Medición del crecimiento económico desde el espacio: un enfoque de país individual

VÍCTOR M. GUERRERO
Departamento de Estadística, ITAM

JUAN A. MENDOZA
Financial Mathematics, University of Chicago

Workshop on Econometrics and Data Science,
Monterrey, N. L., noviembre 8 y 9 de 2018

ÍNDICE

1. Introducción

2. Especificación del modelo de HSW

3. Estimación del crecimiento del PIB de un país

4. Aplicación empírica para México

5. Consideraciones finales

1. Introducción

- La medición adecuada del Producto Interno Bruto (PIB) es uno de los grandes retos que enfrentan las agencias oficiales de estadística, particularmente en **países en desarrollo.**

- Los datos del PIB deben ser creíbles para atraer inversion, elaborar políticas económicas y hacer comparaciones válidas de economías.
  - Hay guías establecidas por la Comisión Estadística de la ONU para mantener el **Sistema de Cuentas Nacionales** (SCN).
  - La incertidumbre en las cifras oficiales de crecimiento del PIB continúa a pesar de que los SCN's se revisan con regularidad.

- Se han propuesto alternativas para calcular el crecimiento económico, como combinar los enfoques de gasto e ingreso para el PIB (Aruoba *et al*., 2013).

- Henderson, Storeygard y Weil (2012) (HSW) combinan datos de luminosidad nocturna, obtenidos de imágenes satelitales, con cifras oficiales de países con ingreso medio, pues los datos de países ricos se consideran confiables.

- Chen y Nordhaus (2011) usan datos de satélites para países con ingresos bajos, cuyos sistemas estadísticos tienden a ser deficientes; mientras que Nordhaus y Chen (2014) encuentran que los datos de luces nocturnas son básicamente **útiles para países en desarrollo.**

- ## Luminosidad y actividad económica

- El Programa Satelital de Defensa Meteorológica de la Fuerza Aérea de Estados Unidos diseñó un sistema de sensores para satélites que permite calcular la luminosidad de las luces nocturnas.

- La *National Oceanic & Atmospheric Administration* (NOAA) aisla la actividad natural de las luces, para brindar conjuntos de datos de luz estable al público.
  - Algunos tipos de **luz excluida** son:
    - incendios forestales,
    - quema de gas,
    - actividad de auroras,
    - efecto de meses de verano (cuando el sol se pone tarde).

