Too much testing, poor testing: Statistical inference in empirical economics

Alfonso Novales
Departamento de Economía Cuantitativa
Universidad Complutense
February 2007

The context

- A fair amount of applied research is devoted to testing parametric restrictions in relatively complex models estimated with relatively sophisticated methods.

- Suggesting a data approach of attempting to show a theory to be either false or true, rather than the alternative of approaching the data searching for evidence consistent with our prior belief.

- Economics is a non-experimental science
  1. So we usually have a single sample available,
  2. The model inputs are non-controlable, thus stochastic,
  3. They contain common information $\Rightarrow$ colinearity, that biases tests towards not rejecting H0: loss of power
  4. *Parameter instability:* variability in the relationship to be estimated
  5. Distribution of standard *t, F* test statistics ?

The questions

- Theory testing
  - Should parametric hypothesis tests be run with the purpose of testing theory?
  - Can we find *conclusive evidence* in the data on a given theoretical model? Can we prove a theory to be true?

- Model testing
  - What is a good model? How can we compare alternative models?
  - Do we want to say that *k*-1 of the available models are (equally) wrong? And that a single model is right?

- What is the role of the data analyst? To produce definite conclusions? To inform the reader on the information contained in the data regarding the question of interest?

- References: D.McCloskey, A.Zellner, S.Ziliak, J. Cohen, Armstrong, Box

Part I : Too much testing

Reason 1: Documenting sample information

- Provide information on descriptive statistics: observed values (discrete vs. continuous), maximum, minimum, sample range, percentiles, mean, median, standard deviation, histogram

- Use non-parametric statistics and graphical procedures, not subject to assumptions on the probability distribution of innovations and valid in small samples

- Analyze association between each explanatory variable and the dependent variable: rank correlation coefficients, contingency tables, … Pay attention to association with discrete variables

- Analyze degree of conditional association

- Examine degree of colinearity by computing measures of association between pairs of explanatory variables

- The evidence relative to the question under study can be obtained by a variety of statistical and graphical methods that do not reduce the sample information to a few estimated statistics.

- Use a variety of methods. Focus on the question, not on the methodology.

Reasons 2 and 3: Hypothesis tests in Economics should often be one – sided

- Model: $s_{t+k} = \alpha + \beta f_t + u_t$
- H0: $\beta=1$ versus: H1: $\beta<1$

- **Reason (Paradox) 2: It does not make sense to test:**

  - Estimate: 1.15 (.30) We do not reject H0 (!!!) [Only one of the two conditions holds]

- **Reason (Paradox) 3: We should not test:**

  - Estimate: .65 (.25) H0 is not rejected (!!!) [Low precision]
  - *Key issue: Relationship between precision in estimation and power in testing*

An illustration:
Testing the Expectations Hypothesis of the term structure of interest rates

Future spot interest rate projected on lagged forward rate

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

H0: β=1 versus: H1: β<1

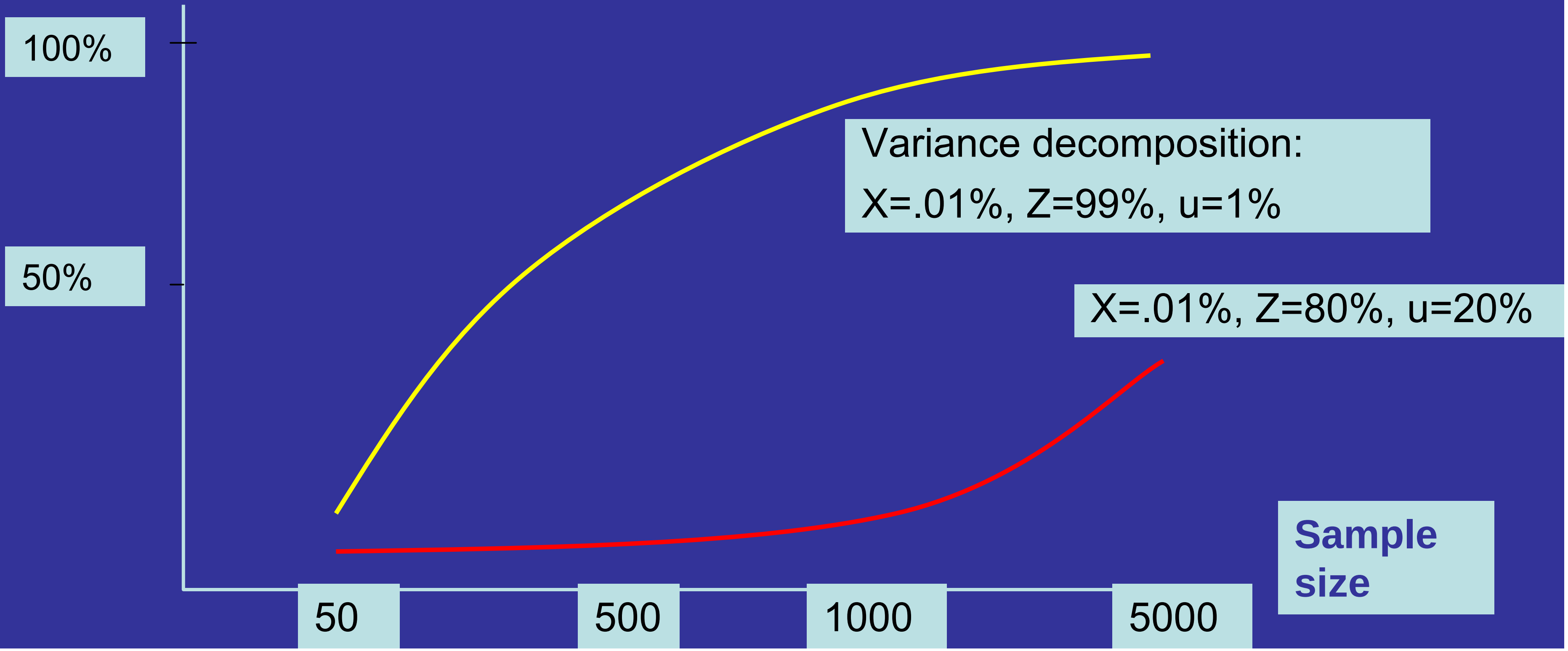
Maturity	Beta	Standard dev.	R2	t stat H0: Beta = 1
1 month	0,96	0,04	0,54	-1,00
3 months	0,93	0,06	0,42	-1,17
6 months	1,10	0,15	0,32	0,67
1 year	1,40	0,24	0,22	1,67
3 years	1,63	0,47	0,15	1,34
5 years	2,40	0,85	0,04	1,65

Reason 4: Too large a sample size

- Simulation experiment

$$y = \beta_1 x + \beta_2 z + u, \;\; Corr(x,z) = Corr(x,u) = Corr(z,u) = 0$$

- How the frequency of rejection of H0: $\beta_x = 0$ changes with T



Reason 5: Too much information reduction / Poor use of sample information

- Reducing the sample information to the numerical value of a very few test statistics
  - without a detailed exam of the information content in residuals from the restricted and the unrestricted models
  - an excessive concentration on hypothesis testing, without exploiting useful non-parametric tests or graphical methods to explore relationships between explanatory and dependent variables or among explanatory variables, which are less influenced by assumptions on distribution theory, lack of stationarity, …, that are simple to implement and have a well-justified probability distribution in small samples

- In spite of all which written conclusions are often very emphatic:
  - "we have *shown* that…."
  - "we have obtained *conclusive evidence* on the relevance of variable X to explain the behavior of variable Y …"

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

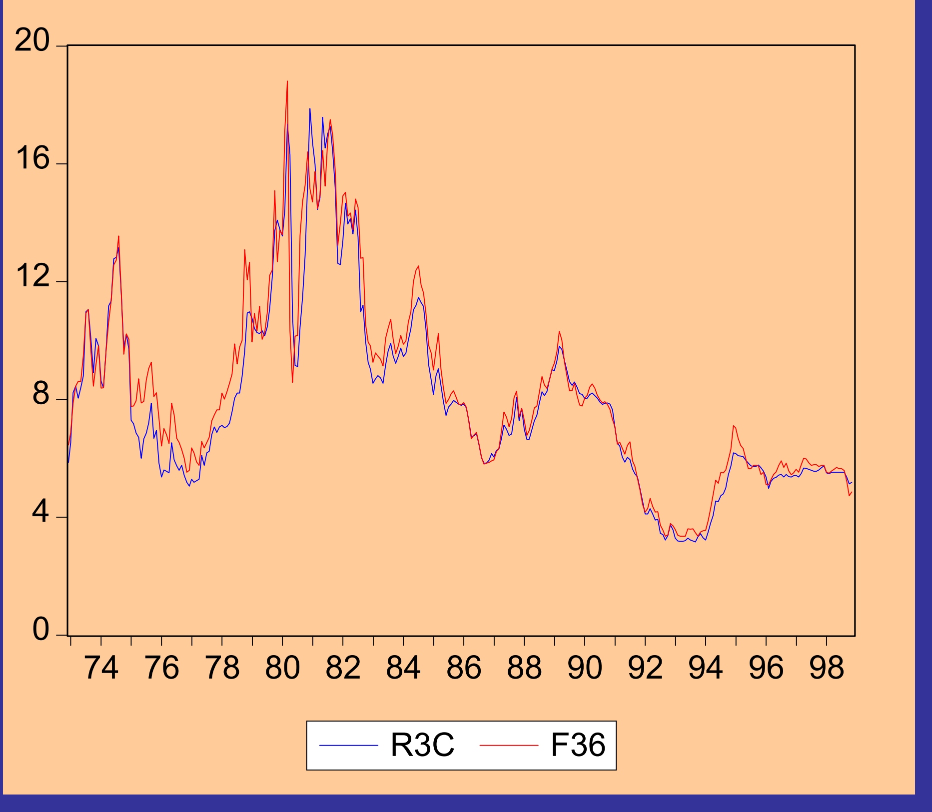
Hipótesis de Expectativas:
El tipo forward como predictor del tipo a corto futuro

