

Métodos Cuantitativos

II

Modelos ARDL y VAR

Alejandro Mosiño - *Universidad de Guanajuato*

v. 2026.05.08

Modelos autorregresivos con rezagos distribuidos (ARDL)

El modelo de regresión lineal

El modelo de regresión lineal en un contexto de series de tiempo es:

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t;$$
$$\varepsilon_t \sim WN(0, \sigma^2)$$

¿Cómo usamos este modelo para hacer predicciones?

Pronóstico condicional (1/2)

Podemos utilizar el modelo de regresión lineal para hacer un pronóstico de y_t *condicional* a los supuestos sobre la variable x_t . Por ejemplo:

$$\mathbb{E}(y_{T+h}|x_{T+h})^* = \beta_1 + \beta_2 x_{T+h}^*$$

donde hemos supuesto que el valor de x_t en el periodo $T + h$ es x_{T+h}^* . Este tipo de análisis se conoce en la práctica como **análisis de escenarios** o **análisis de contingencia**.

Pronóstico condicional (2/2)

Por supuesto, si suponemos que ϵ_t es un ruido blanco gaussiano, entonces la distribución condicional del pronóstico es:

$$y_{T+h} | x_{T+h}^* \sim \mathcal{N}(y_{T+h} | x_{T+h}^*, \sigma^2)$$

Esto nos permite calcular intervalos de confianza para la predicción.

Incertidumbre en los parámetros (1/2)

Naturalmente, el pronóstico que acabamos de realizar no está exento de incertidumbre. Para ver esto, estimamos el modelo por MCO y lo evaluamos en x_{T+h}^* . En este caso, el pronóstico para el periodo $T + h$ es:

$$\hat{y}_{T+h,T} | x_{T+h}^* = b_1 + b_2 x_{T+h}^*,$$

y el residual:

$$\begin{aligned} e_{T+h,T} &= y_{T+h} - \hat{y}_{T+h,T} | x_{T+h}^* \\ &= (\beta_1 - b_1) + (\beta_2 - b_2) x_{T+h}^* + \varepsilon_{T+h} \end{aligned}$$

Incertidumbre en los parámetros (2/2)

Si calculamos la varianza del residual:

$$\mathbb{V}ar(e_{T+h,T}) = \mathbb{V}ar(b_1) + x_{T+h}^{*2} \mathbb{V}ar(b_2) + 2x_{T+h} \mathbb{C}ov(b_1, b_2) + \sigma^2$$

donde los primeros dos términos reflejan la incertidumbre en los parámetros del modelo. Entonces, para construir un intervalo de confianza para el pronóstico que refleje esta incertidumbre:

$$\hat{y}_{T+h} | x_{T+h}^* \sim \mathcal{N}(\hat{y}_{T+h} | x_{T+h}^*, \mathbb{V}ar(b_1) + x_{T+h}^{*2} \mathbb{V}ar(b_2) + 2x_{T+h} \mathbb{C}ov(b_1, b_2) + \sigma^2)$$

Pronóstico no condicional (1/2)

En ocasiones deseamos encontrar el mejor pronóstico posible para y_t . Esto es, un **pronóstico no condicional**. Para hacer esto, es necesario insertar un pronóstico óptimo para x_t , $x_{T+h,T}$, en la ecuación.

$$y_{T+h,T} = \beta_1 + \beta_2 x_{T+h,T}$$

Naturalmente, no siempre contamos con este valor futuro de x_t (tenemos un modelo para y_t , no para x_t).

Pronóstico no condicional (2/2)

Una solución consiste en estimar y_t usando como regresores los valores pasados de x_t : $x_{t-h}, x_{t-h-1}, x_{t-h-2}, \dots$. Por ejemplo, si deseamos pronosticar un periodo hacia adelante, podemos usar el modelo:

$$y_t = \beta_1 + \beta_2 x_{t-1} + \varepsilon_t$$

Por supuesto, es posible incluir más retardos de x_t . También, es importante resaltar que si el objetivo es pronosticar y_t h periodos hacia adelante, las variables x_t deben retardarse al menos h periodos.

