

**SELECCIÓN DE MODELOS DE REGRESIÓN CON  
COEFICIENTES CAMBIANTES BASADA EN LA  
CAPACIDAD PREDICTIVA**

**Patricia Canal**

Trabajo de investigación 012/021

Master en Banca y Finanzas Cuantitativas

Directoras: Eva Ferreira y Susan Orbe

Universidad del País Vasco

Universidad de Castilla-La Mancha

Universidad Complutense de Madrid

Universidad del País Vasco

Universidad de Valencia



## SELECCIÓN DE MODELOS DE REGRESIÓN CON COEFICIENTES CAMBIANTES BASADA EN LA CAPACIDAD PREDICTIVA

### Introducción

Uno de los principales problemas en el área de la Economía y las Finanzas, así como en muchas otras, es el de seleccionar el modelo más adecuado cuando tenemos varias propuestas. En principio puede parecer que introducir todas las variables significativas es una buena opción. Sin embargo, al aumentar el número de regresores nos exponemos a obtener parámetros estimados con mucha varianza debido a la correlación entre las variables (colinealidad). Igualmente, hay que tener en cuenta el coste operativo, que resulta muy elevado si el modelo tiene muchas variables.

Para ayudarnos a tomar una decisión, se han desarrollado numerosas técnicas tales como contrastes de hipótesis o métodos in-sample - Final Prediction Error (Akaike 1969),  $C_p$  de Mallows (1973), AIC (Akaike 1974), BIC (Akaike-Schwartz 1978), Hannan and Quinn (HQ 1979).

Por otra parte, podemos estar especialmente interesados en utilizar un modelo para intentar adelantarnos a lo que va a ocurrir en el futuro con determinadas variables. Para esto conocemos métodos tales como Box-Jenkins (1976), autorregresión vectorial (VAR) o alisado exponencial (Holt-Winters 1960).

Además es necesario que nuestro modelo tenga una buena capacidad predictiva, para lo que es conveniente que los estadísticos utilizados no dependan de la muestra utilizada, ya que entonces únicamente servirían para determinadas muestras.

La predicción del comportamiento de una variable entraña un problema complicado principalmente si pretendemos hacerlo a largo plazo, ya que las variables pueden adoptar entre ellas relaciones muy diversas con el tiempo. Esto puede suponer la necesidad de un modelo diferente según el instante de tiempo que consideremos. Se ha demostrado que la mejor forma de paliar este inconveniente es permitiendo que los coeficientes del modelo sean variables con el tiempo.

Existen múltiples tests sobre capacidad predictiva, como el de Morgan-Granger-Newbold (1977), Meese and Rogoff (1988) o Diebold y Mariano (1995). Estos últimos proponen un test para contrastar la igualdad en capacidad predictiva de dos modelos con variable dependiente común basándose en las diferencias de las imágenes mediante una cierta función de pérdida de los errores teóricos de dichos modelos. Sin embargo, Ferreira y Stute (2009) advierten que generalmente no conocemos la parte sistemática de los modelos y hay que tener en cuenta su estimación; éstos demuestran

que bajo ciertas condiciones el estadístico funciona a pesar de usarlo con los residuos procedentes de la estimación en lugar de con los errores.

En este artículo analizaremos empíricamente el comportamiento del contraste para determinadas condiciones sobre los modelos, continuando así con el trabajo de Ferrerira y Stute para comprobar en qué situaciones el estadístico es válido y por tanto en qué situaciones podemos aplicarlo. Escogemos diferentes funciones de pérdida para otorgar distintos pesos a los errores. Analizaremos cómo influye en los resultados el tamaño muestral, variando el tamaño muestral de estimación y de validación para observar la relación existente entre ambos. Probaremos con distintas distribuciones de los errores y las variables, ya que los errores pueden ser de muy diversos tipos, generalmente asimétricos. Aumentaremos el número de variables que intervienen en los dos modelos incluyendo una en común para comprobar qué resultados nos ofrece el estadístico en caso de que exista cierta correlación entre los modelos.

Dado que en la práctica es difícil que las variables de un modelo tengan una relación lineal con la variable dependiente, un modelo de coeficientes fijos resulta inadecuado en la mayoría de los casos. Por ello, finalmente propondremos un modelo con coeficientes cambiantes en el tiempo.

## Objetivo

Supongamos que tenemos dos modelos de la forma

$$Y = m_i(X_i) + \epsilon_i, i = 1,2$$

y que conocemos  $Y$ ,  $m_1$ ,  $m_2$ ,  $X_1$ ,  $X_2$ . Estamos interesados en descubrir qué covariable,  $X_1$  ó  $X_2$ , predice mejor el comportamiento futuro de  $Y$ . Existen varios tests para contrastar la igualdad de capacidad predictiva de  $X_1$  y  $X_2$ . Es decir:

$$\begin{cases} H_0: E(g(\epsilon_1)) = E(g(\epsilon_2)) \\ H_1: E(g(\epsilon_1)) > E(g(\epsilon_2)) \end{cases}$$

donde  $g$  es una cierta función de pérdida.

Un estadístico para este contraste es el propuesto por Diebold y Mariano (1995)

$$\frac{n^{1/2}\bar{d}}{\hat{\sigma}_d}$$

siendo

$$\bar{d} = n^{-1} \sum_{t=T+1}^{T+n} [g(\epsilon_{1t}) - g(\epsilon_{2t})]$$

y la distribución que sigue dicho estadístico corresponde a una normal estándar.

Sin embargo, en la realidad no solemos disponer de  $m_1$  y  $m_2$ , por lo que Ferreira y Stute proponen analizar si este resultado sigue siendo válido a pesar de tener que estimar la parte sistemática, y demuestran que, bajo ciertas condiciones, el estadístico funciona. En este caso debemos estimar los modelos

