

*"Los libros no están hechos para que uno crea en ellos, sino para ser sometidos a investigación. Cuando consideramos un libro, no debemos preguntarnos qué dice, sino qué significa".*

*Umberto Eco*

## *¿Podemos contrastar empíricamente hipótesis teóricas?*

- Los males del mecanicismo estadístico en Economía
- Falsacionismo versus corroboración
- Significación, Precisión, Potencia
- ¿Dónde está la función de potencia?
- Excesivo resumen de la información
- ¿Cómo interpretar modelos de relación entre variables económicas?

## *¿Cómo podemos avanzar?*

- ¿Realmente existe un sesgo de variables omitidas?
- Tratamiento de la colinealidad
- La información está en el componente no explicado por el modelo
- Una alternativa a la contrastación de hipótesis paramétricas
- Explorando no linealidades
- Evaluando modelos por objetivos

## Primera sesión:

- *Toda hipótesis teórica es falsa en una muestra: la cuestión a dilucidar es si la evidencia muestral es suficientemente contraria a la hipótesis nula del contraste (y favorable a la hipótesis alternativa) como para rechazarla*
- *Rechazar la hipótesis nula requiere 2 condiciones: a) que la evidencia muestral sea contraria a  $H_0$ , y b) que sea favorable a  $H_1$*
- *Una precisión reducida en la estimación  $\Rightarrow$  pérdida de potencia en el contraste*
- *Los modelos econométricos son un instrumento, que debe utilizarse junto con un análisis exhaustivo de la información muestral: descriptivo, gráfico, estadístico*
- *No encontrar evidencia en contra de la hipótesis nula suele interpretarse incorrectamente como haber probado que dicha hipótesis es cierta*
- *Debido a que olvidamos la función de potencia del contraste*

- La desviación típica como unidad de medida
  - Tamaño medio del cambio entre dos observaciones
  - Tamaño de los intervalos de confianza
  - Determina el resultado de un contraste de hipótesis: *¿cómo es de contraria a  $H_0$  la evidencia muestral?*

$$\hat{\beta} = \rho_{xy} \frac{S_y}{S_x}; \quad Var(\hat{\beta}) = \frac{\sigma_u^2}{NS_x^2} \Rightarrow \text{ratio } t : \frac{\hat{\beta}}{\sqrt{Var(\hat{\beta})}} = \frac{\sqrt{N}\rho_{xy}}{\sigma_u / S_y}$$

- Precisión en la estimación
  - Inverso de la desviación típica
  - Necesitamos variabilidad en factores explicativos
  - Tamaño muestral grande (homogeneidad)

- **Modelo:**

$$R_i = \alpha + \beta r_i + u_i = 0.23 + 0.80r_i + \hat{u}_i, \quad u_i \approx N(0,1)$$

$$\text{Supongamos } r_i \approx N(0, \lambda^2) \Rightarrow \text{Var}(\hat{\beta}) = \frac{1}{T\lambda^2}$$

$$\text{Si } T = 100 \Rightarrow DT(\hat{\beta}) = \frac{1}{10\lambda}$$

*intervalo de confianza del 95% :*

$$\lambda = 0.1 \Rightarrow 0.95 = P(-1.2 \leq \beta \leq 2.8)$$

$$\lambda = 1 \Rightarrow 0.95 = P(0.6 \leq \beta \leq 1.0)$$

$$\lambda = 5 \Rightarrow 0.95 = P(0.76 \leq \beta \leq 0.84)$$

## El modelo de regresión múltiple

$$y_i = \alpha + \beta_1 x_i + \beta_2 z_i + u_i, i = 1, 2, \dots, N$$

$$\hat{\beta}_1 = \frac{S_{x_1 y} S_{x_2}^2 - S_{x_2 y} S_{x_1 x_2}}{S_{x_1}^2 S_{x_2}^2 - (S_{x_1 x_2})^2}; \hat{\beta}_2 = \frac{S_{x_2 y} S_{x_1}^2 - S_{x_1 y} S_{x_1 x_2}}{S_{x_1}^2 S_{x_2}^2 - (S_{x_1 x_2})^2}$$

$$\text{Var}(\hat{\beta}_1) = \sigma_u^2 \frac{S_{x_2}^2}{S_{x_1}^2 S_{x_2}^2 - (S_{x_1 x_2})^2} = \frac{\sigma_u^2}{S_{x_1}^2 (1 - \rho_{x_1 x_2}^2)}$$

$$t\text{-ratio}(\hat{\beta}_1) = \frac{S_{x_1 y} S_{x_2}^2 - S_{x_2 y} S_{x_1 x_2}}{\sigma_u S_{x_2} \sqrt{S_{x_1}^2 S_{x_2}^2 - (S_{x_1 x_2})^2}}$$

$$\hat{\beta}_1 = \frac{\rho_{x_1 y} S_y - \rho_{x_2 y} S_y \rho_{x_1 x_2}}{\sqrt{1 - \rho_{x_1 x_2}^2}} \Rightarrow t\text{-ratio}(\hat{\beta}_1) = \frac{S_y}{\sigma_u} \frac{\rho_{x_1 y} - \rho_{x_2 y} \rho_{x_1 x_2}}{\sqrt{1 - \rho_{x_1 x_2}^2}}$$

- ¿Por qué podemos perder precisión en la estimación?
  - Variabilidad temporal en la relación
  - Tamaño muestral reducido,
  - Naturaleza de la muestra,
  - Reducida variabilidad muestral de las variables explicativas
  - Correlaciones entre variables explicativas => podemos estimar efectos globales
- Si las relaciones entre variables económicas varían en el tiempo, entonces nuestras estimaciones recogen efectos “promedio” a lo largo de la muestra

*Simulando un modelo*



$$y_t = \alpha + \beta x_t + \delta z_t + u_t$$

Supongamos :  $u_t \approx N(0,3) \Rightarrow \{\hat{u}_t\}_{t=1}^{t=500}$

$$x_t = 9 + U(0,2); \quad z_t \approx 5 + N(0,1)$$

$$\alpha = 2.5; \quad \beta = 0.85; \quad \delta = -1.20$$

*Finalmente* :  $y_t = 2.5 + 0.85x_t - 1.30z_t + u_t, t = 1, 2, \dots, 100$

Esta es una realización.

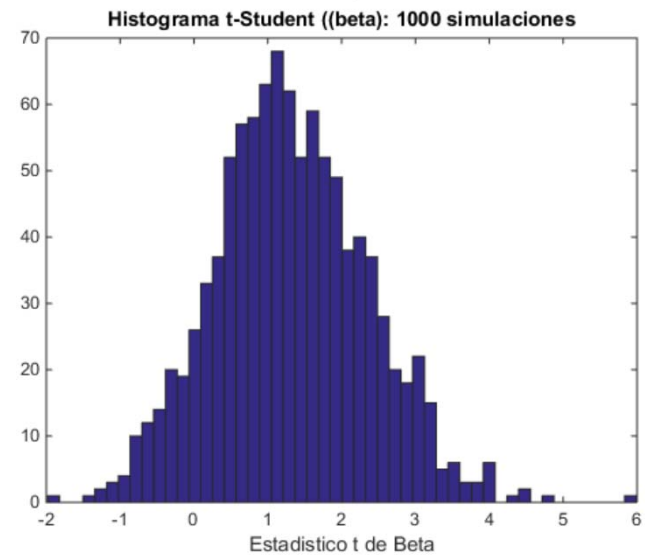
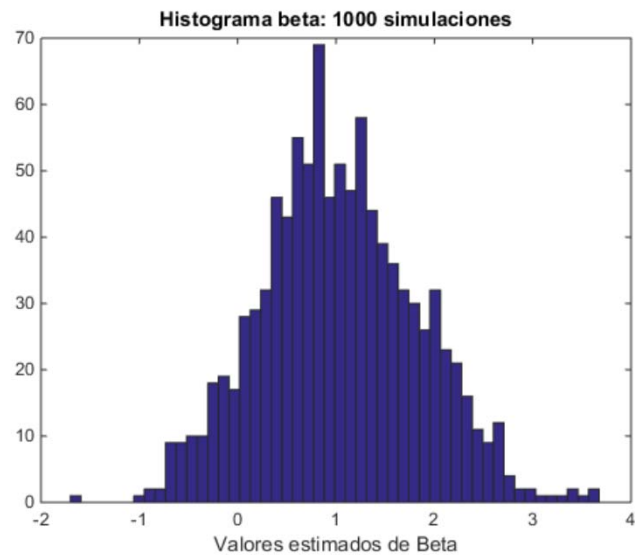
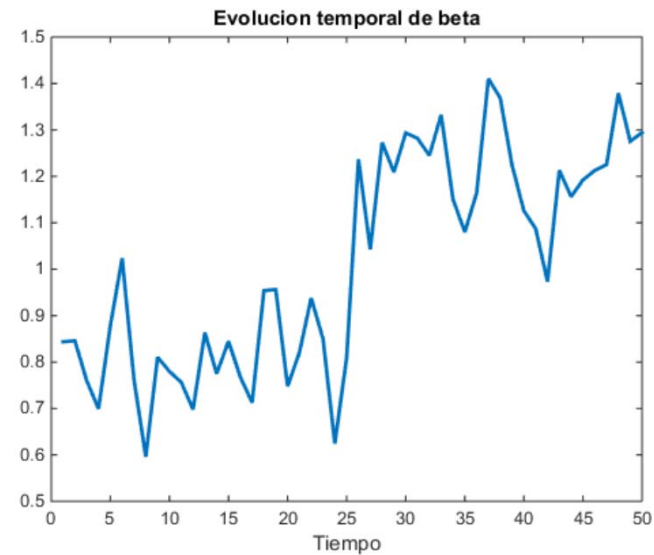
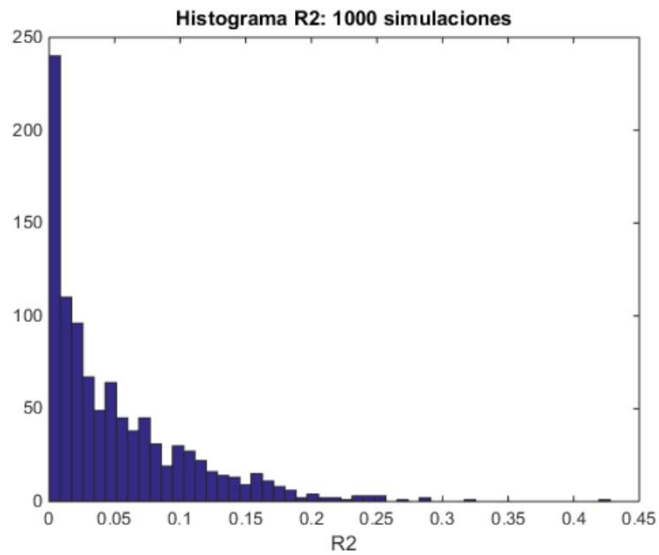
Repetimos el número de realizaciones deseadas.

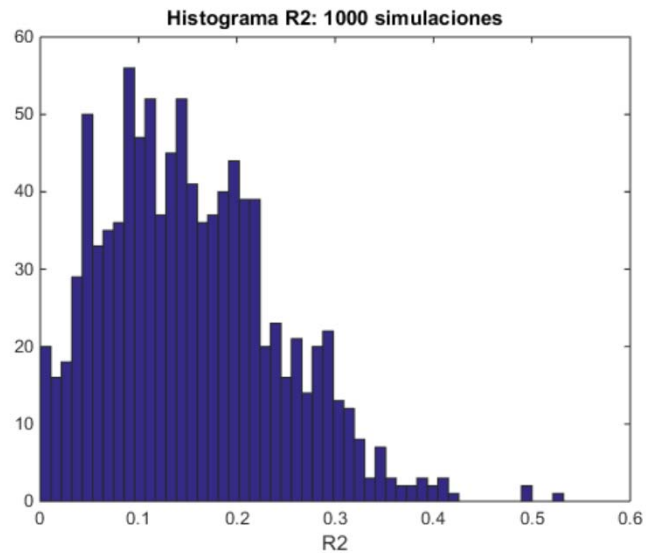
Generalmente, un número alto: 5000

$$y_t = \alpha + \beta_t x_t + u_t, \sigma_u^2 = 5, S_x^2 = 1$$

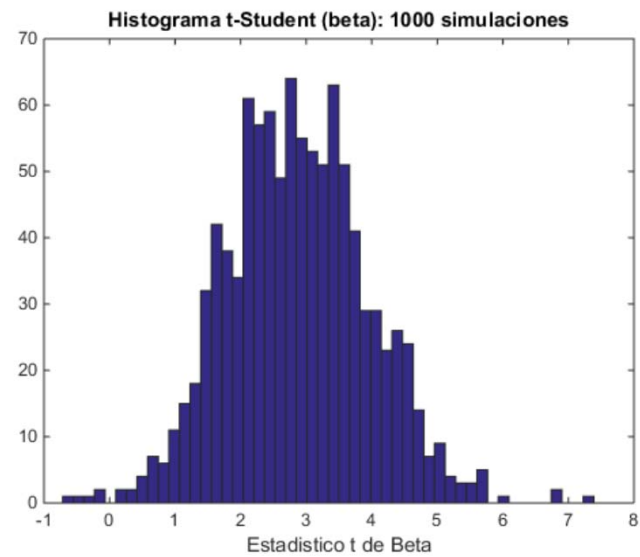
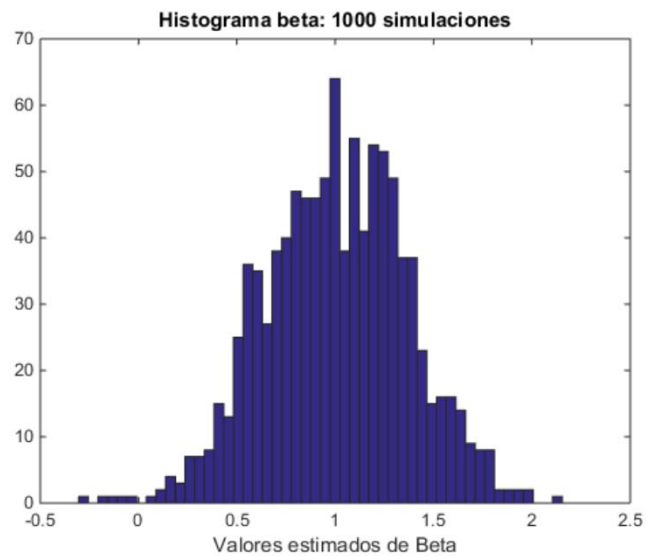
% var( $u$ ) = 96%; % var( $x$ ) = 4%

