



Avances en predicción de series temporales: Comparación de métodos VAR con nuevos desarrollos no lineales

*Hugo Scolnik
Rodrigo Castro
Santiago Laplagne
Daniel Foguelman
Alejandro Danós*

**28 de Mayo de 2018
Ciudad Autónoma de Buenos Aires
Argentina**



SEMINARIO DE INVESTIGACIÓN

INSTITUTO INTERDISCIPLINARIO DE ECONOMÍA POLÍTICA DE BUENOS AIRES
Facultad de Ciencias Económicas | Universidad de Buenos Aires



Agenda

1. **Presentación y contexto**
 2. **VAR y VAR no lineal (NLVAR)**
 3. **NLVAR: Resultados prácticos en un caso de test**
 4. **Resumen de técnicas de optimización no lineal**
 5. **Plataforma de integración de modelos**
-

1. Contexto

¿Quiénes somos?

Hugo Scolnik

Doctor en Matemática,
experto en desarrollo de
modelos, co-autor del Modelo
Mundial Latinoamericano.

Rodrigo Castro

Doctor en Ingeniería,
especialista en simulación y
modelado,
experiencia en modelos globales.

Daniel Foguelman

Lic. en Ciencias en la
Computación,
estudiante doctoral y
científico de datos.

Alejandro Danós

Tesista en Ciencias en
la Computación,
experiencia en
modelos globales.

Santiago Laplagne

Doctor en Matemática.
Experto en desarrollo de
modelos matemáticos.

Instituto de Ciencias de la Computación - ICC-CONICET
Departamento de Computación y Departamento de Matemática
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Contexto

- **Objetivo a mediano plazo**
 - Hacia un repositorio integrado de modelos para simulación de distintos sectores económicos y a distintos niveles de agregación
 - **Capacidad de hacer interactuar varios modelos entre sí**
(idea de “*modelos intersectoriales*”)
 - aceptando diversas técnicas de modelado
 - compartiendo una base de datos unificada, con coherencia entre datos provenientes de distintas fuentes
 - **Algunos objetivos**
 - Testear posibles impactos de decisiones
 - Identificar cursos de acciones posibles según objetivos a alcanzar
 - Contrastar hipótesis antagónicas
 - **Mejorar técnicas de modelado existentes, proponer nuevas**
 - **En el seminario de hoy: Exploración de alternativas al supuesto de linealidad en la construcción de algunos modelos macroeconómicos**
-

2. VAR y VAR no lineal (NLVAR)

Incorporando relaciones no lineales utilizando métodos robustos de optimización no lineal

Introducción al método VAR (vector auto-regression)

	y_1 ↓ PBI	y_2 ↓ Exportaciones	y_3 ↓ Importaciones
...
2014	526	68	66
2015	585	57	60
2016	545	58	56
2017	563	58	67

Para cada variable, se ajusta una fórmula del tipo

$$\tilde{y}_{i,t} = a_{i,1}y_{1,t-1} + a_{i,2}y_{2,t-1} + \dots + a_{i,n}y_{n,t-1} + b_i$$

o en notación matricial, $\tilde{y}_t = Ay_{t-1} + b$.

Pueden usarse también modelos más complejos tomando varios valores hacia atrás (lags) en las fórmulas.

VAR vs. NLVAR

La fórmula lineal para cada variable en el método VAR

$$\tilde{y}_{i,t} = a_{i,1}y_{1,t-1} + a_{i,2}y_{2,t-1} + \cdots + a_{i,n}y_{n,t-1} + b_i$$

se reemplaza por una fórmula no-lineal

$$\tilde{y}_{i,t} = a_{i,1}f_{i,1}(y_{1,t-1}) + a_{i,2}f_{i,2}(y_{2,t-1}) + \cdots + a_{i,n}f_{i,n}(y_{n,t-1}) + b_i,$$

donde las funciones $f_{i,j}$ dependen de parámetros que se calculan por optimización.

Las funciones comúnmente utilizadas son potencias, senos o exponenciales.

VAR vs. NLVAR

Por ejemplo, para el método NLVAR con funciones *potencia*, se reemplazan las fórmulas lineales

$$\begin{cases} \tilde{y}_{1,t} = a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + \cdots + a_{1,n}y_{n,t-1} + b_1 \\ \tilde{y}_{2,t} = a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + \cdots + a_{2,n}y_{n,t-1} + b_2 \\ \vdots \\ \tilde{y}_{n,t} = a_{n,1}y_{1,t-1} + a_{n,2}y_{2,t-1} + \cdots + a_{n,n}y_{n,t-1} + b_n \end{cases}$$

por fórmulas no-lineales

$$\begin{cases} \tilde{y}_{1,t} = a_{1,1}y_{1,t-1}^{c_{1,1}} + a_{1,2}y_{2,t-1}^{c_{1,2}} + \cdots + a_{1,n}y_{n,t-1}^{c_{1,n}} + b_1 \\ \tilde{y}_{2,t} = a_{2,1}y_{1,t-1}^{c_{2,1}} + a_{2,2}y_{2,t-1}^{c_{2,2}} + \cdots + a_{2,n}y_{n,t-1}^{c_{2,n}} + b_2 \\ \vdots \\ \tilde{y}_{n,t} = a_{n,1}y_{1,t-1}^{c_{n,1}} + a_{n,2}y_{2,t-1}^{c_{n,2}} + \cdots + a_{n,n}y_{n,t-1}^{c_{n,n}} + b_n \end{cases}$$

y se calculan los exponentes óptimos utilizando rutinas de optimización.

NLVAR - Estrategia de optimización

Fijados exponentes $\{c_{i,j}\}_{1 \leq i,j \leq n}$, se pueden calcular los coeficientes $\{a_{i,j}\}_{1 \leq i,j \leq n}$ por regresión lineal (mínimos cuadrados) en los datos históricos.

Esto define para cada variable y (i fijo), una función

$$g(c_1, \dots, c_n) = \|y - \tilde{y}\|_2^2$$

donde y son los valores reales históricos e \tilde{y} son los valores obtenidos aplicando la fórmula.

El método NLVAR calcula los valores de (c_1, \dots, c_n) que minimizan la función g .

Ejemplo

Año	PBI	Exportaciones	Importaciones
2010	423	68	57
2011	530	83	74
2012	546	80	68
2013	552	76	74
2014	526	68	66
2015	585	57	60
2016	545	58	56
2017	563	58	67

Fórmula VAR para PBI (1 lag):

$$\text{PBI}_t = 0.13 \cdot \text{PBI}_{t-1} + 0.072 \cdot \text{Expo}_{t-1} - 0.76 \cdot \text{Impo}_{t-1} + 526.18$$

Fórmula NLVAR para PBI con potencias (exponentes acotados entre -2 y 2):

$$\text{PBI}_t = -17630327 \cdot (\text{PBI}_{t-1})^{-2} - 222627 \cdot (\text{Expo}_{t-1})^{-2} - 0.021 \cdot (\text{Impo}_{t-1})^2 + 752$$

Cómo evitar problemas numéricos y sobreajuste

Sin suavización, el método de mínimos cuadrados determina el vector x que minimiza la norma del error:

$$\|y - Ax\|.$$

Para evitar sobreajuste, el método de mínimos cuadrados regularizados determina x que minimiza la fórmula

$$\|y - Ax\| + \lambda \|x\|,$$

donde $\lambda \|x\|$ es un término de penalidad y $\lambda \in \mathbb{R}_{\geq 0}$ es un parámetro de suavizado.

Admite una fórmula cerrada para la solución:

$$w = (A^T A + \lambda I)^{-1} A^T y$$

Cómo evitar problemas numéricos y sobreajuste

Aplicado a NLVAR, para cada variable y , el método de mínimos cuadrados minimiza la suma de los errores

$$\sum_{t=2}^m (y_t - \tilde{y}_t)^2$$

sobre la serie histórica.

En NLVAR- λ , para valores fijos de los exponentes $\{c_j\}_{1 \leq j \leq n}$, se determinan los coeficientes $\{a_j\}_{1 \leq j \leq n}$ por mínimos cuadrados regularizados

$$(a_1, \dots, a_n) = \operatorname{argmin}_{\alpha} \sum_{t=2}^m \left(y_t - \sum_{j=1}^n \alpha_j y_{j,t-1}^{c_j} \right)^2 + \lambda \|(\alpha_1, \dots, \alpha_n)\|$$

Sin suavización vs. Con suavización

Fórmula obtenida sin parámetro de suavizado.

$$\text{PBI}_t = -17630327 \cdot (\text{PBI}_{t-1})^{-2} - 222627 \cdot (\text{Expo}_{t-1})^{-2} - 0.021 \cdot (\text{Impo}_{t-1})^2 + 752$$

Fórmula obtenida con parámetro de suavizado $\lambda = 1$.

$$\text{PBI}_t = 43.39 \cdot (\text{PBI}_{t-1})^{0.36} + 27.39 \cdot (\text{Expo}_{t-1})^{0.47} - 0.017 \cdot (\text{Impo}_{t-1})^2 + 5.25$$

Alternativas para la elección del parámetro de suavizado

1. **Parámetro fijo.** Se elige en base a la experiencia del usuario, por ejemplo $\lambda = 1$.
2. **Parámetro por optimización sobre conjunto de testeo.** Se separan los últimos años de la base de entrenamiento y se elige el parámetro que da mejores predicciones sobre este conjunto de prueba.

3. **Parámetro automático.** Toma λ igual a la raíz cuadrada del menor valor singular. Esto garantiza que la solución este alejada del subespacio de sistemas indeterminados.

The screenshot shows a dialog box titled "Parámetros NLVAR-alfa" with the following settings:

- Suavizado:** Parámetro óptimo
- Parámetro de suavizado:** 0
- Optimización del parámetro de suavizado:**
 - Parámetro de suavizado máximo: 1.0
 - Cantidad de parámetros de suavizado: 11
 - Cantidad de observaciones de testeo: 4
 - Intervalos hacia atrás a utilizar: 3
 - Puntos iniciales para optimización: 2
- Cantidad de valores hacia atrás a predecir:** 0
- Conjunto de variables:** Usar archivo de subconjuntos
- Archivo de subconjuntos:** C:\Users\Santiago\Desktop\Haiti\Haiti\bin\out\Variables.txt
- Botones:** Go! and Cancel

Mejor subconjunto de variables regresoras

Método VAR:

- ▶ Dado un conjunto de variables a predecir, se utilizan todas las variables para predecir cada una de las variables.