$$Y = \hat{m}_i(X_i) + e_i, \quad i = 1,2$$

siendo

$$e_i = Y - \hat{m}_i(X_i) \quad i = 1,2$$

el residuo asociado. De esta forma el estadístico del contraste es

$$\frac{n^{1/2} \bar{d}_1}{\hat{\sigma}_{\bar{d}_1}} \rightarrow N(0,1)$$

siendo

$$\bar{d}_1 = n^{-1} \sum_{t=T+1}^{T+n} [g(e_{1t}) - g(e_{2t})]$$

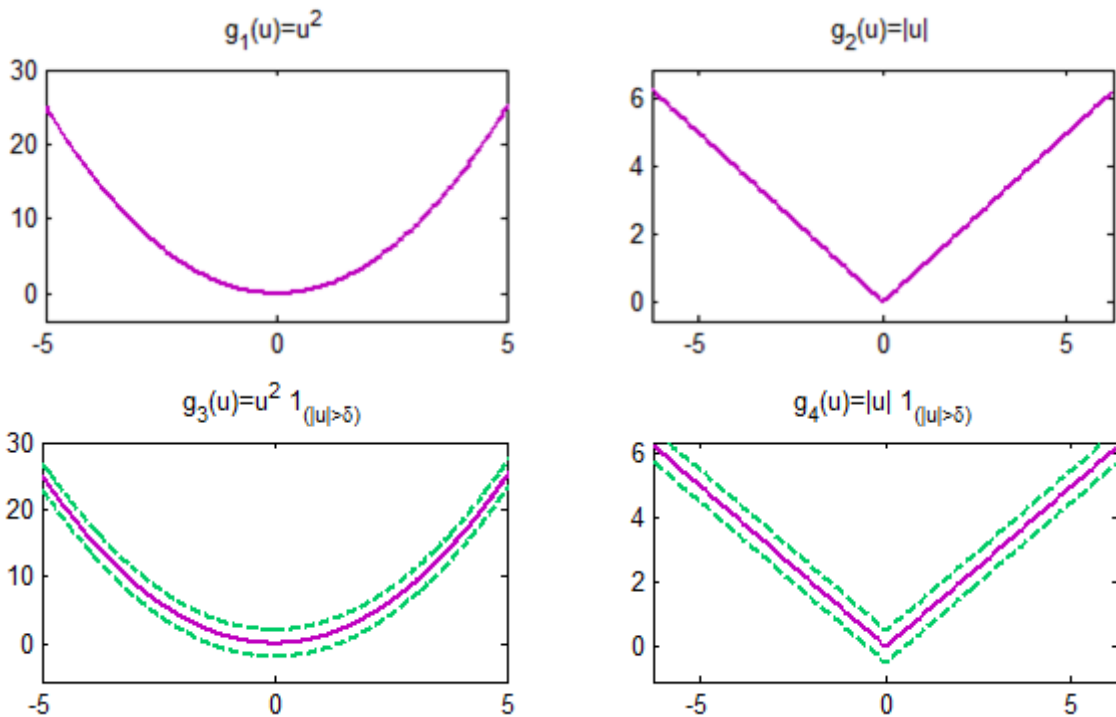
Pretendemos analizar empíricamente si este estadístico que utiliza residuos funciona bien cuando consideramos distintas situaciones para la parte sistemática, así que analizaremos situaciones tales como un mayor tamaño muestral, distintas funciones de pérdida, distintas distribuciones de los errores y las variables o un aumento del número de variables que intervienen en los modelos.

Para comprobar si los resultados que proporciona el contraste son adecuados, es conveniente hallar el porcentaje de veces que se rechaza la hipótesis nula cuando es cierta. Este porcentaje se calcula como el número de veces que el estadístico de prueba supera el valor crítico  $N^{-1}(1 - \text{nivel de significación})$ , para lo que consideraremos los niveles de significación 0.05 y 0.1.

### **Funciones de pérdida y procedimiento**

En primer lugar analizaremos si para algunas funciones de pérdida el tamaño del contraste se acerca al valor teórico.

Consideramos las siguientes funciones de pérdida:



Según elijamos una  $u$  otra estamos dando más peso a residuos grandes o pequeños. Con la primera tratamos a todos los residuos por igual, en contra de lo que ocurre con la segunda. Con las dos últimas descartamos los errores que se encuentran dentro de cierta banda, ya que residuos relativamente pequeños no aportan cambios importantes a los resultados, por lo que nos centramos en los errores mayores.

Una vez propuestas las funciones de pérdida, y siguiendo con el trabajo de Ferreira y Stute, proponemos, para generar los datos, un modelo sencillo

$$Y_t = X_{1t} + X_{2t} + \epsilon_t$$

con el que se generan 1000 muestras de  $T+n$  datos, suponiendo

$$E(X_{1t}) = E(X_{2t}) = E(\epsilon_t) = 0, \quad \text{Var}(X_{1t}) = \text{Var}(X_{2t}) = 1$$

Para los  $T$  primeros datos generados, estimamos los parámetros desconocidos

$$Y = a_i + b_i X_{it} + \epsilon_{it}, \quad \epsilon_{it} \sim N(0,1), \quad t = 1, \dots, T$$

y con los  $n$  últimos datos y los parámetros estimados hallamos los residuos

$$e_{it} = Y - \hat{a}_i - \hat{b}_i X_{it}, \quad i = 1,2, \quad t = T + 1, \dots, T + n$$

con los que procedemos a la parte de validación, en la que calculamos el porcentaje de rechazo de la hipótesis nula.

Los resultados obtenidos son los siguientes:

$g_1$	Usando residuos			Usando errores	
$n (T=50)$	$\alpha=0.05$	$\alpha=0.10$		$\alpha=0.05$	$\alpha=0.10$
10	0.0640	0.1140		0.0690	0.1160
30	0.0630	0.1170		0.0530	0.1100
100	0.0540	0.1120		0.0500	0.1000
200	0.0550	0.1180		0.0350	0.0970
1000	0.1090	0.1630		0.0440	0.1030
2500	0.1710	0.2380		0.0560	0.1000

$g_2$	Usando residuos			Usando errores	
$n (T=50)$	$\alpha=0.05$	$\alpha=0.10$		$\alpha=0.05$	$\alpha=0.10$
10	0.0690	0.1370		0.0660	0.1280
30	0.0540	0.1210		0.0550	0.1120
100	0.0560	0.1030		0.0510	0.0990
200	0.0760	0.1170		0.0480	0.0940
1000	0.0990	0.1510		0.0480	0.0940
2500	0.1500	0.2010		0.0480	0.1030

$g_3, \delta = 0.1$	Usando residuos			Usando errores	
$n (T=50)$	$\alpha=0.05$	$\alpha=0.10$		$\alpha=0.05$	$\alpha=0.10$
10	0.0630	0.1260		0.0060	0.1250
30	0.0690	0.1020		0.0610	0.0970
100	0.0560	0.1130		0.0520	0.1090
200	0.0670	0.1080		0.0460	0.1120
1000	0.1030	0.1650		0.0430	0.0850
2500	0.1450	0.1950		0.0500	0.0850

$g_3, \delta = 0.2$	Usando residuos			Usando errores	
$n (T=50)$	$\alpha=0.05$	$\alpha=0.10$		$\alpha=0.05$	$\alpha=0.10$
10	0.0690	0.1250		0.0720	0.1220
30	0.0540	0.0970		0.0640	0.0870
100	0.0670	0.1120		0.0550	0.1030
200	0.0720	0.1020		0.0470	0.1010
1000	0.1070	0.1740		0.0650	0.1010
2500	0.1570	0.2130		0.0550	0.0920

$g_4, \delta = 0.1$	Usando residuos			Usando errores	
$n (T=50)$	$\alpha=0.05$	$\alpha=0.10$		$\alpha=0.05$	$\alpha=0.10$
10	0.0780	0.1260		0.0770	0.1280
30	0.0470	0.0990		0.0530	0.0940
100	0.0780	0.1140		0.0620	0.1090
200	0.0600	0.1120		0.0500	0.0890
1000	0.1040	0.1760		0.0480	0.1220
2500	0.1490	0.2070		0.0520	0.0990

$g_4, \delta = 0.2$	Usando residuos			Usando errores	
$n (T=50)$	$\alpha=0.05$	$\alpha=0.10$		$\alpha=0.05$	$\alpha=0.10$
10	0.0710	0.1130		0.0720	0.1160
30	0.0660	0.1170		0.0610	0.1130

<b>100</b>	0.0500	0.1020		0.0420	0.1060
<b>200</b>	0.0640	0.1160		0.0480	0.1010
<b>1000</b>	0.1190	0.1600		0.0580	0.0950
<b>2500</b>	0.1530	0.1750		0.0460	0.0930

Como podemos observar, el porcentaje de veces que rechazamos la hipótesis nula es muy similar para valores de  $n$  cercanos a  $T$ , mientras que comienza a alejarse para valores de  $n$  mayores. Esto es debido a que a la hora de estimar nos interesa tener muchos datos para que nuestra estimación sea lo mejor posible. Además, para validar intervienen los residuos, ya que al no conocer  $|Y - m_i(X_i)|$  lo estamos estimando:

$$|Y - m_i(X_i)| = |Y - \hat{m}_i(X_i) + \hat{m}_i(X_i) - m_i(X_i)|$$

La mejor estimación de  $|Y - m_i(X_i)|$  es  $|Y - \hat{m}_i(X_i)|$ , así que debe ocurrir que

$$|\hat{m}_i(X_i) - m_i(X_i)| \rightarrow 0$$

Por tanto, para que no se acumulen demasiados errores,  $n$  debe ser pequeño comparado con  $T$ .

