

MÉTODO «YSIM» PARA AFINAR MODELOS ECONOMETRICOS SIMULACIÓN MONTECARLO SOBRE COEFICIENTES DE REGRESIÓN

Lic. Castillo Manzur¹, Luis Alberto

✉ lcastillo@asfi.gob.bo

RESUMEN

El método *YSIM* incorpora el concepto de “entropía” (término utilizado por Claud Shannon para medir la cantidad de aleatoriedad, o incertidumbre en una población), con el objetivo de “mejorar” la especificación de modelos econométricos. Para ello se utilizan los coeficientes determinados, por ejemplo, a través de MCO², para utilizarlos como valores de entrada para modelos de simulación (con simulación Montecarlo o Hipercúbico Latino), obteniendo por resultado un valor de pronóstico para la variable dependiente (*Y*) denominada: *Y* simulada (*YSIM*).

Para explicar el desarrollo de la metodología se considera como base de análisis modelos de regresión lineal simple, donde los coeficientes encontrados para las variables independientes (o explicativas) son transformados en variables del modelo de simulación (que pueden o no ajustarse a una distribución normal) y cuya variable de pronóstico es la variable dependiente (*Y*). El método implica una pérdida en el nivel de ajuste “*r*”, pero logra una mejora en el nivel de pronóstico del modelo, es decir, los valores pronosticados (con base a $[1, n-1]$) se acercan más a los valores reales (*n*).

El presente trabajo de investigación fue registrado en el Servicio Nacional de Propiedad Intelectual (SENAPI), en fecha 6 de septiembre de 2010, con Resolución Administrativa 1-647/2010, siendo esta la primera vez que se publica en un medio impreso o digital.³

PALABRAS CLAVE

Simulación, Método Montecarlo, econometría, entropía.

ABSTRACT

The *YSIM* method incorporates the concept of “entropy” (a term used by Claud Shannon to measure the amount of randomness, or uncertainty in a population), with the objective of “improving” the specification of econometric models. To do this, the coefficients determined are used, for example, through OLS, to use them as input values for simulation models (with Montecarlo or Latin Hypercube simulation), obtaining as a result a forecast value for the dependent variable (*Y*) called : And simulated (*YSIM*).

To explain the development of the methodology, simple linear regression models are considered as the basis of analysis, where the coefficients found for the independent (or explanatory) variables are transformed into simulation model variables (which may or may not conform to a normal distribution). and whose forecast variable is the dependent variable (*Y*). The method implies a loss in the level of adjustment “*r*”, but achieves an improvement in the level of forecast of the model, that is, the predicted values (based on $[1, n-1]$) are closer to the values real (*n*).

The present research work was registered in the National Intellectual Property Service (SENAPI) on September 6, 2010, with Administrative Resolution 1-647 / 2010, this being the first time it was published in a printed or digital media.

KEYWORDS

Simulation, Monte Carlo Method, econometrics, entropy.

¹ Analista de monitoreo del sistema financiero - Autoridad de Supervisión del Sistema Financiero (ASFI).
Nota: El contenido del presente documento es de responsabilidad del autor y no compromete la opinión de ASFI.

² traducido es método de Mínimos Cuadrados Ordinarios

³ Derecho de autor y derechos conexos. (Septiembre – Octubre / 2010). Obras literarias - Método *YSIM* para afinar modelos econométricos. Obtenida el 20 de octubre de 2017, Servicio Nacional de Propiedad Intelectual, de http://www.senapi.gob.bo/snpPublicacionesGaceta/Publicacion_dac_septiembre_octubre_2010.pdf.

1. ANTECEDENTES

La hipótesis del modelo implica la presencia de un comportamiento aleatorio en las observaciones, que anulan el supuesto de Normalidad.

Dicho comportamiento puede ser “absorbido” parcialmente aplicando el método Montecarlo sobre los coeficientes determinados por MCO (m , b), proceso que se denominará de “afinación”.

Al afinarse el modelo MCO por simulación Montecarlo, los coeficientes de la regresión dejarán de ser los mejores estimadores linealmente insesgados (MELI) y su efecto inmediato será una reducción en el coeficiente de correlación lineal “ r ”.