Método NLVAR- λ :

- ▶ Para cada variable a predecir, se utiliza el mejor conjunto de variables regresoras. Es decir, un subconjunto que tenga el mayor poder explicativo sobre la variable respuesta.

De esta forma, reducimos el sobreajuste y obtenemos fórmulas más robustas.

Mejor subconjunto de variables regresoras

Salida del módulo de selección de variables regresoras.
Criterio: R^2 -ajustado.

Regressions with 9 variable(s) (Adjusted R-squared)

Criterion	Variables
87.0	1 3 5 6 7 8 9 11 12
86.4	1 3 4 5 6 8 9 11 12
86.0	1 2 3 4 5 6 8 9 11
85.7	1 3 4 5 6 7 8 9 11

Regressions with 10 variable(s) (Adjusted R-squared)

Criterion	Variables
87.1	1 2 3 5 6 7 8 9 11 12
86.7	1 3 4 5 6 7 8 9 11 12
86.6	1 2 3 4 5 6 8 9 11 12
86.1	1 3 4 5 6 8 9 10 11 12

Regressions with 11 variable(s) (Adjusted R-squared)

Criterion	Variables
86.9	1 2 3 5 6 7 8 9 10 11 12
86.8	1 2 3 4 5 6 7 8 9 11 12
86.4	1 3 4 5 6 7 8 9 10 11 12
86.3	1 2 3 4 5 6 8 9 10 11 12
85.5	1 2 3 4 5 6 7 8 10 11 12

Regressions with 12 variable(s) (Adjusted R-squared)

Criterion	Variables
86.5	1 2 3 4 5 6 7 8 9 10 11 12

1

Best Regression with 10 variable(s) (Adjusted R-squared)

Variable	Coefficient	Standard Error	t-statistic	p-value
1	0.305	0.0861	3.538	0.0011
2	-0.115	0.1023	-1.121	0.2692
3	-0.725	0.1864	-3.889	0.0004
5	-1.231	0.2001	-6.151	0.0000
6	-0.981	0.2330	-4.212	0.0001
7	0.153	0.1182	1.295	0.2029
8	2.023	0.3612	5.601	0.0000
9	0.215	0.0917	2.349	0.0240
11	1.439	0.2517	5.717	0.0000
12	-0.490	0.2336	-2.097	0.0425

Variables endógenas y exógenas

1. Para explicar variables de la economía argentina, se utilizan variables nacionales (endógenas) e internacionales (exógenas).
2. Los modelos tipo VAR requieren predicciones de todas las variables, tanto endógenas como exógenas.
3. Para predecir las variables exógenas utilizamos modelos de pronóstico de series de tiempo que no dependen de otras variables: ajuste por splines con parámetro óptimo por predicciones parciales.

3. NLVAR

Resultados prácticos

Aplicación a datos de la
economía argentina

Aplicación a datos de la economía argentina

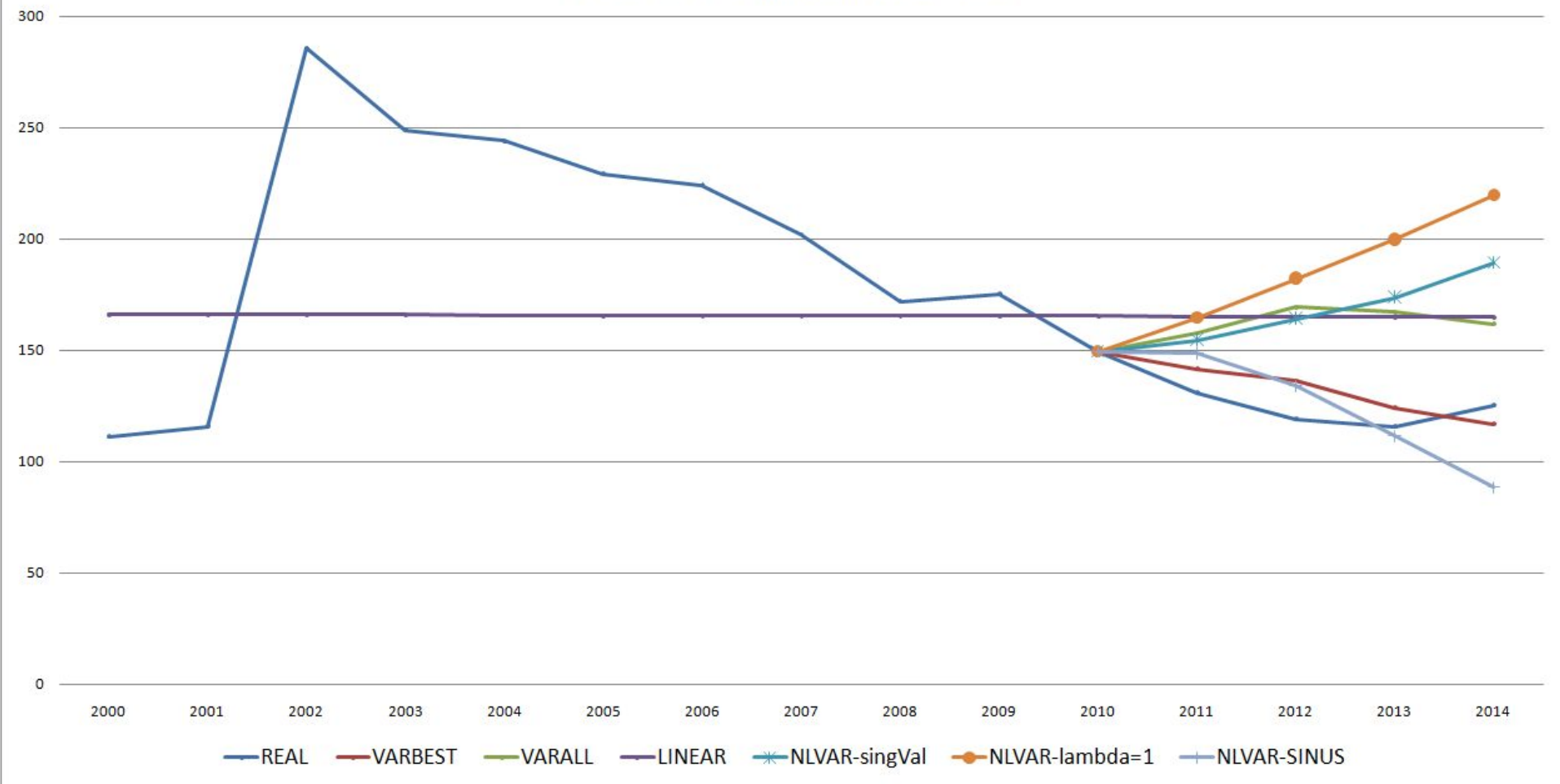
- Utilizamos datos macroeconómicos provistos por el **Instituto Interdisciplinario de Economía Política de Buenos Aires (IIEP, FCE-CONICET)**
 - Conjunto de variables de la economía argentina (endógenas) y variables mundiales (exógenas)
 - Comparamos distintos métodos de predicción y su aplicabilidad a este conjunto de datos
-

VAR Benchmark

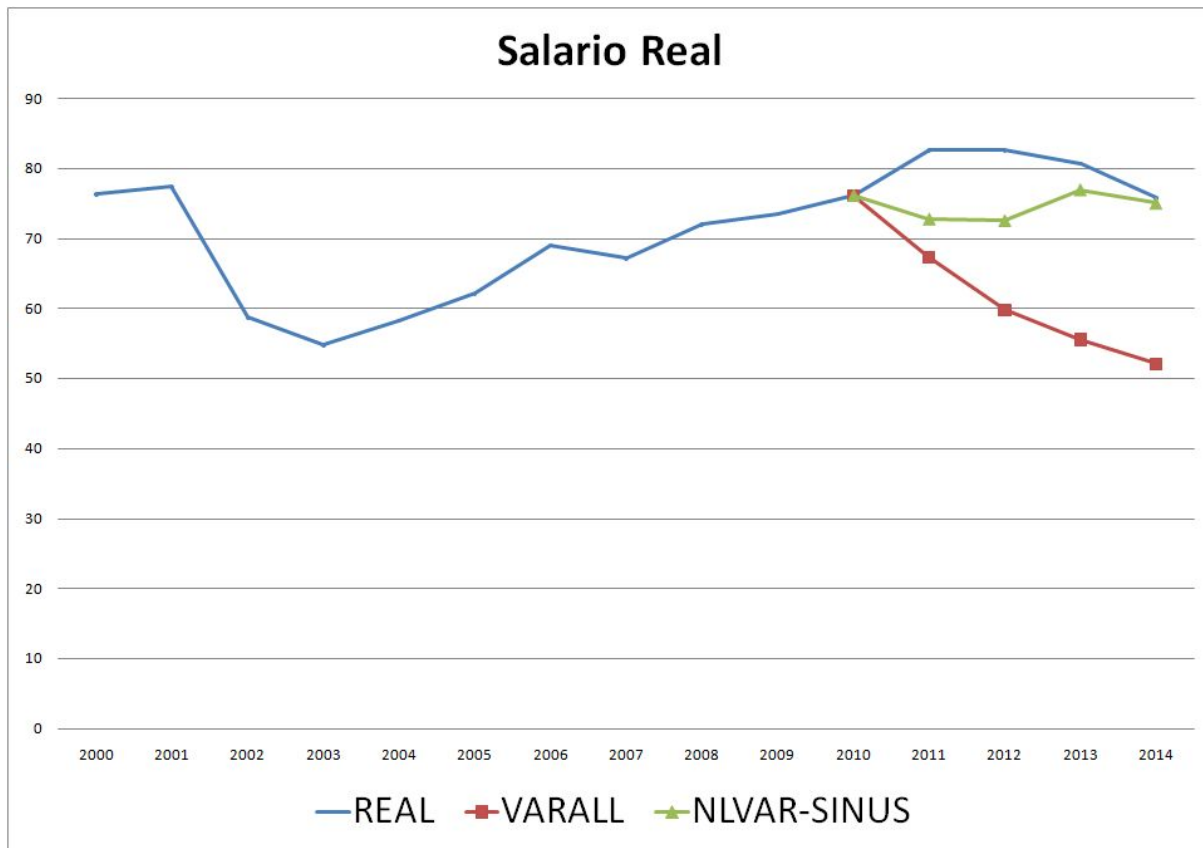
Variable	Descripción	Unidad
SAL_REAL	Salario real Mensual en poder adquisitivo de 1993 1970 = 100	Poder adquisitivo relativo a 1993
TC_LIBRE_REAL	Tipo de cambio libre real	Relación de tipo de cambio dada en ARS\$
TC_IMPO_REAL	Tipo de cambio importación	Relación de tipo de cambio dada en ARS\$
IMPO_VOLUMEN_2004	Importación volumen pesos constantes 2004	Índice de volumen pesos 2004
EXPO_TOT_2004	Exportación total pesos constantes 2004	Índice de volumen pesos 2004
PIB_LCU_BRAZIL	PIB moneda local de Brasil	Expresado en unidad de cambio local constante
PIB_LCU_CHINA	PIB moneda local de China	Expresado en unidad de cambio local constante
TASA_ANUAL_FED	Tasa anual federal EE.UU.	Tasa de interés
Vol_PIB_encaden_Fisher_2004	Volumen PIB encadenado Fisher pesos constantes 2004	Índice de volumen Fisher pesos 2004
TI_2004	Términos de intercambio pesos constantes 2004	Sin unidad
IPC_DEQUIV_MENSUAL	IPC de equivalencia mensual	Índice
EGRE_CORR_PRIMARIO_PROP_PIB	Gastos corrientes primarios proporcional al PIB	Pesos proporcional al PIB
EGRE_CORR_TOT_PROP_PIB	Gastos corrientes totales proporcional al PIB	Pesos proporcional al PIB

