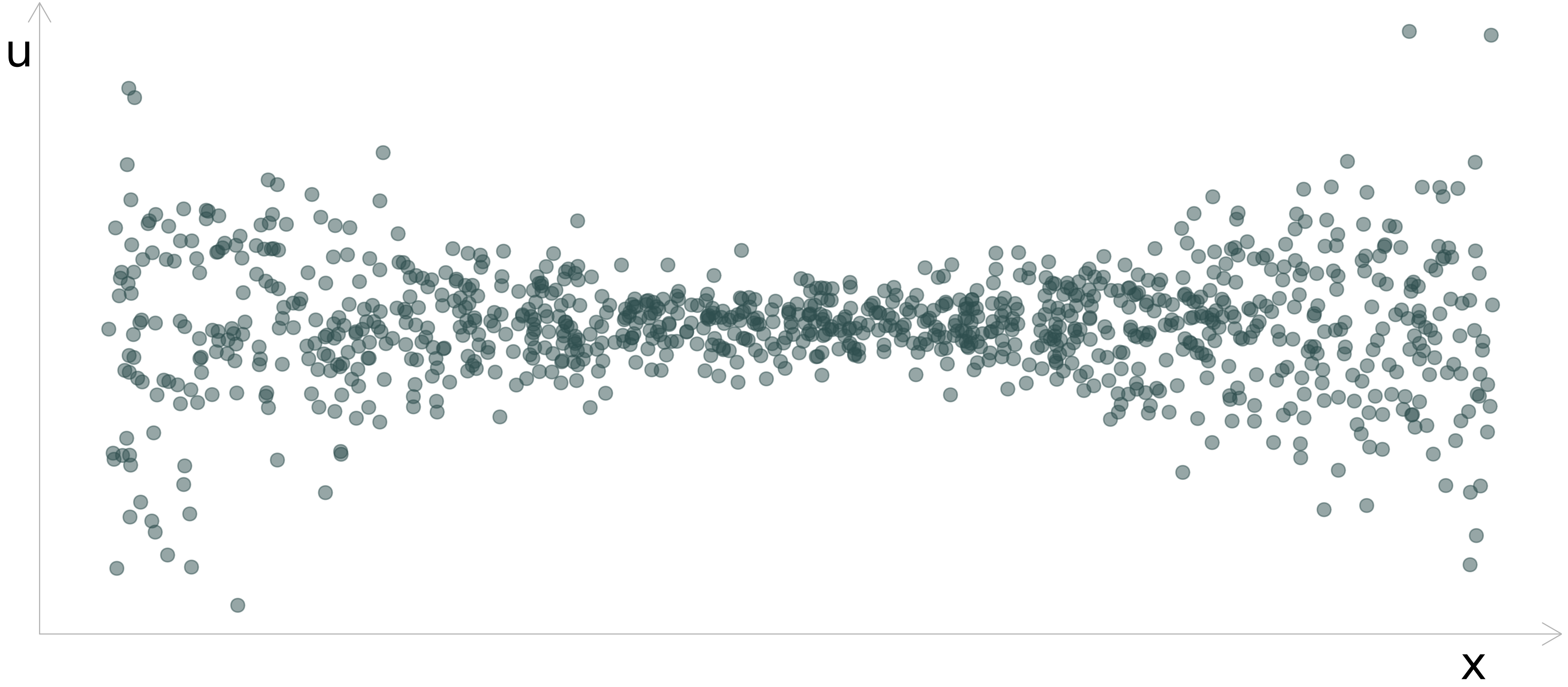
Heterocedasticidad

La varianza de μ_i se incrementa en la medida que las x lo hacen.