Rezagos distribuidos (1/2)

El modelo anterior puede generalizarse como (nota el cambio de notación):

$$y_t = \phi_0 + \sum_{i=1}^{N_x} \gamma_i x_{t-i} + \varepsilon_t$$

En este caso decimos que y_t depende de los **rezagos distribuidos** de x_t . Pero, ¿cuánto vale N_x ?

Nota. ¿Por qué este modelo tendrá este nombre?

Rezagos distribuidos (2/2)

Para elegir el valor de N_x tenemos varias opciones:

- Usar pruebas de significancia estadística.
- Usar los criterios de información (AIC y BIC).
- Revisar el comportamiento de los residuales.

Los modelos con rezagos distribuidos (si las series son estacionarias) pueden estimarse por MCO o por MCG.

Modelo autorregresivo con rezagos distribuidos

Los modelos con rezagos distribuidos pueden generalizarse aun más:

$$y_t = \phi_0 + \sum_{i=1}^{N_y} \phi_i y_{t-i} + \sum_{j=1}^{N_x} \gamma_j x_{t-j} + \varepsilon_t$$

En este modelo hemos agregado retardos de la variable dependiente. Este es un **modelo AR con rezagos distribuidos (ARDL) de orden N_y, N_x** . Este puede estimarse (si las series son estacionarias)) por Máxima Verosimilitud, MCO o MCG.

Ejemplo (1/4)

Consideremos el siguiente modelo:

$$u_t = \phi_0 + \phi_1 u_{t-1} + \phi_2 u_{t-2} + \gamma_1 g_{t-1} + \varepsilon_t$$

Donde:

- u_t es la tasa de desempleo en tiempo t ,
- g_t es la tasa de crecimiento del PIB en tiempo t .

Ejemplo (2/4)

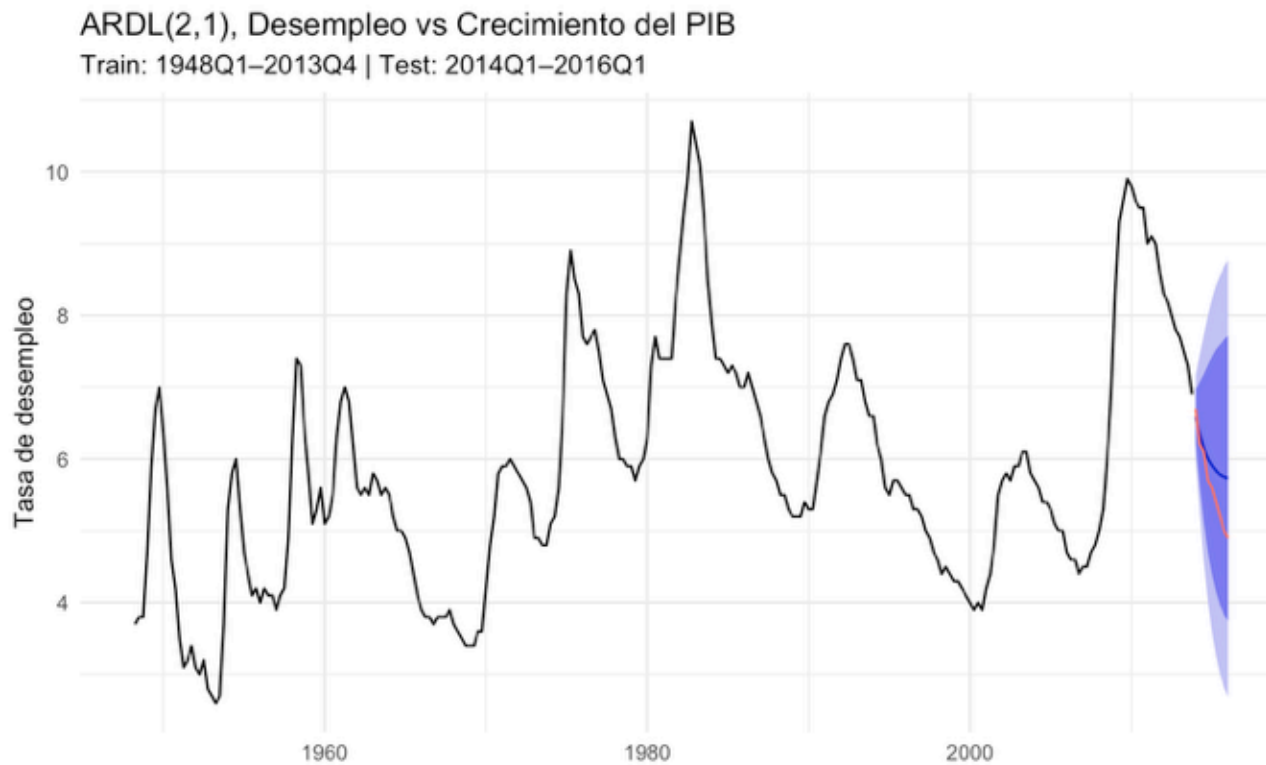
Usando datos de los Estados Unidos tenemos que:

Modelo ARDL(2,1)

term	estimate	std.error	statistic	p.value
(Intercept)	0.3615676	0.0722796	5.002348	0.000001
lag(u, 1)	1.5331023	0.0555538	27.596726	0.000000
lag(u, 2)	-0.5817890	0.0555935	-10.465055	0.000000
lag(g, 1)	-0.0482379	0.0194888	-2.475156	0.013939

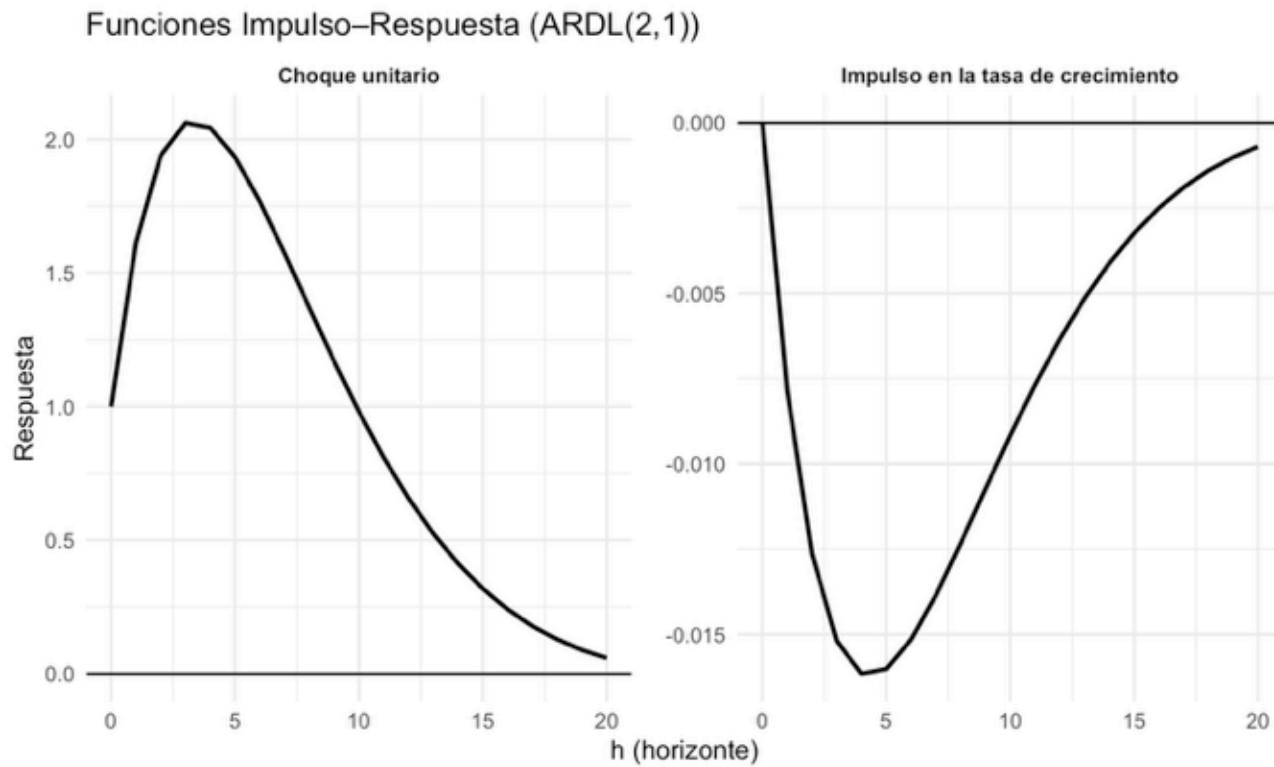
Ejemplo (3/4)

Podemos hacer pronósticos usando modelos autorregresivos con rezagos distribuidos usando las mismas técnicas aprendidas antes (ver modelos ARMA).