Comparación de predicciones

Tasa de Cambio Importaciones_real



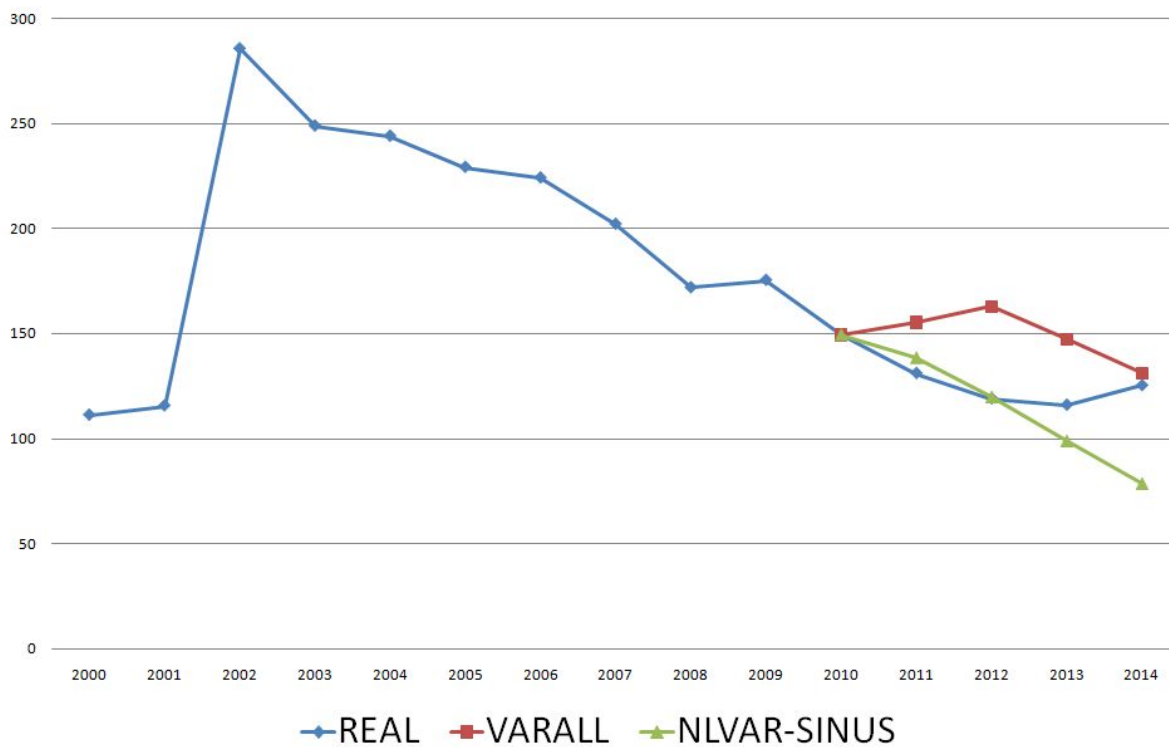
Comparación de predicciones



Errores	VAR	NLVAR-sin
2011	18.52%	11.93%
2012	27.57%	12.16%
2013	31.09%	4.70%
2014	31.23%	1.08%
Prom.	27.10%	7.47%

Comparación de predicciones

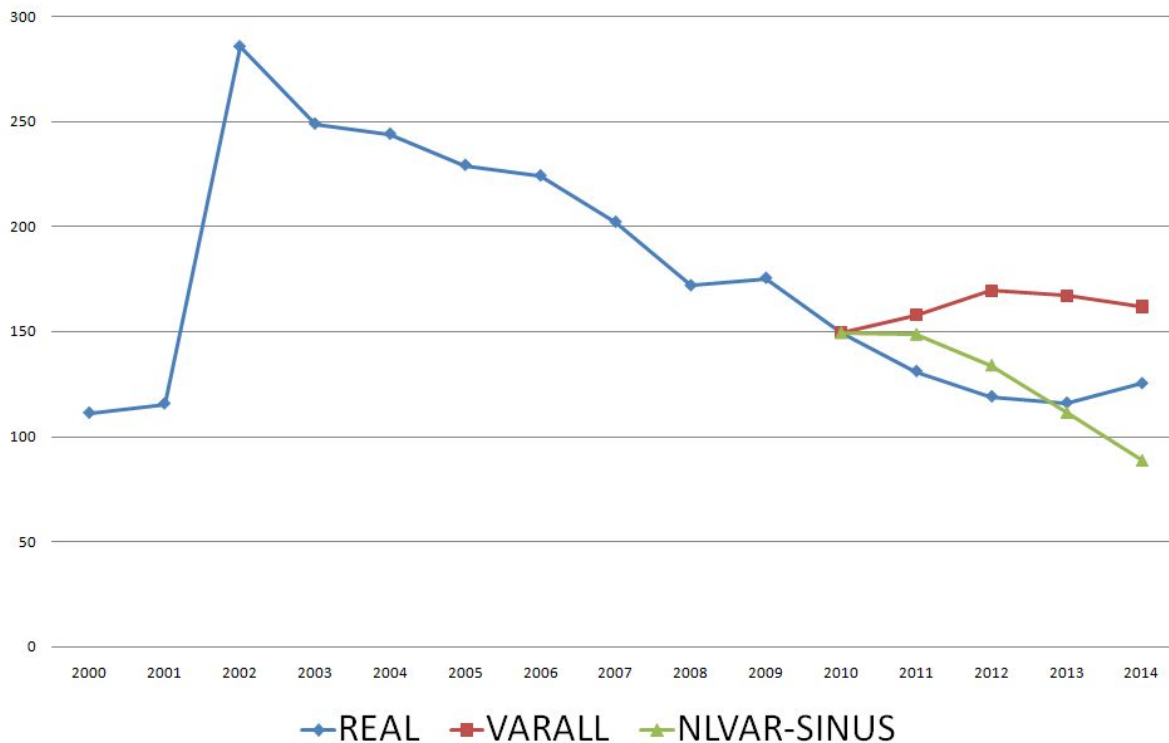
Tasa de Cambio Libre Real



Errores	VAR	NLVAR-sin
2011	18.58%	5.71%
2012	36.83%	0.75%
2013	27.21%	14.62%
2014	4.85%	37.14%
Prom.	21.87%	14.55%