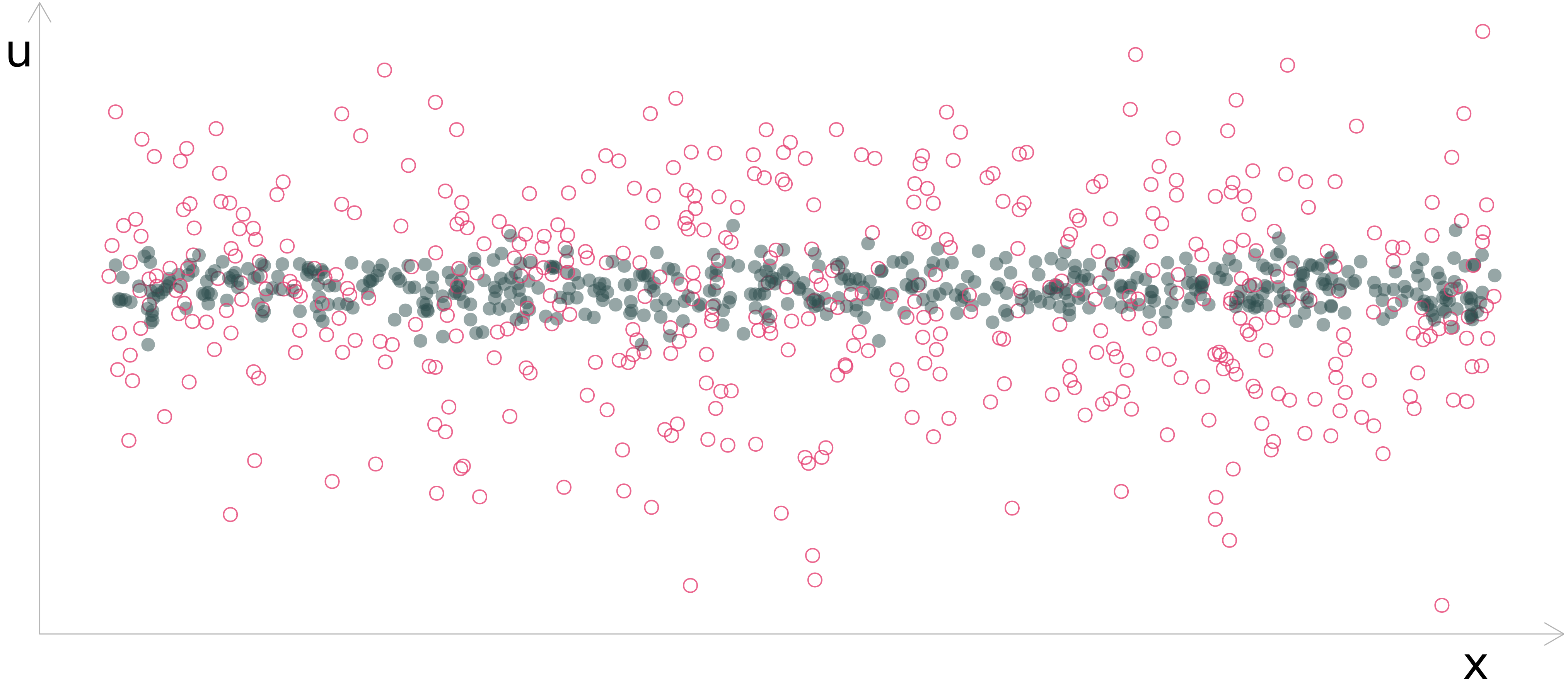
Heterocedasticidad

- Varianza de μ_i se incrementa con los extremos de x



Heterocedasticidad

🚀 Otro ejemplo de heterocedasticidad pero cuando μ_i varia por grupos



Heterocedasticidad

Consecuencias

- Entonces que consecuencias hay cuando es heterocedástico el modelo? es **Sesgado**? es **Ineficienciente**?
- Hay que mirar la insesgadez

Recordeis$_1$: MCO para ser insesgado requiere $\boldsymbol{E}\left[\hat{\beta}_k \middle| X\right] = \beta_k$ para todo k.

Recordeis$_2$: habiamos visto que $\hat{\beta}_1 = \dfrac{\sum_i \left(y_i - \overline{y}\right)\left(x_i - \overline{x}\right)}{\sum_i \left(x_i - \overline{x}\right)^2}$

Esto permite reescribir nuestro estimador como:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_i \left(x_i - \overline{x}\right) u_i}{\sum_i \left(x_i - \overline{x}\right)^2}$$

Heterocedasticidad

Demostración de lo anterior

Asuma que $y_i = \beta_0 + \beta_1 x_i + u_i$

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_i \left(y_i - \overline{y}\right)\left(x_i - \overline{x}\right)}{\sum_i \left(x_i - \overline{x}\right)^2} \\
&= \frac{\sum_i \left(\left[\beta_0 + \beta_1 x_i + u_i\right] - \left[\beta_0 + \beta_1 \overline{x} + \overline{u}\right]\right)\left(x_i - \overline{x}\right)}{\sum_i \left(x_i - \overline{x}\right)^2} \\
&= \frac{\sum_i \left(\beta_1 \left[x_i - \overline{x}\right] + \left[u_i - \overline{u}\right]\right)\left(x_i - \overline{x}\right)}{\sum_i \left(x_i - \overline{x}\right)^2} \\
&= \frac{\sum_i \left(\beta_1 \left[x_i - \overline{x}\right]^2 + \left[x_i - \overline{x}\right]\left[u_i - \overline{u}\right]\right)}{\sum_i \left(x_i - \overline{x}\right)^2} \\
&= \beta_1 + \frac{\sum_i \left(x_i - \overline{x}\right)\left(u_i - \overline{u}\right)}{\sum_i \left(x_i - \overline{x}\right)^2}
\end{aligned}
$$

Heterocedasticidad

$$
\begin{aligned}
\hat{\beta}_1 = \cdots &= \beta_1 + \frac{\sum_i \left(x_i - \overline{x}\right)\left(u_i - \overline{u}\right)}{\sum_i \left(x_i - \overline{x}\right)^2} \\
&= \beta_1 + \frac{\sum_i \left(x_i - \overline{x}\right) u_i - \overline{u} \sum_i \left(x_i - \overline{x}\right)}{\sum_i \left(x_i - \overline{x}\right)^2} \\
&= \beta_1 + \frac{\sum_i \left(x_i - \overline{x}\right) u_i - \overline{u} \left(\sum_i x_i - \sum_i \overline{x}\right)}{\sum_i \left(x_i - \overline{x}\right)^2} \\
&= \beta_1 + \frac{\sum_i \left(x_i - \overline{x}\right) u_i - \overline{u} \left(\sum_i x_i - n\overline{x}\right)}{\sum_i \left(x_i - \overline{x}\right)^2} \\
&= \beta_1 + \frac{\sum_i \left(x_i - \overline{x}\right) u_i - \overline{u} \left(\sum_i x_i - \sum_i x_i\right)}{\sum_i \left(x_i - \overline{x}\right)^2} \\
&= \beta_1 + \frac{\sum_i \left(x_i - \overline{x}\right) u_i}{\sum_i \left(x_i - \overline{x}\right)^2} \quad 😅
\end{aligned}
$$