Los nuevos valores encontrados a través de simulación para (m , b) determinarán la nueva ecuación de ajuste de Y , la cual se denomina *YSIM*.

2. ESPECIFICACIÓN DEL MODELO YSIM

PRIMER PASO - Representación del Modelo MCO

Se tiene la expresión:

$$Y = mX + b \quad (1)$$

Dados los parámetros que determinan la ecuación de ajuste:

$$m = \frac{\sum(xy) - \frac{(\sum x)(\sum y)}{n}}{\sum(x)^2 - \frac{(\sum x)^2}{n}} \quad (2)$$

$$b = \frac{\sum y - m \sum x}{n} \quad (3)$$

Ajustados por el coeficiente de correlación lineal:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (4)$$

SEGUNDO PASO – El modelo MCO afectado por Entropía

Concepto de Entropía según la Teoría introducida por Claud Shanon:

$$Entropía = P \cdot \log_2(P) + Q \cdot \log_2(Q) \quad (5)$$

Donde se tienen los dos casos extremos:

$$Entropía = 1 \cdot \log_2(1) + 0 \cdot \log_2(0) = 0,0$$

(No existe aleatoriedad)

$$Entropía = 0,5 \cdot \log_2(0,5) + 0,5 \cdot \log_2(0,5) = 1,0$$

(Presencia absoluta de aleatoriedad)

Hipótesis del Modelo:

Si:

$$Entropía > 0 \quad (6)$$

Entonces:

No “necesariamente” se cumple el supuesto de Normalidad

$$X \sim N(\mu, \sigma) \quad (7)$$

Nota: No necesariamente significa que puede o no cumplirse el supuesto en cada una de las observaciones de forma independiente.

Método de Simulación Montecarlo

Una estimación de Monte Carlo $\hat{T}$ de (T) es obtenida por generar al azar N muestras $\{x_i\}_{i=1}^N$ desde una distribución f .

TERCER PASO - Método YSIM

1. Para cada x_i se determina su distribución de probabilidades (f).
2. Para cada x_i se establece los límites:
 $x_{min} < \bar{x} < x_{max}$.
3. Cada x_i constituye una variable supuesta en el modelo de simulación.
4. Se establecen como variables de pronóstico los coeficientes MCO: $(m, b, "r")$.
5. Se generan números randómicos (R) para cada x_i sujetos a su comportamiento probabilístico determinado en el paso 1) y a sus restricciones de límites determinados en el paso 2).
6. Cada número generado afecta los coeficientes $(m, b, "r")$, generándose a su vez una nueva base de datos con los valores simulados para los coeficientes de pronóstico.
7. La media aritmética de la base de datos generada por simulación para $(m, b, "r")$, nos da los nuevos valores de la ecuación original obtenida por MCO:

$$Y_{(sim)} = m_{(sim)} X + b_{(sim)} \quad (8)$$

8. Donde la media calculada para R obtenida en el paso 6), muestra:

$$r_{(sim)} < r \quad (9)$$

9. Donde la relación de las diferencias de los valores de pronóstico medidas en términos absolutos muestran:

$$10. [YSIM_{t+1} - X_{t+1}] < [Y_{t+1} - X_{t+1}] \quad (10)$$

Nota: Los resultados de los experimentos que prueban esta hipótesis se muestran en los ejemplos de aplicación 1 y 2.

CUARTO PASO – Restricciones de Información para Determinar las Distribuciones de Probabilidad

Métodos de Ajuste Típicos en Econometría

Si existiera una cantidad mayor de datos históricos a cinco observaciones, entonces se puede utilizar uno de los siguientes métodos de ajuste:

- 1) Anderson-Darling.
- 2) Chi-cuadrado.
- 3) Kolmogorov-Smirnov.

Métodos de Ajuste Basados en Lógica Difusa

Si la información histórica fuese insuficiente, se puede recurrir a un método alternativo basado en lógica difusa como lo es la aplicación de «Tripletas de Confianza».

El método se basa en la adquisición de conocimiento experto para determinar las distribuciones de probabilidad para cada x_i .