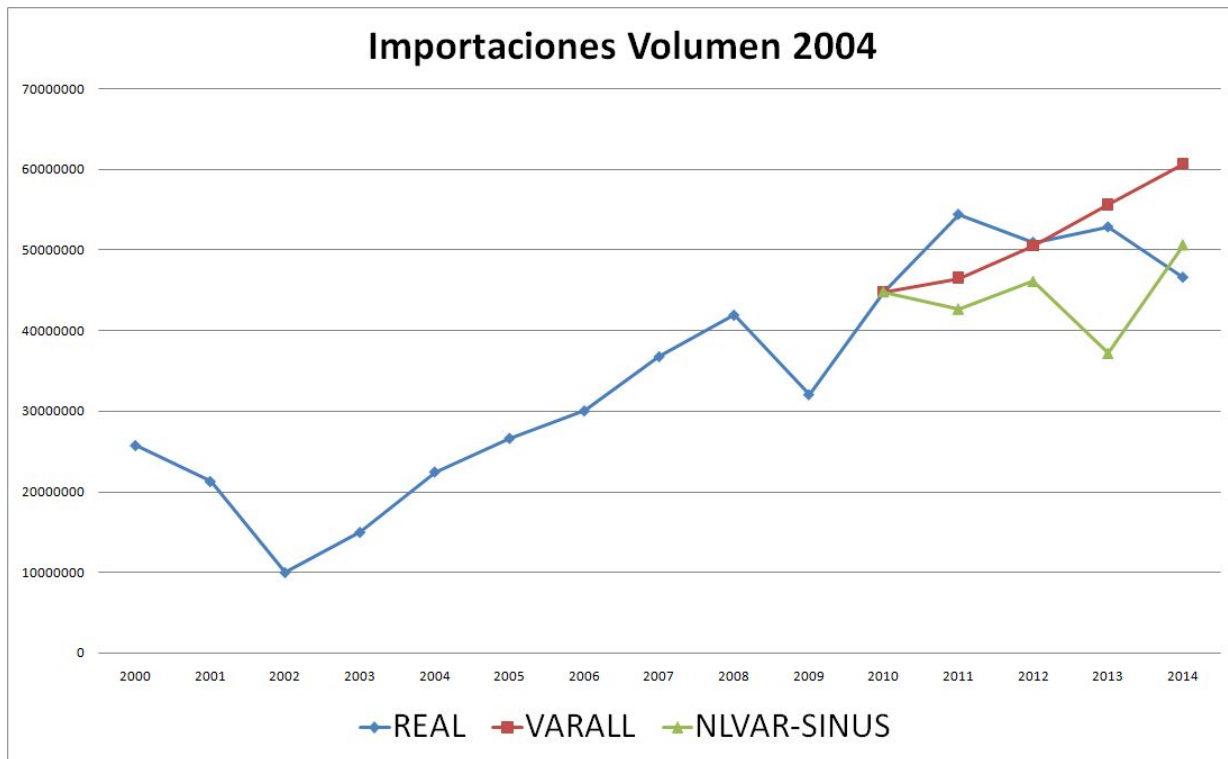
Comparación de predicciones

Tasa de Cambio Importaciones Real



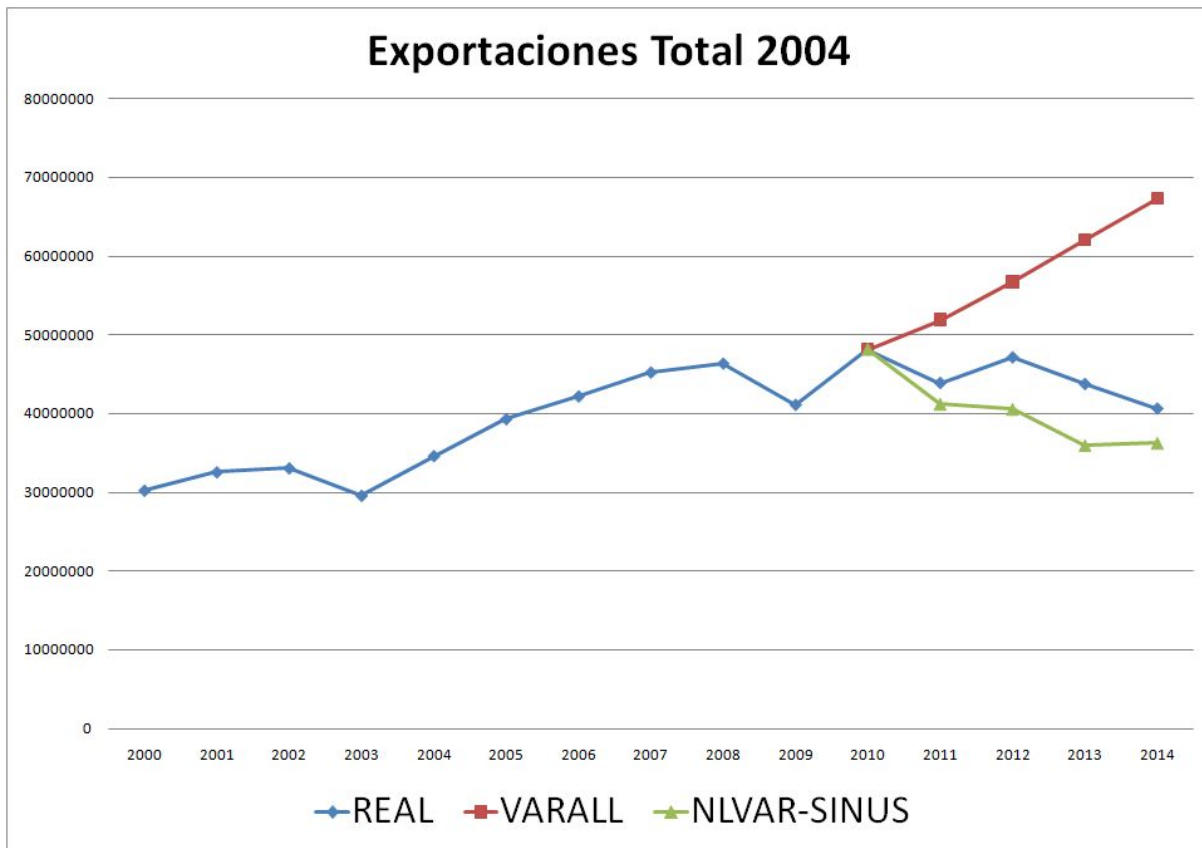
Errores	VAR	NLVAR-sin
2011	20.44%	13.50%
2012	42.37%	12.62%
2013	43.98%	3.91%
2014	29.23%	29.32%
Prom.	34.01%	14.84%

Comparación de predicciones



Errores	VAR	NLVAR-sin
2011	14.48%	21.57%
2012	0.81%	9.51%
2013	5.24%	29.66%
2014	30.27%	8.63%
Prom.	12.70%	17.34%

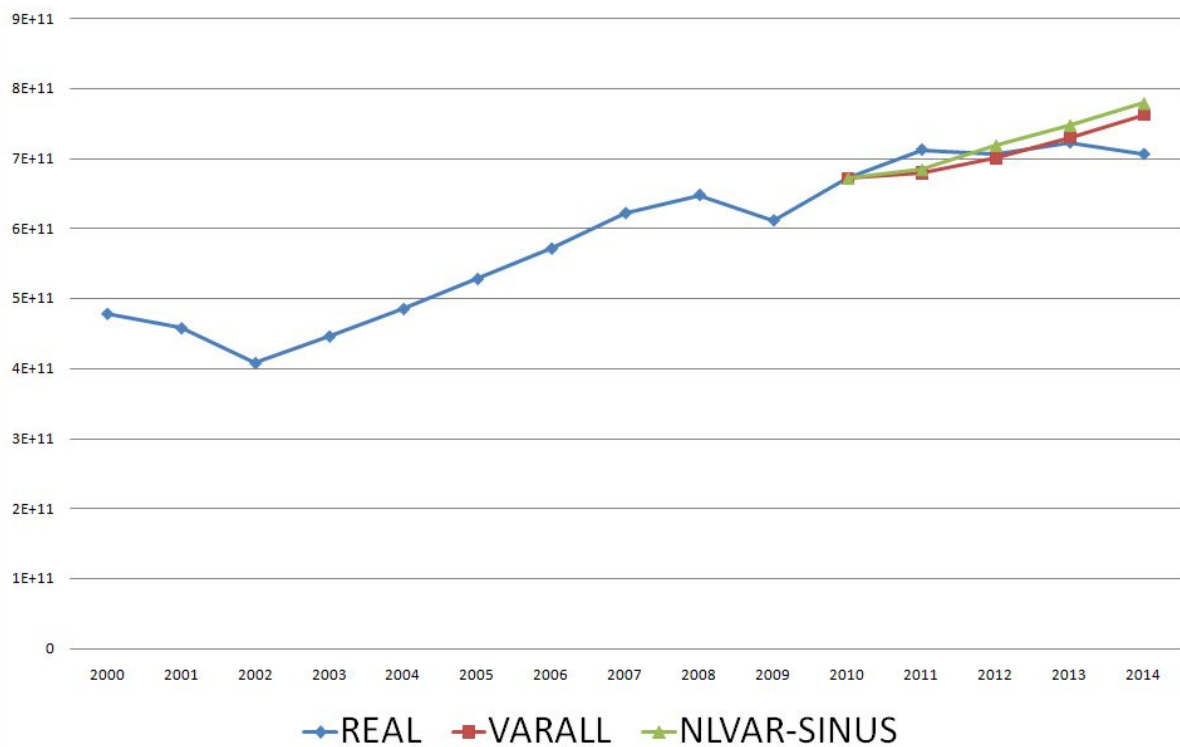
Comparación de predicciones



Errores	VAR	NLVAR-sin
2011	18.16%	6.21%
2012	20.22%	14.03%
2013	41.62%	17.92%
2014	65.44%	10.89%
Prom.	36.36%	12.26%