Ejemplo (4/4)

La representación ARMA del modelo ARDL nos permite calcular su *función de impulso-respuesta*. En el ejemplo tenemos:



Vectores Autorregresivos (VAR)

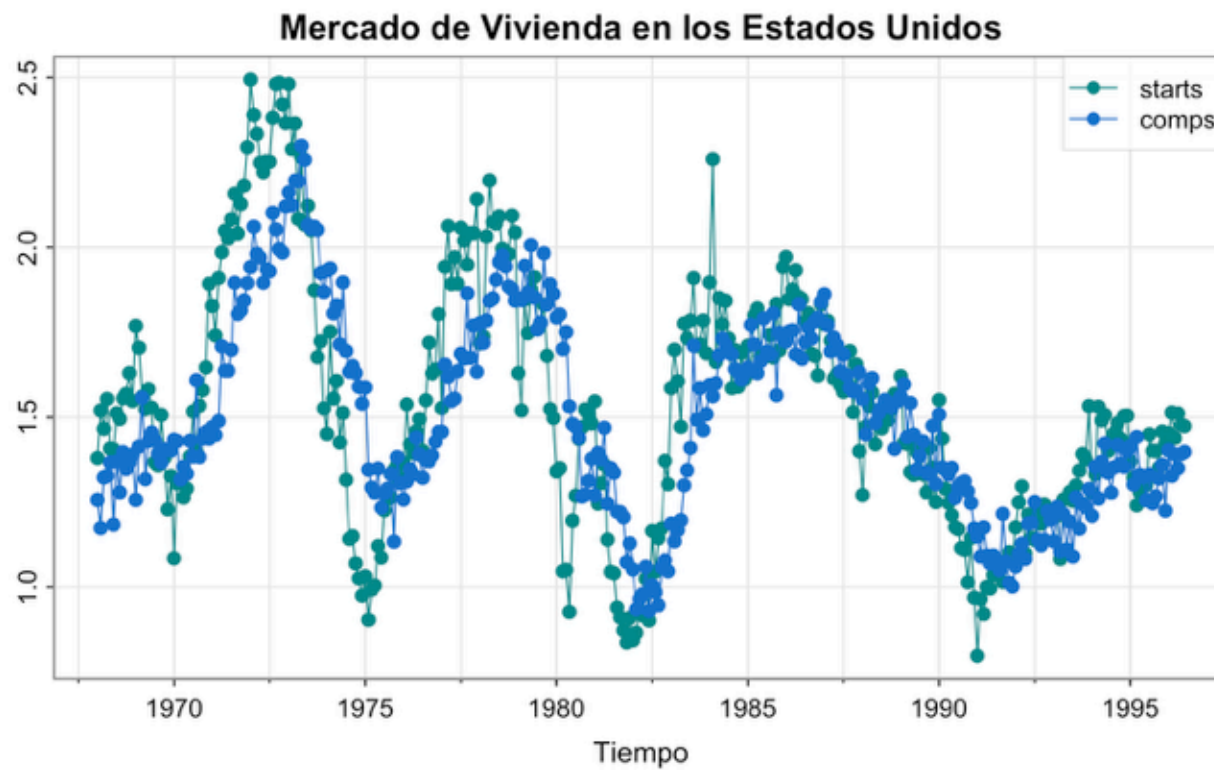
¿Qué son los vectores autorregresivos?

Los **modelos con vectores autorregresivos (VAR)** son una generalización de los *modelos autorregresivos uniecuacionales (AR)* y *de rezagos distribuidos (DL)*. Estos sirven para capturar la evolución y las interdependencias entre múltiples series de tiempo.

Nota. A diferencia de los modelos ARDL, en un VAR todas las variables del sistema son tratadas como endógenas.