A través de este método los expertos consensuan criterios para la determinación de una distribución de probabilidades, común para toda la serie, y se analizarán como casos especiales aquellas que se consideren que tuvieron perturbaciones exógenas, las cuales alteraron su normal desempeño.

Los pasos a seguir para la aplicación de una tripleta de confianza son:

1. Identificar al grupo de expertos del área específica a modelarse.

2. Identificar un grupo de distribuciones de probabilidad explicativas de x_i .
3. Aplicar la metodología de Tripletas de confianza para cada Distribución de Probabilidades identificada en el paso 2).
4. Reevaluar el comportamiento excepcional de las x_i .
5. Valuación e identificación de las distribuciones de probabilidad asociadas a cada x_i .

Ejemplo de aplicación 1 - Experimento sobre el Comportamiento de Ventas en Supermercados⁴

Se tiene la siguiente base de datos que contiene las ventas anuales y mensuales desde el año 1998 al año 2008.

Para armar el modelo MCO se tomaron en cuenta los años 1998 a 2007. El año 2008 sirvió para contrastar los resultados de pronóstico obtenidos tras aplicar el método *YSIM* respecto al modelo original.

Siguiendo los pasos 1 y 2 del método se procedió a determinar las distribuciones de probabilidad y los valores mínimos y máximos para todas las observaciones de la serie.

Para calcular las distribuciones de

Tabla 1
Venta en Supermercados por Mes, según Período Anual 1998-2008.
(Expresado en miles de pesos argentinos)

Mes	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Total	259.645	261.391	260.219	263.658	316.018	384.049	450.021	545.701	673.243	962.637	1.344.560
Enero	20.375	22.670	20.782	22.445	20.945	29.349	34.567	41.356	49.835	65.870	95.833
Febrero	19.594	20.448	19.220	20.230	20.247	28.049	33.841	38.304	48.096	63.437	96.654
Marzo	20.540	22.235	20.803	22.840	23.887	31.305	34.897	41.854	51.649	71.718	107.042
Abril	20.415	20.772	21.687	22.150	22.890	29.704	35.162	41.854	53.425	70.799	102.266
Mayo	21.399	20.556	20.018	21.347	24.308	30.162	35.399	42.888	51.609	69.928	107.429
Junio	19.049	20.372	20.243	21.964	26.503	30.028	35.190	42.408	54.069	76.066	107.943
Julio	21.886	22.433	22.484	22.293	27.069	32.358	39.048	47.843	58.749	79.076	111.952
Agosto	22.015	21.211	21.627	21.607	28.420	33.258	38.162	46.466	57.196	83.348	118.809
Septiembre	21.403	20.167	21.349	20.484	26.641	30.756	37.164	45.583	57.753	86.021	111.251
Octubre	23.135	22.421	21.804	20.889	28.650	33.444	39.920	49.395	58.161	88.688	120.292
Noviembre	21.666	20.591	21.262	20.928	30.716	33.261	37.390	47.531	57.518	89.775	121.102
Diciembre	28.168	27.515	28.940	26.481	35.742	42.375	49.281	58.025	75.183	117.911	143.987

Fuente: Encuesta de Supermercados - Instituto Nacional de Estadística y Censos de la República Argentina INDEC
Elaboración: Propia

probabilidad se utilizó Anderson-Darling. Los valores mínimos y máximos mensuales se multiplicaron por 12 para obtener los valores equivalentes anuales.

Los resultados de aplicar una simulación Montecarlo tres veces, con 10.000, 50.000, 20.000 y 30.000 corridas para *YSIM* 1, 2, 3 y 4 respectivamente:

⁴ Nota: Para la determinación de las distribuciones de probabilidad en el caso 1 se utilizó el software "Crystal Ball" y para el caso 2 se utilizó el "software EasyFit". Para la simulación Monte Carlo en ambos casos se utilizó el software "Crystal Ball".