25% Rechazos :  $H_0: \beta = 0, H_1: \beta \neq 0$



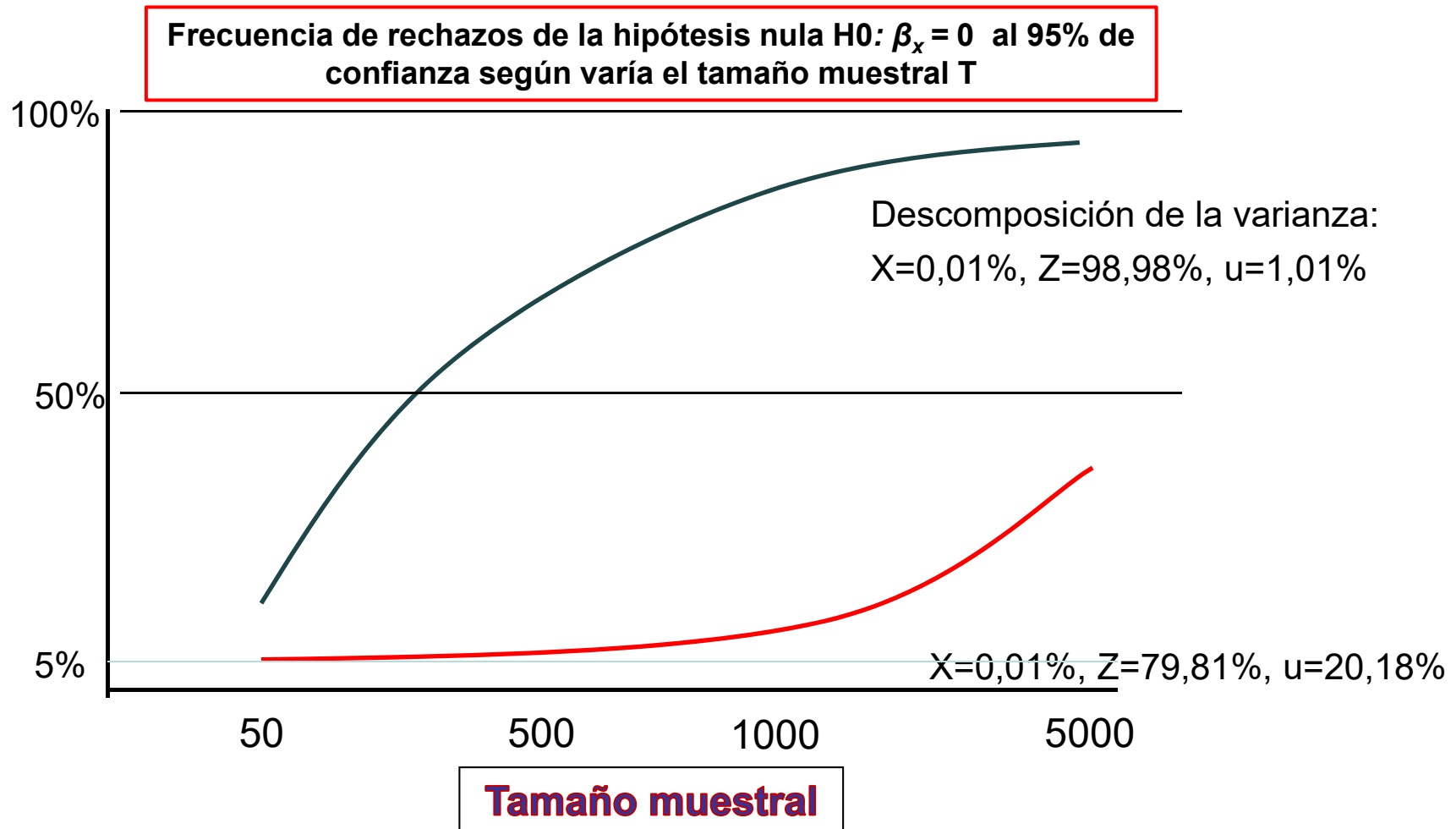


$y_t = \alpha + \beta_t x_t + u_t, \sigma_u^2 = 1, S_x^2 = 1$   
 $\% \text{var}(u) = 50\%; \quad \% \text{var}(x) = 50\%$   
 80% Rechazos :  $H_0: \beta = 0, \quad H_1: \beta \neq 0$

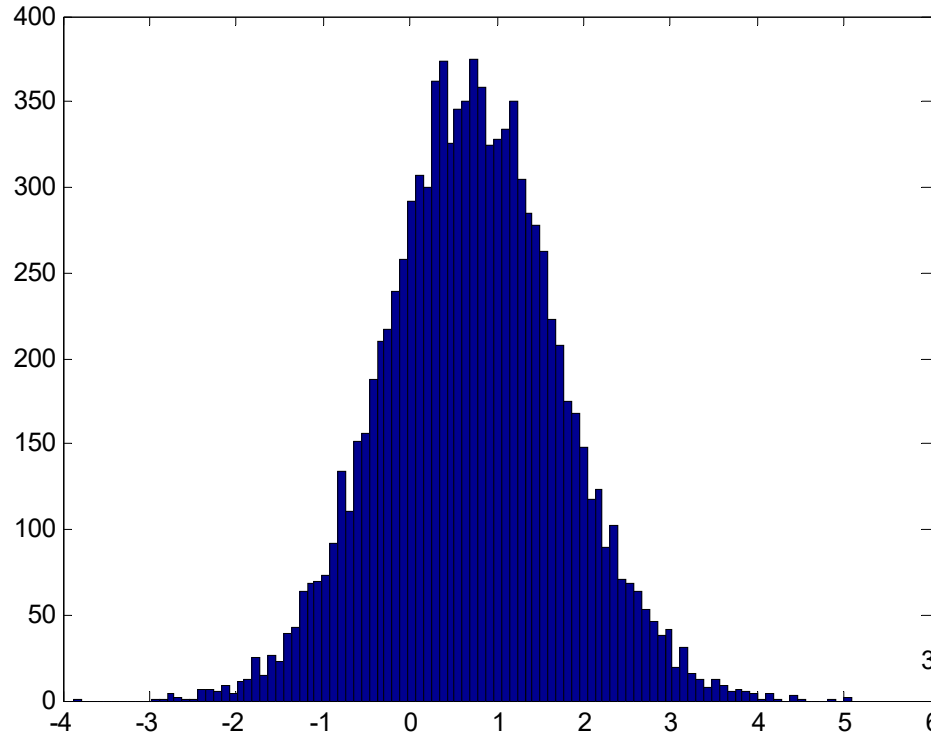


## P5: También puede haber **'Excesiva' precisión**

- Experimento de simulación:  $y = \beta_1 x + \beta_2 z + u$ ,  $Corr(x, z) = Corr(x, u) = Corr(z, u) = 0$



tstat for b=0 in Model 1 with T=50



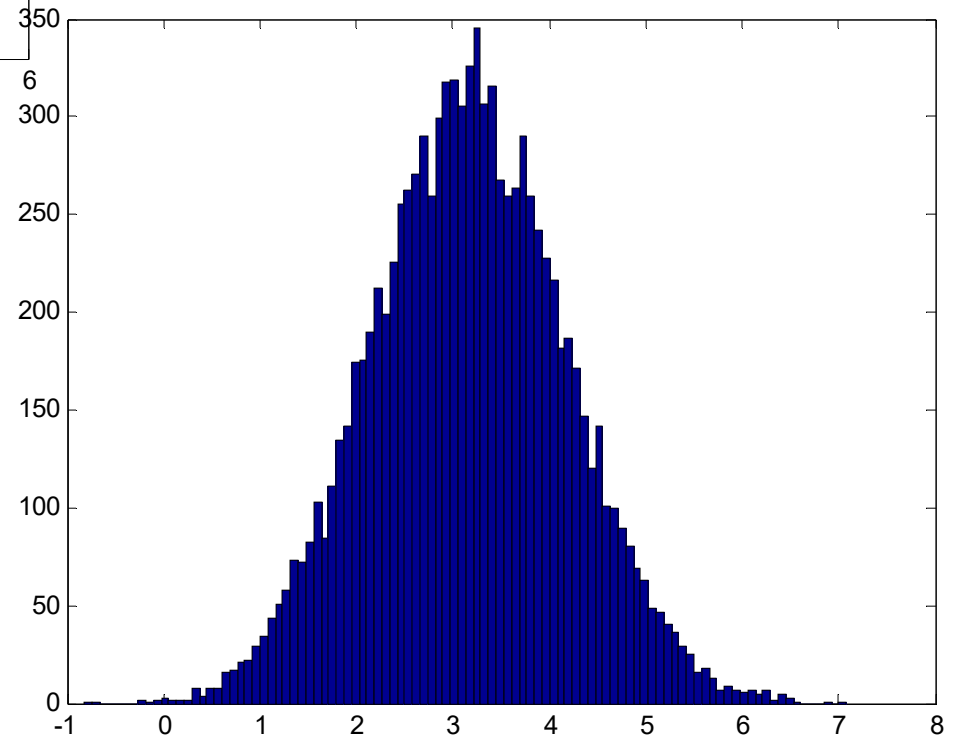
$$y = \beta_1 x + \beta_2 z + u,$$

$$\text{Corr}(x, z) = \text{Corr}(x, u) = \text{Corr}(z, u) = 0$$

Variance decomposition:

X=0,01%, Z=98,98%, u=1,01%

tstat for b=0 in Model 1 with T=1000



- A pesar de todo lo cual, las conclusiones alcanzadas se emiten con carácter demasiado categórico:
  - ✓ “we have shown that....”
  - ✓ “we have obtained conclusive evidence on the relevance of variable X to explain the behavior of variable Y ...”
  - ✓ “an increase of 1% in X leads to a higher contemporaneous increase in Y by 1,5% , followed by an increase of 0,4% next year”

*Caso práctico: Contrastando la Hipótesis de Expectativas en la formación de tipos de interés*

## *Caso práctico: La hipótesis de expectativas en la formación de tipos de interés*

- Los tipos a largo observados en  $t$  son un promedio de los tipos a corto cotizados en  $t$  y las expectativas en  $t$  de tipos a corto en períodos sucesivos:

$$r_t^{6m} = r_t^{3m} + E_t r_{t+3}^{3m} + \pi_t^{3m,6m}$$

- Bajo racionalidad de expectativas, ello implica que el spread largo-corto, el tipo forward continuo, es el mejor predictor del tipo a corto futuro:

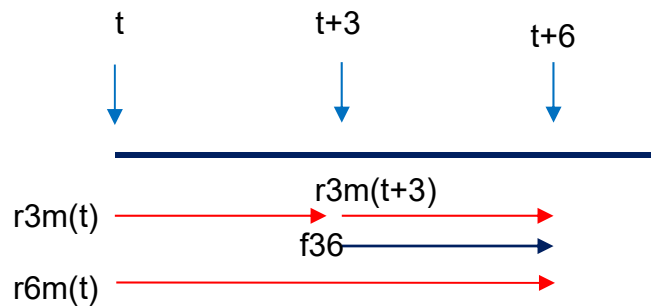
$$r_t^{3m} = \alpha + \beta f_{t-3,t}^{3m,6m} + u_t, \quad H_0 : \alpha = 0, \beta = 1, \quad H_1 : \alpha \neq 0, \beta \neq 1$$

- no hay información adicional en ningún otro aspecto de la estructura temporal de tipos de interés (ETTI) actual
- no hay información adicional en características de ETTI anteriores
- en particular, no la hay en tipos de interés pasados



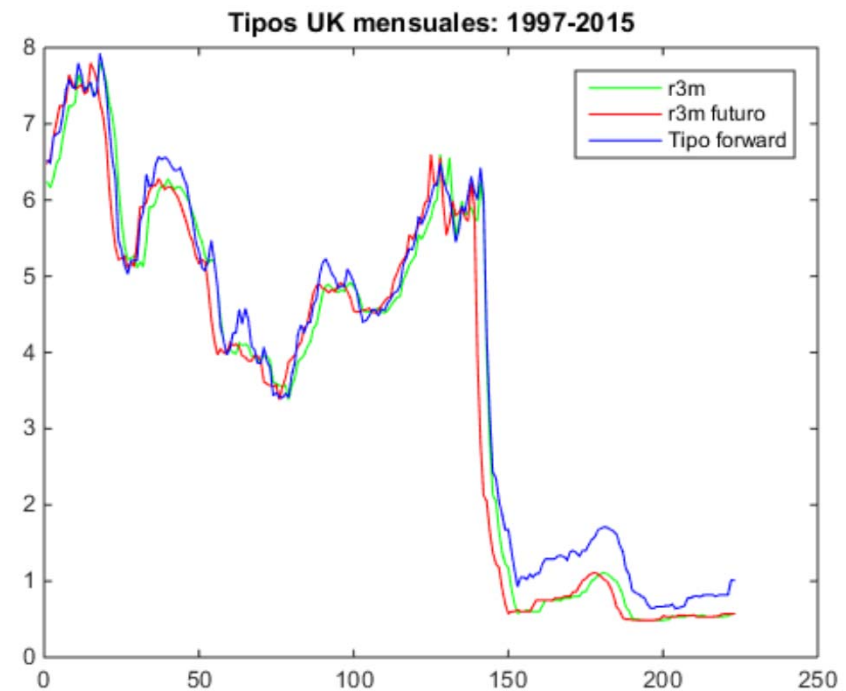
# Hipótesis de Expectativas: El tipo forward como predictor del tipo a corto futuro