Vectores autorregresivos: ejemplo

Como ejemplo, consideramos la relación entre el inicio y el fin de construcción de nuevas viviendas en los Estados Unidos:



Estructura de los VAR

Un VAR involucra k variables. En un vector autorregresivo de orden p con k variables, VAR(p), tenemos que estimar k diferentes ecuaciones. En cada ecuación tenemos como regresores p rezagos de la variable dependiente y p rezagos de todas las demás variables del sistema.

Por ejemplo, para un VAR(1) con dos variables, $y_{1,t}$ y $y_{2,t}$, tenemos:

$$y_{1,t} = \phi_{1,1}^1 y_{1,t-1} + \phi_{2,1}^1 y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = \phi_{1,1}^2 y_{1,t-1} + \phi_{2,1}^2 y_{2,t-1} + \varepsilon_{2,t}$$

$$\varepsilon_{1,t} \sim WN(0, \sigma_1^2)$$

$$\varepsilon_{2,t} \sim WN(0, \sigma_2^2)$$

y:

$$\text{Cov}(\varepsilon_{1,t}, \varepsilon_{2,\tau}) = \begin{cases} \sigma_{12}; & t = \tau \\ 0; & t \neq \tau \end{cases}$$

Estimación de un VAR(p)

Es muy fácil estimar un vector autorregresivo. Simplemente, estimamos N ecuaciones lineales utilizando el método de los MCO. Para elegir el orden del VAR, p , utilizamos los criterios de Akaike y Schwartz.

Nota. Los modelos VAR se construyen, en principio, para variables estacionarias.

- Si las variables son estacionarias, el VAR puede estimarse directamente.
- Si las variables presentan raíz unitaria, usualmente se transforman mediante diferencias.
- Si las variables son no estacionarias *pero cointegradas*, el modelo apropiado es un VECM.

Causalidad de Granger (1/3)

La **causalidad de Granger** evalúa si los valores pasados de una variable y_{2t} ayudan a predecir a otra variable y_{1t} , más allá de la información contenida en los rezagos de y_{1t} misma.

Formalmente, decimos que y_{2t} *no causa a y_{1t} en el sentido de Granger si*:

$$\mathbb{E}(y_{1t} \mid \Omega_{t-1}) = \mathbb{E}(y_{1t} \mid \Omega_{t-1} \setminus y_{2,t-1}, y_{2,t-2}, \dots),$$

donde Ω_{t-1} es el conjunto de toda la información disponible hasta $t - 1$.

Causalidad de Granger (2/3)

Para un VAR(p) bivariado tenemos:

$$y_{1t} = c_1 + \sum_{i=1}^p \phi_{11,i} y_{1,t-i} + \sum_{i=1}^p \phi_{12,i} y_{2,t-i} + \varepsilon_{1t},$$

$$y_{2t} = c_2 + \sum_{i=1}^p \phi_{21,i} y_{1,t-i} + \sum_{i=1}^p \phi_{22,i} y_{2,t-i} + \varepsilon_{2t}.$$

Entonces:

- y_{2t} no causa a y_{1t} en el sentido de Granger *si y solo si* $\phi_{12,1} = \dots = \phi_{12,p} = 0$.
- y_{1t} no causa a y_{2t} en el sentido de Granger *si y solo si* $\phi_{21,1} = \dots = \phi_{21,p} = 0$.

Causalidad de Granger (3/3)

Podemos verificar la causalidad de Granger empíricamente. Para esto, planteamos:

$$H_0 : \phi_{ij,1} = \phi_{ij,2} = \dots = \phi_{ij,p} = 0, \quad i \neq j.$$

Usamos la prueba F:

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(T - k)},$$

donde:

- SSR_R : suma de residuos al cuadrado del modelo restringido (sin rezagos de y_2)
- SSR_{UR} : suma de residuos del modelo no restringido
- $q = p$: número de restricciones.