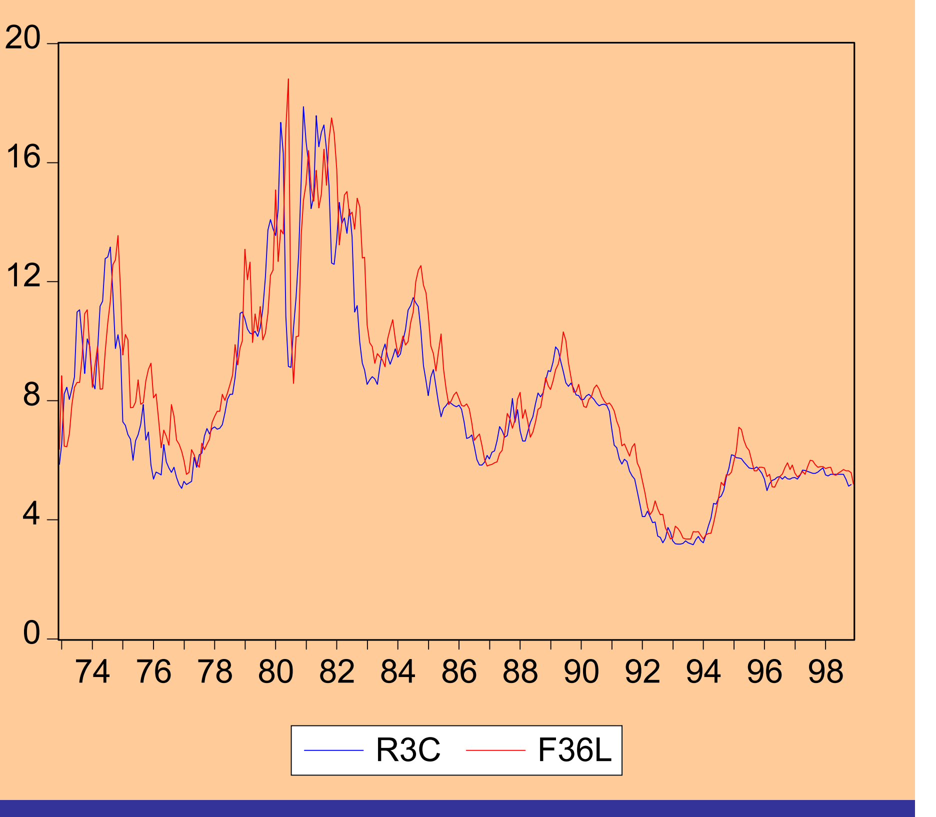
R3C: Tipo contado a 3 meses

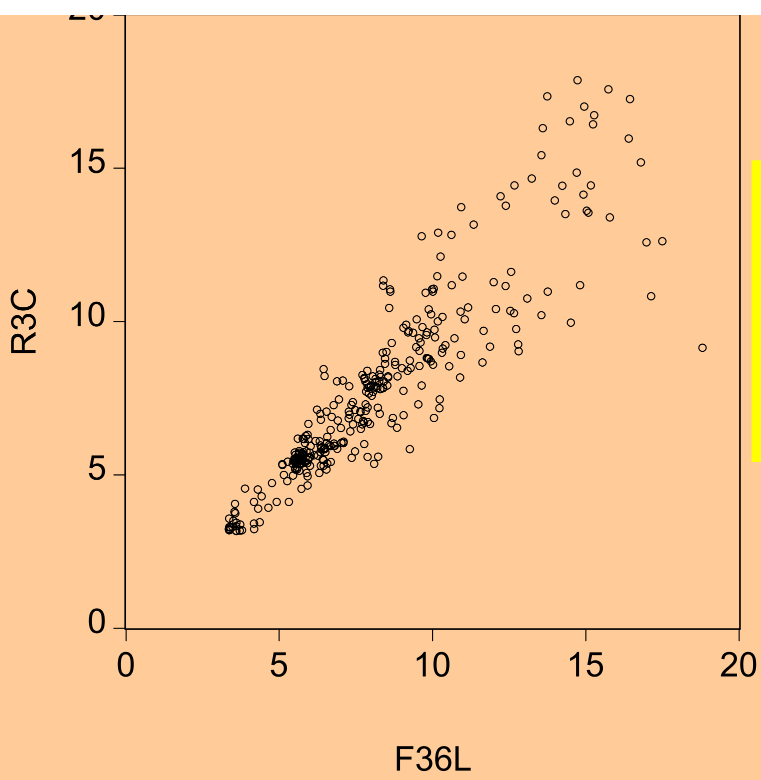
R6C: Tipo contado a 6 meses

F36: Forward 3/6 meses

F36L: Forward retardado 3 meses





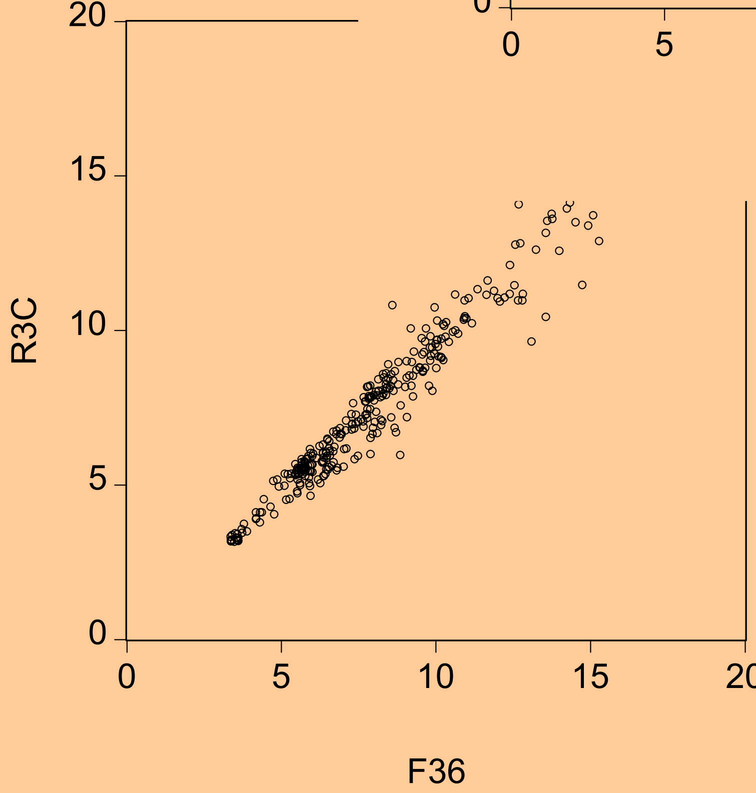


R3C: Tipo contado a 3 meses

R6C: Tipo contado a 6 meses

F36: Forward 3/6 meses

F36L: Forward retardado 3 meses



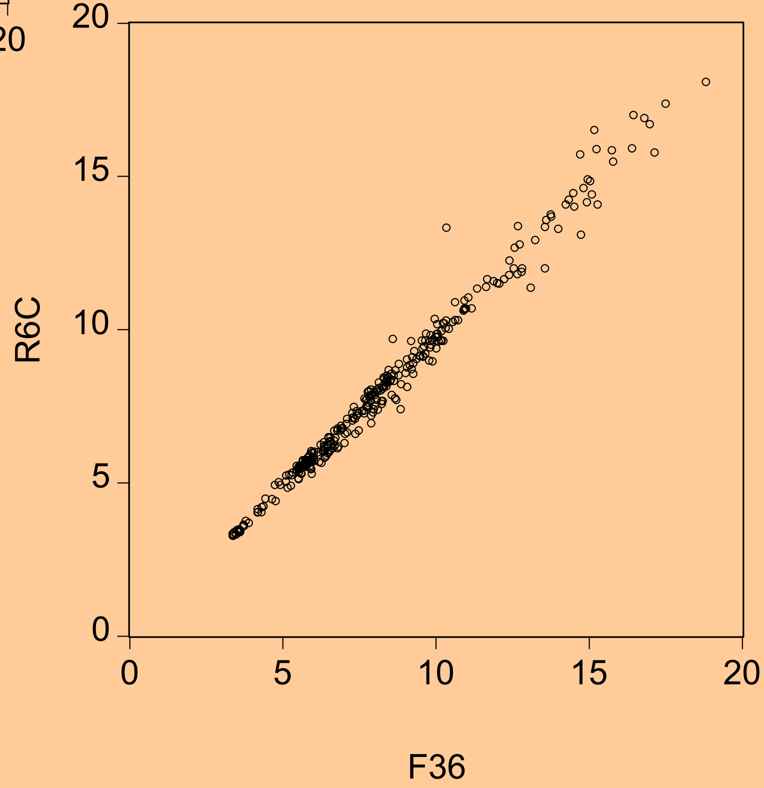


Table 5
Estimated cointegrating relationship:

$$r^{6}_{t+6} = \alpha + \beta f^{6}_{t,t+6} + u_{t+6}, \quad f^{6}_{t,t+6} = 2r^{12}_{t} - r^{6}_{t}$$

	Maturity	λ_{max}/Trace[a]	Parameter estimates: 1984-1998 β[b]	n[c]	H_0: β =1[d]	ADF and Phillips-Perron statistics[e]: $r^{m}_{t}-f^{m}_{t-m,t}$
US	1 m.	15.0* / 22.1**	0.979 (0.010)	4	1.28 (0.26)	-5.6(2) / -10.0
	3 m.	10.9 / 14.1	0.983 (0.019)	12	0.04 (0.84)	-4.8(9) / -5.7
	6 m.	14.8* / 19.0*	0.986 (0.037)	12	0.36 (0.55)	-3.7(12) / -4.3
Yen	1 m.	29.6*** / 31.8***	0.998 (0.008)	9	0.29 (0.59)	-7.2(3) / -13.0
	3 m.	23.8*** / 26.5***	1.029 (0.008)	6	13.56 (0.00)	-5.9(3) / -6.9
	6 m.	17.5*** / 19.4*	1.081 (0.027)	12	12.58 (0.00)	-4.8(2) / -4.1
DM	1 m.	121.9*** / 25.0**	0.918 (0.030)	12	0.02 (0.90)	-3.8(12) / -9.9
	3 m.	12.8 / 16.4	0.996 (0.028)	4	2.20 (0.14)	-3.8(6) / -5.6
	6 m.	11.0 / 13.8	1.071 (0.044)	12	4.46 (0.03)	-4.2(12) / -4.0

Notes: (a) Critical values when testing the existence of *zero cointegrating relationship*, at the 10%, 5% and 1% significance levels, are 13.8, 15.7 and 20.2 for the *Maximum eigenvalue,* and 17.8, 20.0 and 24.7, for the *Trace* statistic (Osterwald-Lenum (1992)).
(b) Numbers in brackets are maximum-likelihood standard errors.
(c) Number of lags used in the *VAR* in first differences.
(d) *Likelihood ratio* statistic to test the null hypothesis that the cointegrating vector is (1,-1).
(e) Critical values for both tests at the 10%, 5% and 1% significance levels are -1.62, -1.94 and -2.57 when no constant is included in the vector autoregression, and -2.57, -2.87 and -3.46 when a constant is included.