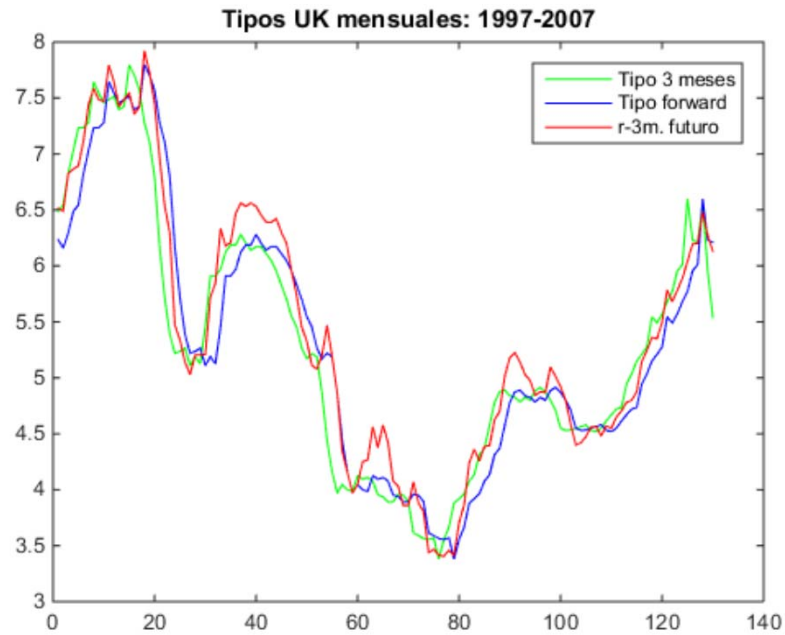
$$(1 + r_t^{6m})^2 = (1 + r_t^{3m})(1 + E_t r_{t+3}^{3m}) \rightarrow 1 + f_t^{3m,6m} = \frac{(1 + r_t^{6m})^2}{1 + r_t^{3m}} \approx 2r_t^{6m} - r_t^{3m}$$



*El tipo forward es un predictor insesgado del tipo de contado futuro :*

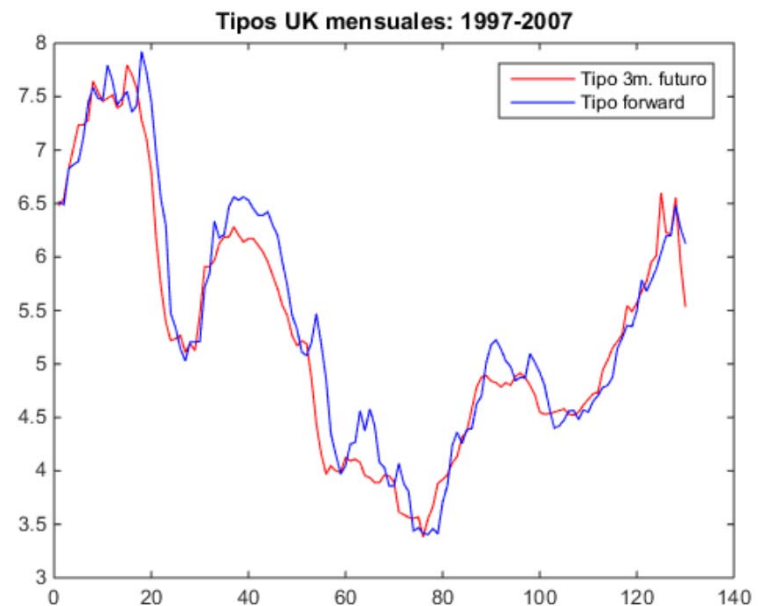
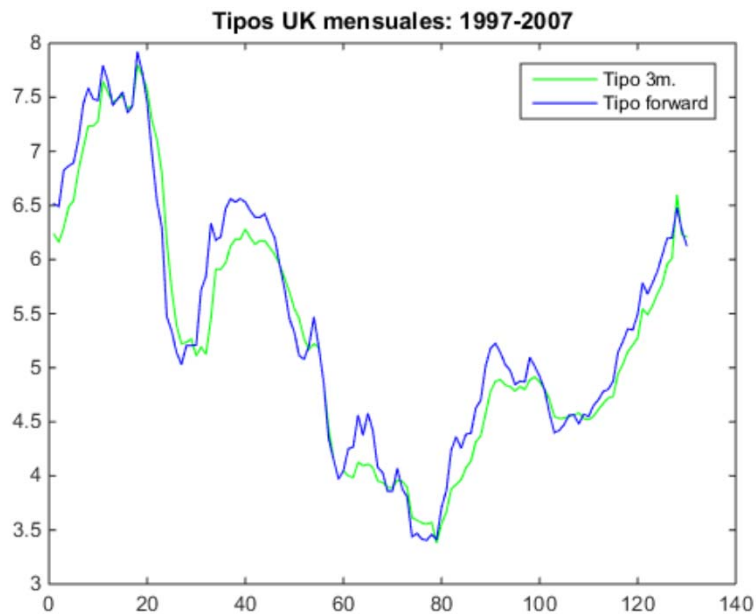
$$r_t^3 = \alpha + \beta f_{t-3}^{3,6} + u_t, \quad H0: \alpha = 0, \beta = 1$$



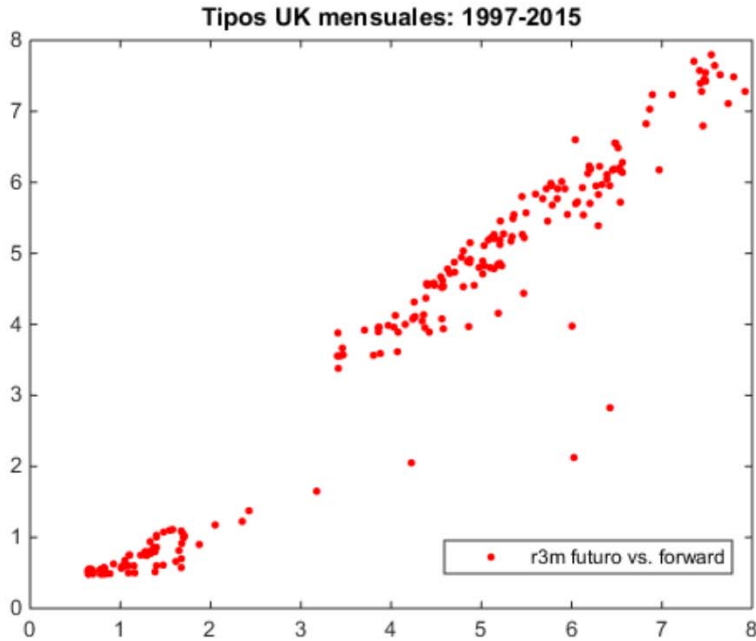
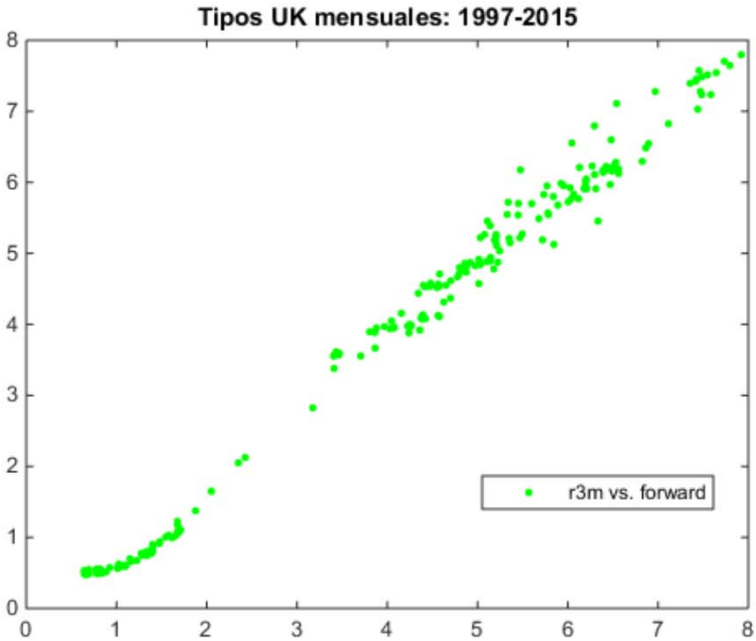


*Lo que no forma parte de la hipótesis de expectativas...  
 ...ni de otras  
 ¿Realmente predicen los futuros? O simplemente replican el mercado?*

ECM:  $r_{3m}(t) - \text{forward}(t-3)$ : 13.37  
 ECM:  $r_{3m}(t-3) - \text{forward}(t-3)$ : 8.39

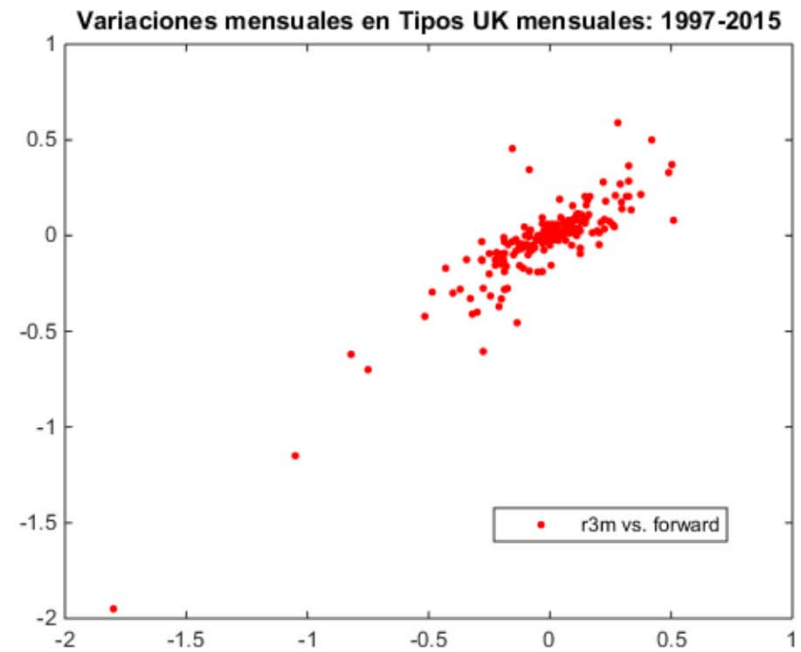
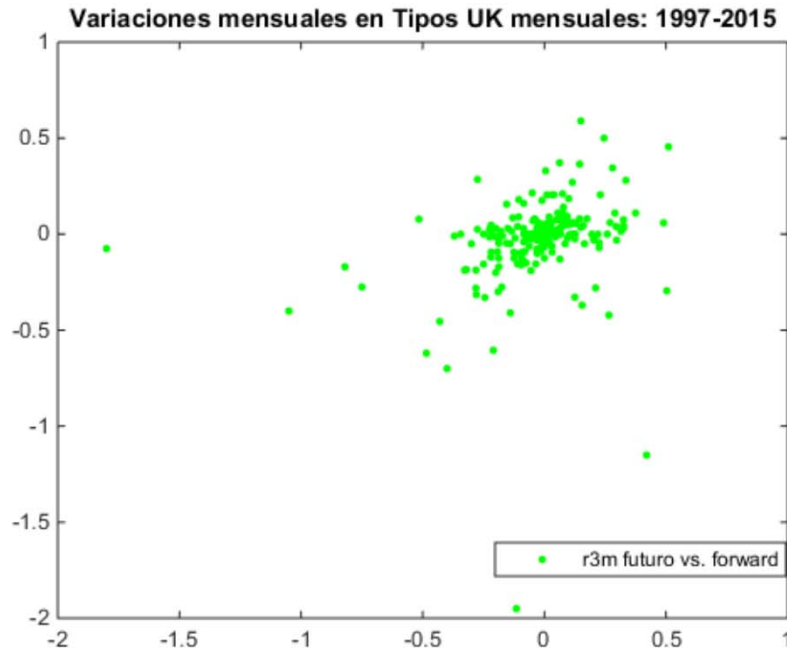


# Tipos de interés mensuales Bank of England: 1997-2015



# Variaciones mensuales en tipos de interés

## Bank of England: 1997-2015



$\text{Corr}(\Delta r3, \Delta r3f, \Delta f36)$

1	0.19	0.87
0.19	1	0.25
0.87	0.25	1

## Estimación datos Bank of England: 1997-2015

$$r_t^{3m} = -0.420 + 1.033 f_{t-3,t}^{3m,6m} + u_t, \quad R^2 = 0.957$$

(0.066)      (0.015)

$$\sigma_u^2 = 0.247, \quad \text{Var}(r_t^{3m}) = 5.734$$

$$0.95 = P(1.004 \leq \beta \leq 1.062)$$

$$0.99 = P(0.994 \leq \beta \leq 1.072)$$

$$r_t^{3m} = -0.253 + 0.595 f_{t-3,t}^{3m,6m} + 0.418 r_{t-3}^{3m} + u_t, \quad R^2 = 0.959, \sigma_u^2 = 0.239$$

(0.087)      (0.145)      (0.142)

Correlación entre ambos conjuntos de residuos: 0.981

$$r_t^{3m} = -0.006 + 0.981 r_{t-3}^{3m} + u_t, \quad R^2 = 0.956, \sigma_u^2 = 0.255$$

(0.062)      (0.014)

Correlación con residuos anteriores: 0.897

$$r_t^{3m} = -0.014 + 0.997 r_{t-1}^{3m} + u_t, \quad R^2 = 0.992, \sigma_u^2 = 0.048$$