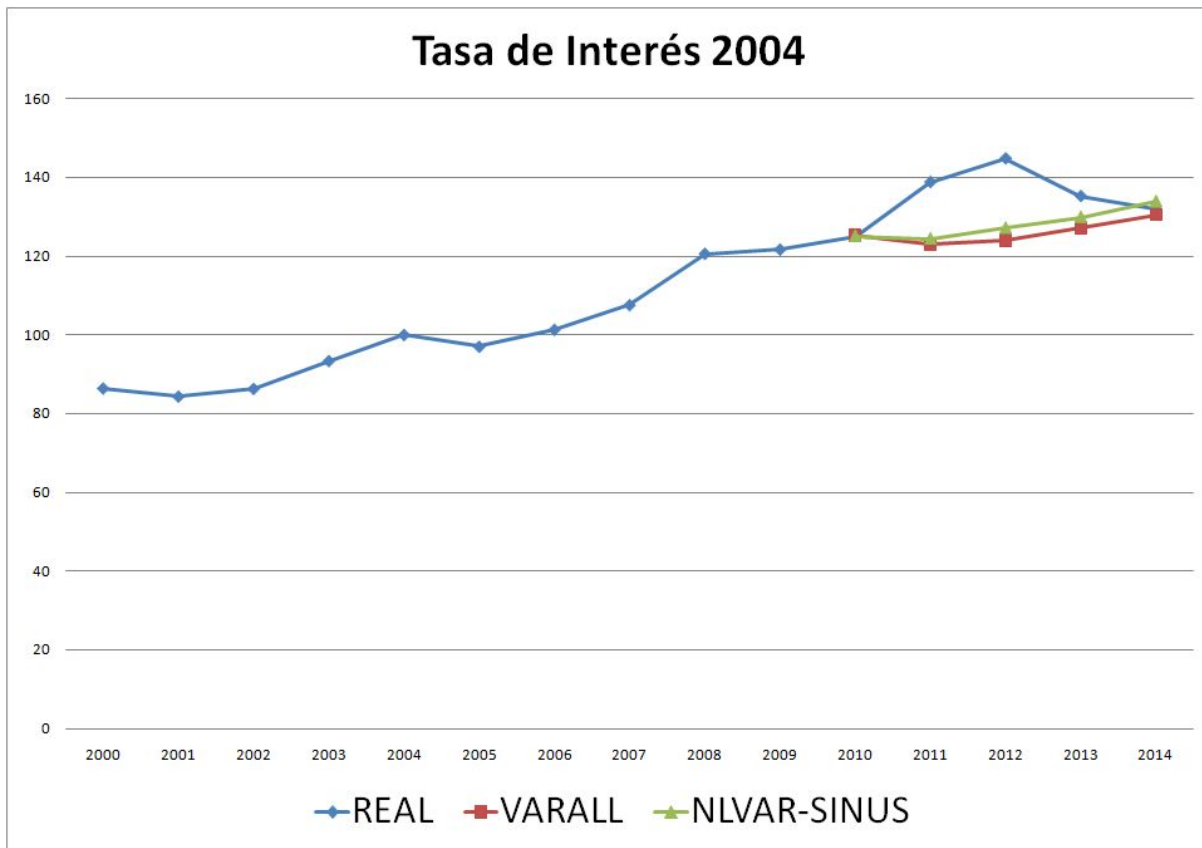
Comparación de predicciones

Volumen PBI Encadenado Fischer 2004



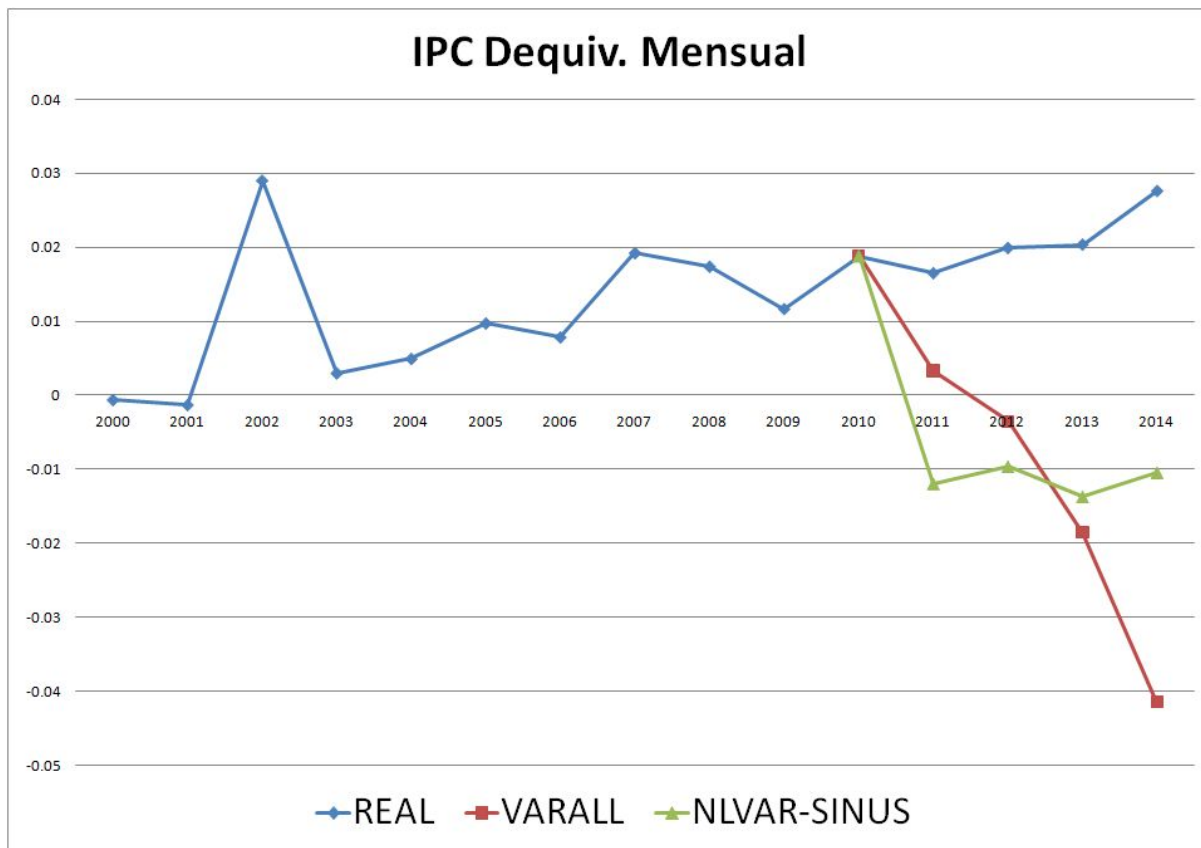
Errores	VAR	NLVAR-sin
2011	4.62%	3.93%
2012	0.93%	1.71%
2013	1.06%	3.55%
2014	7.98%	10.42%
Prom.	3.65%	4.90%

Comparación de predicciones



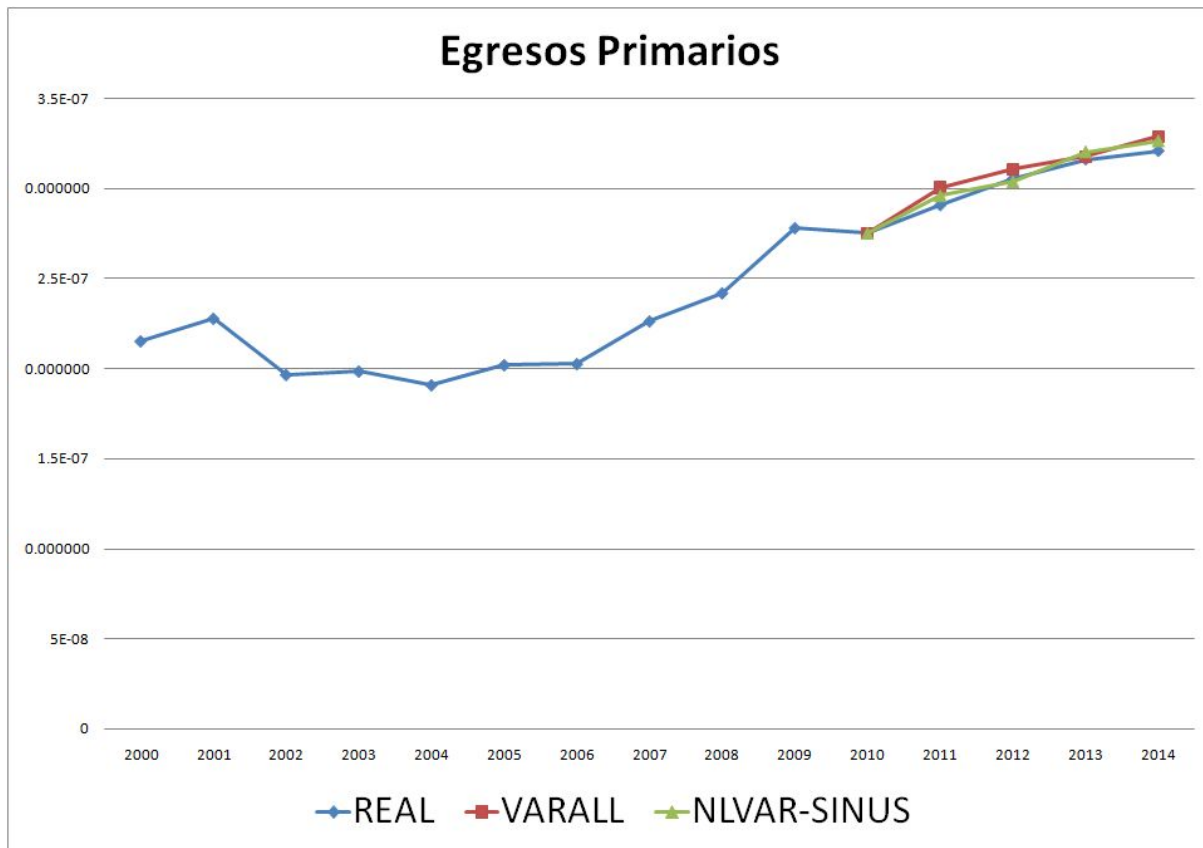
Errores	VAR	NLVAR-sin
2011	11.45%	10.35%
2012	14.38%	12.15%
2013	6.07%	3.92%
2014	1.18%	1.55%
Prom.	8.27%	6.99%

Comparación de predicciones



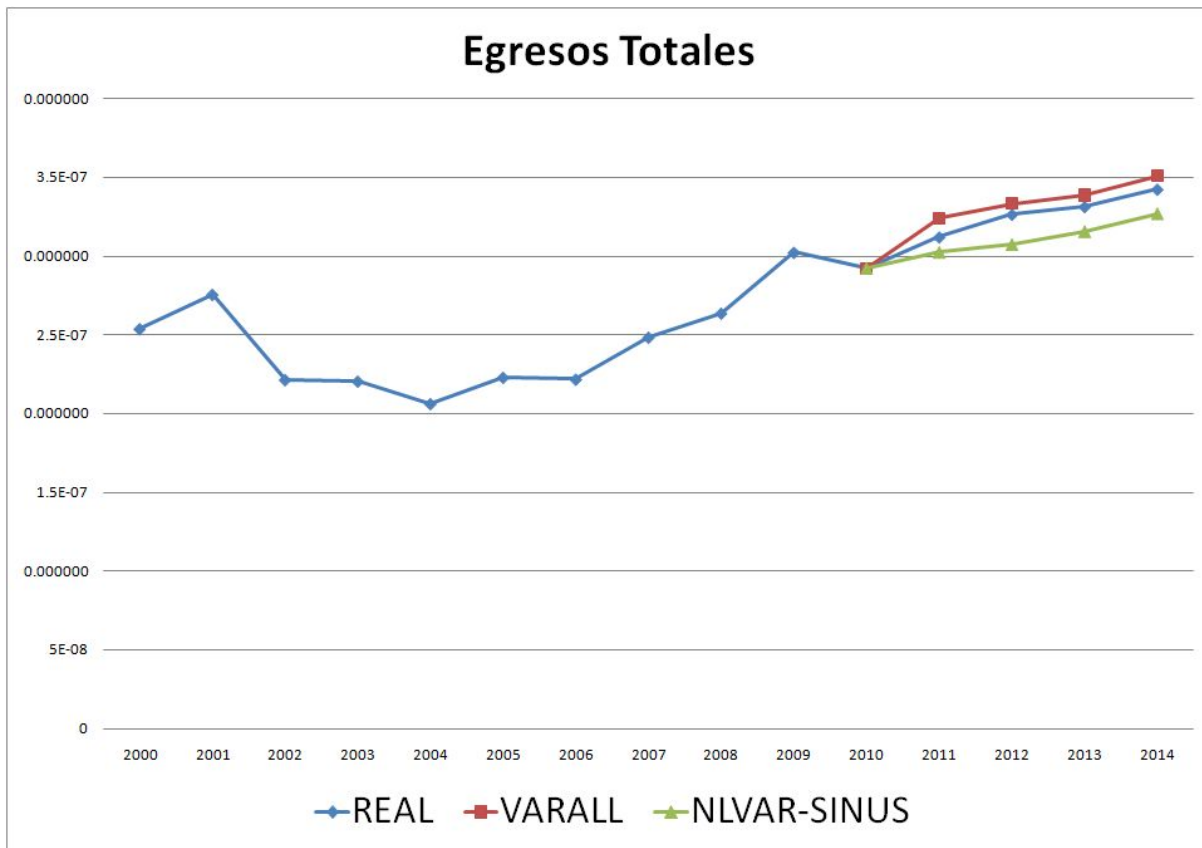
Errores	VAR	NLVAR-sin
2011	79.84%	172.33%
2012	117.74%	148.25%
2013	191.14%	167.23%
2014	250.07%	137.77%
Prom.	159.70%	156.40%