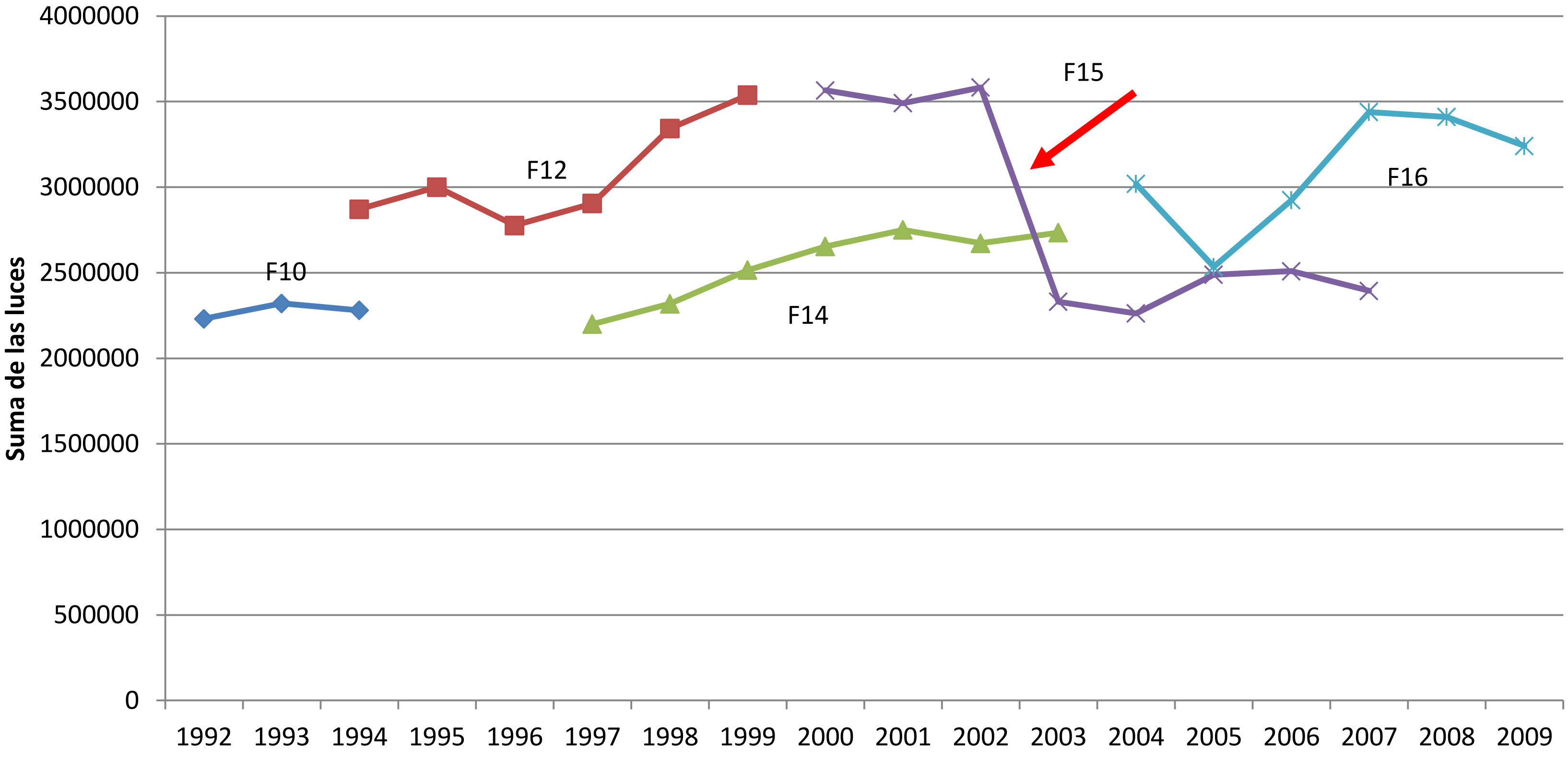
- Los datos son **anuales** y forman una rejilla por latitud y longitud, con la intensidad de la luz medida como un número digital (ND).
  - El ND es un **número entero** en el intervalo [0, 63], con 0 no actividad luminosa y 63 la mayor intensidad registrada.
  - El sistema utilizado registra la **intensidad de la luz terrestre** por pixeles, 14 veces al día, desde 1992.
  - Los valores para los pixeles surgen de un complejo proceso de cálculo, que tiene en cuenta la curvatura terrestre.

- HSW calcularon promedios ponderados de los datos por pixeles dentro de cada país, donde la ponderación para cada pixel es la participación territorial del país.

- La capacidad instalada en el país se asocia con la actividad luminaria.

- Las comparaciones de ND por sí solas cambian al paso del tiempo pues los satélites fallan en sus mediciones.
  - Por ejemplo, en 2003 el satélite **F15 señaló erróneamente una disminución** de actividad luminaria en todos los países.

Satélites empleados para la obtención de la suma de los ND's para México



- El presente trabajo surge de la especificación del modelo que usan HSW, que usa datos de panel para diversos países y se apoya en las calificaciones del Banco Mundial para la calidad de los datos de ingresos.

- La distinción de este trabajo respecto al de HSW es que:

  (i) se usan datos de **series de tiempo y satelitales** para cada país por separado, en lugar de usar datos de panel,

  (ii) **no se requiere calificar la calidad de los datos**, y

  (iii) se basa en un **método sencillo de estimación no-paramétrica**.