### Aumento del tamaño muestral

Dado que hemos considerado un tamaño muestral relativamente pequeño, para observar mejor este efecto repetiremos el proceso considerando tamaños muestrales mayores.

En este caso, obtenemos los siguientes resultados:

$$g_1(u) = u^2$$

n (T=50)	$\alpha=0.05$	$\alpha=0.10$	n (T=250)	$\alpha=0.05$	$\alpha=0.10$	n (T=500)	$\alpha=0.05$	$\alpha=0.10$
<b>10</b>	0.0640	0.1140	<b>50</b>	0.0560	0.1270	<b>100</b>	0.0580	0.0920
<b>30</b>	0.0630	0.1170	<b>150</b>	0.0470	0.1100	<b>300</b>	0.0470	0.0940
<b>100</b>	0.0540	0.1120	<b>300</b>	0.0540	0.1030	<b>1000</b>	0.0520	0.1270
<b>200</b>	0.0550	0.1180	<b>500</b>	0.0460	0.0980	<b>2000</b>	0.0410	0.1000
<b>1000</b>	0.1090	0.1630	<b>1000</b>	0.0480	0.1220	<b>10000</b>	0.0480	0.1150
<b>2500</b>	0.1710	0.2380	<b>3000</b>	0.0690	0.1100	<b>25000</b>	0.0750	0.1340
			<b>5000</b>	0.0760	0.1160			

$$g_2(u) = |u|$$

n (T=50)	$\alpha=0.05$	$\alpha=0.10$	n (T=250)	$\alpha=0.05$	$\alpha=0.10$	n (T=500)	$\alpha=0.05$	$\alpha=0.10$
<b>10</b>	0.0690	0.1370	<b>50</b>	0.0540	0.0930	<b>100</b>	0.0620	0.1080
<b>30</b>	0.0540	0.1210	<b>150</b>	0.0560	0.1050	<b>300</b>	0.0530	0.1150
<b>100</b>	0.0560	0.1030	<b>300</b>	0.0450	0.1090	<b>1000</b>	0.0440	0.1120
<b>200</b>	0.0760	0.1170	<b>500</b>	0.0580	0.0940	<b>2000</b>	0.0530	0.1210
<b>1000</b>	0.0990	0.1510	<b>1000</b>	0.0480	0.0850	<b>10000</b>	0.0640	0.1080
<b>2500</b>	0.1500	0.2010	<b>3000</b>	0.0470	0.1080	<b>25000</b>	0.0680	0.1170
			<b>5000</b>	0.0500	0.1180			



$$g_3(u) = u^2 1_{\{|u|>\delta\}}, \quad \delta = 0.1$$

n (T=50)	$\alpha=0.05$	$\alpha=0.10$	n (T=250)	$\alpha=0.05$	$\alpha=0.10$	n (T=500)	$\alpha=0.05$	$\alpha=0.10$
10	0.0630	0.1260	50	0.0560	0.1160	100	0.0460	0.0970
30	0.0690	0.1020	150	0.0380	0.1120	300	0.0680	0.1060
100	0.0560	0.1130	300	0.0470	0.1050	1000	0.0590	0.0960
200	0.0670	0.1080	500	0.0590	0.0990	2000	0.0560	0.0940
1000	0.1030	0.1650	1000	0.0570	0.1240	10000	0.0620	0.0980
2500	0.1450	0.1950	3000	0.0560	0.1260	25000	0.0600	0.0990
			5000	0.0590	0.1160			

$$g_3(u) = u^2 1_{\{|u|>\delta\}}, \quad \delta = 0.2$$

n (T=50)	$\alpha=0.05$	$\alpha=0.10$	n (T=250)	$\alpha=0.05$	$\alpha=0.10$	n (T=500)	$\alpha=0.05$	$\alpha=0.10$
10	0.0690	0.1250	50	0.0560	0.1060	100	0.0530	0.0930
30	0.0540	0.0970	150	0.0430	0.0930	300	0.0440	0.0930
100	0.0670	0.1120	300	0.0460	0.1150	1000	0.0540	0.1120
200	0.0720	0.1020	500	0.0470	0.1100	2000	0.0510	0.1210
1000	0.1070	0.1740	1000	0.0420	0.1030	10000	0.0550	0.1370
2500	0.1570	0.2130	3000	0.0570	0.0940	25000	0.0600	0.1140
			5000	0.0820	0.1340			

$$g_3(u) = u^2 1_{\{|u|>\delta\}}, \quad \delta = 0.7$$

n (T=50)	$\alpha=0.05$	$\alpha=0.10$	n (T=250)	$\alpha=0.05$	$\alpha=0.10$	n (T=500)	$\alpha=0.05$	$\alpha=0.10$
10	0.0640	0.1230	50	0.0510	0.1190	100	0.0440	0.1160
30	0.0510	0.1280	150	0.0720	0.1040	300	0.0460	0.0990
100	0.0610	0.1120	300	0.0500	0.0980	1000	0.0420	0.0910
200	0.0570	0.1190	500	0.0500	0.1010	2000	0.0580	0.0990
1000	0.0990	0.1610	1000	0.0330	0.1100	10000	0.0500	0.1170
2500	0.1520	0.2160	3000	0.0580	0.1050	25000	0.0660	0.1260
			5000	0.0670	0.1090			

$$g_4(u) = |u| 1_{\{|u|>\delta\}}, \quad \delta = 0.1$$

n (T=50)	$\alpha=0.05$	$\alpha=0.10$	n (T=250)	$\alpha=0.05$	$\alpha=0.10$	n (T=500)	$\alpha=0.05$	$\alpha=0.10$
10	0.0780	0.1260	50	0.0570	0.1050	100	0.0510	0.0850
30	0.0470	0.0990	150	0.0500	0.1000	300	0.0520	0.1100
100	0.0780	0.1140	300	0.0570	0.1050	1000	0.0480	0.0870
200	0.0600	0.1120	500	0.0550	0.1140	2000	0.0530	0.1010
1000	0.1040	0.1760	1000	0.0420	0.1030	10000	0.0690	0.1040
2500	0.1490	0.2070	3000	0.0580	0.1190	25000	0.0710	0.1060
			5000	0.0620	0.1280			

$$g_4(u) = |u| 1_{\{|u|>\delta\}}, \quad \delta = 0.2$$

n (T=50)	$\alpha=0.05$	$\alpha=0.10$	n (T=250)	$\alpha=0.05$	$\alpha=0.10$	n (T=500)	$\alpha=0.05$	$\alpha=0.10$
10	0.0710	0.1130	50	0.0480	0.0990	100	0.0500	0.1010
30	0.0660	0.1170	150	0.0550	0.0870	300	0.0390	0.1020
100	0.0500	0.1020	300	0.0500	0.1180	1000	0.0440	0.0980
200	0.0640	0.1160	500	0.0390	0.1140	2000	0.0570	0.1100
1000	0.1190	0.1600	1000	0.0600	0.1010	10000	0.0560	0.1260

