



# **Google Trends: predicción del nivel de Empleo Agregado en Perú usando datos en tiempo real, 2004-2011**

**Jillie Vanessa Chang Kcomt  
Andrea Katherine Del Río Lazo**

**Octubre 2011**

# Contenido



1. **Objetivo**
2. **Conceptos: ¿qué es Google Trends?**
3. **Literatura**
4. **Google Trends y el empleo**
5. **Metodología**
6. **Resultados**
7. **Conclusiones**

# 1. Objetivo



Analizar si la utilización de los datos en tiempo real de Google Trends durante el periodo 2004-2011 mejora las predicciones del índice de empleo de Lima para empresas de 100 y más trabajadores (IE100).

## 2. ¿Qué es Google Trends?



- Servicio de Google que cuantifica las búsquedas que se realizan a través del buscador.
  
- Reporta un índice semanal del 0 al 100 que se obtiene después de dos procesos:
  1. Normalización
  2. Escalamiento

# 3. Literatura (1/2)



## Variables microeconómicas

- Varian y Choi (2009) : ventas minoristas, ventas de autos, casas y viajes.
- Carrière-Swallow y Labbé (2010): ventas de autos
- Song y Pan (2010) : demanda por cuartos de hoteles

## 3. Literatura (2/2)



### Variables macroeconómicas

- Della Penna y Huang (2009) y Schmidt y Vosen (2009): índice de confianza del consumidor
- Bersier (2010), D'Amuri (2009), Askitas y Zimmermann (2009), Suhoy (2009) y Oleksandr (2010) : desempleo

# 4. Google Trends y empleo (1/5)



## Índice de empleo para empresas de 100 y más trabajadores (IE100)

- Se calcula a partir de la Encuesta Nacional de Variación Mensual de Empleo en Empresas de 10 y más Trabajadores (ENVME) y es elaborada por el MINTRA.
- La ENVME se aplica a empresas y establecimientos con trabajadores sujetos al régimen laboral.
- Rezago en la publicación: aproximadamente 3 meses.

# 4. Google Trends y empleo (2/5)



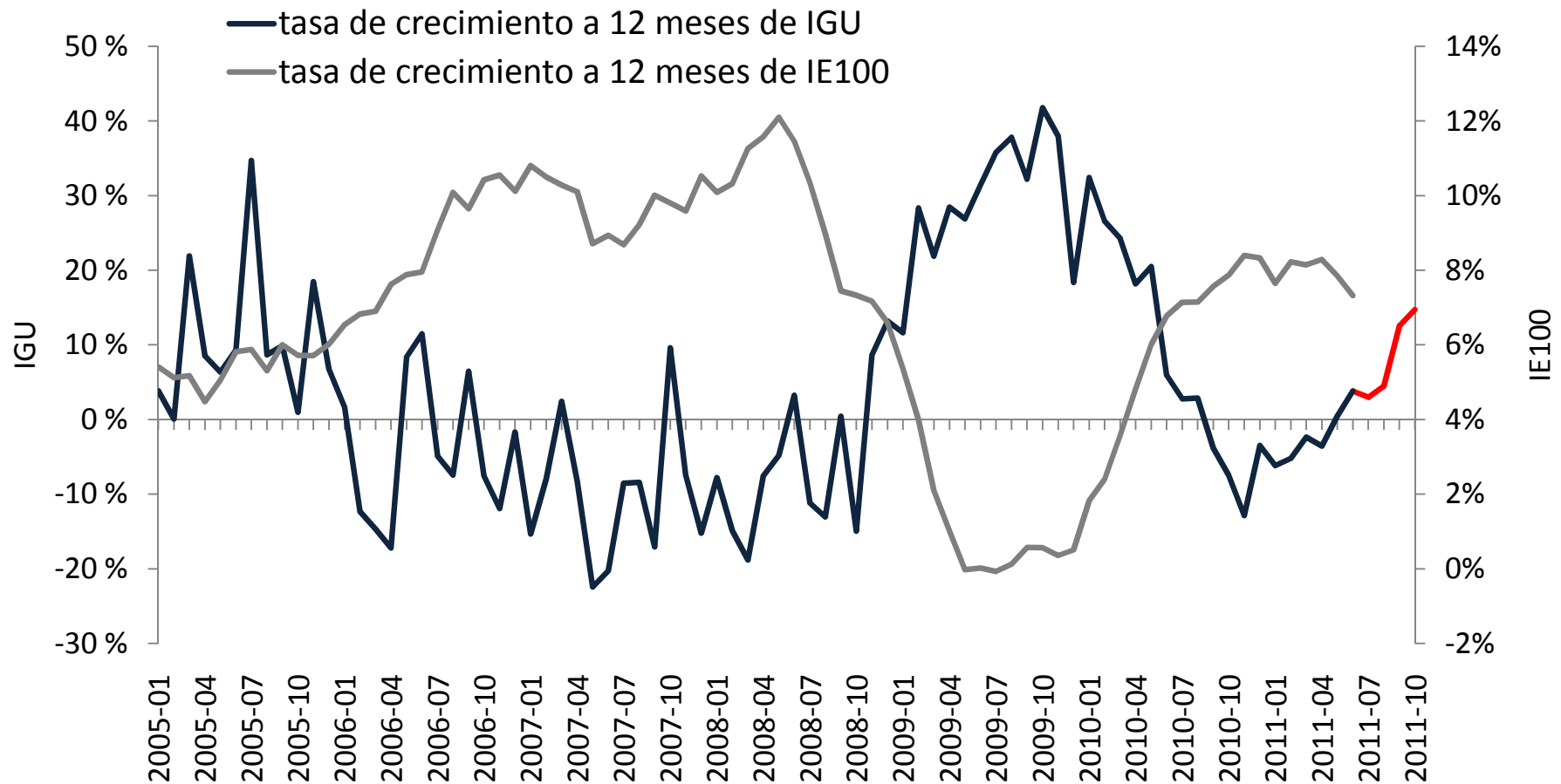
## Índice de Google de desempleo (IGU)