- El método es coherente con el cálculo oficial del PIB, pues sólo emplea fuentes de datos del país en cuestión.

2. Especificación del modelo de HSW

- Para el país j y año t:

  $Y_{j,t}$ es el **PIB verdadero**, expresado en términos reales,

  $Z_{j,t}$ es el **PIB real reportado oficialmente,**

  $X_{j,t}$ es el **valor de las luces,**

  $y_{j,t} = \log(Y_{j,t})$, con log(.) el logaritmo natural; igual para $z_{j,t}$ y $x_{j,t}$.

- El crecimiento se calcula como diferencia logarítmica, o sea,

$$Dy_{j,t} = y_{j,t} - y_{j,t-1} \quad \text{for} \quad t = 2, \ldots, N$$

y **los crecimientos se relacionan** mediante

$$Dz_{j,t} = Dy_{j,t} + \varepsilon_{Dz,j,t},$$

donde $\varepsilon_{Dz,j,t}$ es un error aleatorio con varianza σ^2_{Dz}; y por la **relación**

$$Dx_{j,t} = \beta Dy_{j,t} + \varepsilon_{Dx,j,t},$$

con $\varepsilon_{Dx,j,t}$ otro error con varianza σ^2_{Dx}, no-correlacionado con $\varepsilon_{Dz,j,t}$.

- El crecimiento del PIB oficial se predice con el de las luces, como

$$\widehat{Dz}_{j,t} = \hat{\Psi} Dx_{j,t},$$

donde

$$\hat{\Psi} = \hat{Cov}(Dx, Dz)/\hat{\sigma}^2_{Dx}$$

es el estimador de Mínimos Cuadrados Ordinarios (MCO) del **inverso de la elasticidad de las luces** respecto al ingreso.

- Este estimador es sesgado e inconsistente para $1/\beta$, a menos que $\sigma^2_{Dx} = 0$ (lo que implicaria proporcionalidad perfecta entre $Dx_{j,t}$ y $Dy_{j,t}$) porque

$$\text{plim}(\hat{\Psi}) = \frac{1}{\beta}\left(\frac{\beta^2 \sigma^2_{Dy}}{\beta^2 \sigma^2_{Dy} + \sigma^2_{Dx}}\right).$$

- HSW consideran el crecimiento predicho con $\widehat{\Psi}$ tan solo como una aproximación, la cual proponen combinar con los datos oficiales mediante

$$\widehat{\mathrm{Dy}}_{\mathrm{j,t}} = \lambda \mathrm{Dz}_{\mathrm{j,t}} + (1-\lambda)\widehat{\mathrm{Dz}}_{\mathrm{j,t}} \;\text{ con }\; \lambda \in (0,1).$$

- La **varianza del error de predicción** es

$$\mathrm{Var}(\widehat{\mathrm{Dy}}_{\mathrm{j,t}} - \mathrm{Dy}_{\mathrm{j,t}}) = \lambda^2 \sigma_{\mathrm{Dz}}^2 + (1-\lambda)^2 \frac{\sigma_{\mathrm{Dy}}^2 \sigma_{\mathrm{Dx}}^2}{\beta^2 \sigma_{\mathrm{Dy}}^2 + \sigma_{\mathrm{Dx}}^2},$$

así que el mínimo de esta varianza se obtiene al elegir el valor

$$\tilde{\lambda} = \frac{\sigma_{\mathrm{Dy}}^2 \sigma_{\mathrm{Dx}}^2}{\sigma_{\mathrm{z}}^2\left(\beta^2 \sigma_{\mathrm{Dy}}^2 + \sigma_{\mathrm{Dx}}^2\right) + \sigma_{\mathrm{Dy}}^2 \sigma_{\mathrm{Dx}}^2}.$$

- Dicha $\tilde{\lambda}$ se obtiene al fijar el valor de $\emptyset$, dado por el cociente de "varianza de señal a varianza total" para cada país, o sea,

$$\emptyset = \frac{\sigma^2_{Dy}}{\sigma^2_{Dy}+\sigma^2_{Dz}}.$$

- Para especificar $\emptyset$, HSW usan las calificaciones de calidad relativa de los datos de ingreso nacional que asigna el Banco Mundial y clasifican los países en dos grupos, en escala de 0 a10:
  (1) con **buenas estadísticas** (calificados entre 7 y 10), y
  (2) con **malos datos estadísticos** (calificados debajo de 7).