Vectores autorregresivos: Ejemplo

Estimamos un VAR(4) para el inicio y el fin de construcción de nuevas viviendas en los Estados Unidos.

```
##
## Call:
## lm(formula = y ~ -1 + ., data = datamat)
##
## Coefficients:
## starts.l1    comps.l1    starts.l2    comps.l2    starts.l3    comps.l3    starts.l4
##  0.655743    0.007276    0.235371   -0.118989    0.138673   -0.021141    0.011877
##  comps.l4      const
##  0.002711    0.131220
```

```
##
## Call:
## lm(formula = y ~ -1 + ., data = datamat)
##
## Coefficients:
## starts.l1    comps.l1    starts.l2    comps.l2    starts.l3    comps.l3    starts.l4
##  0.06802    0.20486    0.05966    0.21775    0.03358    0.10504    0.09309
##  comps.l4      const
##  0.18231    0.04524
```

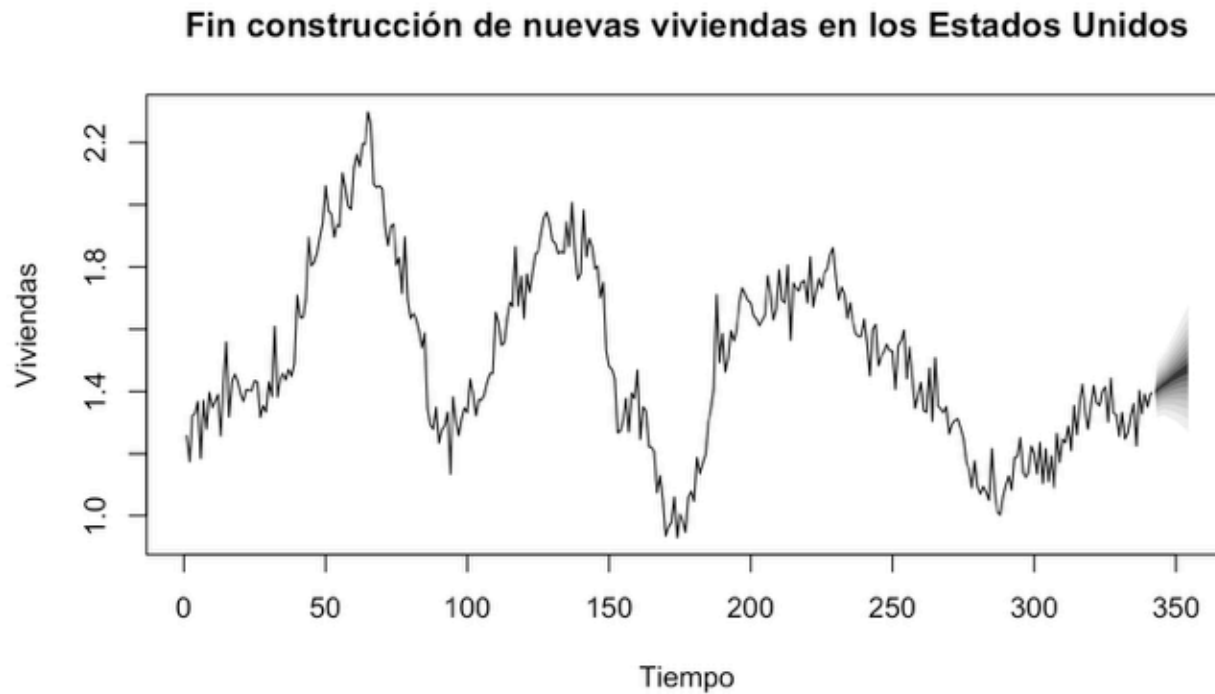
Vectores autorregresivos: predicción (1/2)

Para predecir el futuro usando un modelo $\text{VAR}(\mathbf{p})$, usamos exactamente las mismas técnicas que para los modelos $\text{ARMA}(\mathbf{p}, \mathbf{q})$. Usamos la muestra 1968:01-1991:12 para estimar el modelo y pronosticamos el periodo 1992:01-1996:06. Para el inicio de construcción de nuevas viviendas tenemos:



Vectores autorregresivos: predicción (2/2)

Para el fin de construcción de nuevas viviendas tenemos:



Vectores autorregresivos: causalidad de Granger (1/2)

La siguiente tabla muestra el resultado del test de Granger para el inicio de construcción de viviendas:

```
## $Granger
##
##   Granger causality H0: comps do not Granger-cause starts
##
## data:  VAR object var_fit
## F-Test = 2.3066, df1 = 4, df2 = 658, p-value = 0.05685
##
##
## $Instant
##
##   H0: No instantaneous causality between: comps and starts
##
## data:  VAR object var_fit
## Chi-squared = 11.478, df = 1, p-value = 0.0007043
```