- Selección de palabras claves como: “busco trabajo”, “laborum” “bolsa de trabajo”, “aptitus”, etc.
- Mensualización de cada serie semanal a través de promedios simples.
- Indexación de las series ponderado por desviación estándar.



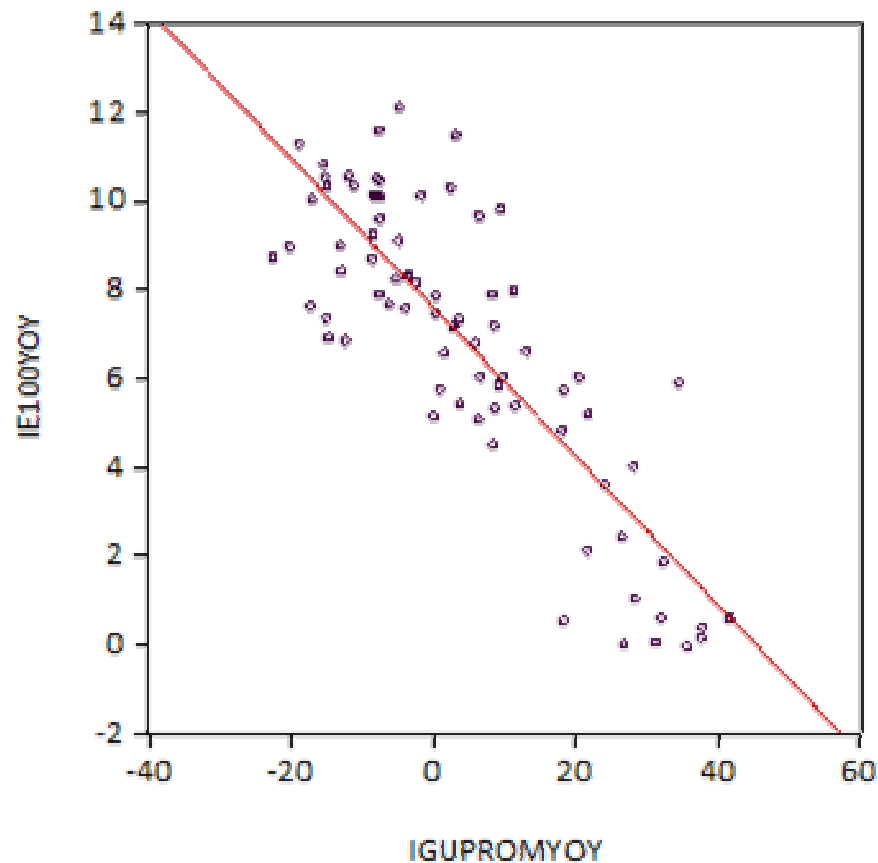
# 4. Google Trends y empleo (4/5)