Part II : Poor testing

Reason 1: Statistical significance versus information content

- Identifying statistical significance of a coefficient with explanatory power or information content of the associated variable,
  - ✓ Forgetting the relationship between precision in estimation and power in hypothesis testing
  - ✓ Possibility of estimating with high precision a numerically small coefficient
  - ✓ Or with low precision a large coefficient, because of sample variation, poor signal to noise ratio, nonlinearities, colinearity, …

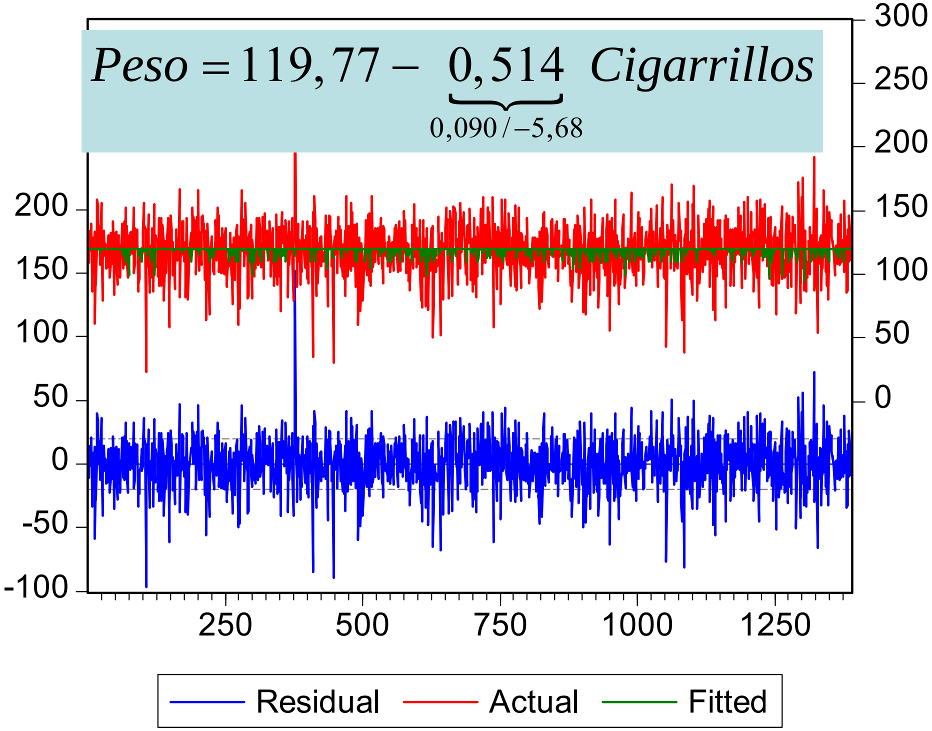
- Leading to serious fallacies and misinterpretations of estimation results
  - ✓ Qualitatively: X explains Y iff its associated coefficient is statistically different from zero
  - ✓ Quantitatively: explanatory power ⇔ high absolute value of Student's *t*
  - ✓ Sign econometrics
  - ✓ Asterisk econometrics

It is relatively easy to find evidence contrary to
H0: *lack of statistical significance of β*
with a large sample even though the associated variable has little information content. **Type I error (?).**

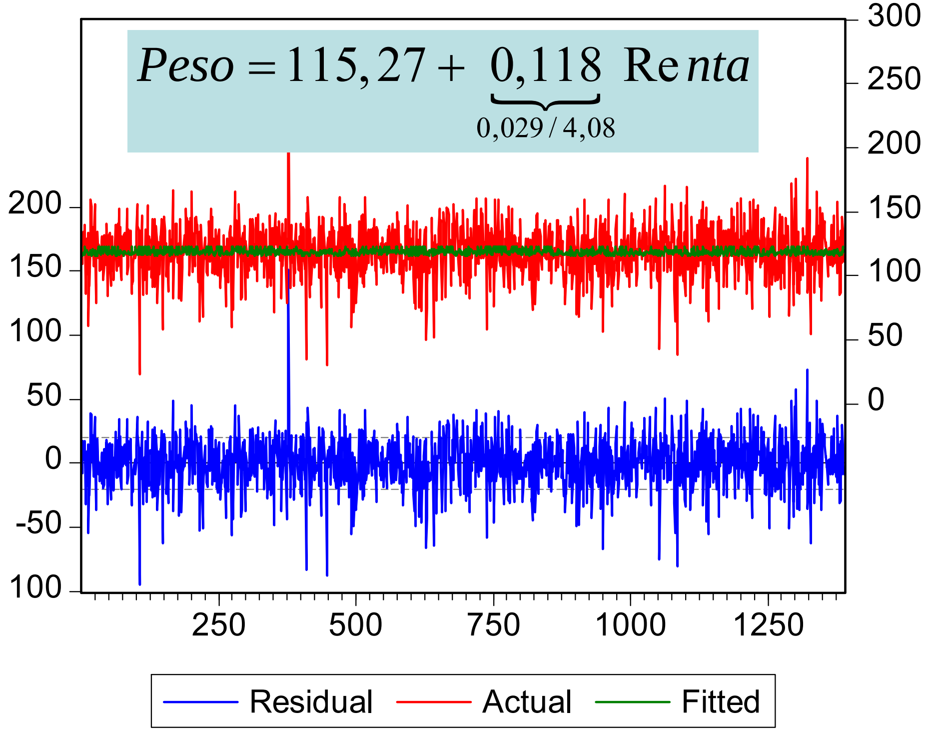
$$t = \frac{\hat{\beta} - \beta_0}{DT(\hat{\beta})} = \frac{\hat{\beta} - \beta_0}{\sqrt{\sigma_u^2 / \sum_1^n (x_i - \overline{x})^2}}$$

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.

$$Peso = 119,77 - \underbrace{0,514}_{0,090\,/\,-5,68}\ Cigarrillos$$



$$Peso = 115,27 + \underbrace{0,118}_{0,029\,/\,4,08}\ Renta$$



Dependent Variable: PESO
Sample: 1 1388 IF EDUCP <> NA AND EDUCM <> NA
Included observations: 1191

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	114.5	3.728	30.71	0.00
CIGS	-0.595	0.110	-5.40	0.00
EDUCP	0.472	0.282	1.67	0.09
EDUCM	-0.370	0.319	-1.15	0.24
FAMINC	0.056	0.036	1.53	0.12
ORDENAC	1.787	0.659	2.71	0.00

R-squared	0.038	Mean dependent var	119.5
Adjusted R-squared	0.034	S.D. dependent var	20.14
S.E. of regression	19.78	Akaike info criterion	8.813

EDUCP: father's education level

EDUCM: mother's education

CIGS: cigarettes/day smoked by mother

FAMINC: family income

ORDENAC: order among siblings

Sample: 1388

Smokers among mothers : 212 (15,2%)

Dependent Variable: PESO
Sample: 1 1388 IF EDUCM<>NA AND EDUCP<>NA
Included observations: 1191

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	117.8	1.21	96.98	0.00
CIGS	-0.632	0.107	-5.87	0.00
ORDENAC	1.763	0.657	2.68	0.00

R-squared	0.032	Mean dependent var	119.52
Adjusted R-squared	0.031	S.D. dependent var	20.14
S.E. of regression	19.82	Akaike info criterion	8.81

Dependent Variable: PESO
Sample: 1 1388
Included observations: 1191
Excluded observations: 197

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	109.0	3.937	27.68	0.00
CIGS	-0.598	0.109	-5.450	0.00
ORDENAC	1.915	0.655	2.923	0.00
RENTA	0.043	0.036	1.183	0.23
EDUCM	-0.328	0.317	-1.033	0.30
EDUCP	0.411	0.281	1.463	0.14
MALE	3.795	1.142	3.321	0.00
WHITE	4.713	1.607	2.931	0.00

R-squared	0.054	Mean dependent var	119.52
Adjusted R-squared	0.048	S.D. dependent var	20.141
S.E. of regression	19.64	Akaike info criterion	8.800

F-statistic: 9.67 , p-value=0

Dependent Variable: PESO
Sample: 1 1388 IF EDUCM<>NA AND EDUCP<>NA
Included observations: 1191

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	111.0	1.942	57.1	0.00
CIGS	-0.627	0.106	-5.87	0.00
ORDENAC	1.913	0.653	2.92	0.00
WHITE	5.364	1.572	3.41	0.00
MALE	3.738	1.141	3.27	0.00

R-squared	0.050	Mean dependent var	119.5298
Adjusted R-squared	0.047	S.D. dependent var	20.14
S.E. of regression	19.66	Akaike info criterion	8.799

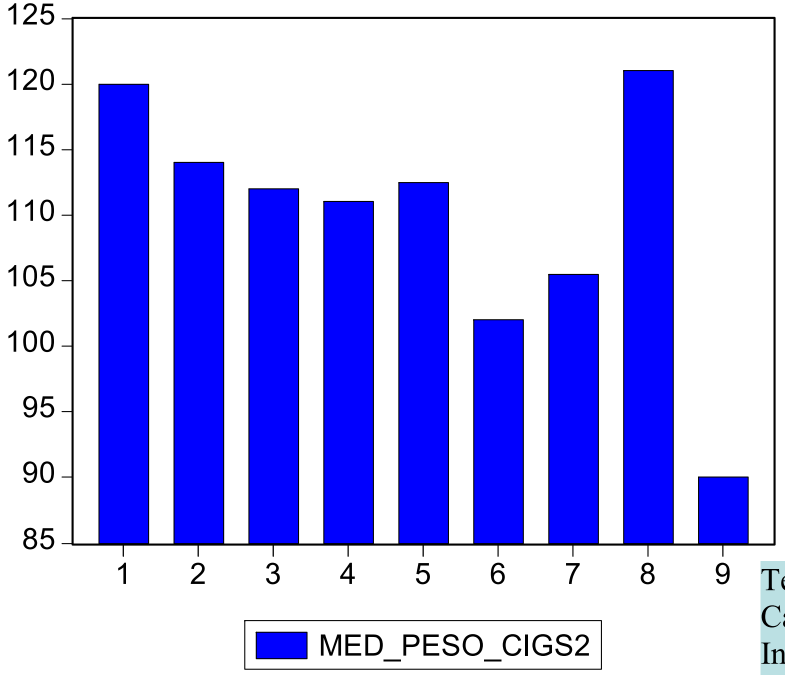
¿What kind of information do we really get out of the t - statistic?