(0.027)      (0.006)

$$\text{Var}(r_t^{3m} - f_{t-3}^{3m,6m}) = 0.025$$

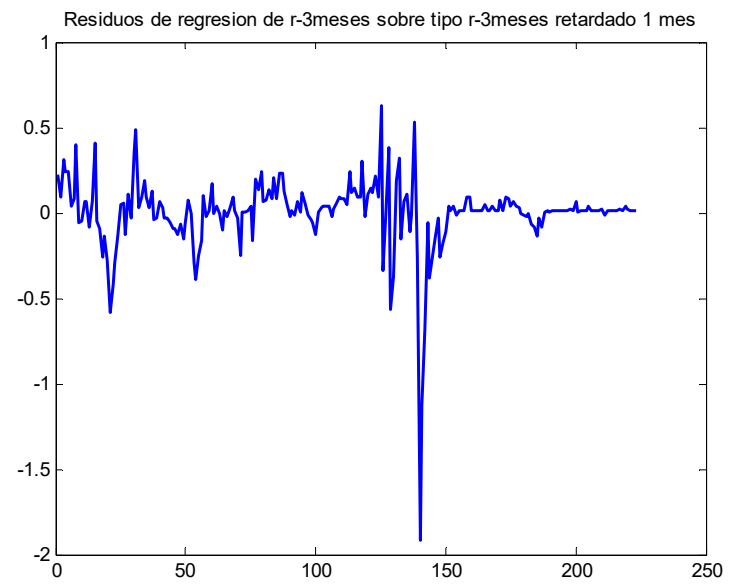
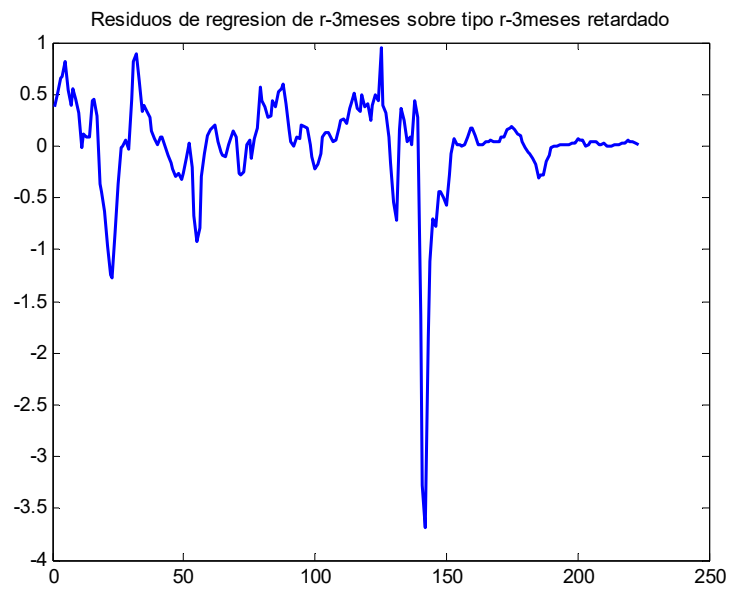
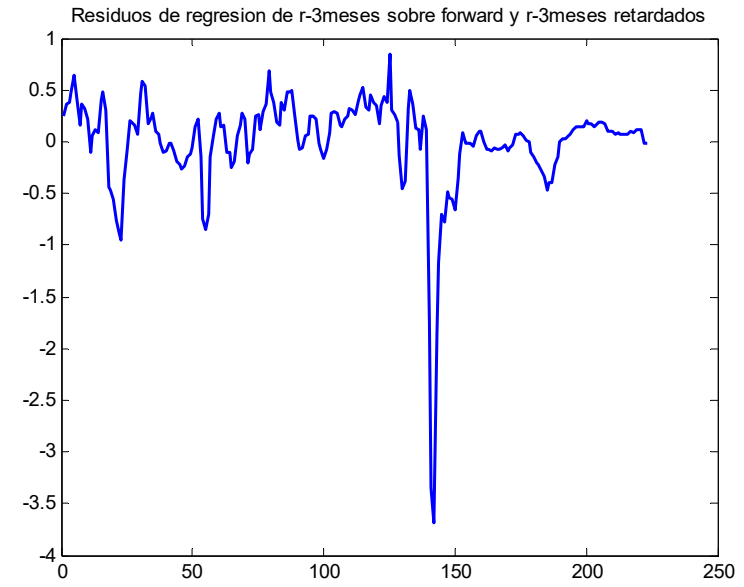
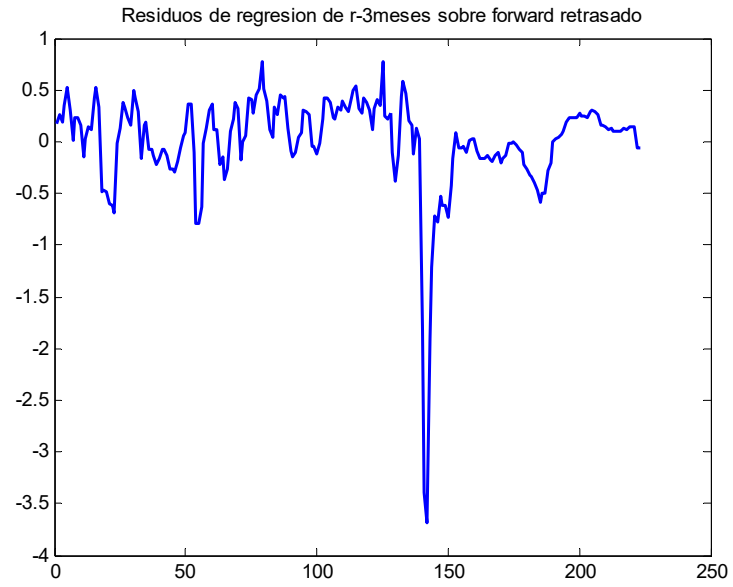
$$\text{Var}(r_t^{3m} - f_t^{3m,6m}) = 0.007$$

$$\text{Var}(r_t^{3m} - r_{t-3}^{3m}) = 0.026$$

¿Es  $\alpha=0, \beta=1$ ? Depende de para qué vayamos a utilizar el modelo

- Contraste de la Hipótesis de Expectativas
- Predicción de tipos de interés

No reduzcamos la cuestión que analizamos a una restricción paramétrica



## *Testing the Expectations Hypothesis of the term structure of interest rates: Error tipo II*

*Future spot interest rate projected on lagged forward rate*

Model:  $r_t^3 = \alpha + \beta f_{t-3}^{3,6} + u_t$ ,  $f_t^{3,6} = 2r_t^6 - r_t^3$

H0:  $\beta=1$  versus: H1:  $\beta<1$

<b>Plazo</b>	<b>Beta</b>	<b>Desviación típica</b>	<b>R2</b>	<b>Estadístico t H0: Beta = 1</b>
<b>1 mes</b>	<b>0,96</b>	<b>0,04</b>	<b>0,54</b>	<b>-1,00</b>
<b>3 meses</b>	<b>0,93</b>	<b>0,06</b>	<b>0,42</b>	<b>-1,17</b>
<b>6 meses</b>	<b>1,10</b>	<b>0,15</b>	<b>0,32</b>	<b>0,67</b>
<b>1 año</b>	<b>1,40</b>	<b>0,24</b>	<b>0,22</b>	<b>1,67</b>
<b>3 años</b>	<b>1,63</b>	<b>0,47</b>	<b>0,15</b>	<b>1,34</b>
<b>5 años</b>	<b>2,40</b>	<b>0,85</b>	<b>0,04</b>	<b>1,65</b>

# ¿Dónde está la función de potencia?

- Excesiva frecuencia de contrastes bilaterales
- Práctico olvido de la **función de potencia de un contraste**: Probabilidad de identificar una hipótesis nula falsa como tal.
- Contra la práctica habitual, no tiene justificación enfatizar hipótesis que no se hayan rechazado, a menos que se haya comprobado la potencia del contraste
- *Cálculo de la función de potencia*: Una vez determinada la región crítica bajo  $H_0$ , se calcula la probabilidad de rechazo para distintas  $H_0$

- ***El p-value de un contraste no es suficiente***

- $p$ -value: probabilidad de extraer una muestra con información más contraria a  $H_0$  que la muestra disponible:

$$P [ \text{muestra} / H_0 ]$$

- Que se interpreta en ocasiones, equivocadamente, como la probabilidad de que  $H_0$  sea cierta a la luz de la información muestral:

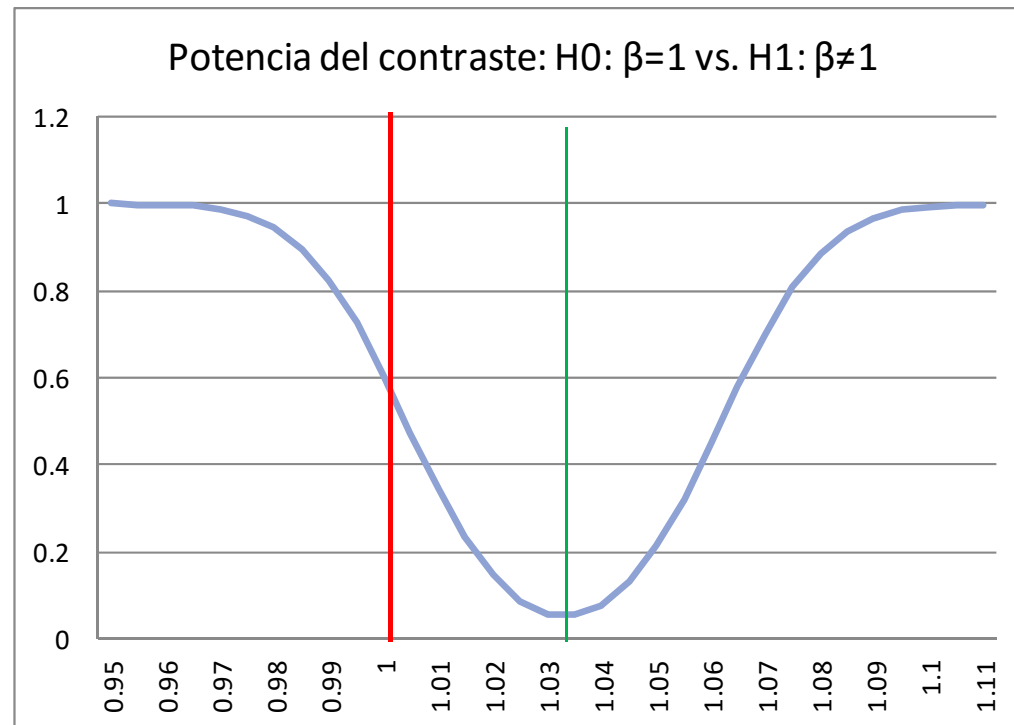
$$P [ H_0 / \text{muestra} ]$$



$$r_t^{3m} = -0.420 + 1.033 f_{t-3,t}^{3m,6m} + u_t, \quad R^2 = 0.957$$

(0.066)      (0.015)

$$\sigma_u^2 = 0.247, \quad \text{Var}(r_t^{3m}) = 5.734$$



*Significación estadística,  
capacidad explicativa, contenido informativo*

*Interpretación de coeficientes individuales en  
regresiones múltiples*

## *Evaluando la capacidad explicativa de una variable*

- Pregunta más frecuente: *Capacidad explicativa*: ¿Para explicar una variable  $Y$ 
  - ✓ ¿es significativa la variable  $X$  ?
  - ✓ ¿es más significativa la variable  $X$  o la variable  $Z$  ?
  - ✓ ¿se ha hecho más significativa  $X$  en la segunda parte de la muestra (o desde la crisis)?
- ¿Cómo se mide el impacto de una variable  $X$  sobre la variable dependiente  $Y$  ?
  - Valor absoluto del coeficiente  $\beta$  : impacto sobre  $Y$  de variación unitaria en  $X$  ¿Cómo es de grande dicha variación?
  - Valor absoluto del estadístico  $t$ -Student
- Pero hay serias dificultades tanto con la lectura directa del coeficiente estimado, como con el (ab)uso habitual del estadístico  $t$

# Significación estadística vs. Relevancia económica

- Estamos identificando significación estadística de un coeficiente con capacidad explicativa o contenido informativo de una variable,
    - ✓ Cualitativamente: X explica a Y si y solo si su coeficiente asociado es significativamente distinto de cero
    - ✓ Cuantitativamente: Capacidad explicativa  $\Leftrightarrow$  valor absoluto elevado de la *t*-Student
    - ✓ Olvidando la relación entre precisión en estimación y potencia en el contraste de hipótesis
    - ✓ Practicando Econometría de signos
    - ✓ Practicando Econometría de asteriscos ¿indicando qué?
  - Estimación numérica y significación. Interpretación habitual de estadísticos *t* :
    - “... tiene signo contrario al correcto, pero no es significativo...”
    - “... tiene el signo esperado según el modelo teórico, pero no explica a la variable Y ...”
- ⇒ **Errónea interpretación habitual de la posible precisión estadística en la estimación de un coeficiente con la relevancia económica de la variable que le acompaña**

## *Las dificultades con el estadístico $t$ para evaluar la capacidad explicativa de $X$ sobre $Y$ :*

- Práctica habitual para juzgar si una determinada variable explicativa debe incluirse en un modelo: *Una variable  $X$  contribuye a explicar el comportamiento de  $Y$  si y sólo si el estadístico  $t$  de Student de su coeficiente excede de 2 en valor absoluto*
- En caso contrario, la variable *no es importante*, y se retira del modelo, que vuelve a estimarse con las variables que tienen una  $t$  superior a 2.

En este uso del estadístico  $t$  hay 3 problemas:

### **Problema 1:**

- Para determinar la significación estadística de un coeficiente, la precisión en la estimación juega un papel fundamental: un estadístico  $t$  puede ser elevado incluso si el coeficiente estimado es pequeño => Rechazaríamos  $H_0$ , o puede ser reducido incluso con un coeficiente estimado alto

### **Problema 2:**

- Equívoca identificación de tres conceptos diferentes:
  - significación estadística de un coeficiente,
  - relevancia económica de la variable asociada
  - capacidad explicativa/contenido informativo de  $X$  sobre  $Y$

### **Problema 3:**

***El concepto de “variable significativa” no existe***

## *... y en comparaciones entre variables o entre submuestras*

- Evaluamos de manera inapropiada la capacidad explicativa relativa de varias variables sobre Y
  - ✓ Comparando sus coeficientes estimados: *“La variable más importante para explicar el comportamiento de Y es Z”*.
  - ✓ Comparando los valores numéricos de sus estadísticos- $t$  : *“La variable más significativa en el modelo que hemos estimado es Z”*
  - ✓ O comparamos el valor numérico del estadístico  $t$  de una misma variable en dos submuestras diferentes: *“La variable X se ha hecho más significativa después de 2007”*
- Pero ninguno de los dos criterios tiene justificación sólida
  - ✓ Por la interpretación de los coeficientes individuales
  - ✓ Por las dificultades asociadas al abuso habitual del estadístico  $t$

# *Interpretación de coeficientes individuales*

¿Qué es un coeficiente estimado alto o bajo?