<b>2500</b>	0.1530	0.1750	<b>3000</b>	0.0450	0.0960	<b>25000</b>	0.0680	0.1150
			<b>5000</b>	0.0690	0.1370			

$$g_4(u) = |u|1_{\{|u|>\delta\}}, \quad \delta = 0.7$$

n (T=50)	$\alpha=0.05$	$\alpha=0.10$	n (T=250)	$\alpha=0.05$	$\alpha=0.10$	n (T=500)	$\alpha=0.05$	$\alpha=0.10$
<b>10</b>	0.0670	0.1040	<b>50</b>	0.0460	0.0940	<b>100</b>	0.0510	0.0930
<b>30</b>	0.0620	0.0990	<b>150</b>	0.0550	0.0930	<b>300</b>	0.0550	0.1140
<b>100</b>	0.0650	0.1160	<b>300</b>	0.0510	0.1130	<b>1000</b>	0.0570	0.1130
<b>200</b>	0.0550	0.1140	<b>500</b>	0.0400	0.1110	<b>2000</b>	0.0560	0.0980
<b>1000</b>	0.0940	0.1650	<b>1000</b>	0.0510	0.1070	<b>10000</b>	0.0600	0.1000
<b>2500</b>	0.1240	0.1870	<b>3000</b>	0.0440	0.0920	<b>25000</b>	0.0600	0.0950
			<b>5000</b>	0.0600	0.1340			

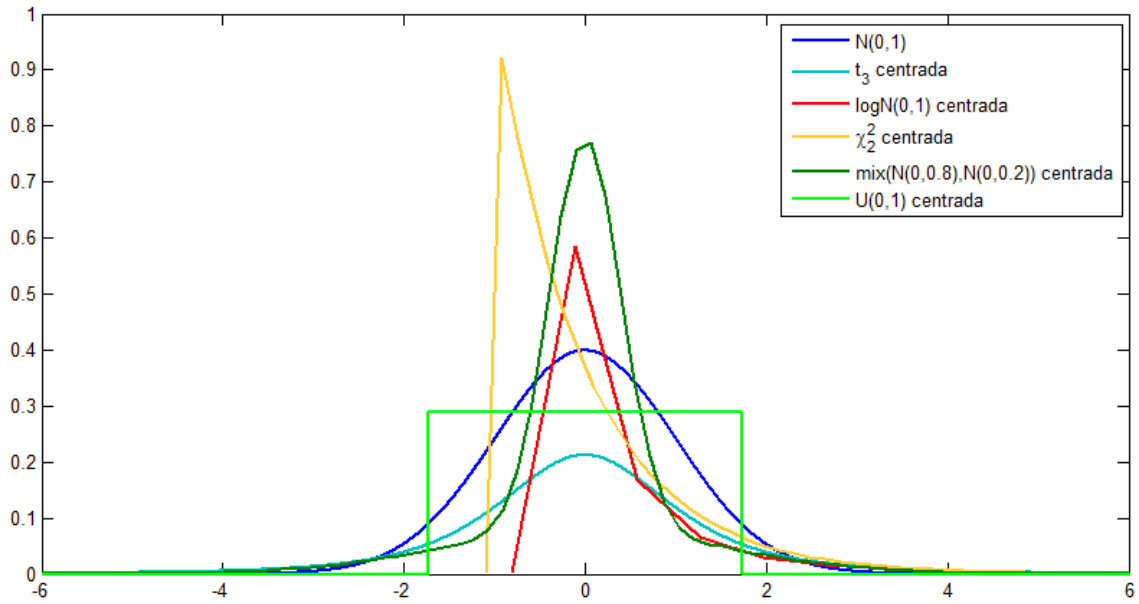
Observamos que a medida que aumentan T y n, los porcentajes de rechazo son más cercanos a los valores teóricos y por tanto podemos admitir un tamaño muestral de validación mayor que el de estimación.

### Distintas distribuciones de los errores

A continuación nos disponemos a probar con distribuciones de los errores distintas de la normal estándar. Proponemos una serie de distribuciones cuya forma difiere de la normal, algunas asimétricas, y las tipificamos para poder compararlas con la misma.

Las distribuciones propuestas son las siguientes:

- T-Student con tres grados de libertad
- Lognormal con  $\mu=0, \sigma = 1$
- Chi-cuadrado con dos grados de libertad
- Mixtura de dos normales de media cero y desviaciones 0.8 y 0.2, con probabilidad de mezcla 0.3
- Uniforme en (0,1)



A continuación exponemos los resultados para las distintas funciones de pérdida consideradas y tamaño muestral grande, variando la distribución de los errores:

$$g_1(u) = u^2$$

n (T=500)	$t_3$		logN(0,1)		$\chi_2^2$	
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
100	0.0480	0.1110	0.0470	0.0980	0.0570	0.0940
300	0.0490	0.0960	0.0500	0.0810	0.0590	0.0910
1000	0.0440	0.1130	0.0480	0.1000	0.0640	0.1090
2000	0.0460	0.1010	0.0460	0.1020	0.0420	0.1120
10000	0.0480	0.1090	0.0630	0.1120	0.0700	0.0930
25000	0.0640	0.1150	0.0440	0.1210	0.0740	0.1090

n (T=500)	mixture*		U(0,1)	
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
100	0.0440	0.1090	0.0390	0.1080
300	0.0480	0.0870	0.0480	0.0960
1000	0.0450	0.1040	0.0570	0.1090
2000	0.0520	0.1010	0.0600	0.1060
10000	0.0570	0.1050	0.0560	0.1090
25000	0.0650	0.1370	0.0610	0.1210

\* Mixture de normales de media 0 y desviaciones 0.8 y 0.2, con probabilidad de mezcla de 0.3.

$$g_2(u) = |u|$$

n (T=500)	$t_3$		logN(0,1)		$\chi_2^2$	
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
100	0.0630	0.1090	0.0440	0.1090	0.0540	0.1060
300	0.0520	0.0950	0.0490	0.1090	0.0550	0.0880
1000	0.0580	0.1020	0.0570	0.1170	0.0550	0.1130
2000	0.0520	0.0950	0.0550	0.1130	0.0600	0.1020

<b>10000</b>	0.0530	0.1140	0.0840	0.1160	0.0930	0.1450
<b>25000</b>	0.0630	0.1360	0.1200	0.1640	0.1100	0.1710

	mixture*		U(0,1)	
n (T=500)	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
<b>100</b>	0.0630	0.1030	0.0600	0.1070
<b>300</b>	0.0600	0.1140	0.0500	0.1180
<b>1000</b>	0.0500	0.1070	0.0540	0.1000
<b>2000</b>	0.0440	0.1030	0.0380	0.1010
<b>10000</b>	0.0600	0.0990	0.0550	0.1000
<b>25000</b>	0.0800	0.1230	0.0640	0.1130

\* Mixture of normals with mean 0 and standard deviations 0.8 and 0.2, with a mixture probability of 0.3.