Heterocedasticidad

$$\begin{aligned}
\boldsymbol{E}\left[\hat{\beta}_1 \middle| X\right] &= \boldsymbol{E}\left[\beta_1 + \frac{\sum_i \left(x_i - \overline{x}\right) u_i}{\sum_i \left(x_i - \overline{x}\right)^2} \middle| X\right] \\
&= \beta_1 + \boldsymbol{E}\left[\frac{\sum_i \left(x_i - \overline{x}\right) u_i}{\sum_i \left(x_i - \overline{x}\right)^2} \middle| X\right] \\
&= \beta_1 + \frac{\sum_i \left(x_i - \overline{x}\right)}{\sum_i \left(x_i - \overline{x}\right)^2} \underbrace{\boldsymbol{E}\left[u_i | X\right]}_{=0} \\
&= \beta_1 \quad 😹
\end{aligned}$$

Ohhh. **MCO se mantiene insesgado** para los β_k.

Heterocedasticidad

- Con insesgadez no hay problema
- Con **ineficiencia** si que lo hay

Cómo así?

La eficiencia y la inferencia de MCO no sobreviven a la heterocedasticidad.

- En presencia de heterocedasticidad, MCO **ya no es el más eficiente** (mejor) estimador lineal insesgado.
- Sería más (eficiente) **ponderar las observaciones** inversamente a la varianza de su u_i.
  - Disminuir la ponderación de los u_i de alta varianza (demasiado difícil para aprender por ahora).
  - Aumentar la ponderación de las observaciones con u_i de baja varianza (más "fiables").
  - Ahora hay que hacer uso de los mínimos cuadrados ponderados (WLS)

Consecuencias: Inferencia

- Intervalos de confianza erróneos
- Problemas para las pruebas de hipótesis (tanto las pruebas t como las F)
- Es de cuidado la inferencia. **Imagine que algo que le dicen y no puede ser testeado**

Preguntas que nos hacemos

Pregunta: ¿Cuál es la definición de heterocedasticidad?

- **R./:**
  Matematicamente: $\text{Var}(u_i|X) \neq \text{Var}\left(u_j|X\right)$ para algunos $i \neq j$.
  Palabras: Existe una relación sistemática entre la varianza de u_i y nuestras variables explicativas.

P: ¿Por qué nos preocupa la heterocedasticidad?

- **R./:** Porque sesga nuestros errores estándar, arruinando nuestras pruebas estadísticas e intervalos de confianza. Además: **MCO** ya no produce el (mejor) estimador más eficiente.

P: ¿La gráfica de (y) contra (x), nos dice algo sobre la heterocedasticidad?

- **R./:** No es exactamente lo que queremos, pero como (y) es una función de (x) y (u), todavía puede ser informativo. Si (y) se vuelve más/menos disperso a medida que (x) cambia, es probable que tengamos **heterocedasticidad**.

Test y pruebas formales:

La **eficiencia** de nuestros estimadores depende de la presencia o no de la heterocedasticidad. Los siguientes autores, tuvieron una idea para su detección y formularon un par de pruebas:

1. Prueba de **Park**.

2. Prueba de **Goldfeld-Quandt**.

3. Prueba de **Breusch-Pagan**.

4. Prueba de **White**.

🎯 Cada una de estas pruebas se centra en el hecho de que podemos **utilizar el residuo OLS** e_i **para estimar la perturbación de la población** u_i.

Test para la heterocedasticidad

Test y pruebas formales: Park

Asume o le da formas **funcionales**[1] a la varianza de μ_i (residuos)

$$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{v_i}$$

Encontrando una estimación logarítmica:

$$Ln\ \mu_i^2 = ln\ \sigma^2 + \beta\ ln\ X + v_i$$

> Debe evaluar si es o no significativo el coeficiente del (β). De esta forma estará detectando heterocedasticidad. *La significancía va con los P-valores*

[1] Recuerde cuando se gráfica `(y=f(x))`

Test y pruebas formales: Goldfeld-Quandt

Se centra en un tipo específico de heterocedasticidad: si la varianza de u_i difiere **entre dos grupos**.[†]

¿Recuerda cómo utilizamos nuestros residuos para estimar el σ^2?

$$s^2 = \frac{\text{SRC}}{n-1} = \frac{\sum_i e_i^2}{n-1}$$

Utilizaremos esta misma idea para determinar si hay evidencia de que nuestros dos grupos difieren en las varianzas de sus perturbaciones, comparando efectivamente s_1^2 y s_2^2 de nuestros dos grupos.

[†]: La prueba G-Q fue una de las primeras pruebas de heterocedasticidad (1965).