Índice de empleo de Lima para empresas de 100 y más trabajadores e Índice de Google de desempleo en Perú, 2004-2011 (tasa de variación a 12 meses)



# 4. Google Trends y empleo (5/5)

Gráfico de dispersión del Índice de empleo en Lima para empresas de 100 y más trabajadores e IGU en Perú, 2004-2011 (tasa de variación a 12 meses)



# 5. Metodología (1/4)



## Metodología 1:

modelo autorregresivo de media móvil (ARMA) y  
modelo autorregresivo de razagos distribuidos  
(ARDL)

## Metodología 2:

Cointegración y modelo de corrección de errores

# 5. Metodología (2/4)

## Modelos ARMA y ARDL

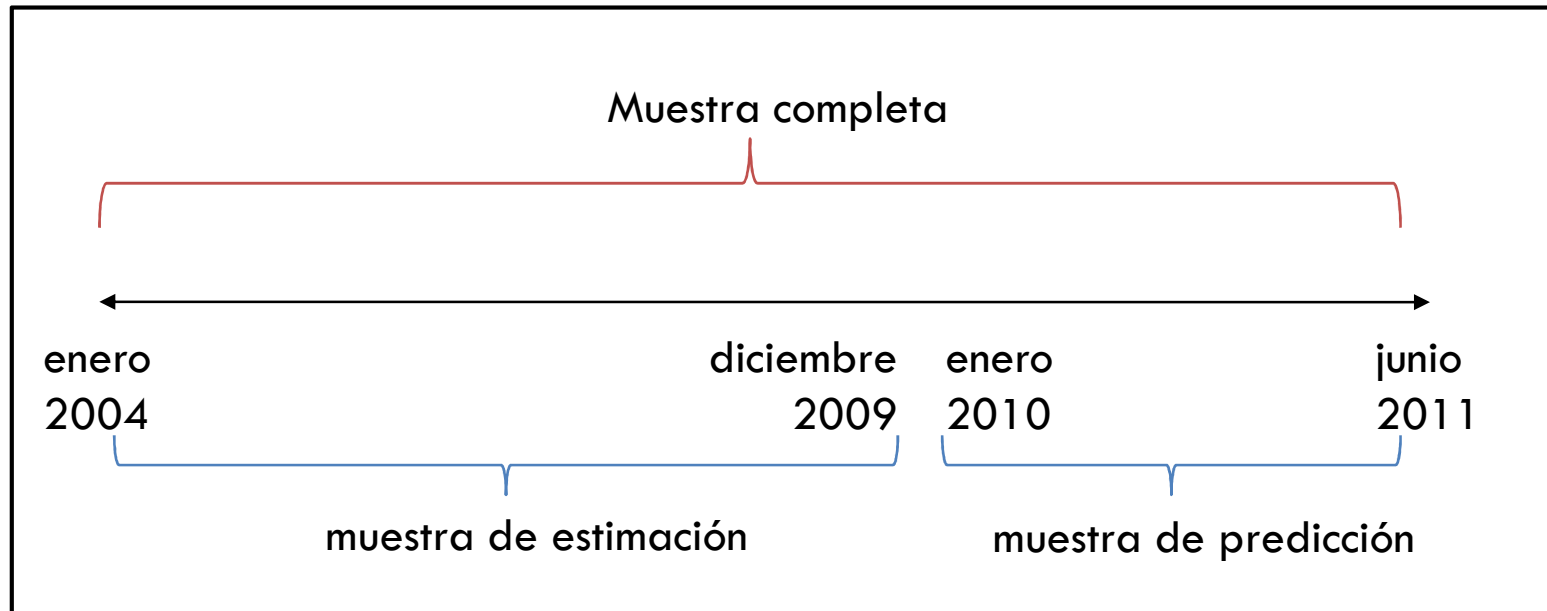
Modelo de referencia (ARMA) :

$$y_t = \beta_0 + \sum_{i=1}^p \beta_i y_{t-i} + \epsilon_t^k + \sum_{i=1}^q \alpha_i \epsilon_{t-i}$$