$$g_3(u) = u^2 1_{\{|u|>\delta\}}, \quad \delta = 0.1$$

	$t_3$		logN(0,1)		$\chi_2^2$	
n (T=500)	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
<b>100</b>	0.0460	0.1210	0.0520	0.1100	0.0470	0.0970
<b>300</b>	0.0480	0.0740	0.0360	0.0970	0.0510	0.0930
<b>1000</b>	0.0570	0.0880	0.0570	0.1000	0.0450	0.0810
<b>2000</b>	0.0630	0.0970	0.0460	0.1090	0.0500	0.1010
<b>10000</b>	0.0490	0.1050	0.0540	0.1120	0.0600	0.0980
<b>25000</b>	0.0610	0.1190	0.0580	0.1230	0.0720	0.1120

	mixture*		U(0,1)	
n (T=500)	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
<b>100</b>	0.0510	0.1070	0.0580	0.1140
<b>300</b>	0.0600	0.0840	0.0590	0.0840
<b>1000</b>	0.0380	0.0990	0.0560	0.0980
<b>2000</b>	0.0460	0.0980	0.0480	0.0980
<b>10000</b>	0.0540	0.0960	0.0690	0.1110
<b>25000</b>	0.0660	0.1230	0.0630	0.1240

\* Mixture of normals with mean 0 and standard deviations 0.8 and 0.2, with a mixture probability of 0.3.

$$g_4(u) = |u| 1_{\{|u|>\delta\}}, \quad \delta = 0.1$$

	$t_3$		logN(0,1)		$\chi_2^2$	
n (T=500)	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
<b>100</b>	0.0430	0.0950	0.0560	0.0940	0.0620	0.1000
<b>300</b>	0.0600	0.1020	0.0520	0.0980	0.0530	0.0950
<b>1000</b>	0.0510	0.1080	0.0580	0.0980	0.0420	0.0900
<b>2000</b>	0.0590	0.1110	0.0610	0.1180	0.0710	0.0990
<b>10000</b>	0.0660	0.1020	0.0930	0.1330	0.0740	0.1580
<b>25000</b>	0.0720	0.1200	0.1230	0.1810	0.1120	0.1620

	mixture*		U(0,1)	
n (T=500)	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
<b>100</b>	0.0500	0.1020	0.0550	0.1070

<b>300</b>	0.0540	0.0940	0.0510	0.0930
<b>1000</b>	0.0450	0.1020	0.0530	0.0900
<b>2000</b>	0.0480	0.1110	0.0410	0.0990
<b>10000</b>	0.0630	0.0880	0.0460	0.1100
<b>25000</b>	0.0580	0.1300	0.0630	0.1010

\* Mixtura de normales de media 0 y desviaciones 0.8 y 0.2, con probabilidad de mezcla de 0.3.

Como podemos observar, en general los resultados son aceptables, si bien la lognormal y la chi-cuadrado parecen responder peor que el resto para los valores de  $n$  más grandes considerados, siendo esta diferencia más apreciable en  $g_2$  y  $g_4$ . A la vista de esto, cuando mantenemos las variables normales estándar y variamos la distribución de los errores, el contraste parece funcionar bien incluso para distribuciones asimétricas.

### Distintas distribuciones de los errores y las variables

A continuación, probamos aplicando el cambio en distribución tanto para las variables como para los errores, utilizando las mismas distribuciones citadas anteriormente y tipificando únicamente los errores.

$$g_1(u) = u^2$$

n (T=500)	$t_3$		logN(0,1)		$\chi_2^2$	
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
<b>100</b>	0.0420	0.1280	0.0280	0.1050	0.0480	0.1280
<b>300</b>	0.0480	0.1100	0.0430	0.0860	0.0430	0.1050
<b>1000</b>	0.0490	0.0990	0.0390	0.1320	0.0520	0.1050
<b>2000</b>	0.0530	0.0930	0.0360	0.1090	0.0450	0.1250
<b>10000</b>	0.0340	0.0840	0.0490	0.1010	0.0400	0.1000
<b>25000</b>	0.0410	0.0910	0.0490	0.1010	0.0640	0.1060

n (T=500)	mixtura*		U(0,1)	
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
<b>100</b>	0.0510	0.0960	0.0450	0.1130
<b>300</b>	0.0510	0.0940	0.0490	0.1050
<b>1000</b>	0.0560	0.1130	0.0610	0.1010
<b>2000</b>	0.0510	0.1110	0.0570	0.1130
<b>10000</b>	0.0510	0.1140	0.0730	0.1180
<b>25000</b>	0.0680	0.1370	0.1040	0.1450

\* Mixtura de normales de media 0 y desviaciones 0.8 y 0.2, con probabilidad de mezcla de 0.3.

$$g_2(u) = |u|$$

n (T=500)	$t_3$		logN(0,1)		$\chi_2^2$	
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
<b>100</b>	0.0420	0.1280	0.0280	0.1050	0.0480	0.1280
<b>300</b>	0.0480	0.1100	0.0430	0.0860	0.0430	0.1050
<b>1000</b>	0.0490	0.0990	0.0390	0.1320	0.0520	0.1050
<b>2000</b>	0.0530	0.0930	0.0360	0.1090	0.0450	0.1250
<b>10000</b>	0.0340	0.0840	0.0490	0.1010	0.0400	0.1000
<b>25000</b>	0.0410	0.0910	0.0490	0.1010	0.0640	0.1060

100	0.0420	0.1070	0.0500	0.1110	0.0570	0.1040
300	0.0510	0.1060	0.0640	0.1400	0.0600	0.1150
1000	0.0610	0.0970	0.0770	0.1370	0.0740	0.1270
2000	0.0470	0.0840	0.1130	0.1740	0.0710	0.1420
10000	0.0700	0.1040	0.2080	0.2680	0.1470	0.1920
25000	0.0740	0.1240	0.3240	0.3160	0.2600	0.3180

	mixture*		U(0,1)	
n (T=500)	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
100	0.0490	0.1250	0.0420	0.0960
300	0.0380	0.1000	0.0580	0.1030
1000	0.0630	0.1090	0.0400	0.0910
2000	0.0570	0.1070	0.0590	0.1020
10000	0.0810	0.1300	0.0620	0.1190
25000	0.1090	0.1600	0.0710	0.1360

\* Mixture of normals of mean 0 and deviations 0.8 and 0.2, with probability of mixture of 0.3.

$$g_3(u) = u^2 1_{\{|u|>\delta\}}, \quad \delta = 0.1$$

	$t_3$		logN(0,1)		$\chi_2^2$	
n (T=500)	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
100	0.0470	0.1210	0.0310	0.1100	0.0610	0.1020
300	0.0440	0.0990	0.0380	0.1150	0.0550	0.1100
1000	0.0350	0.1150	0.0470	0.0890	0.0580	0.0910
2000	0.0420	0.0780	0.0340	0.0980	0.0510	0.0900
10000	0.0450	0.0950	0.0460	0.1090	0.0360	0.0970
25000	0.0420	0.0950	0.0510	0.0960	0.0530	0.1130

	mixture*		U(0,1)	
n (T=500)	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
100	0.0420	0.1130	0.0430	0.1300
300	0.0480	0.1150	0.0540	0.1010
1000	0.0480	0.1220	0.0490	0.1050
2000	0.0530	0.1040	0.0680	0.1000
10000	0.0610	0.1110	0.0710	0.1240
25000	0.0690	0.1260	0.1020	0.1590