Test y pruebas formales: Goldfeld-Quandt

El asunto es mas o menos este

1. Ordenar las observaciones por x
2. Dividir los datos en dos grupos de tamaño n$\star$
    - G_1: El primer tercio
    - G_2: El último tercio
3. Realizar regresiones separadas de y en x para G_1 y G_2
4. Guardar SRC_1 y SRC_2 respectivamente
5. Calcular el **estadístico G-Q**

Test y pruebas formales: Goldfeld-Quandt

La prueba sigue una distribución F (bajo hipótesis nula) con $n^\star - k$ y $n^\star - k$ grados de libertad.[a]

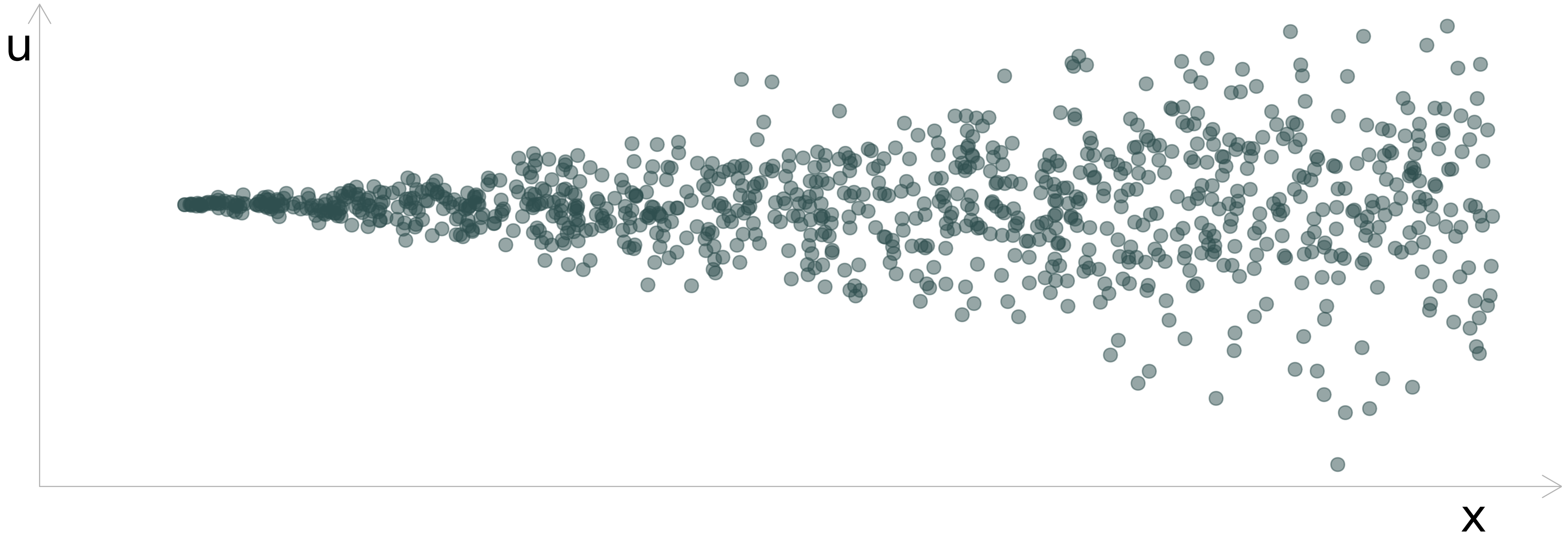
$$F_{(n^\star - k,\, n^\star - k)} = \frac{\text{SRC}_2/(n^\star - k)}{\text{SRC}_1/(n^\star - k)} = \frac{\text{SRC}_2}{\text{SRC}_1}$$

Notas

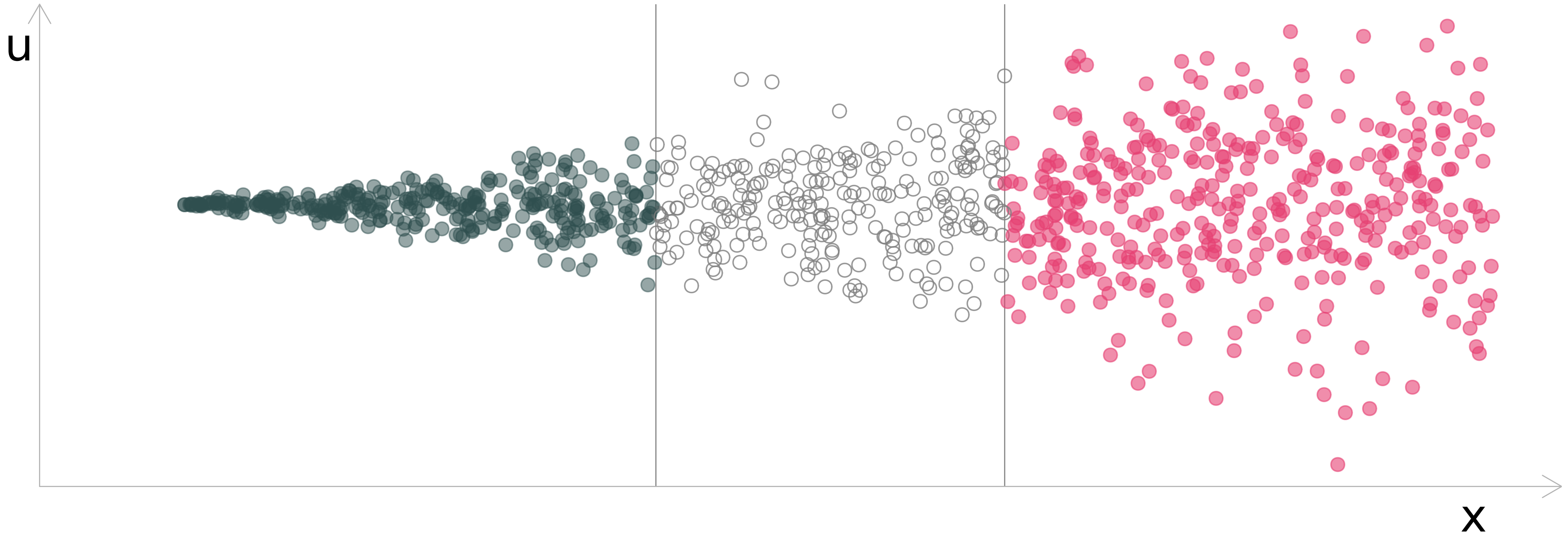
- La prueba G-Q requiere que las perturbaciones sigan distribuciones normales.
- El G-Q asume un tipo/forma muy específico de heterocedasticidad.
- Funciona muy bien si conocemos la forma de heterocedasticidad potencial.