Modelo aumentado (ARDL):

$$y_t = \beta_0 + \sum_{i=1}^p \beta_i y_{t-i} + \epsilon_t + \sum_{i=1}^q \alpha_i \epsilon_{t-i} + \sum_{i=1}^T \gamma_{i,l} IGU_{t-i}$$

# 5. Metodología (3/4)



**Se evaluarán las predicciones dentro y fuera de la muestra. Las Pruebas utilizadas serán:**

- Raíz del Error Cuadrático Medio (RECM)
- Error Absoluto Medio (EAM)
- Prueba Diebold y Mariano(1995) propuesta por Clark y West (2007)

## 5. Metodología (4/4)

- Si las series cointegran, entonces es posible representar la relación como:

$$IE100_t = \beta IGU_t + v_t$$

- De acuerdo con el teorema de representación de Granger, es posible analizar la dinámica de corto plazo a través del siguiente MCE:

$$\Delta IE100_t = \alpha [IE100_{t-1} - \beta IGU_{t-1}] + \sum_{t=1}^m \phi_i \Delta IE100_{t-j} + \sum_{t=1}^m \theta_j \Delta IGU_{t-j} + \varepsilon_t$$
$$\Delta IGU_t = \alpha' [IE100_{t-1} - \beta IGU_{t-1}] + \sum_{t=1}^m \phi'_i \Delta IE100_{t-j} + \sum_{t=1}^n \theta'_j \Delta IGU_{t-j} + \varepsilon'_t$$

## 5. Metodología (5/5)

$$\Delta IE100_t = \alpha[IE100_{t-1} - \beta IGU_{t-1}] + \sum_{t=1}^m \phi_i \Delta IE100_{t-j} + \sum_{t=1}^m \theta_j \Delta IGU_{t-j} + \varepsilon_t$$
$$\Delta IGU_t = \alpha'[IE100_{t-1} - \beta IGU_{t-1}] + \sum_{t=1}^m \phi'_i \Delta IE100_{t-j} + \sum_{t=1}^n \theta'_j \Delta IGU_{t-j} + \varepsilon'_t$$

- IGU será un predictor insesgado si  $\beta = 1$ . Más aún, será útil como indicador líder si  $\alpha \neq 0$  pues su nivel precedería (causa en el sentido de Granger) a IE100. Finalmente, permitirá anticipar la senda futura de IE100 si IGU es fuertemente exógena, es decir, si  $\alpha' = 0 = \phi'_1 = \dots = \phi'_m$ .

## 6. Resultados



# ARMA: Predicción dentro de la muestra

## Parámetros estimados y coeficientes de capacidad predictiva para predicciones dentro de la muestra

*Variable endógena: diferencia del IE100( $y_t$ )*

<i>Variable exógena</i>	<i>Modelo de referencia</i>		<i>Modelo aumentado</i>	
	(1a)	(1b)	(2a)	(2b)
$y_{t-1}$	0.95 (0.03)	1.22 (0.06)	0.87 (0.04)	1.14 (0.06)
$y_{t-3}$		-0.30 (0.06)		-0.32 (0.05)
$DUM_t$	-0.39 (0.17)	-0.38 (0.13)	-0.34 (0.17)	-0.34 (0.11)
$MA(12)$		-0.89 (0.04)		-0.89 (0.03)
$IGU_t$			-0.02 (0.01)	-0.02 (0.01)
<i>Observaciones</i>	77	75	77	75
$R^2_{ajustado}$	0.956	0.98	0.958	0.99
$RECM_{dinámico}$	5.39	0.77	3.50	0.37
$EAM_{dinámico}$	5.14	0.66	3.34	0.31
$RECM_{estático}$	0.74	0.31	0.73	0.28
$EAM_{estático}$	0.63	0.24	0.62	0.23