\* Mixture of normals of mean 0 and deviations 0.8 and 0.2, with probability of mixture of 0.3.

$$g_4(u) = |u| 1_{\{|u|>\delta\}}, \quad \delta = 0.1$$

	$t_3$		logN(0,1)		$\chi_2^2$	
n (T=500)	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
100	0.0530	0.1060	0.0510	0.1120	0.0540	0.1130
300	0.0540	0.1080	0.0660	0.1290	0.0470	0.1050
1000	0.0480	0.0930	0.0850	0.1350	0.0590	0.1180
2000	0.0660	0.1190	0.1060	0.1460	0.0810	0.1550
10000	0.0580	0.1310	0.2110	0.2750	0.1670	0.2200
25000	0.0770	0.1230	0.2930	0.3500	0.2280	0.2980

n (T=500)	mixture*		U(0, 1)	
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
100	0.0540	0.1180	0.0580	0.0990
300	0.0410	0.1070	0.0590	0.0820
1000	0.0480	0.1030	0.0500	0.1010
2000	0.0430	0.1070	0.0450	0.1190
10000	0.0770	0.1320	0.0690	0.1000
25000	0.0960	0.1620	0.0770	0.1160

\* Mixture de normales de media 0 y desviaciones 0.8 y 0.2, con probabilidad de mezcla de 0.3.

De nuevo obtenemos los mismos resultados: a partir de un valor demasiado grande de  $n$  en comparación a  $T$  el porcentaje de rechazo aumenta, principalmente cuando consideramos la lognormal y la chi cuadrado.

### Aumento del número de variables

A continuación analizamos los resultados del estadístico generando los datos mediante un modelo con más variables (en este caso cinco)

$$Y_t = X_1 + X_2 + X_3 + X_4 + X_5 + \epsilon_t$$

y estimando los modelos

$$Y_t = a_1 + b_1 X_{1t} + c_1 X_{2t} + d_1 X_{3t} + \epsilon_{1t}$$

$$Y_t = a_2 + d_2 X_{3t} + e_2 X_{4t} + f_2 X_{5t} + \epsilon_{2t}$$

de forma que éstos tengan alguna variable en común, con lo que la correlación entre ambos no sea nula.

$$g_1(u) = u^2$$

n (T=500)	Usando residuos		Usando errores	
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
100	0.0580	0.0960	0.0500	0.0920
300	0.0490	0.1150	0.0500	0.1190
1000	0.0460	0.0940	0.0400	0.0840
2000	0.0570	0.0720	0.0540	0.0720
10000	0.0440	0.1190	0.0370	0.0940
25000	0.0650	0.1200	0.0420	0.0920

$$g_2(u) = |u|$$

n (T=500)	Usando residuos		Usando errores	
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
100	0.0520	0.0960	0.0530	0.0960

<b>300</b>	0.0420	0.1090		0.0440	0.1070
<b>100</b>	0.0530	0.0980		0.0580	0.0990
<b>2000</b>	0.0630	0.1020		0.0690	0.0970
<b>10000</b>	0.0620	0.1330		0.0460	0.1120
<b>25000</b>	0.0820	0.1430		0.0590	0.0840

$$g_3(u) = u^2 1_{\{|u|>\delta\}}, \quad \delta = 0.1$$

	Usando residuos			Usando errores	
<b>n (T=500)</b>	$\alpha=0.05$	$\alpha=0.10$		$\alpha=0.05$	$\alpha=0.10$
<b>100</b>	0.0400	0.1020		0.0380	0.1080
<b>300</b>	0.0650	0.0910		0.0700	0.0940
<b>100</b>	0.0580	0.0930		0.0480	0.0880
<b>2000</b>	0.0620	0.0850		0.0530	0.0850
<b>10000</b>	0.0690	0.1020		0.0530	0.0850
<b>25000</b>	0.0880	0.1260		0.0590	0.0810

$$g_4(u) = |u| 1_{\{|u|>\delta\}}, \quad \delta = 0.1$$

	Usando residuos			Usando errores	
<b>n (T=500)</b>	$\alpha=0.05$	$\alpha=0.10$		$\alpha=0.05$	$\alpha=0.10$
<b>100</b>	0.0460	0.0970		0.0460	0.0960
<b>300</b>	0.0530	0.1060		0.0530	0.1060
<b>100</b>	0.0590	0.1130		0.0590	0.1090
<b>2000</b>	0.0440	0.1010		0.0410	0.1020
<b>10000</b>	0.0640	0.1090		0.0480	0.0950
<b>25000</b>	0.0770	0.1330		0.0560	0.0870

Para valores de  $n$  cercanos a  $T$  los resultados siguen siendo buenos, pero a partir de cierto  $n$  aumenta el tamaño del contraste.

### Coefficientes variables

Dadas las características sencillas del modelo con el que hemos creado los datos hasta ahora, realizaremos un estudio empírico sobre cómo afecta a los resultados del estadístico el hecho de considerar coeficientes variables, aproximándonos de esta forma a un modelo que refleja en mayor medida situaciones reales.

Proponemos entonces el siguiente modelo:

$$Y_t = a + b_t X_{1t} + c_t X_{2t} + \epsilon_t$$

con

$$a = cte, \quad b_t = bZ_{1t}, \quad c_t = cZ_{2t}$$

donde  $Z_{1t}$  y  $Z_{2t}$  serán coeficientes cuyo carácter iremos cambiando para analizar distintos casos. Por tanto, los modelos a estimar son



$$Y_t = a + bZ_{1t}X_{1t} + \epsilon_{1t}$$

$$Y_t = a + cZ_{2t}X_{2t} + \epsilon_{2t}$$

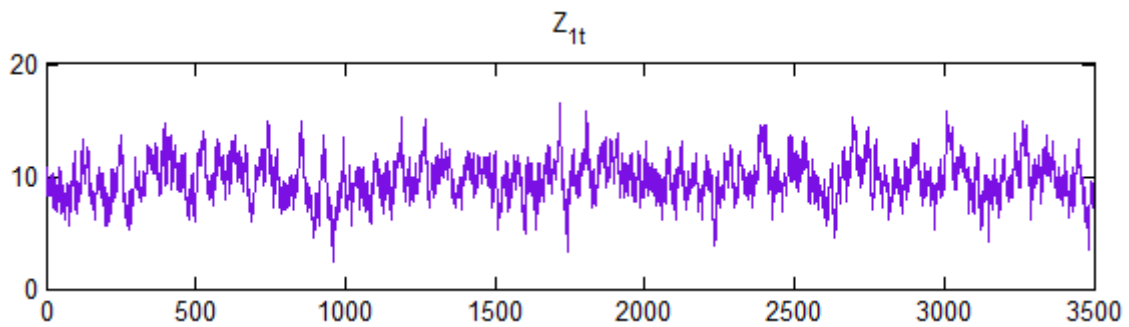
En primer lugar consideramos que las variables  $X_1$  y  $X_2$  son dos normales de media cero y varianza ocho, y los errores siguen una distribución normal de media cero y varianza el 10% de la varianza de la parte sistemática. Esta forma de generar los errores la mantendremos hasta el final. Si asumimos que  $Z_1$  y  $Z_2$  son la unidad, nos encontramos en un caso similar a los anteriores, de forma que tomando  $T=2500$  y  $n=1000$ , el porcentaje de rechazo de la hipótesis nula cuando es cierta es próximo al nivel de significación considerado 0.05. La función de pérdida considerada de ahora en adelante será  $g(u) = u^2$ .