- La elasticidad de las luces respecto al ingreso (la β del modelo de HSW) la estimaron como 1.15 con datos panel de 113 países..
  – El valor de λ (deducido de β) es el mismo para todos los países con *buenos datos*.

3. Estimación del crecimiento del PIB de un país

- En este caso se supone un **modelo de señal más ruido** para el crecimiento del PIB oficial,

$$\mathrm{Dz}_t = \mathrm{Dy}_t + \eta_t \quad \text{para} \quad t = 2, \dots, N,$$

con η_t **el ruido** que **distorsiona la señal** (**el verdadero crecimiento del PIB**) Dy_t; $E(\eta_t) = 0$, $\mathrm{Var}(\eta_t) = \sigma^2_\eta$ y $\mathrm{Cov}(\eta_t, \eta_{t'}) = 0$ si $t \neq t'$.

  - También $\mathrm{Cov}(\mathrm{Dy}_t, \eta_t) = 0$ y $\mathrm{Var}(\mathrm{Dy}_t) = \sigma^2_{\mathrm{Dy}}$ para toda t.

- Se supone también una relación de elasticidad constante entre luces nocturnas e ingreso, o sea, $X_t = KY_t^\beta$ con K una constante positiva y β la elasticidad de las luces respecto al ingreso. Así surge el **modelo**

$$\mathrm{Dx}_t = \beta \mathrm{Dy}_t + \varepsilon_t \quad \text{para} \quad t = 2, \dots, N,$$

donde $\{\varepsilon_t\}$ son errores no-autocorrelacionados, con $E(\varepsilon_t) = 0$, $\mathrm{Var}(\varepsilon_t) = \sigma^2_\varepsilon$, $\mathrm{Cov}(\mathrm{Dy}_t, \varepsilon_t) = 0$ y $\mathrm{Cov}(\eta_t, \varepsilon_t) = 0$.

- # Mejor Estimación Lineal e Insesgada

- Al usar notación matricial y suponer que β, σ_ε^2 y $\alpha = \sigma_\eta^2/\sigma_\varepsilon^2$ son conocidos, se usa Mínimos Cuadrados Generalizados (MCG) para obtener el **Mejor Estimador Lineal e Insegado** (MELI) de D**y**,

$$\widehat{\mathrm{D}\mathbf{y}} = \lambda \mathrm{D}\mathbf{z} + (1 - \lambda)\widetilde{\mathrm{D}\mathbf{z}}$$

con $\widetilde{\mathrm{D}\mathbf{z}} = \beta^{-1}\mathrm{D}\mathbf{x}$ y $\lambda = \frac{\alpha^{-1}}{\alpha^{-1}+\beta^2} \in (0,1)$.

  – Éste es básicamente el resultado que obtuvieron HSW, pero ahora deducido con un enfoque estadístico diferente, que:
    (1) hace uso eficiente de ambas fuentes de información,
    (2) justifica la combinación lineal del crecimiento oficial con el de la intensidad de las luces, y
    (3) muestra cómo la ponderación λ se relaciona con α y β.

- Además, la matriz de varianza-covarianza del vector de errores de predicción se obtiene también por MCG y resulta ser

$$\mathrm{Var}(\widehat{\mathrm{D}\mathbf{y}} - \mathrm{D}\mathbf{y}) = \sigma_\varepsilon^2(\alpha^{-1} + \beta^2)^{-1}\mathrm{I}_{N-1}$$

donde I_{N-1} es la matriz identidad de dimensión (N-1).