Comparación de predicciones



Errores	VAR	NLVAR-sin
2011	3.37%	1.83%
2012	1.73%	0.50%
2013	0.48%	1.33%
2014	2.47%	1.78%
Prom.	2.01%	1.36%

Comparación de predicciones



Errores	VAR	NLVAR-sin
2011	3.87%	3.05%
2012	2.13%	5.80%
2013	2.16%	4.77%
2014	2.41%	4.54%
Prom.	2.65%	4.54%

4. Optimización no lineal

Breve reseña del estado actual
de la disciplina desde un punto
de vista práctico

$\min f(x_1, \dots, x_n)$ sujeta eventualmente a restricciones de la forma

El problema es $g_i(x_1, \dots, x_n) \geq 0 \quad i = 1, \dots, k_1$

$h_j(x_1, \dots, x_n) = 0 \quad j = 1, \dots, k_2$

Posibilidades:

- Variables continuas y funciones diferenciables
- La función objetivo tiene discontinuidades o no es diferenciable en ciertos puntos
- Hay variables enteras (eventualmente booleanas)
- La función es calculable pero no tiene una representación analítica (por ejemplo en algunos modelos dinámicos)
- Etc.

Según el caso hay distintos algoritmos.

Qué queremos ? Obtener resultados (o sea robustez) y poco tiempo de CPU

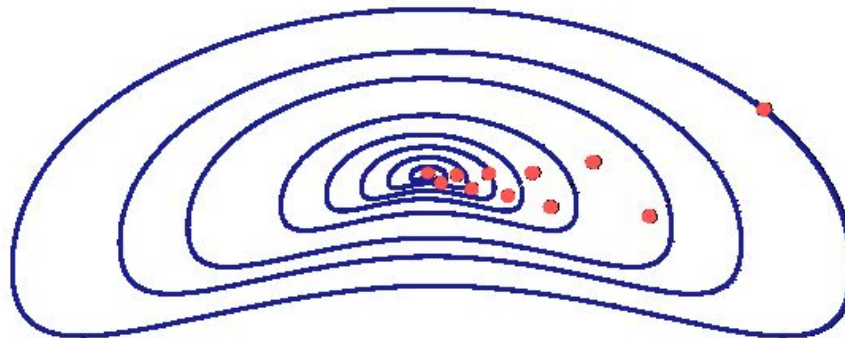
Recordemos el desarrollo en serie de Taylor que es el paradigma de la optimización no lineal por razones que mencionaremos

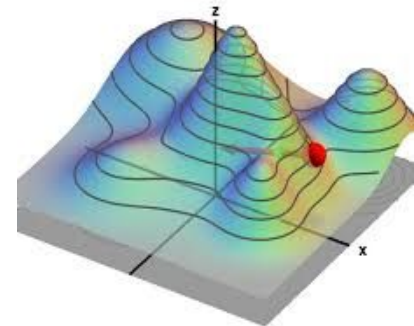
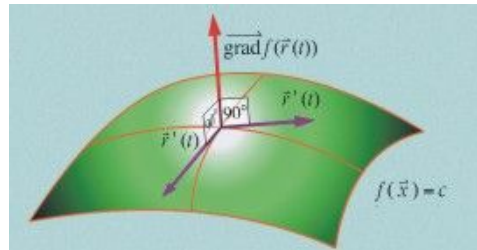
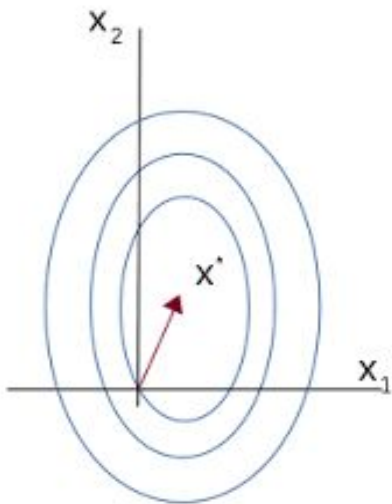
Sea $x \in \mathbb{R}^n$, h un incremento, $f(x+h) = f(x) + h\nabla f(x) + \frac{1}{2} \langle h, H(x)h \rangle + \dots$

si no consideramos los términos de segundo orden la variación máxima de la función estaría dada por el gradiente pues

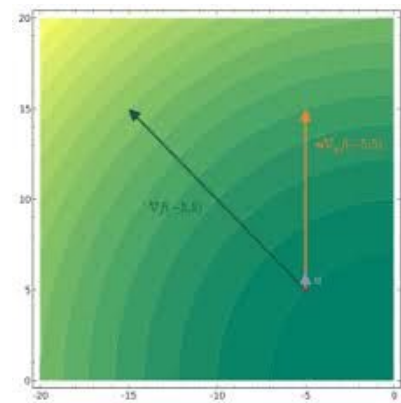
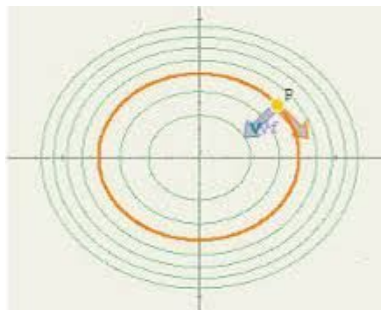
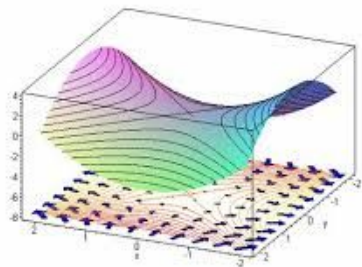
$$\frac{f(x+h) - f(x)}{h} = \nabla f(x)$$

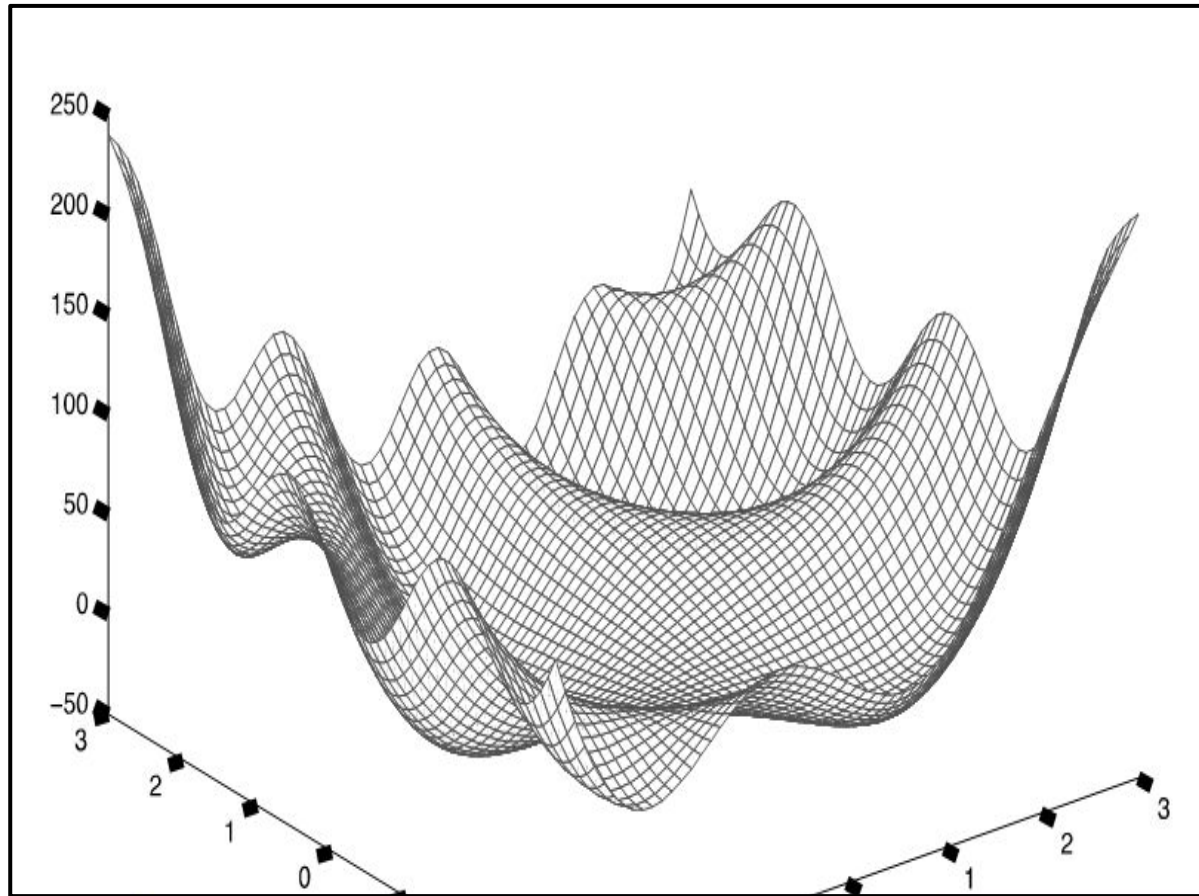
Pero el gradiente siempre es ortogonal a las curvas de nivel y entonces