Tabla 2.
Venta en Supermercados – Valores Mínimos y Máximos
(En miles de pesos argentinos)

Series	1	2	3	4	5	6	7	8	9	10
Distribución de probabilidad	Exponential Distribution	Exponential Distribution	Exponential Distribution	Exponential Distribution	Exponential Distribution	Exponential Distribution	Exponential Distribution	Exponential Distribution	Exponential Distribution	Exponential Distribution
Min. Anual	228.588	242.004	230.640	242.760	242.964	336.588	406.092	459.648	577.152	761.244
Max. Anual	338.016	330.180	347.280	317.772	428.904	508.500	591.372	696.300	902.196	1.414.932
Min. Mens.	19.049	20.167	19.220	20.230	20.247	28.049	33.841	38.304	48.096	63.437
Max. Mens.	28.168	27.515	28.940	26.481	35.742	42.375	49.281	58.508	75.183	117.911

Fuente: Encuesta de Supermercados - Instituto Nacional de Estadística y Censos de la República Argentina INDEC

Elaboración: Propia

Los resultados de aplicar una simulación Montecarlo tres veces, con 5.000 corridas son los siguientes:

Ejemplo de aplicación 2 - Experimento sobre el Comportamiento de Ingresos de una Empresa de Transporte

Se tiene la siguiente base de datos que contiene las ventas anuales y mensuales desde el año 1994 al año 2008.

Tabla 3.
Venta en Supermercados – Resultados Simulación YSIM

Coefficientes	MCO	YSIM-1	YSIM-2	YSIM-3
m	68.269	72.959	72.583	73.255
b	62.177	73.959	75.362	73.146
r	83,03%	83,49%	86,56%	86,44%

Elaboración: Propia

Los resultados obtenidos cumplen con la condición $r_{(sim)} < r$, donde $r = 89\%$.

Los resultados proyectados armando las respectivas ecuaciones MCO y YSIM son los siguientes:

Tabla 4.
Venta en Supermercados – Valores Pronosticados

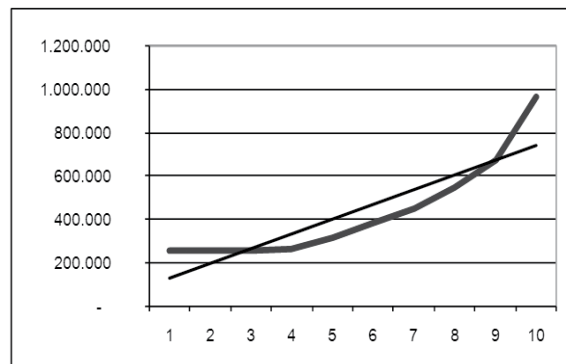
	Valor Real 2008	MCO	YSIM-1	YSIM-2	YSIM-3
Valor	1.344.560	813.139	876.509	876.509	876.509
Diferencia	0	531.421	468.051	468.051	468.051

Elaboración: Propia

Los valores de pronóstico obtenidos y contrastados con el valor real obtenido la gestión 2008 demuestran que se cumple con la condición $[YSIM_{t+1} - X_{t+1}] < [Y_{t+1} - X_{t+1}]$.

Es decir, el valor obtenido por YSIM se acerca más a la realidad que el obtenido por MCO.

Figura 1.
Curva de Ajuste por MCO - Venta en Supermercados



Elaboración: Propia

Tabla 5.
Pasajeros Mensuales - Pagos en Líneas de Superficie. Serie 1994-2008
(Expresado en pesos argentinos)

