



**VAES**

Viceministerio de Análisis  
Económico y Social

# Seminario

- Economía, Finanzas y Ciencias Sociales -

#somoseconomía #somosplanificación #somosdesarrollo



MINECONOMIARD | <https://mepyd.gob.do/>

# Proyectando desempleo en República Dominicana con *Google Trends*

Autor: Reyna Gomera

Viceministerio de Análisis Económico y Social (VAES)  
Dirección de Análisis Macroeconómico (DAM)

Noviembre, 2024

# INTRODUCCIÓN

1

Se parte de la necesidad de información oportuna sobre la tasa de desempleo, dado los rezagos en la publicación de las estadísticas nacionales y el requerimiento de datos para la toma de decisiones a nivel público y privado.

2

Diversas investigaciones, bajo el supuesto de que las necesidades y preferencias de los individuos serían reveladas mediante una simple búsqueda en la red (Ettredge et al., 2005) han incorporado datos de internet para evaluar cómo mejoran los pronósticos de corto plazo del desempleo (Choi y Varían 2009b, 2012; D'Amuri y Marcucci 2009; Askitas y Zimmermann 2009; Tuhkuri 2015; Burop y Montes 2019).

3

El objetivo de este estudio es evaluar si en el contexto nacional con alta informalidad y un acceso a internet relativamente menor a países desarrollados, se observaría una mejora de la capacidad predictiva de modelos convencionales de series de tiempo del desempleo ante la inclusión de índices de *Google Trends*.

4

Los resultados obtenidos confirman una mejora en la capacidad predictiva del desempleo, aunque la ganancia es inferior relativo a lo documentado en países desarrollados.

# DATOS

## Desempleo abierto

1

Personas no ocupadas con disponibilidad para laborar, como porcentaje de fuerza de trabajo.

2

Procedente de la **Encuesta Nacional de Fuerza de Trabajo (ENFT)** y **Encuesta Nacional Continua de Fuerza de Trabajo (ENCFT)** divulgada por el Banco Central de la República Dominicana (BCRD).

3

Desagregación temporal en base al método de Chow y Lin (1971) de la serie calculada en base a ENFT y empalme lineal de las series.

**Tasa de desempleo  
abierto:**



**Frecuencia:**  
Trimestral



**Periodo:**  
T1 2004- T1 2024

## DATOS

### Índices de *Google Trends*

1

Refleja la frecuencia con la cual los usuarios efectúan la búsqueda de un término, relativo al total de búsquedas que se realizan en la plataforma.

2

El índice se calcula:

$$I_{i,t} = \left\{ \frac{\frac{K_{i,t}}{G_{i,t}}}{\max_t \frac{K_{i,t}}{G_{i,t}}} \right\} \times 100$$

Donde:

$K_{i,t}$  es la cantidad de búsquedas realizadas sobre un término en el lugar  $i$  en el tiempo  $t$ .

$G_{i,t}$  total de búsquedas realizadas en el lugar  $i$  en el tiempo  $t$ .

3

La información está disponible desde 2004. Los datos son publicados con un rezago de aproximado de 72 horas.

## DATOS

### Índices de Google Trends

4 Los términos de búsqueda se dividieron en dos grupos:

- En un primer grupo se evaluaron términos estándares de búsquedas de trabajo (p. ej. “empleo”, “empleo rd”, “trabajo”, “vacantes”, “oferta laboral”).
- En una segunda agrupación se contemplaron los portales de empleo disponibles para República Dominicana como p. ej. “aldaba”, “Tu Empleo RD”, “Chiripas”, “LinkedIn”, “Concursa Map”, entre otros.

5 En total se evaluaron 30 términos de búsqueda, de las cuales ocho (8) fueron preseleccionados en base a su popularidad y correlación con la tasa de desempleo.

6 En total se seleccionaron los siguientes cinco índices: “**empleo**”, “**trabajo**”, “**liquidación**”, “**LinkedIn**” y “**aldaba**”, utilizando el algoritmo Forward Stepwise Selection.



## ASPECTOS METODOLÓGICOS

### Modelo *benchmark* (MOD0)

- El modelo *benchmark* identificado sigue la representación ARIMA (Box y Jenkins, 1996). En la etapa de identificación se utilizan las herramientas tradicionales como la visualización de la serie, el correlograma y las pruebas de raíces unitarias.
- Inicialmente, la forma del modelo base es un ARIMA (1,1,0). Se requirió la inclusión de un impulso:

$\zeta_{02/20} = 1$  cuando  $t = 2020Q2$  y cero (0) en caso contrario.