Linear correlation coefficients between dependent variable (weight) and residuals from alternative models

	PESO	R1	R2	R3	R4	R5	R6	R7	R8
PESO	1.0								
PESO_CIGS	0.990	1.0							
PESO_RENTA	0.994	0.987	1.0						
PESO_CIGSRENTA	0.987	0.996	0.992	1.0					
PESO_CIGSEDUCP	0.984	0.998	0.984	0.996	1.0				
PESO_CIGSEDUPM	0.984	0.997	0.983	0.995	0.999	1.0			
PESO_TODAS	0.980	0.993	0.983	0.994	0.995	0.996	1.0		
PESO_TODAS_FIC	0.972	0.985	0.975	0.986	0.987	0.988	0.992	1.0	
PESO_TODAS_NOORD	0.976	0.989	0.978	0.990	0.991	0.991	0.989	0.996	1.0

In spite of which there is sample evidence on the fact that smoking during pregnancy affects child birth weight: *difference in median weights* …

… that boys weight more than girls and white babys weight more than comparable non-white babys



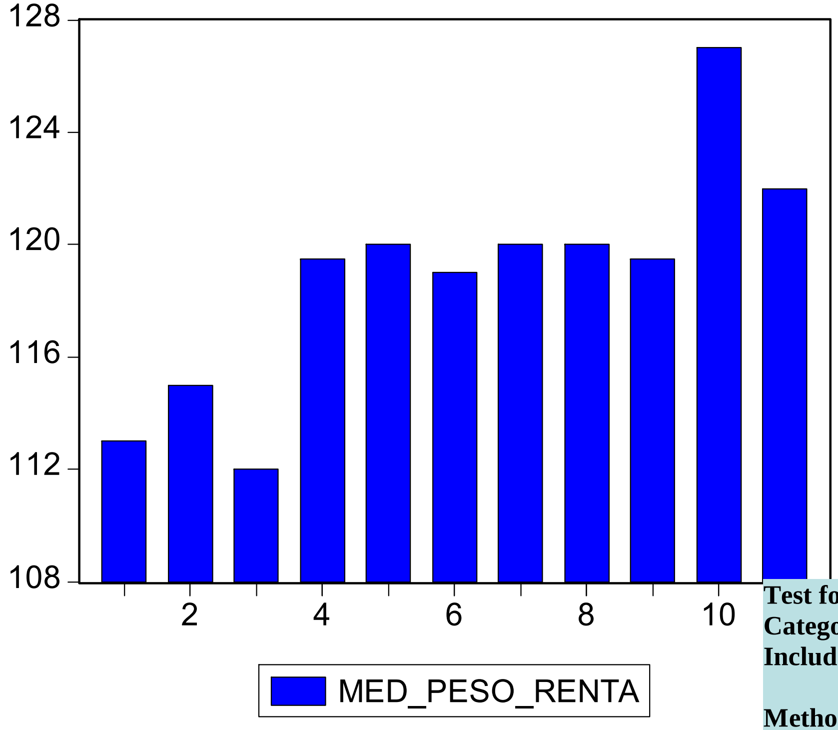
¿Does the number of cigarettes smoked during pregnancy affect the weight of the baby at birth?

Test for Equality of Medians of PESO
Categorized by values of CIGS
Included observations: 1388

Method	df	Value	Probability
Adj. Med. Chi-square	8	23.41541	0.0029
Kruskal-Wallis (tie-adj.)	8	38.35477	0.0000

Category Statistics

CIGS	Count	Median	> Overall Median	Mean Rank	Mean Score
[0, 5)	1199	120.0000	599	719.8499	0.062291
[5, 10)	35	114.0000	13	591.3143	-0.266882
[10, 15)	60	112.0000	18	545.0833	-0.342382
[15, 20)	19	111.0000	4	468.3421	-0.550108
[20, 25)	62	112.5000	18	526.9758	-0.428783
[30, 35)	5	102.0000	1	455.1000	-0.549822
[40, 45)	6	105.5000	2	462.7500	-0.612905
[45, 50)	1	121.0000	1	743.0000	0.087637
[50, 55)	1	90.00000	0	99.00000	-1.466366
All	1388	120.0000	656	694.5000	-2.76E-05



¿Is it higher the weight of the babys born to higher rent families?

Test for Equality of Medians of PESO
Categorized by values of RENTA
Included observations: 1388

Method	df	Value	Probability
Adj. Med. Chi-square	10	16.92791	0.0760
Kruskal-Wallis (tie-adj.)	10	32.52589	0.0003

Category Statistics

RENTA	Count	Median	> Overall Median	Mean Rank	Mean Score
[0, 5)	101	113.0000	37	609.0792	-0.192424
[5, 10)	129	115.0000	54	628.7907	-0.174702
[10, 15)	118	112.0000	47	579.6271	-0.295473
[15, 20)	178	119.5000	81	689.1938	-0.007684
[20, 25)	151	120.0000	74	728.0861	0.075848
[25, 30)	137	119.0000	66	698.4307	0.009026
[30, 35)	136	120.0000	66	727.2537	0.075535
[35, 40)	94	120.0000	46	716.1223	0.067456
[40, 45)	84	119.5000	40	695.7857	0.011219
[45, 50)	68	127.0000	43	846.9706	0.393375
[65, 70)	192	122.0000	102	741.5339	0.110268
All	1388	120.0000	656	694.5000	-2.76E-05

Reason 2: Where is the power function?

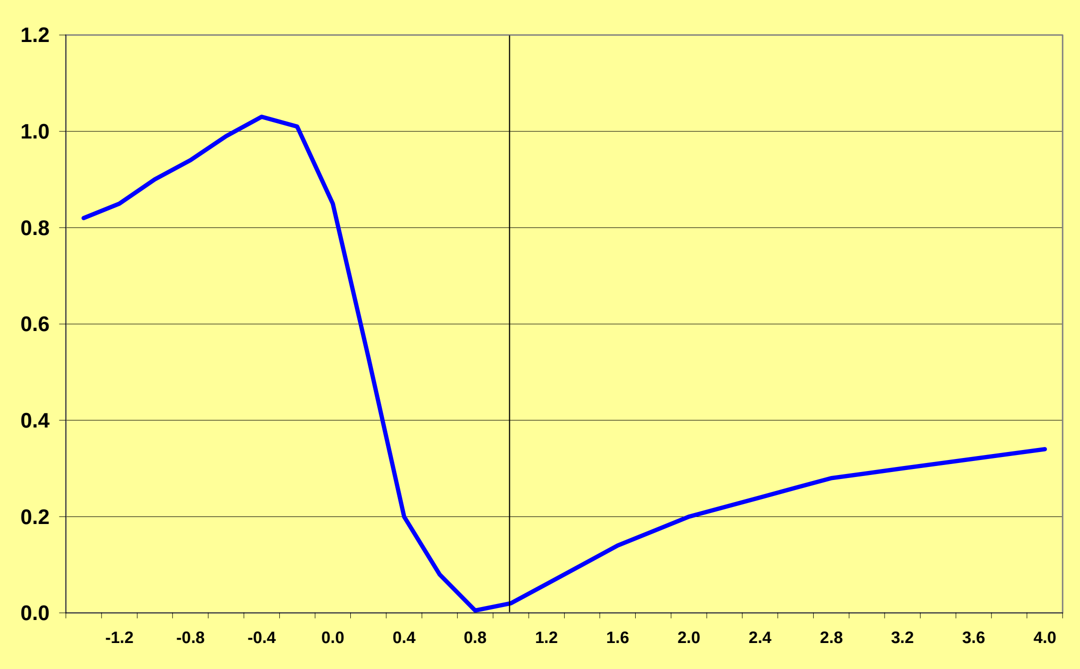
- Lack of rejection of the null hypothesis is usually identified with having shown that the hypothesis is true,
  - ✓ As a consequence of completely ignoring the power function of the test
  - ✓ And forgetting about the fact that rejecting the null hypothesis requires sample evidence in favor of the alternative hypothesis
  - ✓ Surprisingly low frequency of one-sided tests