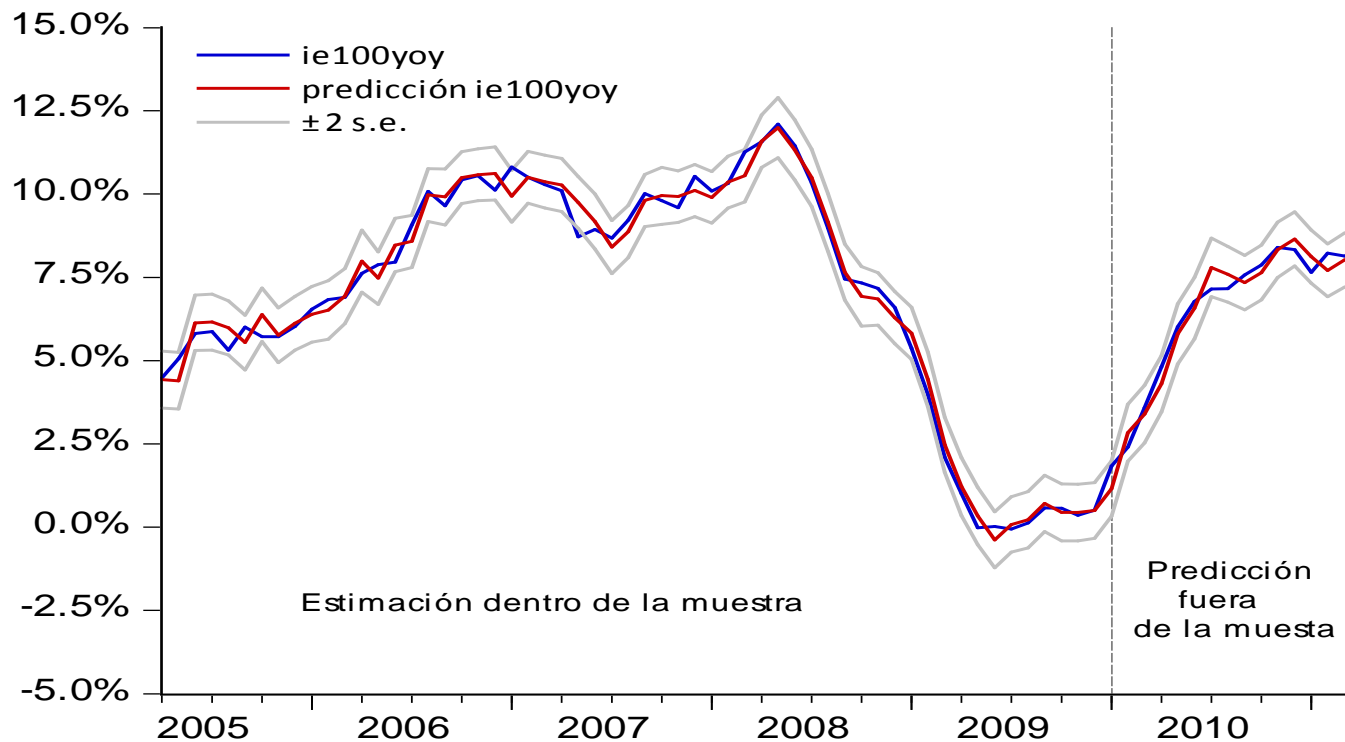
*Notas:*

- Nivel de significancia al 0.05
- Todas las variables se encuentran en cambios porcentuales a 12 meses (“tasa YoY”)
- Todos los modelos cuentan con constante

# Predicción fuera de la muestra (1/2)

Predicción dinámica: ECM y RECM ↓ 20% - 60%  
Predicción dinámica: ECM y RECM ↓ 5% y 30%