Además, bajo estos  $Z_1$  y  $Z_2$  fijos, si tornamos la distribución de  $X_2$  en una chi cuadrado con cuatro grados de libertad de forma que controlemos la varianza<sup>1</sup> y la centramos, el porcentaje de rechazo se mantiene próximo a 0.05.

Para empezar a variar  $Z_1$  y  $Z_2$ , comenzamos suponiendo que ambos siguen un proceso autorregresivo de segundo orden con deriva de la siguiente forma:

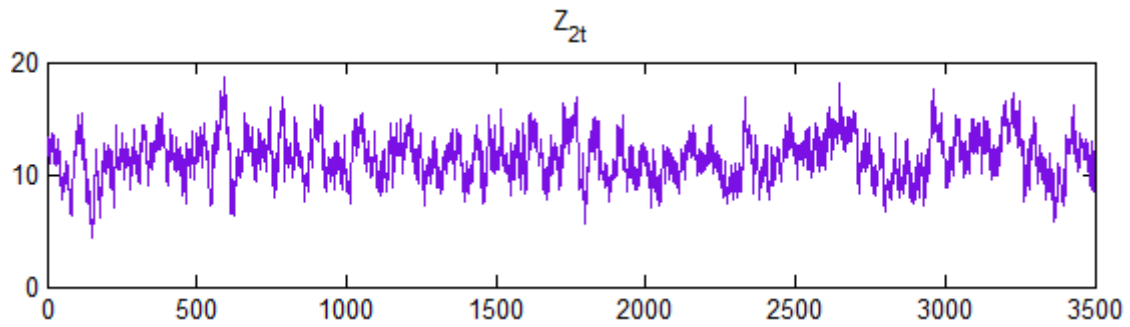
$$Z_{1t} = 1 + 0.6Z_{1,t-1} + 0.3Z_{1,t-2} + a_{1t}, \quad a_{1t} \sim N(0,1)$$

$$Z_{2t} = 1 + \sqrt{\frac{17.8}{105}}Z_{2,t-1} + 0.5Z_{2,t-2} + a_{2t}, \quad a_{2t} \sim N(0,1)$$



<sup>1</sup> Para hallar la varianza de los errores y escoger los parámetros b y c recurrimos a la siguiente expresión, teniendo en cuenta que todas las variable son creadas de forma independiente:

$$Var(XY) = Var(X)Var(Y) + E^2(Y)Var(X) + E^2(X)Var(Y)$$



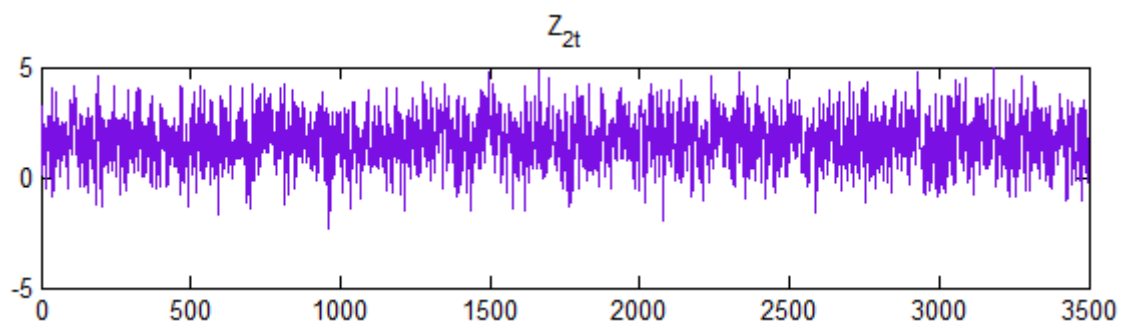
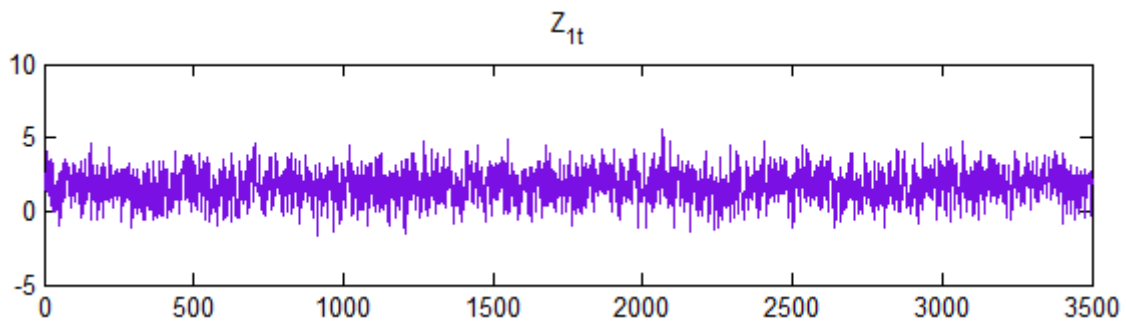
Por tanto la media y la varianza de estos procesos son 10 y  $0.7/0.169$  respectivamente. Siendo  $X_1$  y  $X_2$  normales de media cero y varianza ocho y realizando las operaciones oportunas para obtener la igualdad de varianzas para la hipótesis nula, nuestro modelo para generar los datos es

$$Y_t = 2 + \sqrt{19} Z_{1t} X_{1t} + \sqrt{17.6} Z_{2t} X_{2t} + \epsilon_t$$

siendo  $Z_1$  y  $Z_2$  los autorregresivos citados. El porcentaje de rechazo en este caso es de 0.2320. Este resultado no es muy deseable, posiblemente sea debido a la correlación serial que introducimos con los autorregresivos, por lo que nuestra siguiente prueba consistirá en disminuir la persistencia de los mismos. Los nuevos procesos considerados son:

$$Z_{1t} = 1 + 0.25Z_{1,t-1} + 0.15Z_{1,t-2} + a_{1t}, \quad a_{1t} \sim N(0,1)$$

$$Z_{2t} = 1 + 0.3Z_{2,t-1} + 0.1Z_{2,t-2} + a_{2t}, \quad a_{2t} \sim N(0,1)$$



así que nuestro modelo generador de datos es

$$Y_t = 2 + \sqrt{121.9104} Z_{1t} X_{1t} + \sqrt{111.99} Z_{2t} X_{2t} + \epsilon_t$$

Obtenemos con esto un porcentaje de rechazo en torno a 0.18, que a pesar de seguir sin ser aceptable nos sugiere que la persistencia ejerce bastante influencia sobre el estadístico.

La siguiente prueba realizada consiste en considerar que  $X_1$  y  $X_2$  son normales de media cero y varianza ocho y  $Z_1$  y  $Z_2$  son chi cuadrado con cuatro grados de libertad centradas. En este caso obtenemos un porcentaje de rechazo en torno a 0.05. Además, si permutamos la distribución de  $X_1$  y  $Z_1$  de forma que la primera sea una chi cuadrado y la segunda una normal, obtenemos el mismo resultado. En ambos casos las constantes b y c deben ser iguales.