Mes	1994	1995	1996	1997	1998	1999	2000	2001	2002
Total	38.295.555	53.484.941	69.811.767	80.578.712	84.081.493	83.456.459	81.731.509	74.548.729	65.670.656
Enero	2.560.000	2.950.935	4.683.988	5.437.294	6.065.250	5.921.603	5.929.720	5.763.749	4.813.276
Febrero	2.548.000	3.094.000	4.770.892	5.470.048	5.925.237	5.924.419	6.151.227	5.784.833	4.713.223
Marzo	3.135.000	3.865.000	5.679.833	6.537.537	7.308.380	7.431.868	7.474.277	6.771.654	5.253.359
Abril	3.202.000	3.578.600	5.738.490	7.200.800	7.142.304	7.113.718	6.776.013	6.337.013	5.342.742
Mayo	3.449.000	3.522.864	6.166.159	7.126.818	7.083.804	7.160.955	6.865.277	6.589.197	5.650.058
Junio	3.073.000	4.453.158	5.469.692	6.357.319	6.831.504	6.862.637	6.615.927	6.237.491	5.209.623
Julio	3.028.000	4.873.620	5.942.268	7.037.629	7.303.532	6.900.187	6.828.412	6.231.929	5.705.310
Agosto	3.467.000	5.270.977	6.034.344	6.718.538	7.288.592	6.982.995	7.268.806	6.687.900	5.746.607
Septiembre	3.400.000	5.306.000	5.797.556	7.315.140	7.282.320	7.392.283	7.180.180	6.322.493	5.770.525
Octubre	3.381.000	5.548.596	6.826.135	7.576.028	7.630.230	7.277.601	7.209.607	6.424.079	6.011.818
Noviembre	3.642.000	5.668.117	6.558.448	6.997.877	7.268.182	7.518.477	7.017.920	6.417.480	5.860.311
Diciembre	3.410.555	5.353.074	6.143.962	6.803.864	6.952.158	6.969.716	6.414.143	4.980.911	5.593.804

Mes	2003	2004	2005	2006	2007	2008
Total	68.749.774	69.730.096	71.436.043	75.475.209	75.792.088	73.207.048
Enero	4.894.731	4.919.115	5.049.899	5.316.521	5.775.671	5.666.141
Febrero	4.818.443	5.274.726	5.286.834	5.341.250	5.630.726	5.651.774
Marzo	5.544.541	6.213.868	6.161.201	6.402.657	6.648.405	5.901.421
Abril	5.886.981	5.388.241	6.201.873	6.076.478	6.076.328	6.271.339
Mayo	5.990.287	5.699.965	6.206.230	6.423.772	6.554.966	6.347.938
Junio	5.684.817	5.758.258	5.892.332	6.061.815	3.193.453	5.820.680
Julio	5.980.490	5.672.766	5.912.278	6.367.450	6.249.410	6.398.106
Agosto	5.831.138	5.824.440	5.986.551	6.662.303	6.437.591	6.354.677
Septiembre	6.130.947	6.317.520	6.187.033	6.829.102	6.443.961	6.321.444
Octubre	6.310.794	6.246.100	6.136.629	6.769.880	6.774.532	6.499.962
Noviembre	5.862.974	6.258.509	6.250.511	6.950.006	6.807.417	6.106.330
Diciembre	5.813.631	6.156.588	6.164.672	6.273.975	6.199.628	5.867.236

Fuente: Comisión Nacional de Regulación del Transporte Público (CNRT República Argentina)
Elaboración: Propia

Para armar el modelo MCO se tomaron en cuenta solo los años 1994 a 2007. El año 2008 sirvió para contrastar los resultados de pronóstico obtenidos tras aplicar el método *YSIM* respecto al modelo original.

Siguiendo los pasos 1 y 2 del método, se procedió a determinar las distribuciones de probabilidad y los valores mínimos y

máximos para las observaciones de la serie.

Para calcular las distribuciones de probabilidad se utilizó Anderson-Darling. Los valores mínimos y máximos mensuales se multiplicaron por 12 para obtener los valores equivalentes anuales.

Tabla 6.
Pasajeros Mensuales – Ajuste Distribuciones de Probabilidad

Nº	Distribution	Min Anual	Min Mens	Max Anual	Max Mens
1	Weibull	30.576.000	43.704.000	2.548.000	3.642.000
2	Beta	35.411.220	68.017.404	2.950.935	5.668.117
3	Logistic	56.207.856	81.913.620	4.683.988	6.826.135
4	Normal	65.247.528	90.912.336	5.437.294	7.576.028
5	Triangular	71.102.844	91.562.760	5.925.237	7.630.230
6	Triangular	71.102.844	90.221.724	5.921.603	7.518.477
7	Beta	71.156.640	89.691.324	5.929.720	7.474.277
8	Beta	59.770.932	81.259.848	4.980.911	6.771.654
9	Beta	56.558.676	72.141.816	4.713.223	6.011.818
10	Triangular	57.821.316	75.729.528	4.818.443	6.310.794
11	Beta	59.029.380	75.810.240	4.919.115	6.317.520
12	Normal	60.598.788	75.006.132	5.049.899	6.250.511
13	Logistic	63.798.252	83.400.072	5.316.521	6.950.006
14	Beta	67.568.712	81.689.004	5.630.726	6.807.417