**Predicción de IE100 usando el modelo (2b)**



# ARMA: Predicción fuera de la muestra (2/2)

## Resultados de la prueba Diebold y Mariano

	(a)	(b)
<i>Observaciones</i>	18	18
<i>Estadístico DM</i>	2.58	2.43
<i>P value</i>	0.02	0.03

*Notas:*

- *H<sub>0</sub>: modelos 1 y 2 son similares en términos de predicción*
- *H<sub>a</sub>: el modelo 2 predice mejor que el modelo 1*

# Cointegración y modelo de corrección de errores (1/4)

## ADF a residuos de Largo plazo

<i>ADF sobre los residuos de <math>IE100_t = \beta IGU_t + \varepsilon_t</math></i>	
<i>Con constante</i>	0.00
<i>Estadístico t (Ho: intercepto nulo)</i>	0.94
<i>Sin constante ni tendencia</i>	0.00

*Notas:*

*- Nivel de significancia al 0.05*

*- Los valores mostrados son las probabilidades asociadas*

## Metodología de Johansen

<i>Estadístico Traza</i>	<i>Probabilidad</i>
<i>No existe vector de cointegración</i>	0.00
<i>Hasta 1 vector de cointegración</i>	0.15
<i>Estadístico Valor propio máximo</i>	
<i>No existe vector de cointegración</i>	0.00
<i>Hasta 1 vector de cointegración</i>	0.15
<i>¿Cointegran?</i>	<b>Sí</b>

*Notas: Solo intercepto en la ecuación de cointegración y no tendencia en el VAR*

# Cointegración y modelo de corrección de errores (2/4)

**Relación de cointegración:**

$$IE100_t = 104.11 - 1.03 IGU$$

(4.32)    (0.09)

**¿IGU es un predictor insesgado de la tendencia?**

**Modelo con restricción**

<i>Restricción: <math>B(1,1) = 1, B(1,2) = 1</math></i>	<i>Probabilidad</i>
<i>Chi - square</i>	0.09
<i>Probabilidad</i>	0.77

Al 5% de significancia no se rechaza la Ho

# Cointegración y modelo de corrección de errores (3/4)

## Exogeneidad débil (metodología de Johansen)

	<i>Ecuación de IE100</i>	<i>Ecuación de IGU</i>
Velocidad de ajuste	-0.15	-0.34
<i>Estadístico t</i>	-2.48	-2.16
<b><i>Conclusión</i></b>	<b><i>IE100 no es Débilmente exógena</i></b>	

## Modelo VEC con restricción

<i>Restricción: <math>B(1,1) = 1, B(1,2) = 1</math> , <math>A(2,1)=0</math></i>	<i>Probabilidad</i>
<i>Chi – square</i>	6.05
<i>Probabilidad</i>	0.05

# Cointegración y modelo de corrección de errores (4/4)

- En este contexto, la prueba de causalidad en el sentido de Granger indica que IE100 no causa en el sentido de Granger a IGU. Por lo cual se puede asumir que IGU es fuertemente exógena y, por tanto, un buen predictor de IE100.

## Causalidad a lo Granger

<i>Hipótesis nula</i>	<i>Probabilidad</i>
<i>IGU no causa a IE100</i>	0.02
<i>IE100 no causa a IGU</i>	0.17



# 7. Conclusiones





# **Google Trends: predicción del nivel de Empleo Agregado en Perú usando datos en tiempo real, 2004-2011**

**Jillie Vanessa Chang Kcomt  
Andrea Katherine Del Río Lazo**

**Octubre 2011**

### Ejemplo de indexación para IGU

fecha	promedio mensual de "busco trabajo"	promedio mensual de "bolsa de trabajo"
2011-04	16.5	49.5
2011-05	17.2	50.8
Promedio (A)	16.85	50.15
Desviación Estándar (B)	0.49	0.92
1/Desviación Estándar (C)	2.02	1.09
Suma de la inversa de las desviaciones estándar (2.02+1.09) (D)	3.11	
ponderadores (C/D)	0.65	0.35

$$\text{IGU(indexado) para el mes 04} = \sum \alpha_i x_{i,t=4} = 16.5 * 0.65 + 49.5 * 0.35$$