Esto sugiere que todas las variables deberían tener media cero, así que repetimos los casos en que consideramos procesos autorregresivos, esta vez sin deriva, de forma que la media de ambos procesos sea nula. Esto es, generamos los datos con el modelo:

$$Y_t = 2 + b Z_{1t} X_{1t} + c Z_{2t} X_{2t} + \epsilon_t$$

siendo  $b=c$ ,  $X_1$  y  $X_2$  normales de media cero y varianza ocho,  $\epsilon$  normal de media cero y varianza la correspondiente según hemos mencionado anteriormente.

En primer lugar consideramos

$$Z_{1t} = 0.6Z_{1,t-1} + 0.3Z_{1,t-2} + a_{1t}, \quad a_{1t} \sim N(0,1)$$

$$Z_{2t} = \sqrt{\frac{17.8}{105}} Z_{2,t-1} + 0.5Z_{2,t-2} + a_{2t}, \quad a_{2t} \sim N(0,1)$$

De esta forma obtenemos que el porcentaje de rechazo es 0.3460.

En segundo lugar proponemos los otros dos autorregresivos que tienen menor persistencia:

$$Z_{1t} = 0.25Z_{1,t-1} + 0.15Z_{1,t-2} + a_{1t}, \quad a_{1t} \sim N(0,1)$$

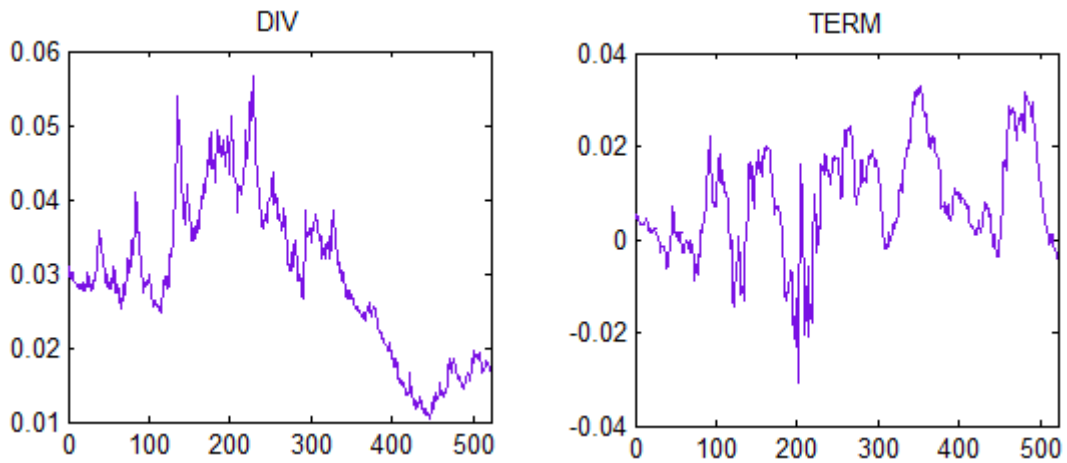
$$Z_{2t} = 0.3Z_{2,t-1} + 0.1Z_{2,t-2} + a_{2t}, \quad a_{2t} \sim N(0,1)$$

Efectivamente, el porcentaje de rechazo obtenido en este caso es de 0.05.

Todos los resultados anteriores sugieren que el estadístico de Diebold y Mariano, bajo coeficientes cambiantes, únicamente funciona si las variables fluctúan en torno a una media y los coeficientes tienen poca persistencia.

Dado que en la práctica estos coeficientes pueden ser una variable cualquiera, hemos obtenido dos series de 522 datos mensuales correspondientes al dividend yield

(DIV) y al term spread (TERM) desde julio de 1963 hasta diciembre de 2006<sup>2</sup>. Ambas series tienen media cero.



Probamos considerando que  $T=250$  y  $n=272$ ,  $Z_{1t}$  y  $Z_{2t}$  son el dividend yield, y mantenemos  $X_{1t}$  y  $X_{2t}$  como las normales de costumbre:

$$Y_t = 2 + b DIV_t X_{1t} + c DIV_t X_{2t} + \epsilon_t$$

con  $b=c$ . Obtenemos un porcentaje de rechazo de 0.05. Si repetimos el proceso siendo  $Z_{1t}$  y  $Z_{2t}$  el term spread,

$$Y_t = 2 + b TERM_t X_{1t} + c TERM_t X_{2t} + \epsilon_t$$

con  $b=c$ , el porcentaje de rechazo resultante es de 0.048, así que en este caso el estadístico funciona adecuadamente.

---

<sup>2</sup> Los datos para el dividend yield se han obtenido del CRSP (Center for Research in Security Prices) y los del term spread proceden de bonos de la página del FRED (Federal Reserve Economic Data).

## **Conclusión**

Diebold y Mariano proporcionan un estadístico para contrastar la igualdad de capacidad predictiva de dos modelos dados que explican una variable dependiente común. Sin embargo se basan únicamente en errores teóricos. Ferreira y Stute demuestran que bajo ciertas condiciones este estadístico sigue funcionando cuando utilizamos los residuos asociados a los modelos una vez estimada su parte sistemática.

Nuestros resultados muestran que la elección de la función de pérdida no afecta significativamente a los resultados que ofrece el estadístico. Además, observamos que el tamaño muestral utilizado para estimar debe ser similar al utilizado para validar, y que a medida que aumentan los mismos el contraste es más fiable, aunque funciona bien aún con muestras pequeñas.

Por otra parte, el estadístico responde bien tanto si cambiamos la distribución de los errores como la de las variables, a pesar de que debe tenerse más cuidado cuando tratamos con una lognormal o una chi cuadrado, ya que los resultados parecen ser algo menos fiables en estos casos.

La introducción de más variables en los modelos con una variable en común no altera los resultados, por lo que en este caso el estadístico también funciona bien.

Sabemos que el estadístico distingue si dos conjuntos tienen la misma capacidad predictiva si las variables tienen la misma varianza y bajo coeficientes constantes y ciertas características funciona bien. Sin embargo, dado que la única teoría sobre este estadístico es para coeficientes fijos, entramos en el marco de los coeficientes cambiantes. Las pruebas empíricas sugieren que al suprimir la hipótesis de coeficientes fijos, para que el estadístico funcione los coeficientes y las variables deben fluctuar en torno a una media y se debe tener cuidado con la persistencia.

Es decir, si queremos utilizar este estadístico teniendo coeficientes cambiantes, se debe analizar la persistencia de las variables y las betas, y en caso de que ésta aumente demasiado se debe ajustar el estadístico a las nuevas características y probar las propiedades asintóticas del mismo.

## **Referencias**

Box, G. E. P. and Jenkins, G. M. (1976), "Time Series Analysis: Forecasting and Control", San Francisco: Holden-Day.

Clements, Michael P. and Harvey, David I. (2007), "Forecast combination and encompassing".

Diebold, F. X. and Mariano, R. (1995), "Comparing Predictive Accuracy", Journal of Business and Economic Statistics, 13, 253-265.

Ferreira, E. and Stute W. (2009), "Testing for Differences in Predictive Accuracy"

Gujarati, Damodar N. (2003), "Econometrics", 4<sup>th</sup> edition.

Inoue, A., Kilian, L. (2004), "In-sample or out-of-sample tests of predictability: which one should we use?" Econometric Reviews 23, 371-402.

Schinasi, Garry J. and Swamy, P.A.V.B. (1987), "The out-of-sample forecasting performance of Exchange Rate Models When Coefficients Are Allowed to Change".

Mariano, Roberto S. and Preve, Daniel (2012), "Statistical Tests for Multiple Forecast Comparison".