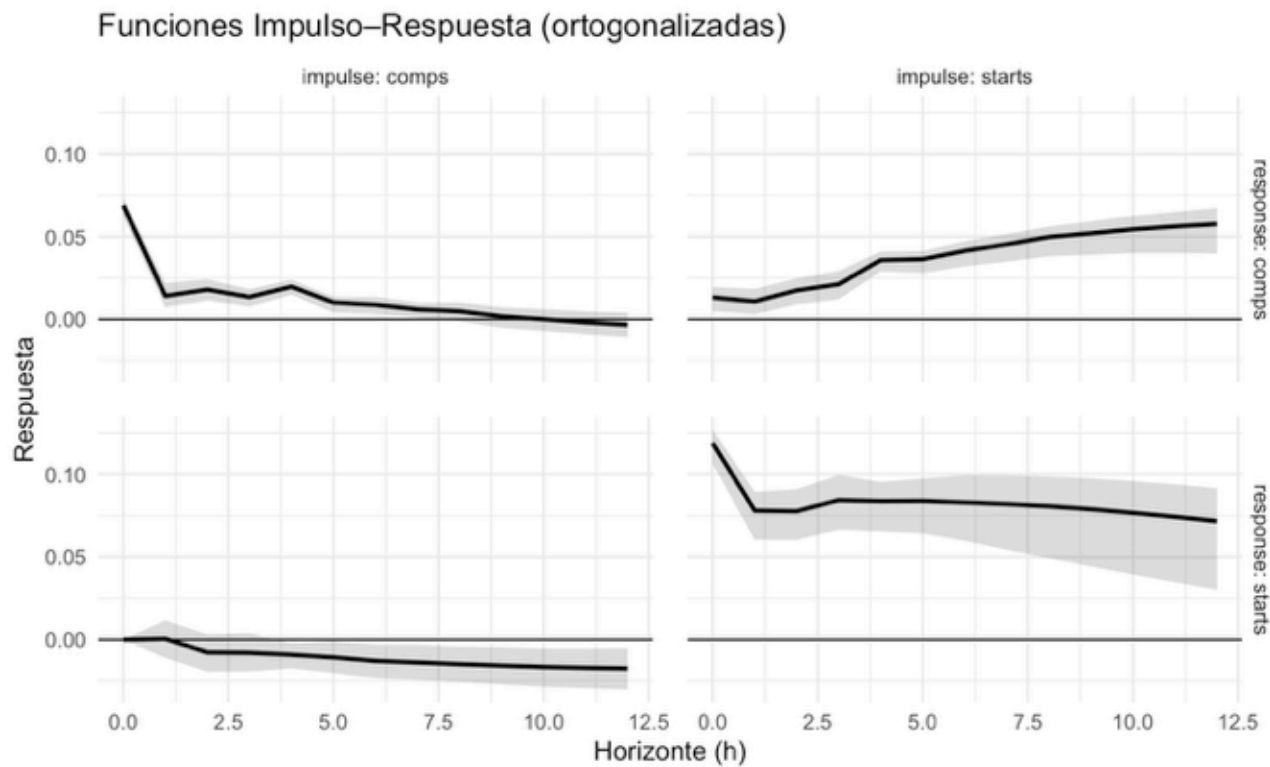
Vectores autorregresivos: causalidad de Granger (2/2)

El test de Granger para el fin de construcción de viviendas:

```
## $Granger
##
##   Granger causality H0: starts do not Granger-cause comps
##
## data:   VAR object var_fit
## F-Test = 32.125, df1 = 4, df2 = 658, p-value < 2.2e-16
##
##
## $Instant
##
##   H0: No instantaneous causality between: starts and comps
##
## data:   VAR object var_fit
## Chi-squared = 11.478, df = 1, p-value = 0.0007043
```

Funciones de impulso-Respuesta

Las funciones de impulso-respuesta para nuestro ejemplo son:



Descomposición de la varianza

Otra forma de caracterizar la dinámica de un VAR —muy relacionada a las funciones impulso-respuesta, es la **descomposición de la varianza**. Esta nos permite analizar el error de predicción en cada variable a un determinado horizonte de tiempo. Este error se atribuye a la incertidumbre acerca de la evolución futura en cada una de las variables.