- En definitiva, la forma del modelo ARIMAX (1,1,0) es la siguiente:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \gamma_1 \zeta_{02/20} \quad (1)$$

- Donde,  $y_t$  corresponde al logaritmo de la tasa de desempleo abierto.
- Se presentan cuatro especificaciones adicionales. La relevancia de cada uno de los modelos dentro de muestra es evaluada utilizando los criterios de información de Akaike (1974) y Schwarz (1978). Mientras para evaluar el desempeño fuera de muestra se utiliza el RMSE y MAPE.

# ASPECTOS METODOLÓGICOS

## Modelos con índices *Google Trends*

### MOD1

Incluye el índice de *Google Trends* del término “trabajo”.

01

03

### MOD3

Contiene tanto el índice de “trabajo” como el de “líquidación”.

### MOD2

Considera el portal de empleo con mayor popularidad durante toda la serie “aldaba”.

02

### MOD4

A partir de las cinco series seleccionadas, se calculan dos componentes que en total explican el 85.76 %, y se incluyen a la especificación base.

04



# RESULTADOS

## Comparación de modelos dentro de muestra

- La estimación del mejor modelo base (MOD0) se presentan en la tabla 1. La inclusión del impulso resulta ser oportuna, con un AIC y BIC inferior a modelos que no contemplan dicha intervención.
- Se observa una **reducción del criterio Akaike para los modelos MOD1 y MOD3** (ver tabla 1).
- Entre estas dos alternativas, el MOD1 resulta **ser el modelo más adecuado dada la significancia de los estimadores y los criterios de información**.
- Para MOD2 y MOD4, no se observa una mejoría significativa dentro de muestra respecto al modelo base.

**Tabla 1.**  
Estimación de los resultados de los modelos ARIMAX para el desempleo de República Dominicana

	MOD0	MOD1	MOD2	MOD3	MOD4
$y_{t-1}$	-0.3704*** (0.1137)	-0.3577*** (0.1158)	-0.3667*** (0.1151)	-0.3596*** (0.11172)	-0.3684*** (0.1165)
$X_t^{trabajo}$		-0.2886** (0.1384)		-0.2857** (0.1415)	
$X_t^{aldaba}$			0.0202 (0.0276)		
$X_t^{liquidacion}$				0.0015 (0.0148)	
$X_t^{PC1}$					0.0541** (0.0275)
$X_t^{PC2}$					0.0023 (0.0205)
$\zeta_{02/20}$	-0.7055*** (0.0784)	-0.6433*** (0.0814)	-0.7002*** (0.0784)	-0.6440*** (0.0816)	-0.6588*** (0.0797)
Desviación estándar de las innovaciones	0.0094	0.0088	0.0093	0.0088	0.0088
Criterio Akaike (AIC)	-116.6739	-118.8766	-115.2071	-116.8871	-116.6257
Criterio Schwarz (BIC)	-110.0598	-110.0578	-106.3883	-105.8637	-105.6022

Nota: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Fuente: Elaboración propia.

# RESULTADOS

## Comparación de modelos fuera de muestra

- Para evaluar cada modelo, en términos de su capacidad predictiva, se procedió a dividir la muestra en dos partes, 80 % para la estimación de los parámetros y 20 % para pruebas fuera de muestra.
- Debido a que la muestra de prueba resulta pequeña, para garantizar la robustez de los resultados se realizó la prueba de Diebold y Mariano (1995), que contrasta la precisión de dos métodos distintos de proyección. **Existe evidencia de que tres de las cuatro alternativas mejoran.**
- De acuerdo con los indicadores de evaluación, la mejoría en el pronóstico de la tasa de desempleo abierto oscila entre 0.25 % y 6.54 %.

Los modelos **MOD1** y **MOD3** presentan la mayor disminución del RMSE y MAPE, con una reducción de más del 6 % respecto al modelo base.