¿Cómo medimos el impacto de X sobre Y?

**Problema 1:** la estimación numérica del coeficiente depende crucialmente de la colinealidad entre variables explicativas

**Problema 2:** el efecto de  $x_t$  sobre  $y_t$  depende no sólo de  $\beta$ , sino también de la volatilidad de  $x$

- El problema no se resuelve completamente con parámetros-elasticidades (variables en logaritmos)
- Siendo necesario utilizar el rango de variación muestral observado, la desviación típica muestral, el rango intercuartílico, etc.
- Tener en cuenta la variabilidad muestral en las variables explicativas
  - ✓ Efecto sobre Y de cambios en el valor de X : desde  $\min(X)$  a  $\text{Max}(X)$ , o desde el percentil 10% al percentil 90% de X (aunque esto ignora colinealidad)
  - ✓  $100 \cdot \beta \cdot R(x)/R(y)$        $100 \cdot \beta \cdot DT(x)/DT(y)$

# Estrategias de diversificación en las exportaciones manufactureras – REA 2003

	Indice de dispersión geográfica de exportaciones	Empleo <50	Empleo (50,100)	Empleo (100,200)	Empleo >200	Segundas líneas productos	Participación capital extranjero	Ratio importador	Concent. industrial	I+D/Ventas	Publicidad/ Ventas	Concent. provincial
<b>Medias muestrales</b>												
Media	0,243	0,203	0,349	0,221	0,226	4,375	0,094	0,114	0,185	0,005	0,009	0,175
Beta			0,026 (4,29)	0,057 (7,25)	0,099 (13,2)	0,004 (2,56)	0,047 (6,53)	-0,015 (1,96)	-0,006 (1,54)	0,051 (2,35)	1,302 (6,44)	0,087 (6,79)
Producto						0,018		-0,002	-0,001	0,000	0,012	0,015
Mínimo	0	0	0	0	0	0	0	0	0,036	0,001	0,001	0,052
Máximo	0,560	1	1	1	1	82	1	0,724	0,810	0,083	0,086	0,745
Producto			0,026	0,057	0,099	0,328	0,047	-0,011	-0,005	0,004	0,111	0,060
<b>Desviaciones típicas</b>												
	0,170					5,480		0,200	0,140	0,010	0,060	0,180
						0,022		-0,003	-0,001	0,001	0,078	0,016

**Taking into account sample variation**

Positive size effect. ✓

Positive effect from the number of products (✓) and from foreign capital participation (≠).

Negative effect from import coefficient (≠)

No effect from market concentration (✓)

Positive effect from the other two product differentiation variables: Advertising (✓) and R&D expenditures (≠)

Competitors in the region of production (✓)



*Trends in income inequality and its impact on economic growth, F. Cingano, OCDE 2014*

Variable dependiente: Crecimiento del PIB sobre 5 años				
	decilas 1 y 8	decilas 2 y 8	decilas 3 y 8	decilas 4 y 8
Desigualdad por abajo (1)	-0,032 (0,018)	-0,083 (0,029)	-0,132 (0,047)	-0,198 (0,084)
Desigualdad por arriba (2)	-0,054 (0,723)	-0,377 (0,465)	-0,233 (0,395)	-0,085 (0,441)
(1): Ratio entre renta disponible media y la renta de la decila				
(2): Ratio entre renta la renta de la decila y la disponible media				

# *El contenido informativo de una variable sobre otra es un concepto condicional*

La interpretación habitual ignora que **el contraste de capacidad explicativa es siempre condicional** en el resto de las variables incluidas como explicativas (y, por supuesto, en la muestra considerada, la frecuencia de datos utilizada, etc.)

- El contenido informativo de X sobre Y es una propiedad de la variable X en un determinado modelo
- Pero a menudo comparamos la capacidad explicativa de X sobre Y en modelos diferentes !!!
- No podemos contrastar el contenido informativo o la capacidad explicativa de X sobre Y en un modelo de regresión múltiple
  - Para ello, necesitamos un modelo de regresión simple
  - En la regresión múltiple únicamente podemos contrastar el contenido informativo que X **añade** a las demás variables incluidas en el modelo sobre Y
- la colinealidad juega un papel central en esta apreciación
- ... y también en la interpretación de los coeficientes estimados

## *6. Tratamiento de la colinealidad*

- El crecimiento monetario podría aparecer como significativo en una regresión simple de la tasa de inflación, pero no en una regresión que ya incluye los tipos de interés como variable explicativa
- Que la tasa de crecimiento monetario contenga información relevante acerca de la tasa de inflación no implica necesariamente que su adición a cualquier modelo que explique la tasa de inflación aporte capacidad explicativa adicional.
- Porque... la colinealidad entre variables explicativas condiciona que una variable adicional aporte información. La colinealidad requiere un tratamiento adecuado, que permita discutir la adición de capacidad explicativa por parte de una variable.

## *Tratamiento de la colinealidad*

- En presencia de colinealidad, no podemos interpretar los coeficientes individuales, si bien podemos medir con precisión la capacidad explicativa conjunta del vector de variables X
- Los coeficientes de correlación entre Y y X son realmente inferiores a lo medido, pues incorporamos junto con la información contenida en X, sus correlaciones con otras variables Z que también explican a Y.
- Estrategia: Establecer un ranking de importancia entre las variables explicativas. Para ello, pueden utilizarse análisis individuales con respecto a Y: regresiones simples, coeficientes de correlación, nubes de puntos
- Estimamos una regresión de cada una de las variables explicativas sobre todas las que les preceden en dicho ranking, y nos quedamos con los residuos. Las variables así construidas, están incorrelacionadas entre sí  $\Rightarrow$  Podemos interpretar directamente los coeficientes estimados
- Es cierto que no estimamos coeficientes para las variables originales, pero es todo lo que podemos estimar con precisión y permitiendo una interpretación incuestionable.
- Siempre podemos intentar recuperar los coeficientes de las variables originales, utilizando las regresiones auxiliares.

## Omitted variable bias: are we getting it right?

$$Ventas_t = 247,6 + 2,204 Pub_t - 1,464 P_t$$

(0,545)                      (0,649)

$$Ventas_t = 96,0 + 3,224 Pub_t$$

(0,375)

An increase of 6.200 euros in advertising = 1 standard deviation

Corr(Pub,P) = -0,829  $\Rightarrow$  decrease of -0,829 standard deviations in price = 4,327 euros.

Total effect on sales: (6,200)(2,204) - (1,464)(-4,327) = 20,0 thousand euros

Single variable model: (3,224) (6,200) = 20,0 thousand euros

- The simple linear regression measures the *global effect (direct+indirect)* on Y of a change in either one of the X
  - It is a biased estimation of the *ceteris paribus individual effect (indirect)*, which is seldom interesting
  - But it is an unbiased estimate of the global effect, the one on which we are usually interested
- While the multiple regression provides a biased estimation of that global effect, because of collinearity
- What is the question ?

Dependent Variable: RENTA  
 Sample: 1 83  
 Included observations: 58  
 Excluded observations: 25

**T1**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.699	0.168	45.74	0.00
OIL	1.130	0.246	4.59	0.00
INTER	-0.434	0.446	-0.97	0.33
PREM	-0.344	0.339	-1.01	0.31
INFLAC	0.032	0.370	0.08	0.93
OVER	-0.143	0.287	-0.49	0.61
DEMOC	0.218	0.180	1.21	0.23
ESTAB	-0.258	0.204	-1.26	0.21
EFICA	0.158	0.308	0.51	0.61
REGULA	0.632	0.207	3.04	0.00
LEYES	0.273	0.235	1.16	0.25
CONTROL	0.353	0.297	1.18	0.24
R-squared	0.674	Mean dependent var	7.577	
Adjusted R-squared	0.596	S.D. dependent var	0.896	
S.E. of regression	0.569	Akaike info criterion	1.894	

**Cleaning “non-significant variables” and re-estimating the model**

	RES_T1	RES_T4	RES_T5
RES_T1	1.00		
RES_T4	0.949	1.00	
RES_T5	0.768	0.801	1.00

Dependent Variable: RENTA  
 Sample: 1 83  
 Included observations: 70  
 Excluded observations: 13

**T4**

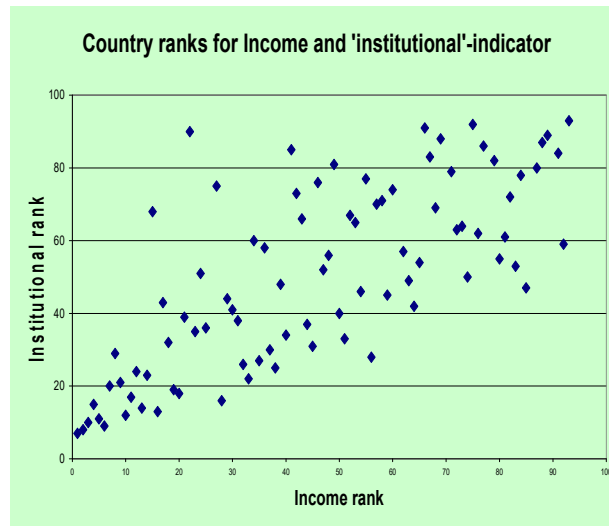
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.652	0.111	68.79	0.00
OIL	1.093	0.204	5.34	0.00
INTER	-0.748	0.338	-2.21	0.03
REGULA	0.717	0.133	5.36	0.00
CONTROL	0.677	0.126	5.34	0.00
R-squared	0.656	Mean dependent var	7.544	
Adjusted R-squared	0.635	S.D. dependent var	0.923	
S.E. of regression	0.557	Akaike info criterion	1.737	

Dependent Variable: RENTA  
 Sample: 1 83  
 Included observations: 82  
 Excluded observations: 1

**T5**

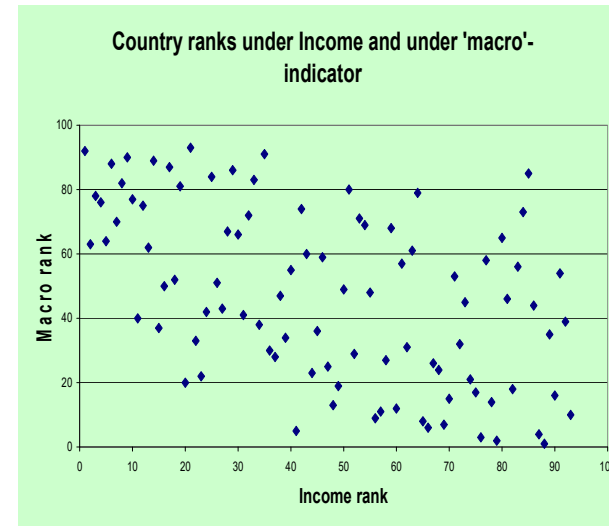
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.376	0.081	90.56	0.00
OIL	0.700	0.233	2.99	0.00
REGULA	0.946	0.116	8.15	0.00
R-squared	0.477	Mean dependent var	7.470	
Adjusted R-squared	0.464	S.D. dependent var	0.943	
S.E. of regression	0.690	Akaike info criterion	2.133	

## Factores económicos vs. Factores institucionales como determinantes del crecimiento



Rank corr: .71 (9.2)

Utilizando institutions\macro:  
Rank corr: .50 (4.9)



Rank corr: -.51 (-5.7)

Utilizando macro\institutions:  
Rank corr: -.11 (-1.0)



Dependent Variable: RENTA. Method: Least Squares  
 Sample: 1 93. Included observations: 87 Excluded observations: 6

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.734	0.072	108.48	0.00
INSTITUTIONS	1.826	0.166	10.98	0.00
MACRO_INSTITUTIONS	-0.211	0.144	-1.46	0.15

R-squared	0.594	Mean dependent var	7.688
Adjusted R-squared	0.584	S.D. dependent var	1.029
S.E. of regression	0.664	Akaike info criterion	2.052
Sum squared resid	37.018	Schwarz criterion	2.137
Log likelihood	-86.28	F-statistic	61.35
Durbin-Watson stat	2.25	Prob(F-statistic)	0.000

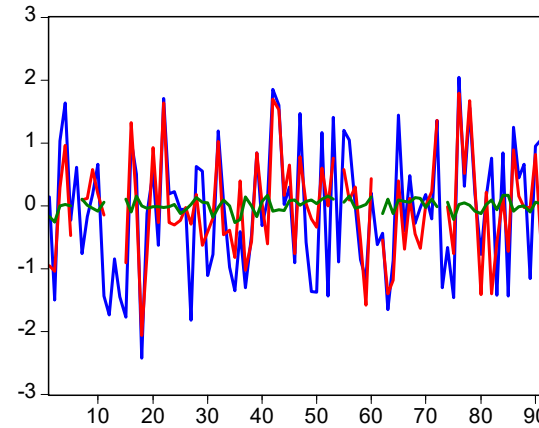
## Orthogonalize explanatory variables

Dependent Variable: RENTA Method: Least Squares  
 Sample: 1 93. Included observations: 87 Excluded observations: 6