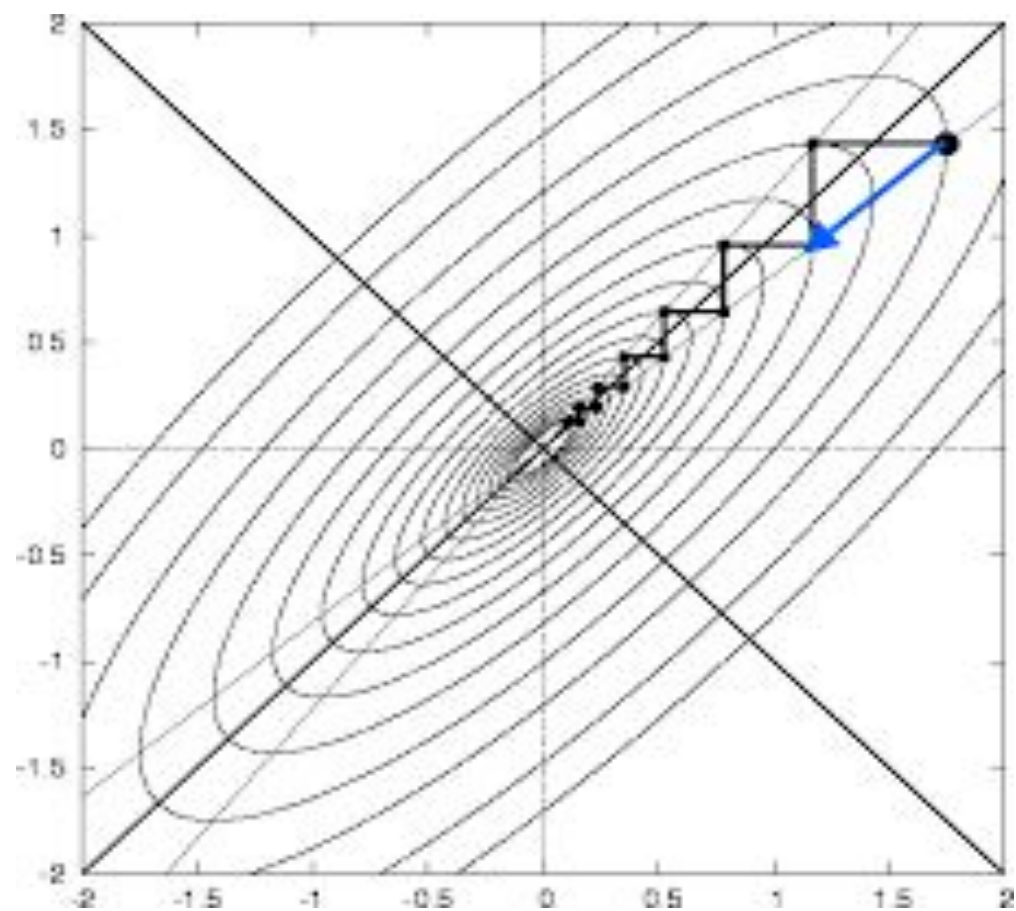




$\theta = 0$
 $u = (-0.91, -0.42)$ $u = (6.7, 1.1)$ $D_u f(u) = 2.00$
 $\|\nabla f(u)\| = 2.00$ $\|\nabla f(u)\| = 2.00$







Cuando las curvas de nivel son esferas concéntricas el método del gradiente es óptimo. Si no lo son, surgen dos posibilidades:

- 1) Los ejes de los elipsoides son los autovectores de la cuadrática aproximante (surgida por ejemplo del método de Newton) de la función a optimizar. Si partimos de ellos el método del gradiente optimiza en un paso.
- 2) Habrá un cambio de variables que consiga esferas concéntricas ? Sí, lo hay y es el método de Newton

$$x^{k+1}(\lambda) = x^k - \lambda H^{-1}(x^k) \nabla f(x^k), \quad k = 0, \dots$$

donde x^0 es el punto inicial

Toda dirección $p(x^k)$ tal que forma un ángulo agudo con $-\nabla f(x^k)$ es una dirección de descenso, o sea que existe $\lambda > 0$ tal que

$$f(x^k + \lambda p(x^k)) < f(x^k)$$

Si $H(x^k) > 0$ entonces $p(x^k) = -H^{-1}(x^k) \nabla f(x^k)$ es de descenso

¿Alcanza para ser útil? No necesariamente.

Desde la publicación del trabajo seminal

VARIABLE METRIC METHOD FOR MINIMIZATION by W. C. Davidon (Argonne National Laboratory, 1959)

se generaron muchos algoritmos que evitaban el cálculo de las derivadas segundas pues se las consideraba una fuente de errores, etc. El primer algoritmo práctico de los métodos de métrica variable o Quasi-Newton fue

R. FLETCHER AND M. J. D. POWELL, A rapidly convergent descent method for minimization, *Comput. J.*, 6 (1963), pp. 163-168.

Luego se llegó a la fórmula BFGS (Broyden-Fletcher-Goldfarb-Shanno, 1970) que es la más robusta de las Quasi Newton porque es la que tiene autovalores más grandes dentro de una familia de infinitas fórmulas (ver por ejemplo Nocedal, Jorge; Wright, Stephen J. (2006), *Numerical Optimization* (2nd ed.), Berlin, New York: [Springer-Verlag](https://www.springer.com))

Hoy en día, y debido a la existencia de la diferenciación simbólica, es preferible usar los métodos con derivadas segundas con trust regions.

La estructura en común de una función y sus derivadas permite optimizar la eficiencia computacional (Masao Iri, José Mario Martínez)

Opciones:

- * Tradicional: minimizar $f(x^{k+1}(\lambda))$ (costoso, puede que no exista el mínimo)
- * Buscar satisfacer condiciones más débiles que aseguran convergencia global (por ejemplo las condiciones de Wolfe)

Lo mejor es la rutina GSRCH desarrollada por M.J.D.Powell

Cuando $H(x^k)$ no es definido positivo una posibilidad es usar Levenberg-Marquardt

$H(x^k) + \mu I$ pues si s_1, \dots, s_n son los autovalores de $H(x^k)$, los de $H(x^k) + \mu I$ son $s_1 + \mu, \dots, s_n + \mu$ (con μ mayor que el valor absoluto del menor autovalor todos los $s_i + \mu$ son positivos)

El problema es que si $\mu \rightarrow \infty$ entonces $p(x^k) \rightarrow -\nabla f(x^k)$

Al variar μ cambia el tamaño de la región y eso da lugar a la teoría de trust regions

Ver por ejemplo:

Classics in Applied Mathematics

Numerical Methods for Unconstrained Optimization and Nonlinear Equations,

[J. E. Dennis, Jr.](#) Rice University, Houston, Texas

[Robert B. Schnabel](#) University of Colorado, Boulder, Colorado

Chapter 6, SIAM, 1996

El problema de encontrar el parámetro óptimo μ lo resolvimos en el paper:

A Curvilinear Search Using Tridiagonal Secant Updates for Unconstrained Optimization

J. E. Dennis, Jr., N. Echebest, M. T. Guardarucci, J. M. Martínez, H. D. Scolnik, and C. Vacchino

SIAM Journal on Optimization 1991, Vol. 1, No. 3, pp. 333-357

Implementado en el software CURVI que también maneja restricciones intervalos de las variables :

CURVIF sólo usa valores de la función

CURVIG usa valores de la función y del gradiente

CURVIH usa valores de la función, del gradiente y del hessiano

Nota: estas rutinas pueden fallar en problemas mal escalados si el método de la minimización a lo largo de la curva no converge. En esos casos devuelven un parámetro de error y conviene recomenzar desde otro punto inicial.

5. Plataforma para integración de herramientas

Distintos métodos de modelado usados en econometría

Distintas herramientas en distintos lenguajes

Integración adoptando un único lenguaje: Python

En proceso de diseño de una interfaz web para generación y uso de modelos

Automatización de simulaciones

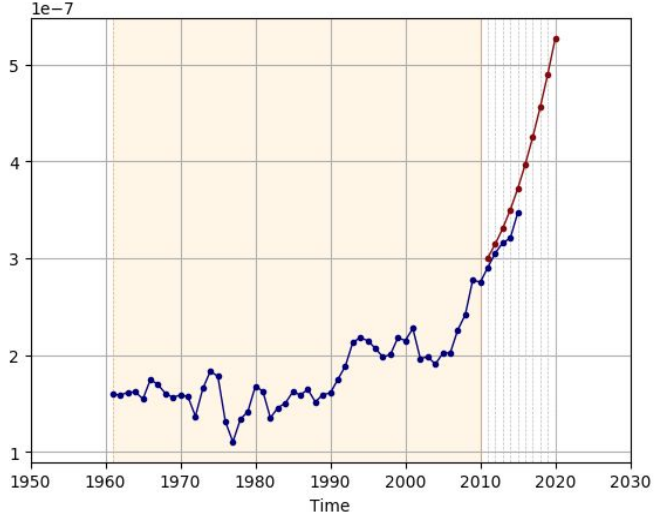
Archivo de configuración (formato JSON)

```

{
  "csv_path"      : "csvs/var_benchmark.csv",
  "normalize_data" : true,
  "years_to_predict" : 10,
  "lags_list"     : [1,2],
  "diffs_list"    : [0,1],
  "training_window" : [1991,2008],
  "models": [
    { "name": "SM_VAR"           , "args": {} } ,
    { "name": "Pyflux_VAR"      , "args": {} } ,
    { "name": "PartialPredictions" , "args": {} } ,
    { "name": "custom_model" ,
      "args": {
        "model_per_case": [ {
          "lag"      : 1,
          "diff"     : 0,
          "formulas": {
            "C.0": "lambda df: 2*df['C.0'].iloc[-1]",
            "C.1": "lambda df: 3*df['C.0'].iloc[-1]",
            "C.2": "lambda df: 4*df['C.0'].iloc[-1]",
            "C.3": "lambda df: 5*df['C.0'].iloc[-1]"
          }
        }
      ]
    }
  ]
}

```

Gráficos



Análisis

Method	lag	aic	bic	RMS: Col.1	RMS: Col.2	...
NLVAR	1	-2063.804	-1547.325	1.13E-08	1.09E-08	...
NLVAR	1	-2217.440	-1390.375	8.54E-09	7.93E-09	...
VAR	1	-1847.490	-1336.661	1.13E-08	1.09E-08	...
VAR	2	-1932.575	-1114.813	8.54E-09	7.93E-09	...
Partial Predicts	1	-1623.275	-1118.172	1.13E-08	1.09E-08	...
Partial Predicts	2	-1814.613	-1006.398	8.54E-09	7.93E-09	...