[a]: Goldfeld y Quandt sugirieron $n^\star$ de $(3/8)n$. La parte de (k) hace referencia al número de parámetros estimados (*p.e*, $(\hat{\beta}_j)$'s).

Test y pruebas formales: Goldfeld-Quandt



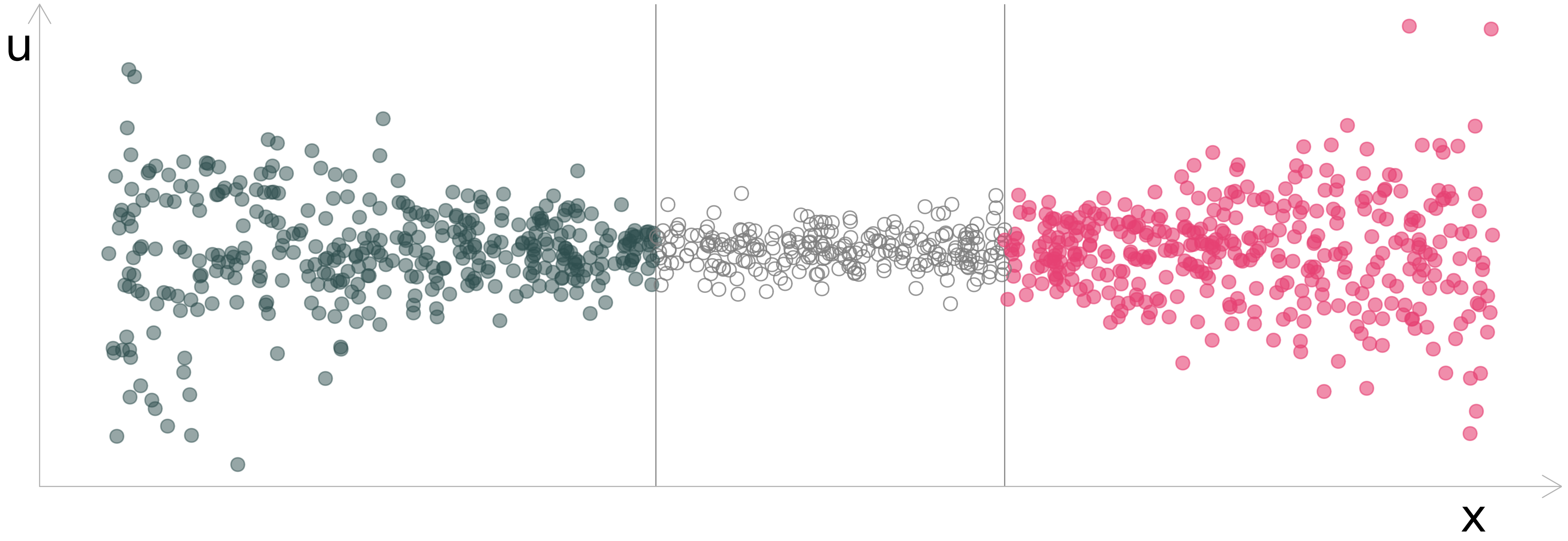
Test y pruebas formales: Goldfeld-Quandt



$$F_{375,\,375} = \frac{\mathrm{SRC}_2 = 18,203.4}{\mathrm{SRC}_1 = 1,039.5} \approx 17.5 \implies p\text{-valor} < 0.001$$

$\therefore$ Rechazamos H_0: $\sigma_1^2 = \sigma_2^2$ y por ende, concluimos que el modelo tiene problemas de **Heterocedasticidad**

Test y pruebas formales: Goldfeld-Quandt



$$F_{375,\,375} = \frac{\text{SRC}_2 = 14,516.8}{\text{SRC}_1 = 14,937.1} \approx 1 \implies p\text{-valor} \approx 0.609$$

$\therefore$ No rechazamos H_0: $\sigma_1^2 = \sigma_2^2$ incluso cuando la **heterocedasticidad** esta presente.


TRAVIS
¿QUÉ FUE ESO?

Test y pruebas formales: Breusch- Pagan

Breusch y Pagan (1981) intentaron resolver este problema de ser demasiado específicos con la forma funcional de la heterocedasticidad.