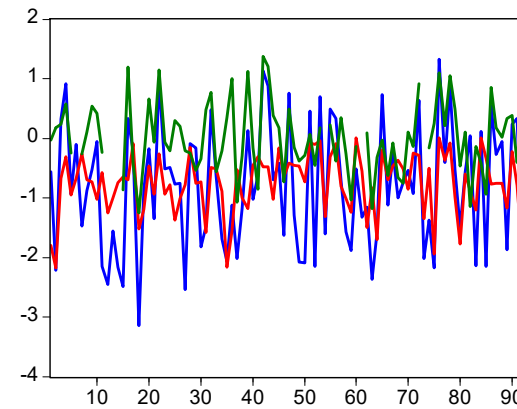
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.446	0.125	67.285	0.000
MACRO	-0.879	0.119	-7.329	0.000
INSTITUTIONS_MACRO	1.663	0.200	8.307	0.000

R-squared	0.594	Mean dependent var	7.688
Adjusted R-squared	0.584	S.D. dependent var	1.029
S.E. of regression	0.664	Akaike info criterion	2.052
Sum squared resid	37.018	Schwarz criterion	2.137
Log likelihood	-86.28	F-statistic	61.35
Durbin-Watson stat	2.25	Prob(F-statistic)	0.000

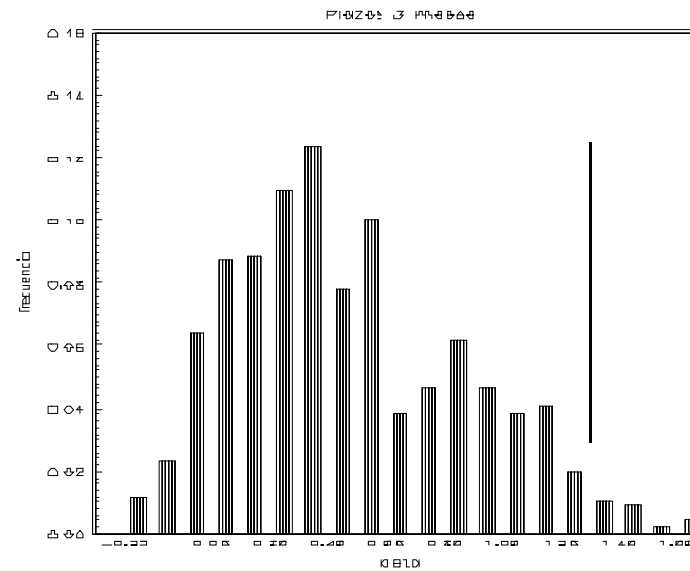
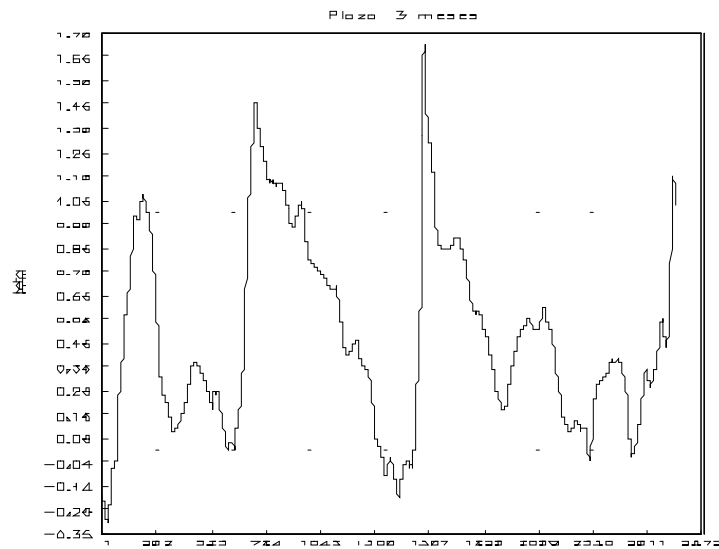


— RENTA\_CORR — EXP\_INST — EXP\_MACRO\_INST



— RENTA\_CORR2 — EXP\_MACRO — EXP\_INST\_MACRO

# *Coeficiente estimado por MCO recursivos del tipo forward a 3 meses sobre el contado futuro*



## *Explorar relaciones no lineales y cambios en las relaciones*

- *Modelo potencial*

$$r_t = \beta_0 + \beta_1 f_{t-s,s}^\gamma + u_t,$$

- *Modelos de cambio de régimen con probabilidades exógenas*

$$r_t = \beta_0 + \beta_1 f_{t-s,s} + u_t, \quad u_t \sim \text{N}(0, \sigma_1^2), \text{ si } \sigma_{r_t}^2 > K$$

$$r_t = \alpha_0 + \alpha_1 f_{t-s,s} + u_t, \quad u_t \sim \text{N}(0, \sigma_2^2), \text{ si } \sigma_{r_t}^2 < K$$

*Contrastar propiedades que no deberían observarse si la teoría es correcta*

*Información útil:  
lo que el modelo no explica*

## *Después de estimar: validación del modelo:*

- Examinar siempre el gráfico de valores de Y junto con los residuos obtenidos,
  - En su evolución temporal, si se trata de datos temporales, y siempre en forma de nube de puntos,
  - y el coeficiente de correlación entre Y y los residuos
  - La reducción en desviación típica al pasar de Y a los residuos
  - Examinar los residuos en grupos de observaciones, especialmente en datos transversales: calidad del ajuste en distintas submuestras (MUY IMPORTANTE)

## *Evitar un excesivo resumen de la información muestral*

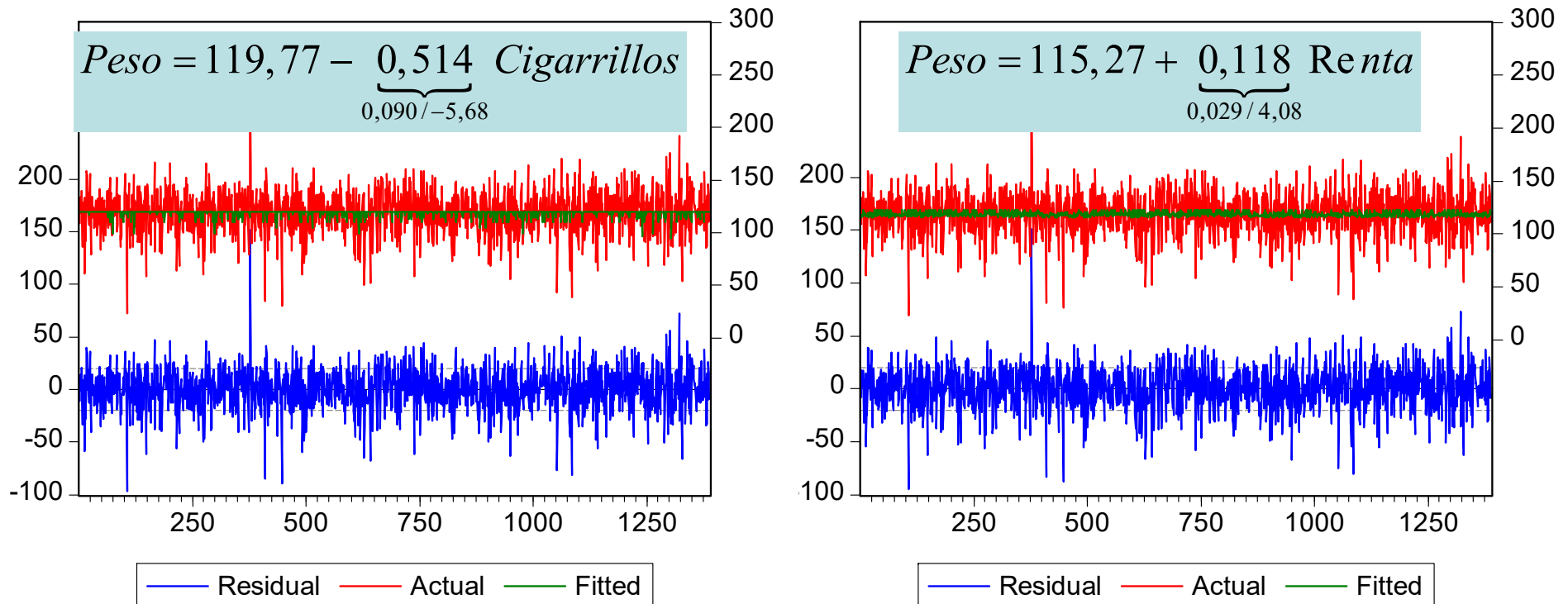
### **Examinar los residuos en detalle**

- Cuando explican dos modelos “lo mismo”?
- Representación gráfica de los datos, los valores ajustados y los residuos
  - They contain complete information on the estimated model
  - Avoiding too much information reduction
  - Search for local information content: a specific but short period of time, a particular set of observations in cross-section data sharing some common characteristic: omitted variables.
- Discover subsamples of poor model fitting
- How to compare alternative models:
  - How much similarity between sets of residuals should we expect to find?
  - What is explained by a model which is not explained by others?
- ***Are we missing most of the story?***

It is relatively easy to find evidence contrary to  $H_0$ : *lack of statistical significance of  $\beta$*  with a large sample even though the associated variable has little information content. **Type I error (?)**.

$$t = \frac{\hat{\beta} - \beta_0}{DT(\hat{\beta})} = \frac{\hat{\beta} - \beta_0}{\sqrt{\sigma_u^2 / \sum_1^n (x_i - \bar{x})^2}}$$

Example (Wooldridge 2003): Explain child birth weight in terms of: cigarettes smoked by the mother per day during pregnancy, family income, education of father and mother, birth order of the child among siblings.



## *S2. Alternative approach to testing parametric constraints*

- *The constraints being tested cannot be rejected on the basis of the sample evidence if and only if the residuals of the restricted and unrestricted models contain the same information regarding the question which is the ultimate object of research*
- *Variable  $x_t$  adds information to vector  $z_t$  to explain  $y_t$  iff the residuals of the model that explains  $y_t$  by vector  $z_t$  are significantly different from those of the model that explains  $y_t$  by  $(z_t, x_t)$ .*
- What matters is not so much the statistical test, but the detailed comparison between the two sets of residuals.
- Different approaches to testing for equality of residuals:
  - Nonparametric point-to-point tests for equality of residual sets
  - Tests for reduction in absolute size / variance in subsamples of largest residuals or in the whole sample.
  - Any appropriate test?



*Contrastando el contenido informativo de X sobre Y mediante la comparación de residuos de los modelos restringido y sin restringir*