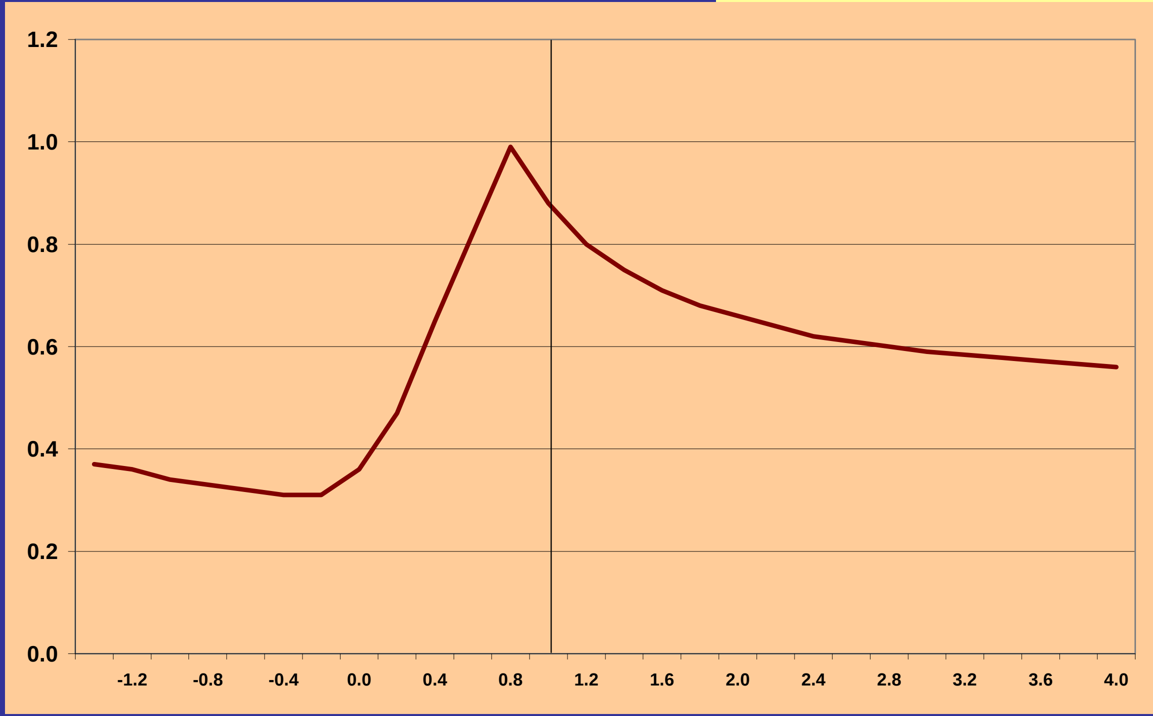
- Providing *p* - values is not enough

- ***Much better*** : graph displaying changes in an appropriate objective function under changes in parameter values

Testing for a unit elasticity of income in the demand for money function

Values of LRT statistic
H0: a given income elasticity





p-values for LRT statistic for income elasticity values

Reason 3: Ignoring the fact that *explanatory power* or *information content* is a conditional concept

- Significance test is a ***conditional*** test: test for information content in a variable additional to that contained in variables already included in the model (Simulation example)

Reason 4: Inappropriately comparing *relative* explanatory power

- Comparing estimated coeficients
- Comparing values of *t* - statistics for different variables
- Or for the same variable between different subsamples
  - ✓ *The more significant variable in the estimated model is Z*
  - ✓ *The more important variable to explain Y is Z*
  - ✓ *Variable X has become more significant after 1998*

Ejemplo: factores institucionales vs. factores macroeconómicos

- Factores macroeconómicos:
  - ✓ porcentaje de años con inflación > 20%
  - ✓ porcentaje de años con tipo e interés real negativo
  - ✓ Sobrevaluación del tipo de cambio
  - ✓ Prima en el mercado negro de divisas
  - ✓ correlación no muy elevada entre factores: alrededor de 0,33
  - ✓ Se construyen componentes principales

- Variables de calidad institucional:
  - ✓ elecciones democraticas, derechos políticos y civiles, independencia de medios de comunicación
  - ✓ estabilidad política
  - ✓ eficacia del gobierno
  - ✓ calidad de la regulacion
  - ✓ cumplimiento de leyes
  - ✓ control de corrupción
  - ✓ correlación elevada entre indicadores: 0,49 a 0,85
  - ✓ Se construye el PROMEDIO de todos ellos

Reason 5: Cleaning "non-significant variables" and re-estimating the model

T1

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

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

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

T5

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

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

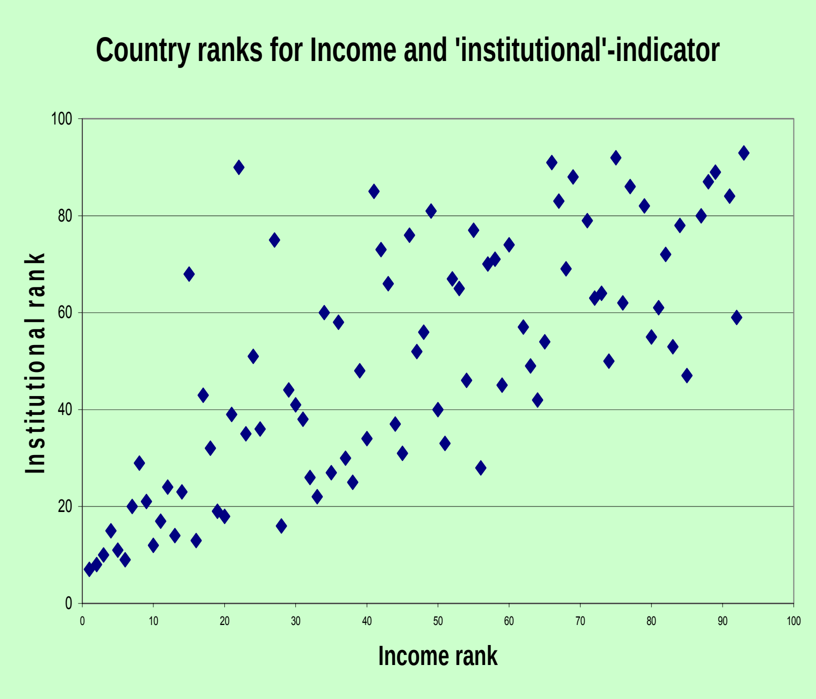
T4

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

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

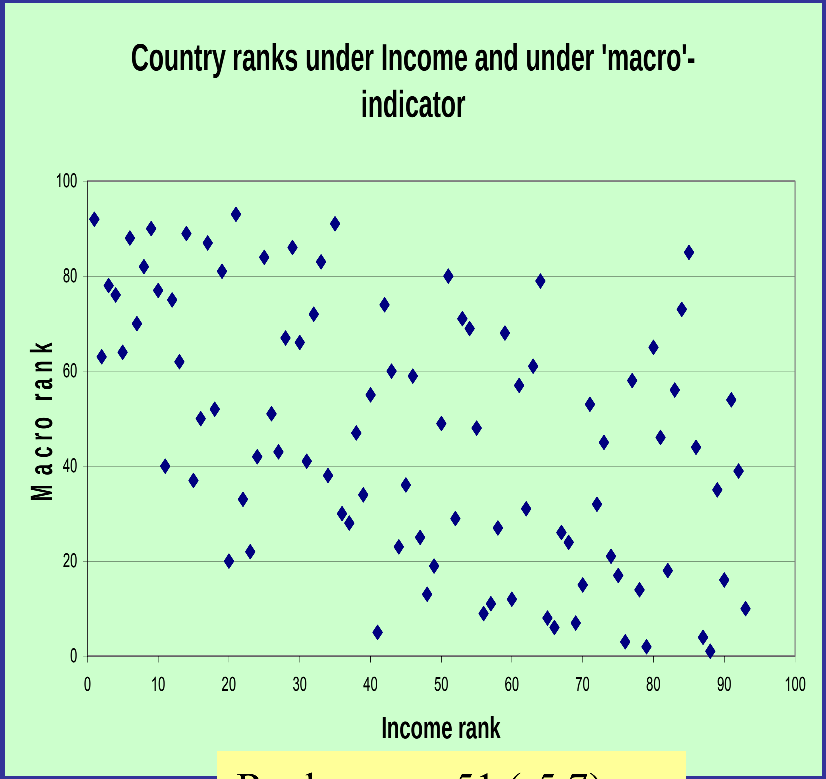
Factores económicos vs. Factores institucionales como determinantes del crecimiento



Rank corr: .71 (9.2)

Utilizando institutions\macro:

Rank corr: .50 (4.9)



Rank corr: -.51 (-5.7)

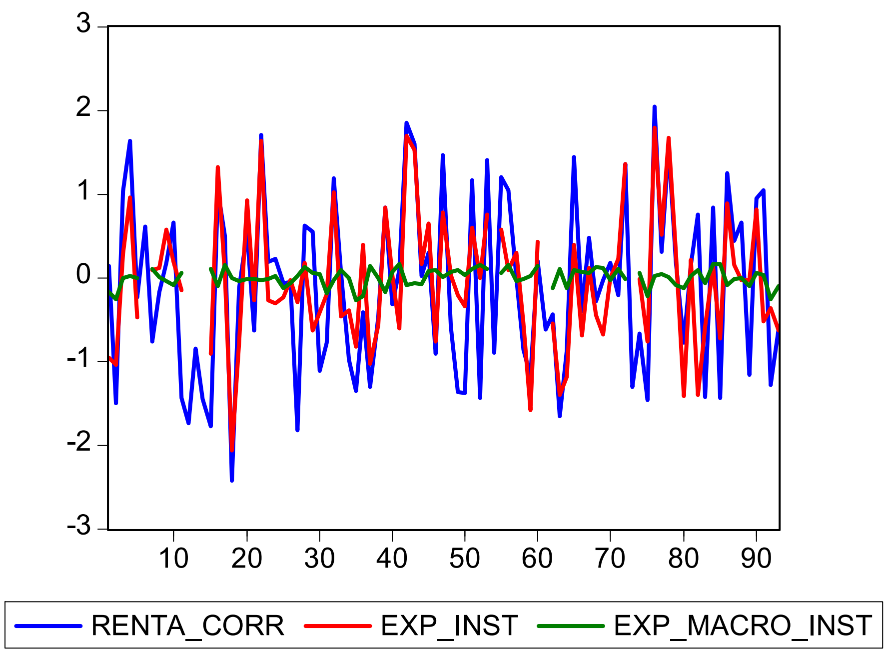
Utilizando macro\institutions:

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

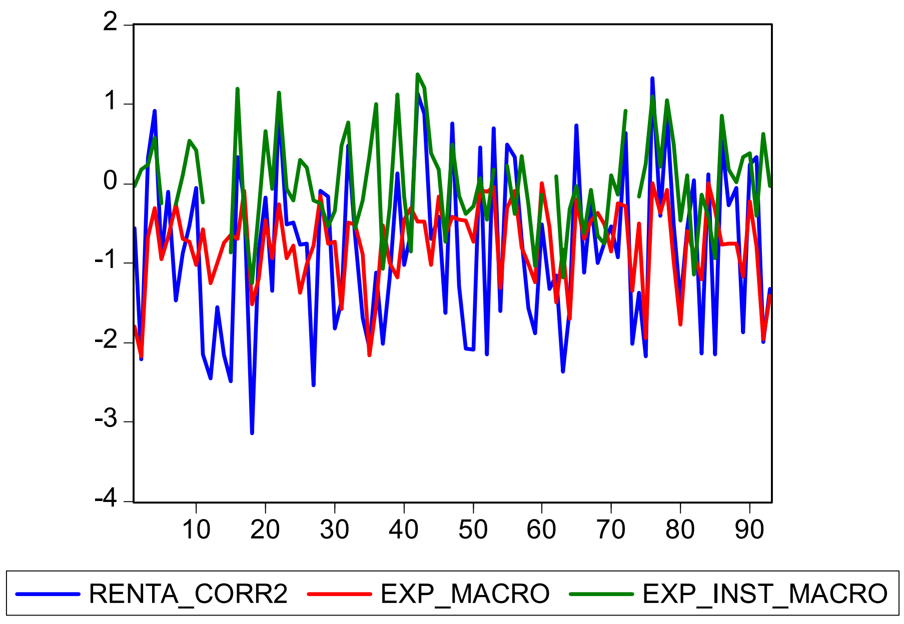


Orthogonalize explanatory variables

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



Reason 6: Interpretating *individual* coefficients as *individual* effects

- ***Three main difficulties***

1. Taking proper account of sample variability in explanatory variables
   - Effect on Y of changes in the value of X from min(X) to Max(X), or from the 10% to the 90% percentile of X (that ignores colinearity, anyway)

2. Omitted variable bias

3. Collinearity
   - Requires some structural interpretation. Cholesky decomposition as an example
   - What type of questions can we give an answer to?

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
Medias muestrales												
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	0,057	0,099	0,004	0,047	-0,015	-0,006	0,051	1,302	0,087
			(4,29)	(7,25)	(13,2)	(2,56)	(6,53)	(1,96)	(1,54)	(2,35)	(6,44)	(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
Desviaciones típicas												
	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

Ejemplo 1: Taking into account sample variation

Efecto tamaño positivo. ✓

Efecto positivo del número de líneas de producto (✓) y de la participación de no residentes en el capital (≠).

Efecto negativo del coeficiente importador (≠)

Sin efecto ni de la concentración de mercado (✓), ni del grado de diversificación geográfica (≠)

Efecto positivo de las otras dos variables de diferenciación del producto: gastos en publicidad (✓) e I+D (≠).

Ejemplo 2: Omitted variable bias

$$Ventas_t = 247{,}6 + \underbrace{2{,}204}_{0{,}545} Pub_t - \underbrace{1{,}464}_{0{,}649} P_t$$

$$Ventas_t = 96{,}0 + \underbrace{3{,}224}_{0{,}375} Pub_t$$

Aumento de 6.200 euros en publicidad = 1 desviación tìpica

Corr(Pub,P) = -0,829 $\Rightarrow$ descenso de -0,829 desviaciones típicas en precio = 4,327 euros.

Efecto total sobre Ventas: (6,200)(2,204) + (1,464)(4,327) = 20,0 mil euros

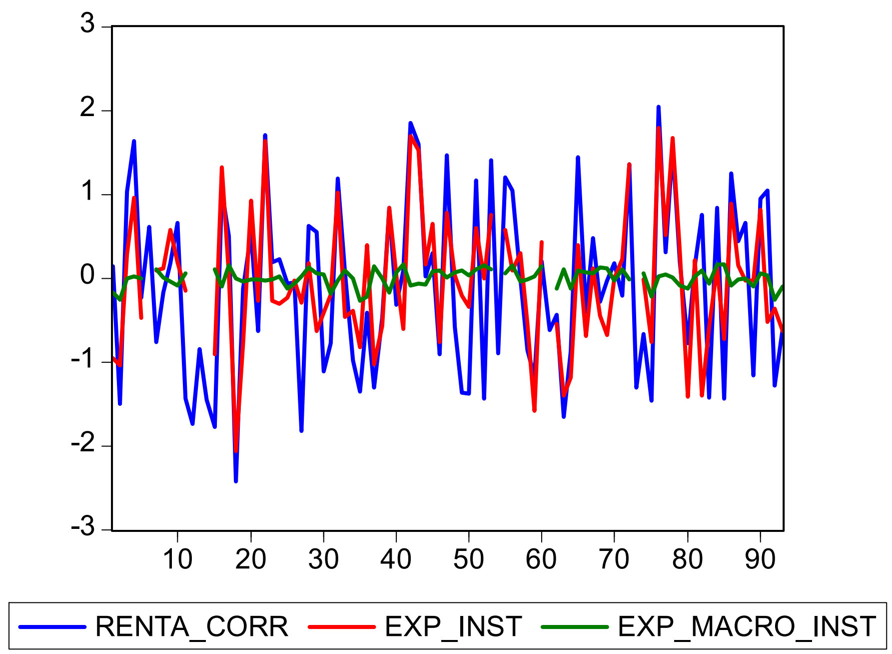
Modelo individual: (3,224) (6,200) = 20,0 mil euros

- La regresión individual (simple) mide el efecto global sobre Y de una variación en una de las X
  - Es una estimación sesgada del impacto de una variación *ceteris paribus* (que pocas veces tiene interés conocer)
  - Pero es una variación insesgada del efecto global, en el que generalmente estamos interesados
- Mientras que la regresión múltiple proporciona una estimación sesgada de dicho efecto global, debido a la colinealidad
- What is the question ?

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

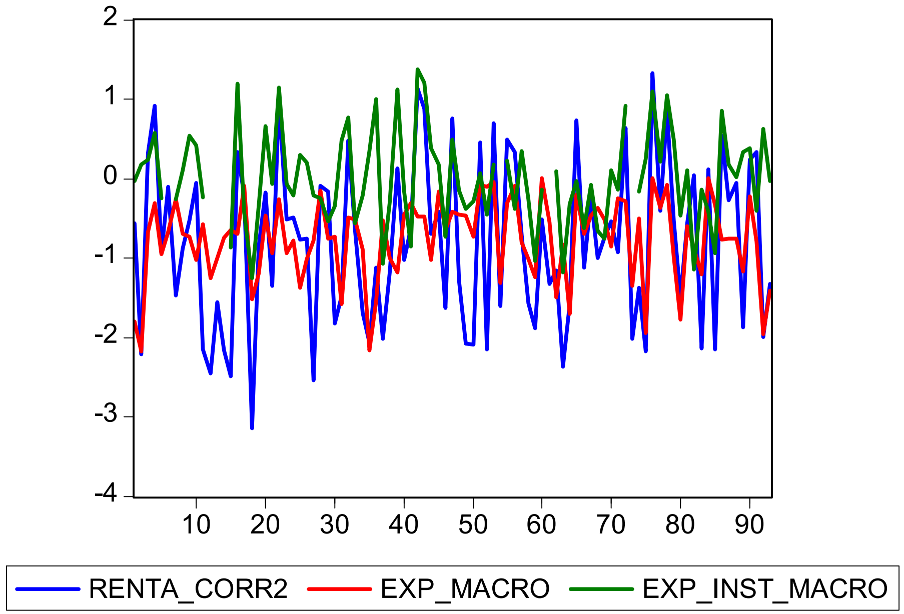


Orthogonalize explanatory variables

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



Reason 7 : Possible parameter variation

(*for a future seminar*)

- Explore regime switching models, nonlinear models
- But start always with simple constant parameter, linear models
- Estimate in subsamples (with time series as well as with cross-section data)

Part III : Poor use of numerical methods to estimate non-linear models
(for a future seminar)

How good are the estimates on which we base our inference?

Part IV : A suggestion

IV.1 Alternative approach to *model evaluation*:

Examine the residuals in detail

- **Graphical representations** of actual data, fitted data and residuals
- 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**
- 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?

Subsamples of model deterioration

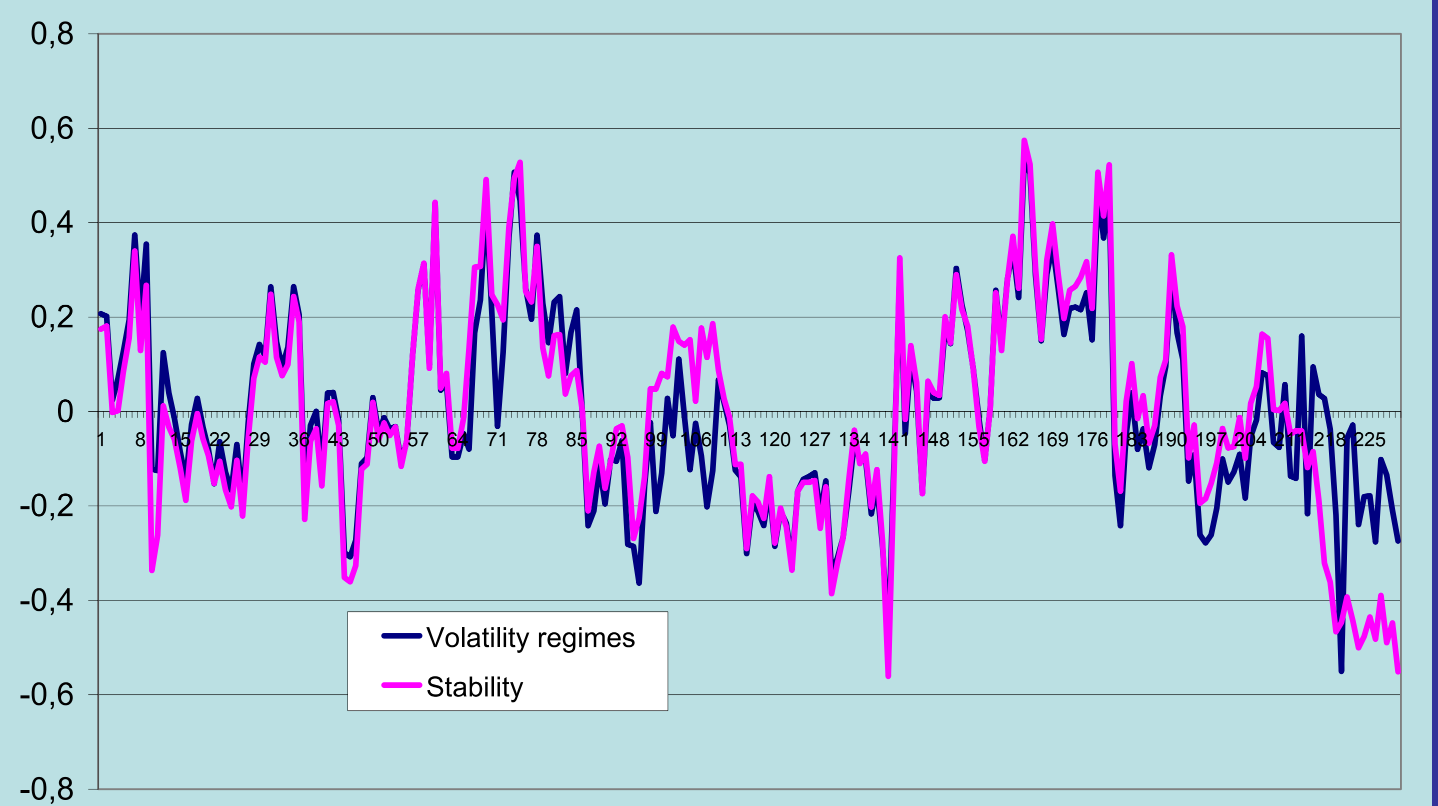
Consumo de tabaco explicado por la renta y el precio