- Permite que los datos muestren si/cómo la varianza de u_i se correlaciona con X.
- Si σ_i^2 se correlaciona con X, entonces tenemos heterocedasticidad.
- Hacen una regresión de e_i^2 sobre $X = [1,\ x_1,\ x_2,\ \ldots,\ x_k]$ y prueba la significancia conjunta.

Test y pruebas formales: Breusch- Pagan

El asunto es mas o menos este

1. Estimar una regresión y con un intercepto y las variables $x_1, x_2, \ldots, x_k$.
2. Guardar los residuos e.
3. Hacer una regresión ahora de e^2 con cada una de las variables $x_1, x_2, \ldots, x_k$.

$$e_i^2 = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \cdots + \alpha_k x_{ki} + v_i$$

Luego, vamos a guardar el R^2 y Hacer la prueba de hipótesis H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$

Test y pruebas formales: Breusch- Pagan

El estadístico de B-P[2] es:

$$\text{LM} = n \times R_e^2$$

donde R_e^2 es el R^2 de la regresión

$$e_i^2 = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \cdots + \alpha_k x_{ki} + v_i$$

Bajo la hipótesis nula H_0, LM se distribuye asintóticamente como χ_k^2.

Este estadístico de prueba pone a prueba H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$.

Rechazar la hipótesis nula implica una evidencia de **heterocedasticidad**.

[2]: Esta forma específica del estadístico de la prueba proviene en realidad de Koenker (1981).

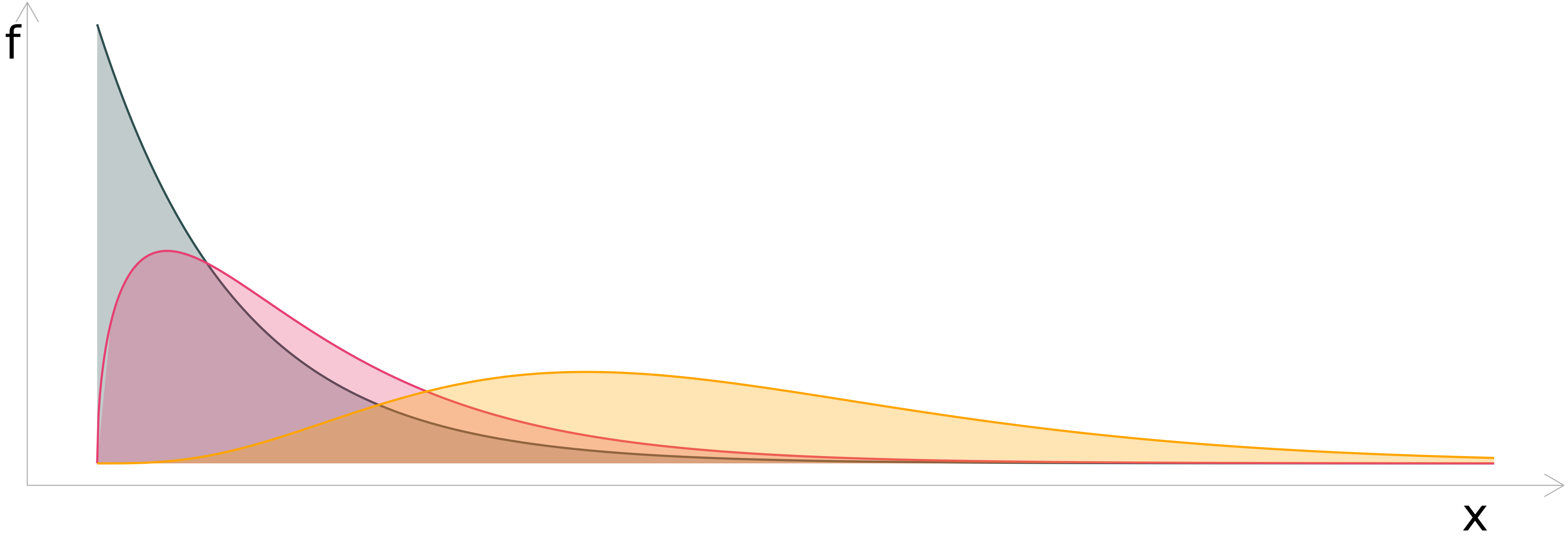
La distribución χ^2

Acabamos de mencionar que bajo la hipotesis nula, el estadístico de B-P se distribuye como una variable aleatoria χ^2 con k grados de libertad.

La distribución χ^2 es sólo otro ejemplo de una distribución común (con nombre) (como la distribución Normal, la distribución t y la misma F).