Irrelevant Z-variable: x:50%, z:0%; u:50%

Relevant Z-variable: x:42%, z:15%, u:42%

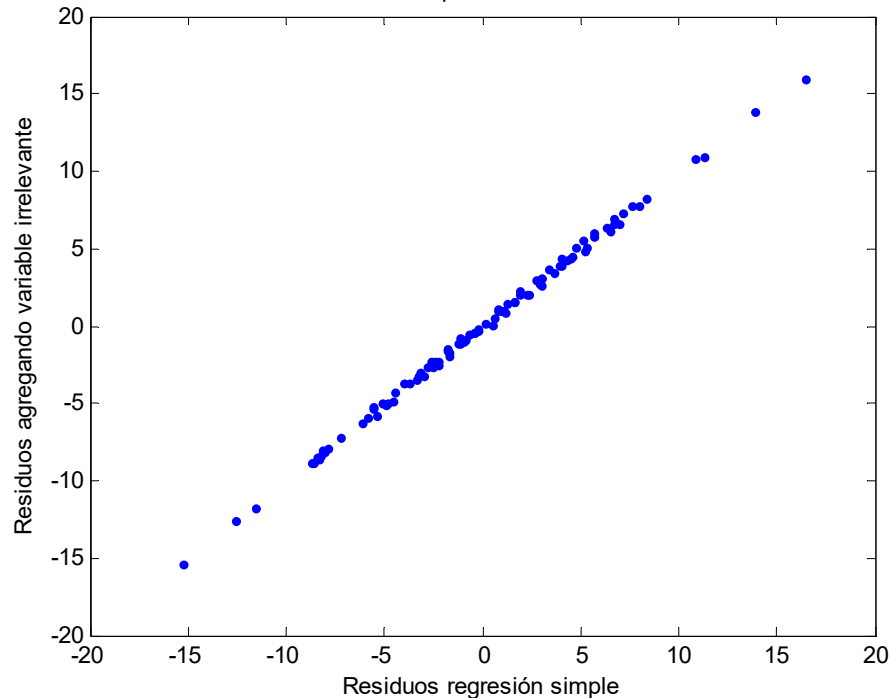
**Correlations:**

	Y	X	Z	u
Y	1.00	0.78	-0.06	0.74
X		1.00	-0.01	0.15
Z			1.00	-0.07
u				1.00

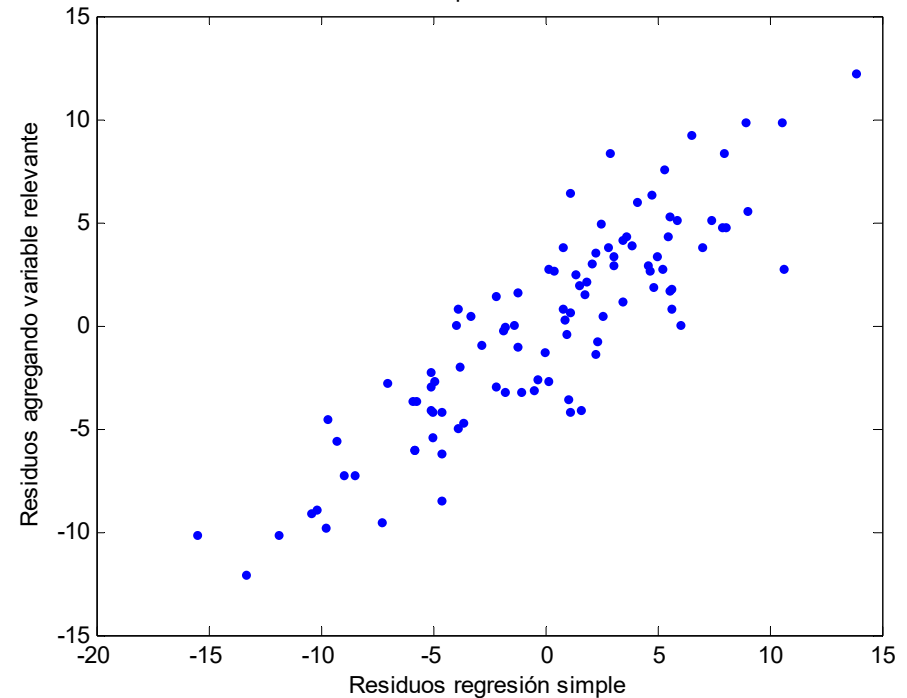
**Correlations:**

	Y	X	Z	u
Y	1.00	0.66	0.33	0.71
X		1.00	-0.04	0.08
Z			1.00	-0.01
u				1.00

Variable explicativa irrelevante

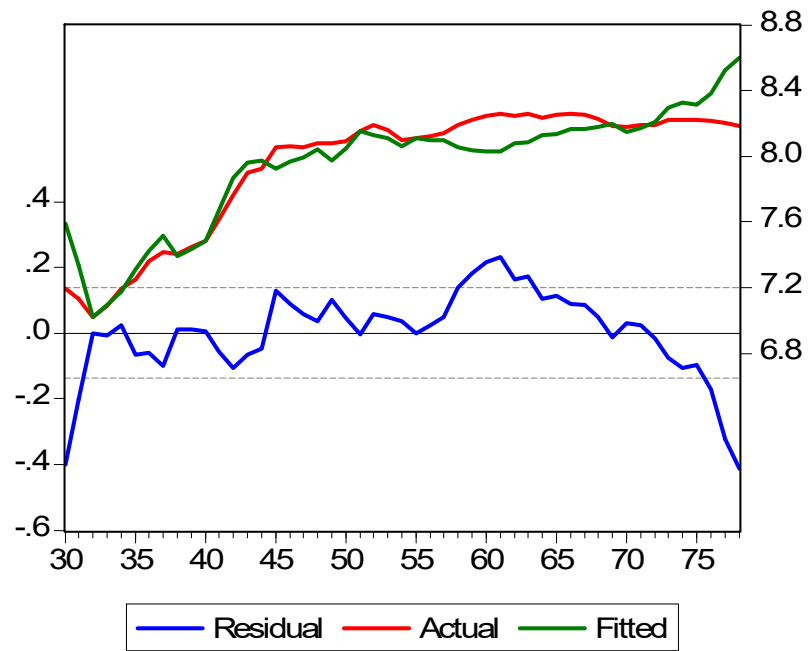


Variable explicativa relevante



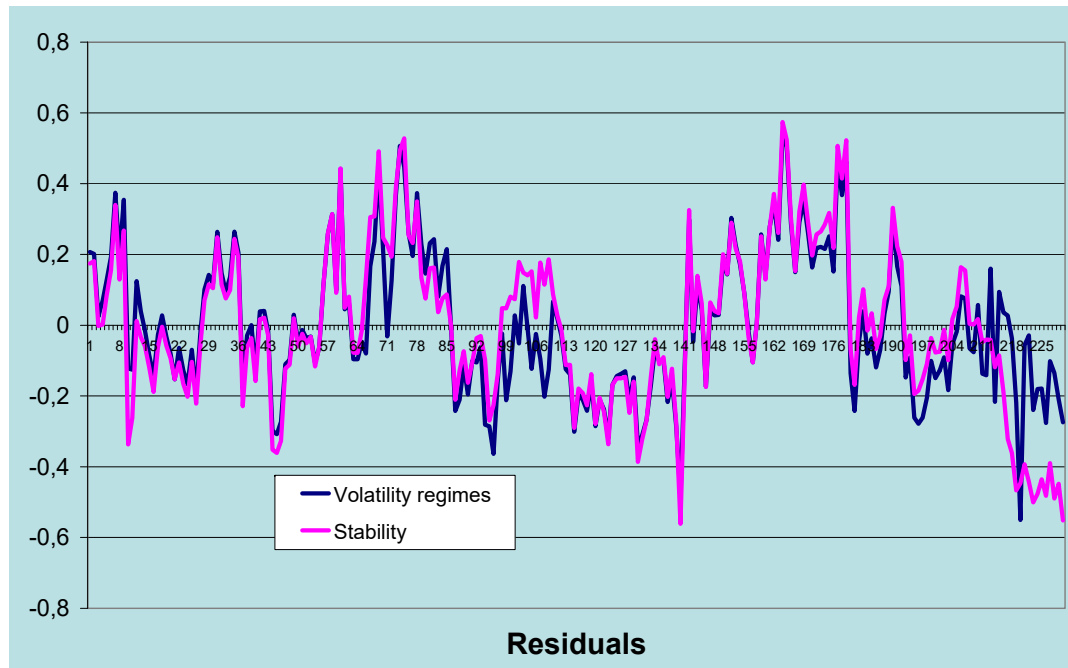
# Subsamples of model deterioration

*Consumo de tabaco explicado por la renta y el precio*



What is explained by a model which is not explained by competing models?

*Explaining credit spreads by regime switching models*



*Pero ¿no es esto lo que ya venimos  
haciendo?*

# *Lo que NO debe hacerse:*

1. *No identificar significación estadística y contenido informativo/capacidad explicativa*
2. *No contrastar la significación estadística de un coeficiente. Evaluar el contenido informativo en una variable como un concepto condicional, y evaluarlo de modo apropiado*
3. *No resumir excesivamente la información muestral*
4. *No apoyarse exclusivamente en estadísticos tipo  $t$  y  $F$  al evaluar restricciones paramétricas o al comparar modelos alternativos*
5. *No comparar el contenido informativo de distintas variables examinando sus coeficientes o sus estadísticos  $t$*
6. *No practicar econometría de signos o de asteriscos*
7. *No poner mucho énfasis en el valor numérico de coeficientes individuales estimados en regresiones múltiples*
8. *No contrastar demasiado...*
  - *... y si se contrasta, y no se rechaza la hipótesis nula, no olvidar analizar la potencia del contraste*
9. *No olvidar nunca a qué pregunta queremos responder con la investigación*
10. *No emitir conclusiones categóricas*

# Lo que debe hacerse:

1. *Comenzar definiendo con claridad la (única) cuestión objeto de análisis. No perder dicha referencia a lo largo del trabajo empírico*
2. *Analizar los datos desde distintos puntos de vista, utilizando una variedad de procedimientos estadísticos*
  - *Describiendo en detalle las características de los datos : rango muestral, histogramas, diagramas stem and leaf, evaluación no paramétrica de asociación entre variables*
  - *Diseño de contrastes paramétricos y no paramétricos acerca de:*
    - ✓ *características que deberían observarse si la hipótesis en estudio fuese correcta*
    - ✓ *características que **no** deberían observarse en los datos si la hipótesis en estudio **no** fuese correcta*
  - *Contrastar hipótesis paramétricas mediante comparación punto a punto de residuos de la regresiones restringida y sin restringir*
3. *Proporcionar evidencia acerca de variabilidad paramétrica y no-linealidades*
4. *Evaluar la relevancia práctica de cada variable explicativa*
5. *Comparar modelos (evaluar restricciones) desde el punto de vista de su impacto sobre la cuestión que se analiza: predicción, medición de riesgo, gestión de carteras...*
  - *Qué explica un modelo que no es explicado por otros modelos? Residuos*
  - *Cómo de distintos son dos modelos alternativos?*
  - *Simulación de modelos*
6. *Tratar la colinealidad y tener en cuenta la variabilidad muestral de cada variable*

*FIN (por ahora)*

- ***What is the final question?***

- *what is the implication of asymmetric effects on conditional variance on VaR calculations?*
- *is there any forecasting improvement from imposing long-run cointegration relationships?*
- *is time between trades weakly exogenous with respect to volume?*



### S3. *How can we decide on information content?*

- *Variable  $x_t$  adds information to vector  $z_t$  to explain  $y_t$  iff the residuals of the model that explains  $y_t$  by vector  $z_t$  are significantly different from those of the model that explains  $y_t$  by  $(z_t, x_t)$ .*
- We move from testing for statistical significance to testing for equality of sets of residuals
- A threshold of  $\rho = .95$  between residuals from restricted and unrestricted models would have solved the problems when testing with large samples
- Questions that are often asked:
  - *Does Z add any information to X to explain Y ?*
  - *Is it X or Z that contains more information on Y?*
    - Look at residuals from single variable projections
    - Orthogonalize using different orderings
    - Are we interpreting correctly the *omitted variable bias*?

## ¿Se cumple una determinada hipótesis teórica?

- Potencia del contraste: capacidad de discriminar hipótesis falsas
  - Potencia en el diseño: ¿Cuántas teorías alternativas implicarían un determinado hecho empírico?
  - Seleccionar cuidadosamente la caracterización empírica de la hipótesis teórica
  - Potencia estadística: Si establecemos un contraste formal, no dejemos de contrastar hipótesis muy diferentes
  - Y restricciones paramétricas que no deberían cumplirse si nuestra hipótesis nula fuese cierta
  - La curva de potencia del contraste es obligada cuando no se rechaza una hipótesis
- 
- Excesivo uso de la teoría estadística de contrastación de hipótesis: *Too much testing*
  - Que se aplica frecuentemente con poco rigor: *Poor testing*

## *Si contrastamos ...*

1. Especificar contrastes unilaterales siempre que sea posible. Tener en cuenta el efecto del tamaño muestral.
2. Examinar evidencia a favor de una hipótesis comparando residuos de modelos restringido y sin restringir. No contrastar mucho.
3. La capacidad explicativa es condicional o adicional. Correcta interpretación de contrastes de significación
4. ¿Cómo contrastar el contenido informativo de una variable explicativa?
5. Contrastes de hipótesis son sólo una de las herramientas. No utilizar mecánicamente:
  - ✓ Tener presente la relación entre precisión y potencia
  - ✓ Distinguir entre significación estadística y contenido informativo

## *Causalidad, evaluación de políticas*

# *Ciencia no experimental*

- La Economía es una ciencia no experimental:
  - Disponemos de única muestra,
  - Factores explicativos son estocásticos, no controlables,
  - Contienen información común: colinealidad: condiciona los contrastes de hipótesis hacia no rechazar  $H_0$ : pérdida de potencia
  - Variabilidad paramétrica y de modelo: estimamos promedios
  - Incertidumbre de modelo, de parámetros, muestral
  - Posible incumplimiento de supuestos subyacentes a la distribución de los estadísticos de contraste habituales  $t$ ,  $F$  :
- El uso de un estimador insesgado no debe ser primordial en Economía
- Significa que, si estimásemos con muchas muestras, el promedio de las estimaciones numéricas obtenidas con dicho método de estimación sería próximo al verdadero valor (desconocido) del coeficiente
- Pero en Economía no experimentamos  $\Rightarrow$  disponemos de una sola muestra
- Estaríamos a salvo si la varianza del estimador fuese muy reducida, pero solo sabemos aspectos parciales:
  - MCO es de mínima varianza entre los estimadores lineales (bajo supuestos)
  - Sería el de menor varianza bajo Normalidad

# *¿Están justificadas las propiedades de los estimadores en que nos basamos?*

- Consistencia: *al aumentar el tamaño de la muestra, la distribución de probabilidad del estimador tiende a concentrarse alrededor del verdadero valor (desconocido) del parámetro*
- La consistencia es más apropiada, al ser propiedad de una sola muestra, según aumenta su tamaño
- ¿Cuándo es una muestra suficientemente grande? Homogeneidad muestral
- ¿Es mejor utilizar datos frecuentes?
  - Mayor variabilidad del componente no explicado
  - Variables de distinta frecuencia de observación (MIDAS)
- Conviene tener un enfoque pragmático, y utilizar un procedimiento de estimación que se comporte “aparentemente” bien de acuerdo con nuestros fines y supuestos
- Parece más prometedor prestar atención a:
  - la variabilidad temporal o muestral de los coeficientes,
  - la existencia de distintos regímenes en la relación, etc.,
  - y tratar estas cuestiones aun con un estimador relativamente no eficiente, que utilizar un estimador supuestamente eficiente y utilizar toda la muestra de una sola vez
- Y evitar las posibles situaciones de inconsistencia
  - Modelo dinámico con autocorrelación
  - Simultaneidad
  - Errores en las variables

## *Significación estadística vs. Relevancia económica*

- Es perfectamente concebible que podamos medir con bastante precisión el efecto de una variable poco importante cuantitativamente  $\Rightarrow$  estadístico  $t$  por debajo de 2. Sin embargo, su impacto es numéricamente despreciable ¿queremos decir que es una variable relevante?
- Y también podemos medir con poca precisión una variable cuantitativamente importante ¿queremos decir que es una variable irrelevante?

***P10: Errónea interpretación habitual de la posible precisión estadística en la estimación de un coeficiente con la relevancia económica de la variable que le acompaña***

# ¿Es $H_0$ cierta?

- El olvido de la función de potencia conduce a *identificar equivocadamente la ausencia de rechazo con la demostración de que la hipótesis nula es correcta*,
  - olvidando que el rechazo de la hipótesis nula requiere evidencia muestral favorable a la hipótesis alternativa
  - y sin considerar posibles alternativas
- Lo cual se justifica por el excesivo afán por probar la falsedad o veracidad de una hipótesis, más que por evaluar la verosimilitud de la misma
- Contra la práctica habitual, no tiene justificación enfatizar hipótesis que no se hayan rechazado, a menos que se haya comprobado la potencia del contraste



*Recordemos:*

*“No rechazar una hipótesis no equivale a haberla probado cierta”*

Paradojas:

Estadístico  $t > 2$  en restricciones de dudoso cumplimiento

Estadístico  $t < 2$  en restricciones sin incumplimiento evidente

En contrastes de significación estadística:

Estadístico  $t > 2$  sin verdadero contenido informativo

Estadístico  $t < 2$  con contenido informativo

## *Alternativa a la contrastación de hipótesis*

- No reducir la información muestral a unos pocos estadísticos. No perder la información visual que proporcionan unos buenos gráficos
- Examinar la información contenida en los residuos de los modelos estimados con y sin las restricciones
- La evidencia relativa a la cuestión en estudio puede obtenerse mediante una variedad de procedimientos estadísticos y gráficos que no reducen la información muestral a un sólo número (análisis extensivo).
- El dogmatismo es el peor enemigo del avance del conocimiento. No utilizemos exclusivamente “el mejor método”. No son buenos los métodos, sino el modo en que se utiliza y se interpretan los resultados que proporcionan
- Utilizar una variedad de procedimientos metodológicos para responder a una determinada pregunta: métodos estadísticos y métodos econométricos, gráficos y numéricos, paramétricos y no paramétricos, no sujetos a supuestos acerca de la distribución de probabilidad y válidos en muestras finitas
- Examinar la validez de restricciones paramétricas que no deberían observarse si la hipótesis nula fuese cierta
- , sin utilizar las ventajas ofrecidos por los contrastes no-paramétricos o por métodos gráficos que exploren las posibles relaciones entre variable dependiente y variables explicativas o entre estas últimas. A estos enfoques no están influidos por supuestos acerca de la distribución de probabilidad de las variables, su estacionariedad, etc.. Son simples de utilizar y tienen una distribución de probabilidad conocida en muestras pequeñas.