Resultados intermedios

- SM_VAR
 - normed_diff_0_lag_1
 - plots
 - residuals
 - EXPO_TOT_2004.svg
 - EGRE_CORR_TOT_PROP_PIB.svg
 - ...
 - predictions
 - ...

Graficación automática

Gráficos tan completos como se deseen:

Al momento:

- Ventana de Entrenamiento
- Ventana de Testing
- Para cada variable:
 - Proyecciones
 - Residuos
 - Ajuste

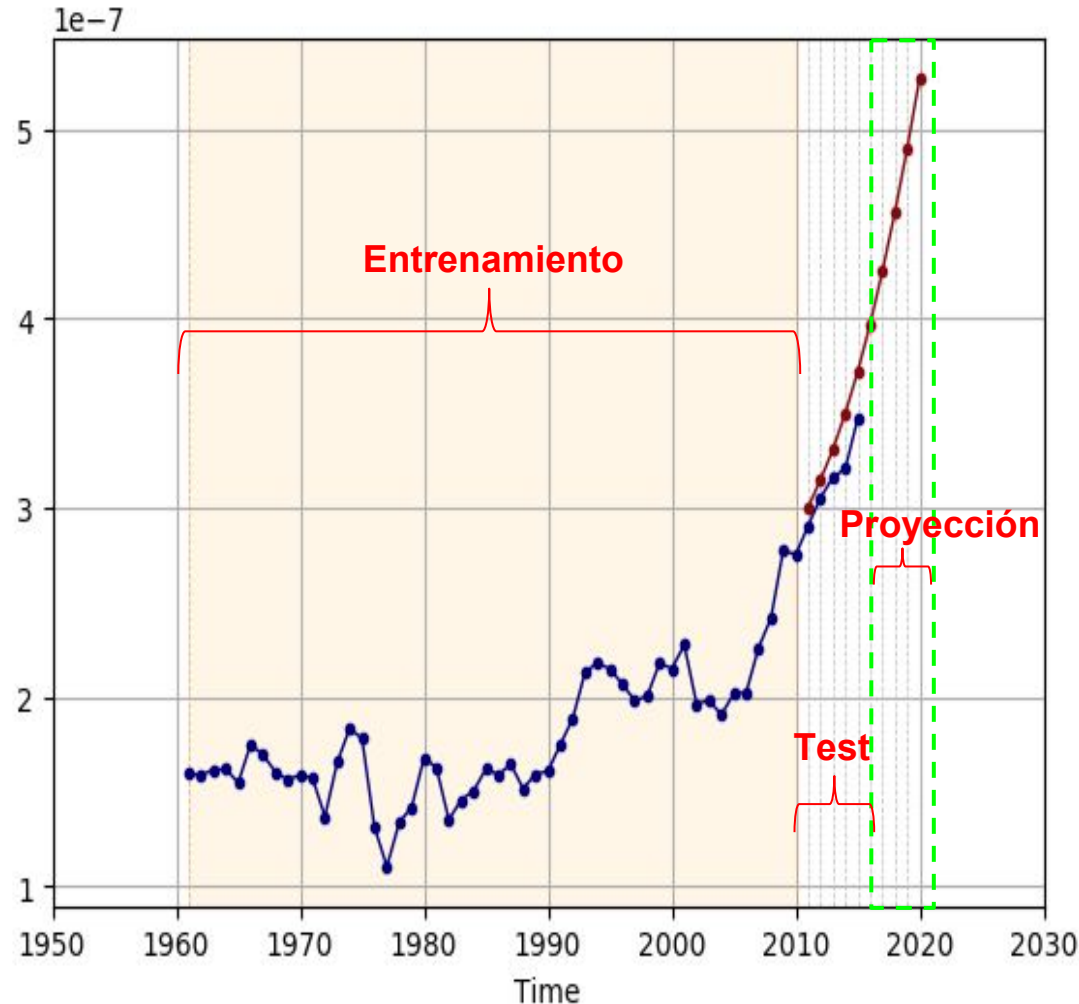
En desarrollo:

- Misma variable según distintos modelos
- Misma variable y mismo método (e.g. VAR, VECM, NLVAR, etc.) para distintas configuraciones (e.g. differences, lags, etc.)

Especificación estructurada de gráficos (mediante JSON)

- el usuario puede definir nuevos gráficos ad-hoc aplicables a simulaciones pasadas.

Variable: EGRE_CORR_PRIMARIO_PROP_PBI
Modelo: VAR, Lags 1, StatsModel (Python)



Comparación analítica automática

Posibilidad de **automatizar la elección del “mejor” modelo** bajo criterios del usuario.

Ejemplos:

- El modelo que tenga mejor AIC
- Ídem pero BIC
- El modelo que minimice el Error Cuadrático Medio (ECM) de la variable “Salario Real”
- Ídem pero además se pide que el ECM de “Exportación Total” no supere un valor deseado
- El que mejor se acerque a los valores reales en la ventana de entrenamiento, bajo algún criterio
- Etc.

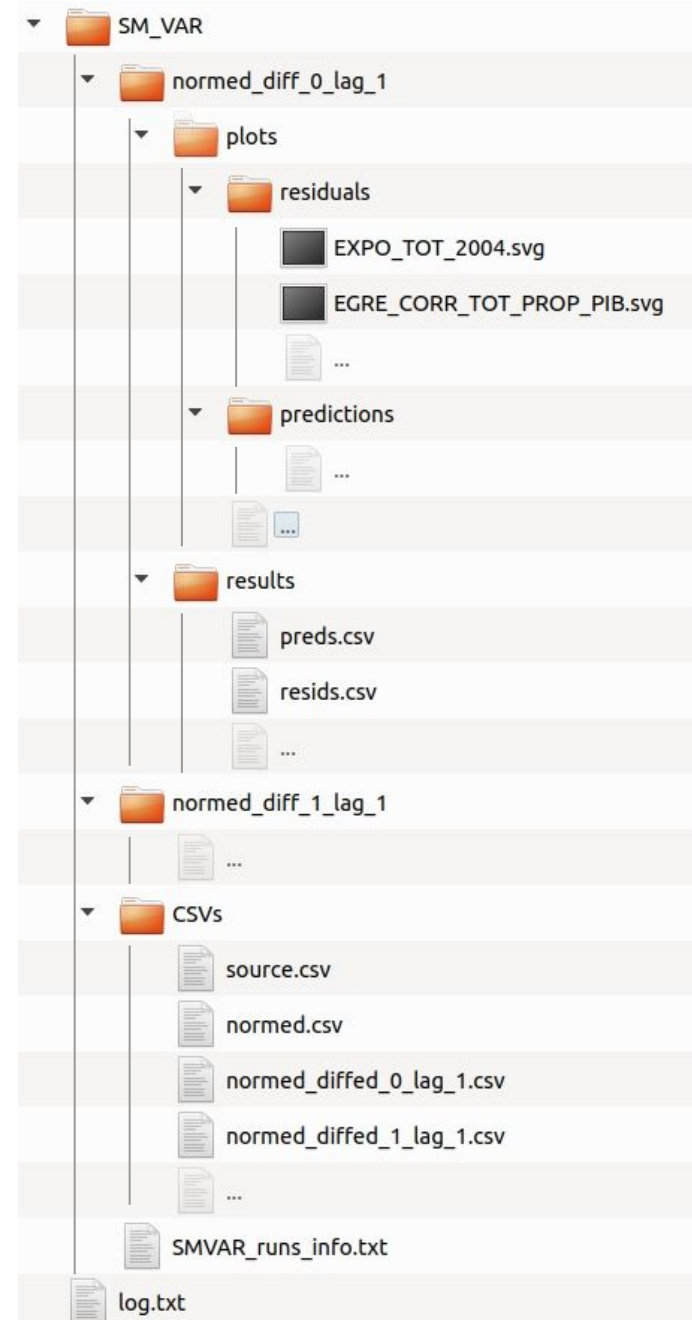
Tabla comparando modelos (datos “de juguete”)

Method	lags	AIC	BIC	RMSE: var.1	RMSE: var.2	...
NLVAR-si n	1	-2063	-1547	1.13E-8	3004	...
NLVAR-λ	1	-2217	-1390	8.54E-9	3050	...
VAR	1	-1847	-1336	1.13E-8	3100	...
VAR	2	-1932	-1114	8.54E-9	2900	...
Partial Predicts	1	-1623	-1118	1.13E-8	2989	...
Partial Predicts	2	-1814	-1006	8.54E-9	2800	...
...

Resultados intermedios automáticos

Persistencia de resultados intermedios:

- CSVs:
 - Salidas del modelo:
 - Predicciones, residuos, ajuste, etc.
 - En el espacio de los datos transformados (normalizados y diferenciados)
 - Anti-transformados (desnormalizados e integrados) al espacio original de los datos
 - Entradas de los modelos: datos originales, normalizados, para cada nivel de diferencia y lags
- Plots de todas las variables para las salidas del modelo



¡Gracias!

hugo@dc.uba.ar
rcastro@dc.uba.ar
slaplagne@dm.uba.ar
dfoguelman@dc.uba.ar
adanos@dc.uba.ar

¿Preguntas?