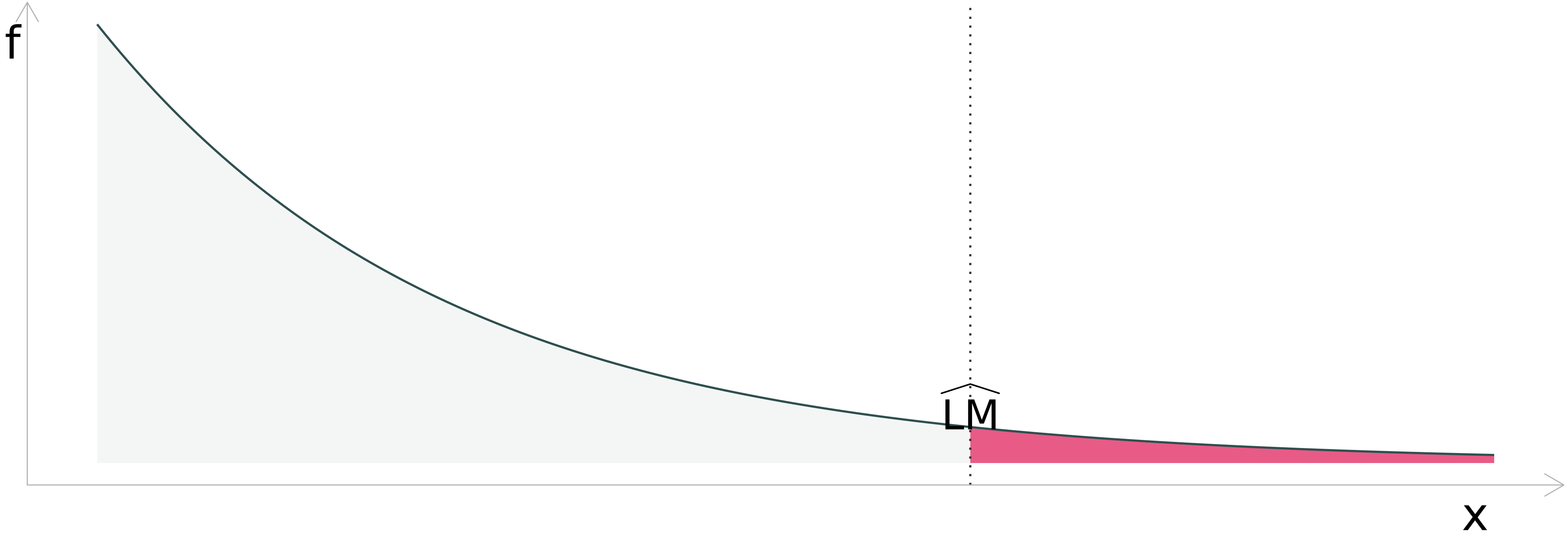
La distribución χ^2

Miremos tres ejemplos de ella, χ^2_k: $k = 1$, $k = 2$, and $k = 9$



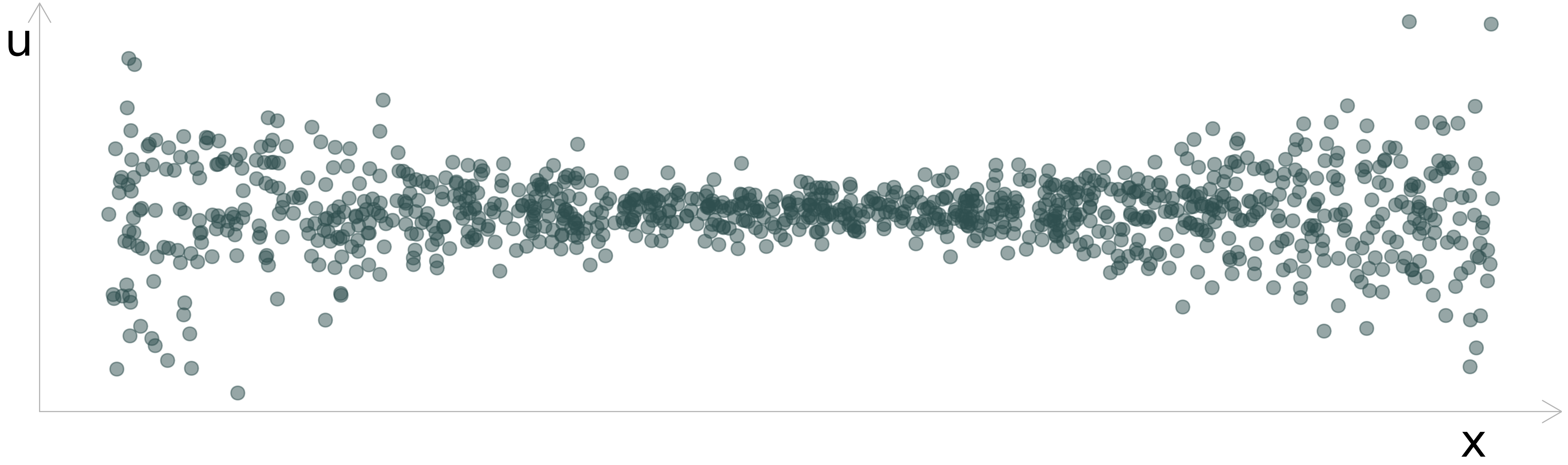
La distribución χ^2

El test de B-P no debe **caer** en el extremo de la distribución $\widehat{\mathrm{LM}}$ bajo el contraste de $H_0 : \sigma_i^2 = \sigma_j^2$



Test y pruebas formales: Breusch- Pagan

El test o prueba de Breusch-Pagan **es sensible a la forma funcional**.



$$e_i^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_{1i} \qquad \widehat{\text{LM}} = 1.26 \qquad p\text{-valor} \approx 0.261$$

$$e_i^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_{1i} + \hat{\alpha}_2 x_{1i}^2 \qquad \widehat{\text{LM}} = 185.8 \qquad p\text{-valor} < 0.001$$

Test y pruebas formales: White

Se debe:

1. Hacer regresión y obtener residuales.

2. Estimar la siguiente regresión auxiliar:

$$\mu_i^2 = \alpha_0 + \underbrace{\alpha_1 x_1 + \alpha_2 x_2}_{\text{Explicativas}} + \underbrace{\alpha_3 x_1^2}_{\text{Var. al Cuadrado}} + \cdots + \underbrace{\alpha_5 x_1 x_2}_{\text{Interacción}} + v_i$$

3. Probar que H_0 es homocedastico. Con estadístico $nR^2 \sim \chi^2 \; g.l.$

Para esto:

> Evaluar si nR^2 es $<$ al *estadístico critico* χ^2 y NO rechazar la hipótesis nula.