Documento Acrobat

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

Explaining credit spreads by regime switching models

Documento Acrobat



Residuals

Excessive summary of sample information

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

$\Delta r_t = \mu_t + \sigma_t \Delta z_t + J_t \Delta n_t$; $\Delta z_t \sim N(0,1)$; $J_t \sim N(\theta_t, \psi^2)$; Δz_t y J_t son independientes.; $\sigma_t^2 = \omega_0 + \omega_1 \left[\Delta r_{t-1} - E_{t-2}\left(\Delta r_{t-1}\right)\right]^2$; μ_t y θ_t están definidos en la siguiente tabla.

	$\mu_t = \alpha\left(r_{t-1} - r_{t-1}^*\right)$		$\mu_t = \alpha_1\left(r_{t-1} - r_{t-1}^*\right)^+ + \alpha_2\left(r_{t-1} - r_{t-1}^*\right)^-$	
Parámetros	Modelo *i* $\theta_t = \theta$	Modelo *ii* $\theta_t = \gamma\left(r_{t-1} - r_{t-1}^*\right)$	Modelo *iii* $\theta_t = \theta$	Modelo *iv* $\theta_t = \gamma\left(r_{t-1} - r_{t-1}^*\right)$
α	-0.046 (-6.601)	-0.0344 (-6.365)	-------	--------
α_1	--------	--------	-0.0465 (-6.244)	-0.0388 (-7.832)
α_2	-------	--------	-1.0928 (-263.81)	-0.9943 (-61.837)
ω_0	0.00003 (5.123)	0.00003 (7.189)	0.0004 (10.127)	0.0004 (12.833)
ω_1	0.6240 (5.922)	0.6092 (5.542)	0.6651 (6.338)	0.6482 (6.053)
θ or γ	-0.0087 (-1.196)	-0.4152 (-8.179)	-0.009 (-1.170)	-0.2758 (-4.406)
ψ	0.2101 (12.939)	0.1903 (14.249)	0.2003 (12.294)	0.1930 (13.035)
δ_0	0.1209 (7.841)	0.1128 (6.917)	0.1025 (7.005)	0.1023 (7.209)
δ_1	0.5260 (10.201)	0.5271 (11.546)	0.5500 (10.714)	0.5347 (10.431)
δ_2	0.5257 (8.838)	0.5076 (9.830)	0.5233 (8.486)	0.5306 (8.460)
δ_3	0.1421 (3.582)	0.1309 (4.239)	0.1344 (3.493)	0.1476 (3.568)
δ_4	0.8307 (10.345)	0.8060 (10.055)	0.8763 (10.372)	0.8847 (10.998)
δ_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 δ_0, δ_1, δ_2, δ_3, δ_4 y δ_5 designan respectivamente las probabilidades de salto; (i) en los

Models with the same measure of fit: R2, log-likelihood. Do they explain the same?

What is the difference between models with statistically significant differences in their log-likelihood values?

What is the final question? How does the answer to it change from alternative models?

IV.2 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*

- What matters is not so much the 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 and in variance in subsamples of largest residuals or in the whole sample.
  - Any appropriate test?

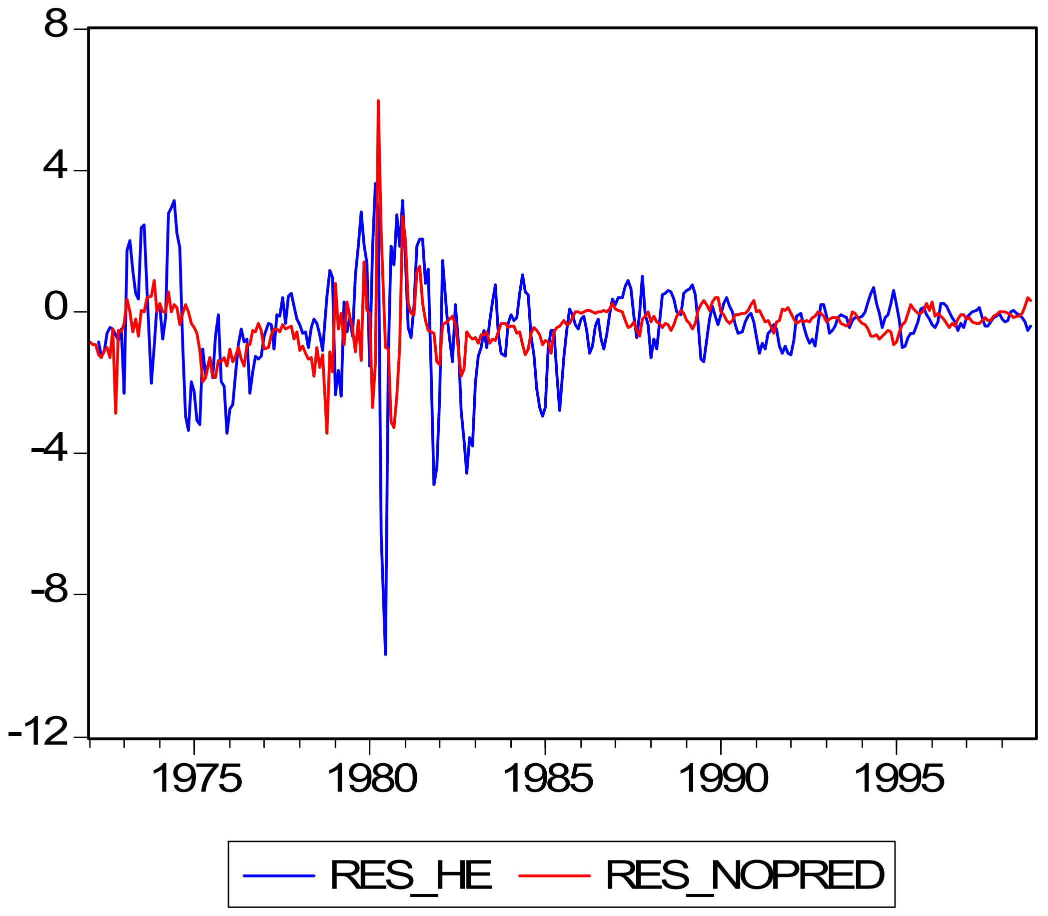
Residual comparison:

Res_HE: Imposing Expectations Hypothesis: Forward = Future spot rate

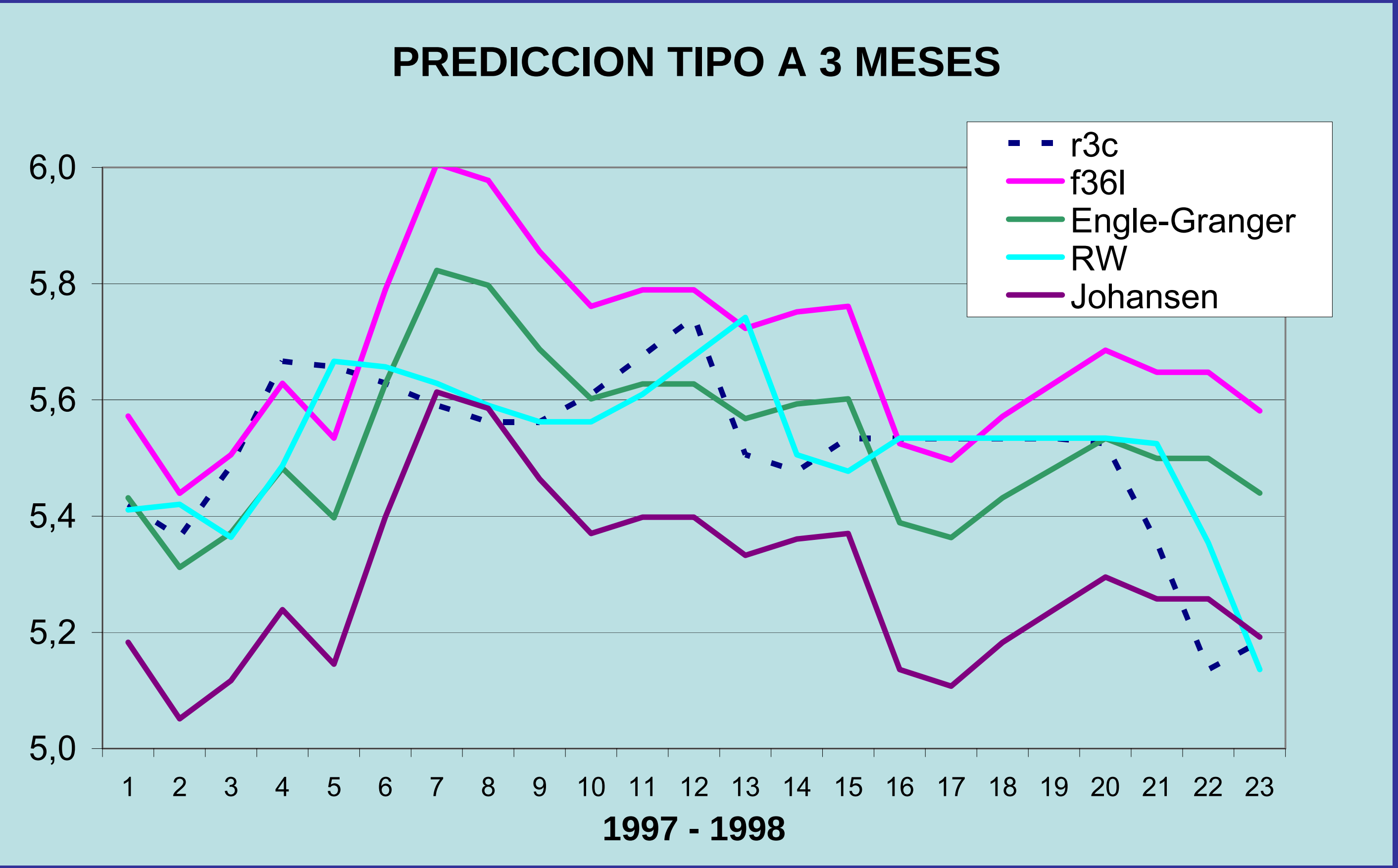
Res_nopred: Forward = Current spot rate

Sample: 1972:01 1998:12

	RES_HE	RES_NOPRED
Mean	-0.402173	-0.391333
Median	-0.303047	-0.309666
Maximum	3.612166	5.966989
Minimum	-9.662312	-3.438893
Std. Dev.	1.378541	0.769801
Skewness	-1.248191	1.320760
Kurtosis	10.64936	19.63791
Jarque-Bera	863.2612	3783.967
Probability	0.000000	0.000000
Sum	-128.6954	-125.2265
Sum Sq. Dev.	606.2194	189.0375



PREDICION DEL TIPO A 3 MESES					
		1997		1997 - 1998	
MODELO	PREDICTOR	MAE	RMSE	MAE	RMSE
H. Expectativas	Tipo forward retardado	0,167	0,212	0,185	0,234
Engle-Granger	MCO	0,115	0,147	0,125	0,156
Johansen	MV	0,258	0,296	0,238	0,276
Camino aleatorio	Tipo contado anterior	0,054	0,074	0,062	0,094
Incapacidad predictiva	Tipo forward actual	0,178	0,212	0,169	0,210



IV.3 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).*

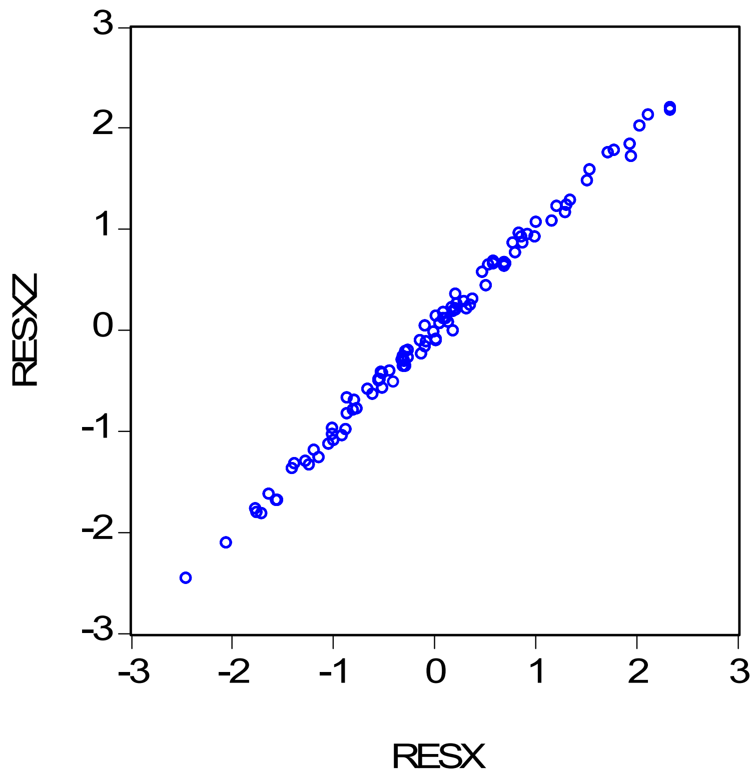
- We move from testing for statistical significance to testing for equality of sets of residuals

- Questions that can be asked:
  - *Does Z ad 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*?

Testing for additional information content by comparing residuals from restricted and unrestrcted models

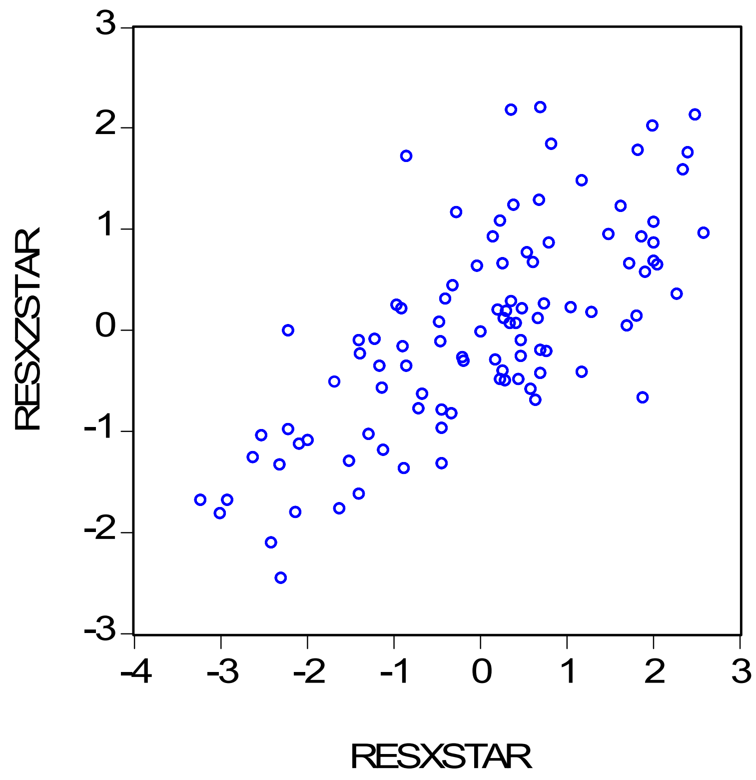
Irrelevant Z-variable

	Y	X	Z	U
Y	1.000000	0.887576	-0.186365	0.401713
X	0.887576	1.000000	-0.170068	-0.065306
Z	-0.186365	-0.170068	1.000000	-0.065612
U	0.401713	-0.065306	-0.065612	1.000000



Relevant Z-variable

	YSTAR	X	Z	U
YSTAR	1.000000	0.785513	0.287400	0.360849
X	0.785513	1.000000	-0.170068	-0.065306
Z	0.287400	-0.170068	1.000000	-0.065612
U	0.360849	-0.065306	-0.065612	1.000000



IV.4 Evaluation of restrictions on the basis of their final implication

- Compare alternative models using loss functions defined on the value of the final objective function under each model

- What a difference does it make a given restriction for the final question under analysis? What implications do alternative models have for the final question? ***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?

But ¿are we not doing all this already?

Do's

- Start by examining the data: sample range, histograms, stem and leaf diagrams, nonparametric evaluation of association between variables
- And by identifying the question to be analyzed
- Think of (additional) information content as a conditional concept, and evaluate it appropriately
- Interpret individual coefficients appropriately. Deal with colinearity and take into account sample variability.
- Test parametric constraints by point-to-point comparison of restricted and unrestricted residuals
- Provide information on possible parameter variation
- Compare models (Look at constraints) from the point of view of their impact on the question of interest: inflation forecasting, VaR computation, risk management, fiscal policy evaluation
- What type of questions can we answer to on the basis of the available sample information ?
- What does a model explain that it is not explained by others ?

Dont's

- Do not identify statistical significance and information content
- Do not rely exclusively on *t*- and *F*-statistics when evaluating parametric restrictions or comparing models
- Do not compare the information content of variables by looking at their coefficients or their t-statistics
- Do not practice *sign* or *asterisk* econometrics
- Do not summarize excessively the sample information
- Do not test for statistical significance
- Do not test too much …
- … and if you do, never forget about the power of the test

Part V : Implications for Statistics/Econometrics teaching

1. Economics is a non-experimental science

2. Focus on single-sample properties. Unbiasedness vs. consistency

3. Run lots of simulations. Estimate empirical distributions. Analyze size and power of tests statistics under different experimental designs

4. Use actual data
   - Graphs depicting relationship btween pairs of variables
   - Descriptive statistics: univariate and multivariate. Sample range, variability, test for equality of medians under classification

5. Emphasize non-parametric statistics

6. Heteroskedasticity, autocorrelation, colinearity are situations always present in the data, unless some evidence is presented otherwise. They should not be ocnsiderd problems or violations. How should we deal with them?
   - Difference between constructive and non-constructive tests.
   - Colinearity is harder than autocorrelation, which is, in turn, harder than heteroskedasticity
   - Do we gain much with Feasible GLS?

7. What are the required assumptions ?

8. Use extended expression for Variance-covariance matrix of OLS estimator

9. Estimate direct vs. second-order effects, individual vs. global effects

10. Multicolinearity chapter (?)
    - Discuss how Var(β) depends on Corr(X,Z)
    - Partial correlation coefficients
    - Relationship between coefficients in simple and in multiple models

11. Emphasize detailed examination of residuals
    - Compare residuals with dependent variable
    - Compare residuals from alternative models
    - Constrained and unconstrained models

12. Do not test alternative models by comparing a single one statistic
    - Loglik, t-stat, …

13. Discuss in detail parameter variation
    - Detection and treatment

14. Simple discussion of limited dependent variables

The (provisional) End