# Excesivo resumen de la informacion muestral

El modelo más general, modelo ARCH-Poisson-Gausiano se especifica como sigue:

$$\Delta r_t = \mu_t + \sigma_t \Delta z_t + J_t \Delta n_t; \quad \Delta z_t \sim N(0,1); J_t \sim N(\theta_t, \psi^2); \Delta z_t \text{ y } J_t \text{ son}$$

independientes.;  $\sigma_t^2 = \omega_0 + \omega_1 [\Delta r_{t-1} - E_{t-2}(\Delta r_{t-1})]^2$ ;  $\mu_t$  y  $\theta_t$  están definidos en la siguiente tabla.

Parámetros	$\mu_t = \alpha(r_{t-1} - r_{t-1}^*)$		$\mu_t = \alpha_1(r_{t-1} - r_{t-1}^*)^+ + \alpha_2(r_{t-1} - r_{t-1}^*)^-$	
	Modelo i $\theta_t = \theta$	Modelo ii $\theta_t = \gamma(r_{t-1} - r_{t-1}^*)$	Modelo iii $\theta_t = \theta$	Modelo iv $\theta_t = \gamma(r_{t-1} - r_{t-1}^*)$
$\alpha$	-0.046 (-6.601)	-0.0344 (-6.365)	-----	-----
$\alpha_1$	-----	-----	-0.0465 (-6.244)	-0.0388 (-7.832)
$\alpha_2$	-----	-----	-1.0928 (-263.81)	-0.9943 (-61.837)
$\omega_0$	0.00003 (5.123)	0.00003 (7.189)	0.0004 (10.127)	0.0004 (12.833)
$\omega_1$	0.6240 (5.922)	0.6092 (5.542)	0.6651 (6.338)	0.6482 (6.053)
$\theta$ or $\gamma$	-0.0087 (-1.196)	-0.4152 (-8.179)	-0.009 (-1.170)	-0.2758 (-4.406)
$\psi$	0.2101 (12.939)	0.1903 (14.249)	0.2003 (12.294)	0.1930 (13.035)
$\delta_0$	0.1209 (7.841)	0.1128 (6.917)	0.1025 (7.005)	0.1023 (7.209)
$\delta_1$	0.5260 (10.201)	0.5271 (11.546)	0.5500 (10.714)	0.5347 (10.431)
$\delta_2$	0.5257 (8.838)	0.5076 (9.830)	0.5233 (8.486)	0.5306 (8.460)
$\delta_3$	0.1421 (3.582)	0.1309 (4.239)	0.1344 (3.493)	0.1476 (3.568)
$\delta_4$	0.8307 (10.345)	0.8060 (10.055)	0.8763 (10.372)	0.8847 (10.998)
$\delta_5$	0.3896 (2.741)	0.3569 (3.080)	0.4368 (2.843)	0.4373 (3.367)
Log-L [ SIC ]	2186.46 [2147.01]	2222.14 [2182.69]	2237.55 [2194.51]	2256.05 [2218.01]

Resultados de la estimación por máxima verosimilitud con desviaciones típicas robustas de Bollerslev y Wooldridge (1992). Estadístico  $t$  entre paréntesis. Log-L es el logaritmo de la función de verosimilitud

Los coeficientes  $\delta_0, \delta_1, \delta_2, \delta_3, \delta_4$  y  $\delta_5$  designan respectivamente las probabilidades de salto: (i) en los

1. Modelos que tienen igual valor de la medida de ajuste: R2, log-likelihood ¿explican lo mismo?

2. ¿Cuál es la diferencia entre modelos que tiene distinto R2?

3. ¿A que cuestión queremos responder? ¿Cómo cambia esa respuesta entre distintos modelos?

- volatilidad
- saltos
- VaR

## *Un (importante) inciso*

- ¿Se puede estimar por mínimos cuadrados si el término de error tiene autocorrelación?
- ¿Y si tiene autocorrelación, o ambas cosas a la vez?
- ¿Y en presencia de colinealidad?
- La ausencia de eficiencia estadística significa que existe otro estimador más eficiente, pero no implica inconsistencia
- La construcción de un estimador teóricamente más eficiente (MCG, por ej.) precisa hacer determinados supuestos y estimar parámetros adicionales, lo cual introduce error muestral en las estimaciones de los coeficientes de interés, posiblemente impidiendo la ganancia de eficiencia que se buscaba.
- El estimador de Whitney-West
- Estimación por Método Generalizado de Momentos

*A modo de resumen ...*

# 1. *Antes de estimar: análisis descriptivo de las variables y sus asociaciones*

- Herramientas gráficas:
  - Histogramas de frecuencias: ¿sobre qué rango de valores tenemos información? Ejemplo: muestra de hombres y mujeres, o de distintos niveles educativos, o de razas, etc.
  - Gráficos de cajas
  - Nubes de puntos con la variable dependiente (contenido informativo)
  - Nubes de puntos entre cada par de variables explicativas (colinealidad/posible pérdida de precisión)
- Estadísticos:
  - Estadísticos básicos: mediana, media, moda (en variables discretas), desviación típica, máximo, mínimo, rango intercuartílico. En ocasiones, asimetría y curtosis.
  - Tipo de distribución: Normal, etc.
  - Medias o Medianas condicionales de la variable
  - Matriz de correlaciones entre variables explicativas y con la variable dependiente. No estrictamente válido para variables discretas, especialmente si son de naturaleza cualitativa
- Asociación entre cada variable explicativa y el indicador a explicar.
  - Tests no paramétricos: test de signos, Wilcoxon, test de Pearson sobre discretización
  - Medida de asociación con variables discretas
  - Gráficos: nubes de puntos, evolución temporal
- Asociación entre pares de variables explicativas (colinealidad/precisión)

## *Distinción importante: variables discretas vs. variables continuas*

- En ocasiones, la diferencia entre variables discretas y continuas no es clara.
- Variable discreta: toma un número reducido de valores. En ocasiones, una variable continua se trata como discreta, mediante agrupación.
- Las variables ficticias son discretas. No requieren un tratamiento especial. Sus coeficientes son fáciles de interpretar.
- Otras variables interesantes son las cualitativas no numéricas que, a veces, convertimos en cualitativas asignándoles valores numéricos *arbitrarios*
- Situación especial: cuando es la variable dependiente quien es discreta: modelos probit, logit

## *Asociación entre X discreta e Y*

- Contraste de igualdad de medianas de Y para distintos valores de una variable discreta X
  - Análisis de discriminación salarial: Si el salario de hombres y mujeres de igual fuese similar, el salario mediano en ambos grupos no debería aparecer como estadísticamente diferente
  - El contraste de igualdad de medianas de Y (salario) para los subgrupos definidos por los valores de la variable discreta (experiencia, nivel educativo) no se ve afectado por el significado real de los valores numéricos asignados a la variable discreta, que pudo ser cualitativa originalmente.
  - El rechazo de la hipótesis nula de igualdad de medianas puede venir unido a evidencia de asociación positiva o negativa entre las medianas de Y y las marcas de clase de X, lo que sugerirá *capacidad explicativa* en X.
  - Si el número de clases es elevado, podemos agrupar primero en un conjunto menos numeroso.



## *No son problemas, sino situaciones a tratar:*

- Los llamados “problemas del modelo de regresión” o “problemas con los datos”...
  - Heterocedasticidad
  - Autocorrelación
  - Multicolinealidad exacta, que no existe
- ... no son tales problemas, sino situaciones *siempre presentes*, que pueden tratarse satisfactoriamente
  - La heterocedasticidad sólo afecta a la precisión de la estimación
  - La autocorrelación afecta a la precisión de la estimación y a la calidad de las previsiones.
  - Pero no es claro cuanto nos acercamos al verdadero valor de los coeficientes al utilizar un método más eficiente (MCG)

## *Las verdaderas dificultades son:*

1. La omnipresente colinealidad, que afecta a la precisión de los estimadores ya la interpretación de los coeficientes estimados
2. La regresión espúrea: evidencia aparente de estrecha relación, donde no la hay, producida por la no estacionariedad de variables dependiente y explicativas, no cointegradas
3. La variación paramétrica a lo largo de la muestra, tanto en datos temporales como en datos de panel

# *Estimación*

1. Estimar modelos para cuantificar impactos (bajo determinados supuestos: estabilidad, invariancia de otras variables explicativas), pero intuyendo ya la mayor parte de los resultados cuantitativos.
2. Indagar acerca de posible inestabilidad paramétrica y, de haberla, utilizar algún procedimiento que permita variación temporal
3. Al interpretar coeficientes individuales:  
Tratar adecuadamente la colinealidad: regresiones simples y múltiples, efectos parciales y efectos totales  
Medición del impacto, de acuerdo con la variabilidad muestral de la variable
4. Evitar la econometría de signos y la econometría de asteriscos

## *Bad statistical practice in applied economics is not specific of regression methods*

- Some non-regression examples:
  - ✓ *Qualitative/Limited dependent variable* models in one or more levels, *Technological efficient frontier* models (loglinear-type specifications)
    - with a large list of explanatory variables whose statistical significance and quantitative effect on Y is individually analyzed
    - Making statements on which variables explain Y and which ones do not
    - If their significance or explanatory power has increased or decreased between different samples, or when adding an additional variable into the model
    - Which variable is the most significant
  - ✓ Strong emphasis on testing an economic theory on the basis of existence of cointegration relationships
    - At fixed significance levels
    - Searching for a single 'best' test, instead of deepening into understanding short- and long relationships
    - ... and the effect of parameter variation
    - Is there any forecasting improvement from imposing long-run relationships?

*7. Predicción:  
Constrastes de capacidad predictiva*

# *The general background*

## **Problem I: Evaluate forecasts from a single model: Are forecasts useful?**

- Indicators: correlation between forecast and data, ability to predict direction of change, lack of serial correlation of order  $h-1$ , or a standard symmetric function like square or absolute value, possibly as a percentage.

## **Problem II: Compare the forecasting ability of alternative models, based on a comparison of forecast errors over a period of time**

- Choice of a loss function defined on forecast errors
- Two models are said to predict equally or have the same predictive ability if the **average** loss is not very different from each other
- Even though, period by period, they could predict very differently
  - ✓ *A model fits the increases and misses the decreases, and the other model does the opposite, by a similar global amount*
  - ✓ *A model fits the first part of the forecast period and misses completely the second part, and the other model does the opposite*
  - ✓ *A model makes a single large mistake, while the other runs into many small forecast errors*

## *The usual choice of loss function*

- Under a square loss function, Mean Square Error (MSE), the two models are compared on the basis of the *variance* of the two sets of forecast errors
- Under the Mean Absolute Error (MAE) criterion, we compare them on the basis of their *average* absolute or percent size.
- In a few cases, an indicator function has been chosen as loss function: has the forecast the same sign than the data? *Proportion* of right signs
- What is the loss function for qualitative forecasting?
- Some criticism on the choice of loss function
  - Excessive summary of sample information
  - Functional form appropriate for the final goal
    - *In some applications, what matters is the number of errors, but not so much their size*
    - *Sometimes, a large error is more important than many small errors*
    - *In some cases, positive errors are more important than negative errors of same size (or the other way around)*
    - *Or an economic loss function (implied volatility forecasting)*

*Can we formally test?:  
The Diebold-Mariano (DM) framework*

- Loss function for Model  $i$ :  $f_{it} = f(y_t, \hat{y}_{it})$
- Null hypothesis:  $H_0 : E[f_{it}] = E[f_{jt}]$
- Define the difference function:  $d_t = f_{it} - f_{jt} = f(y_t, \hat{y}_{it}) - f(y_t, \hat{y}_{jt})$

- Asymptotic  $N(0,1)$  test statistic:  
$$S_1 = \frac{\bar{d}}{\sqrt{\frac{2\pi\hat{h}_d(0)}{N}}} = \frac{N^{-1}\sum_{t=1}^N (f_{it} - f_{jt})}{\sqrt{\frac{2\pi\hat{h}_d(0)}{N}}}$$



# Remarks on Diebold-Mariano

- Contributions of DM:
  - ✓ It allows for non-Gaussian, non-zero mean, serially correlated and contemporaneously correlated forecast errors
  - ✓ General loss functions (non-symmetric, non-differentiable)
  - ✓ Under a quadratic loss: Tests for significance of MSE differences
  - ✓ Everything centers around estimation of the variance of  $\bar{d}$
  - ✓ How many forecasts do we need to be able to discriminate between competing models?
- Limitations in the implementation of the DM test:
  - ✓ Extensive use of quadratic loss function, even in DM Monte Carlo analysis: continuity, symmetry
  - ✓ Compensation of forecast errors. Tests based on the average loss differential and its estimated variance
  - ✓ Excess sensitivity to extreme forecast errors: DM size and power decrease dramatically under an extreme forecast error

- Dar un solo número
- Incertidumbre de diversos tipos

# *Provincias+sector+año*