Test y pruebas formales: White

Ejemplo: Considere el siguiente modelo $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$

Paso 1: Estimar el modelo; obtener residuos (e).

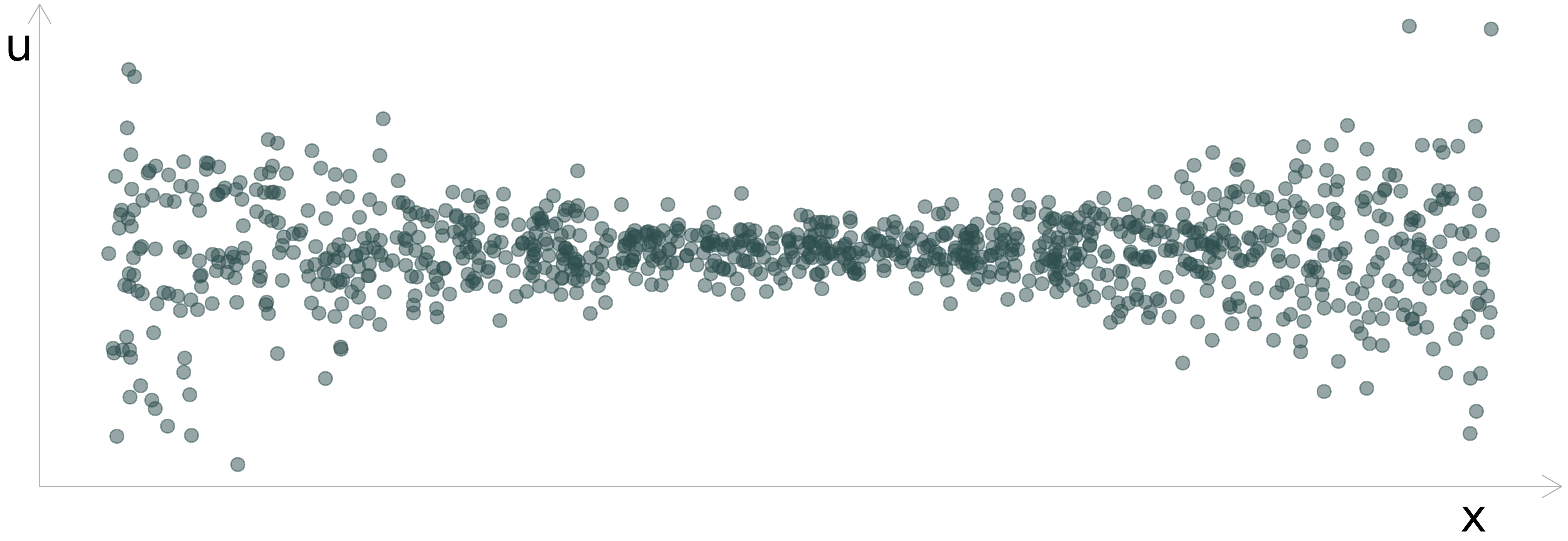
Paso 2: Regresión de e^2 con variables explicativas, al cuadrado y sus interacciones.

$$\begin{aligned} e^2 = & \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1^2 + \alpha_5 x_2^2 + \alpha_6 x_3^2 \\ & + \alpha_7 x_1 x_2 + \alpha_8 x_1 x_3 + \alpha_9 x_2 x_3 + v \end{aligned}$$

Guardamos el R^2 de la ecuación anterior (lo llamamos R_e^2).

Paso 3: Testear H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_9 = 0$ usando $\mathrm{LM} = nR_e^2 \overset{\mathrm{d}}{\sim} \chi_9^2$.

Test y pruebas formales: White



Y tenemos lista esta parte

$$e_i^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_{1i} + \hat{\alpha}_2 x_{1i}^2 \qquad \widehat{\text{LM}} = 185.8 \qquad p\text{-value} < 0.001$$

Una prueba de ojitos 🧠

- Sea la siguiente tabla con estadisticos de **White** determine que modelo tiene **Heterocedasticidad**

Estadístico	P-Valor	Parámetros	Heterocedasticidad
9.47	0.0012	1	
6.45	0.0918	2	
2.17	0.1891	2	
4.23	0.1191	1	

Bibliografía

📘 Gujarati, D. N., & Porter, D. C. (2011). *Econometria Básica*. Ed. Porto Alegre: AMGH..

📘 Stock, J. H., Watson, M. W., & Larrión, R. S. (2012). *Introducción a la Econometría*.

📘 Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Gracias por su atención!

Alguna pregunta adicional?

Carlos Andres Yanes Guerra

cayanes@uninorte.edu.co

keynes37