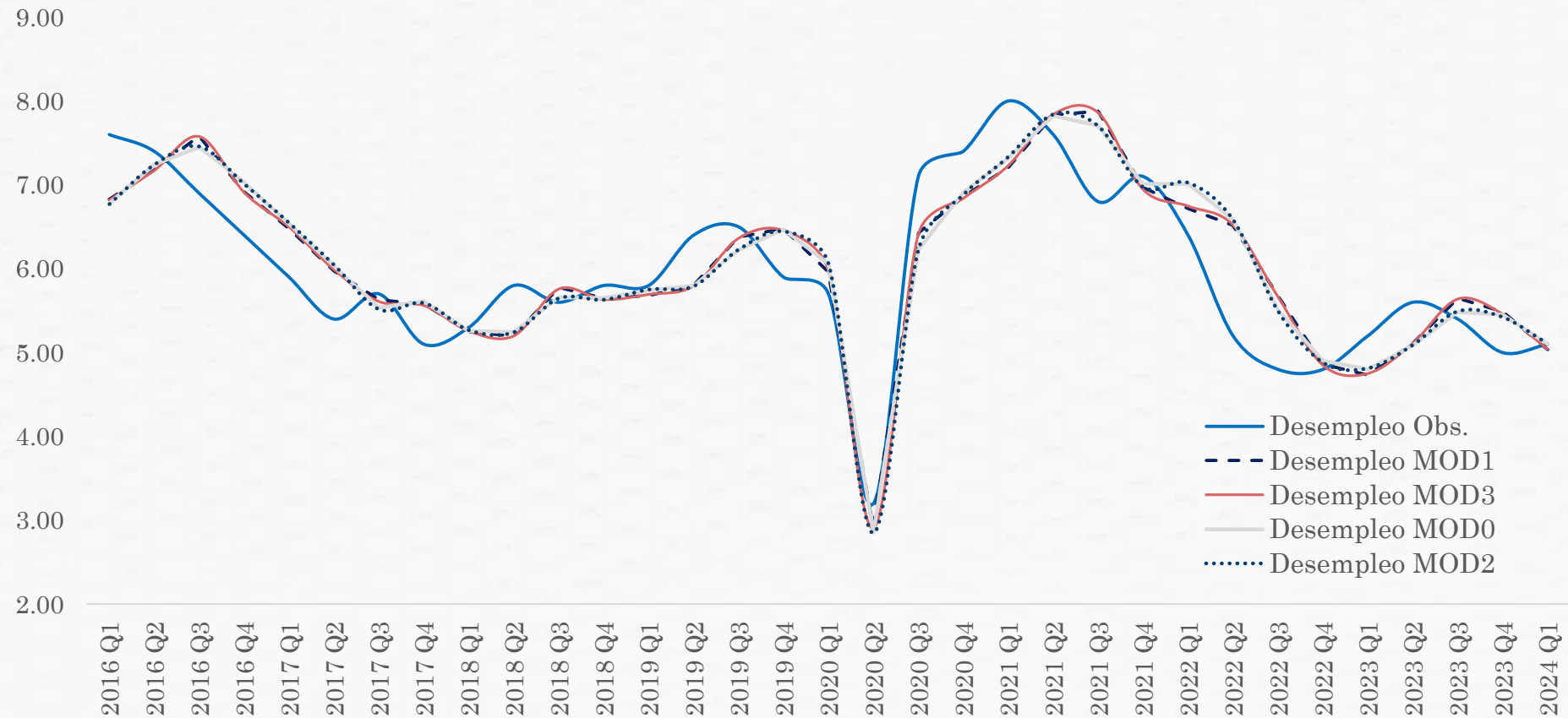
**Tabla 2.**  
Medidas de evaluación de capacidad predictiva fuera de muestra

	MOD0	MOD1	MOD2	MOD3	MOD4
RMSE	0.2859	0.2677	0.2843	0.2678	0.2749
MAPE	0.1488	0.1391	0.1479	0.1391	0.1484
Prueba DM (valor p)		0.0336	0.0348	0.0335	0.6093

Fuente: Elaboración propia.

# RESULTADOS

## Comparación de resultados



Fuente: Elaboración propia.

## CONCLUSIONES

1

Al comparar el desempeño dentro y fuera de muestra de los modelos que incluyen los índices de Google con el de referencia, se obtiene evidencia de que la información contribuye a mejorar el pronóstico de desempleo.

2

Los modelos que incluyen los índices referentes a la búsqueda de términos como “trabajo” o “liquidación” presentan un RMSE y MAPE significativamente menor al del modelo base.

3

Los hallazgos se alinean con las evidencias de investigaciones similares, aunque la ganancia es inferior relativo a países desarrollados. Esta dinámica resulta esperable debido, en primer lugar, a las características del mercado laboral dominicano, y el menor acceso a internet relativo a esos países.

4

En próximas versiones podrían evaluarse nuevas metodologías de pronóstico como Random Forest o regresión LASSO, así como la incorporación de otras variables macroeconómicas de alta frecuencia.



**VAES**

Viceministerio de Análisis  
Económico y Social

# Seminario

- Economía, Finanzas y Ciencias Sociales -

#somoseconomía #somosplanificación #somosdesarrollo



MINECONOMIARD | <https://mepyd.gob.do/>