Elaboración: Propia

Los resultados de aplicar una simulación Montecarlo tres veces, con 10.000, 50.000, 20.000 y 30.000 corridas para YSIM 1, 2, 3 y 4 respectivamente:

Tabla 7.
Pasajeros Mensuales – Resultados Simulación YSIM

Coef.	MCO	YSIM-1	YSIM-2	YSIM-3	YSIM-4
m	892.852	892.852	887.386	887.673	887.672
b	62.611.087	2.640.442	62.650.932	62.645.435	62.645.434
r	30,45%	30,33%	30,26%	30,27%	30,27%

Elaboración: Propia

Los resultados obtenidos cumplen con la condición, donde $r = 30,45\%$.

Los resultados proyectados armando las respectivas ecuaciones MCO y YSIM son los siguientes:

Tabla 8.
Pasajeros Mensuales – Valores Pronosticados

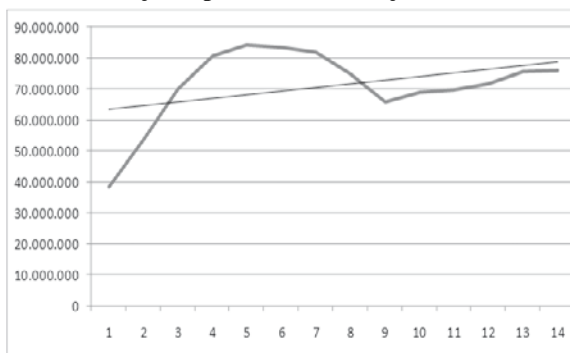
	Valor Real 2008	MCO	YSIM-1	YSIM-2	YSIM-3	YSIM-4
Valor	73.207.048	79.667.259	76.003.868	76.033.223	75.961.719	75.960.526
Diferencia	0	-6.460.211	-2.796.820	-2.826.1785	-2.754.671	-2.753.478

Elaboración: Propia

Los valores de pronóstico obtenidos y contrastados con el valor real obtenido de la gestión 2008 demuestran que se cumple con la condición $[YSIM_{t+1} - X_{t+1}] < [Y_{t+1} - X_{t+1}]$

Al igual que en el anterior caso, el valor obtenido por YSIM se acerca más a la realidad que el logrado por MCO.

Figura 2.
Curva de Ajuste por MCO - Pasajeros Mensuales



Elaboración: Propia

3. CONCLUSIONES

El cálculo de los coeficientes (m , b) obtenidos por el método *YSIM*, ocasiona una pérdida del nivel de ajuste (r); sin embargo, al

haberse absorbido parte de la entropía con la aplicación del método Montecarlo, las diferencias entre los valores reales y los de pronóstico son menores en *YSIM* que en *Y*.

Si bien el método *YSIM* se ha aplicado a modelos MCO, también podría aplicarse a diferentes modelos econométricos: MCG, ARIMA, ARMA, MANCOVA, PROBIT, LOGIT y otros.

El método planteado requiere mayores estudios teóricos, sin embargo, su potencial es prometedor.

BIBLIOGRAFÍA

- VisiRex 2.0 for Windows. (SF). What is Inductive Rule Extraction? Obtenida el 16 de mayo de 2010, de <http://cormactech.com/visirex/faq.html>.
- Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN). (2005). Método de Monte Carlo. Obtenida el 11 de febrero de 2010, de http://www.exa.unicen.edu.ar/catedras/inv_op/apuntes/Apunte_Teorico_MC_2005.pdf.
- John E. Hank & Arthur G. Reitsch, (1996). “*Pronósticos en los Negocios*”. México: Prentice Hall, 5ta Edición.