- Una **solución factible** requiere estimar σ_ε^2, β y α, como sigue.
  - Por MCG, el estimador de la varianza es

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{\lambda}\,\hat{\beta}^2}{N-3}\left[\hat{\alpha}^{-1}\left(\mathrm{D}\mathbf{z} - \widetilde{\mathrm{D}\mathbf{z}}\right)'\left(\mathrm{D}\mathbf{z} - \widetilde{\mathrm{D}\mathbf{z}}\right)\right]$$

de manera que, a menor valor de $\hat{\lambda}$, menor varianza estimada del error de predicción.

- Para estimar β se usa la ecuación

$$Dz_t = \beta^{-1}(Dx_t - \varepsilon_t) + \eta_t = \beta_1 Dx_t + \gamma_t,$$

donde

$$\beta_1 = \beta^{-1} \text{ y } \gamma_t = \eta_t - \beta^{-1}\varepsilon_t,$$

con $E(\gamma_t) = 0$ y $Var(\gamma_t) = \sigma_\varepsilon^2(\alpha + \beta^{-2})$.

  – Como Dx_t y γ_t están correlacionados, el estimador de MCO, $\hat{\beta}_1$, **no es insesgado**.

- Para estimar β_1 sin sesgo, se hace uso del siguiente modelo para el crecimiento del PIB

$$E(Dz_t|D\mathbf{x}) = \beta_1 Dx_t \quad \text{para} \quad t = 2, \ldots, N.$$

- Al promediar sobre todas las observaciones del crecimiento, se obtiene

$$E[\textstyle\sum_{t=2}^{N} Dz_t/(N-1)|\, D\mathbf{x}] = \beta_1 \textstyle\sum_{t=2}^{N} Dx_t/(N\text{-}1).$$

- De forma que β_1 puede estimarse con el cociente de promedios de los crecimientos del PIB oficial y de las luces, o sea,

$$\hat{\beta}_1 = \frac{\sum_{t=2}^{N} Dz_t/(N-1)}{\sum_{t=2}^{N} Dx_t\,/(N-1)} = \frac{z_N - z_1}{x_N - x_1}.$$

- Dicho estimador es insesgado, dado D**x**, si los errores en el modelo para Dz_t tienen media cero.
- Pero sólo utiliza el crecimiento de largo plazo de las variables (**no es confiable** para estimar el crecimiento anual del PIB).
- Su interpretación como inverso de la elasticidad de las luces se pierde.

- La propuesta es **usar la mediana**, en lugar del **promedio** de Dx, o sea, si $\mathrm{med}(\mathrm{Dx})$ es la mediana del crecimento de las luces, entonces

$$\hat{\beta}_1 = (\mathrm{Dz})_{\mathrm{med}(\mathrm{Dx})} / \mathrm{med}(\mathrm{Dx}),$$

con $(\mathrm{Dz})_{\mathrm{med}(\mathrm{Dx})}$ el crecimiento del PIB que corresponde a $\mathrm{med}(\mathrm{Dx})$.

- Este estimador es insesgado dado D**x**, y es robusto ante la presencia de mediciones satelitales erróneas.

- Para α se usa la relación $\lambda = \alpha^{-1}/(\alpha^{-1} + \beta^2)$, de forma que una vez obtenido $\hat{\beta}^2$, **α se estima** como

$$\hat{\alpha} = \frac{1-\lambda}{\lambda\hat{\beta}^2}$$

 al elegir un valor apropiado de $\lambda \in (0, 1)$.

- Como la varianza estimada del error de predicción disminuye con λ, se elige el menor valor de λ que cumple con el siguiente Teorema.

  – **Teorema de Tchebysheff**. Sea W una variable aleatoria con media finita μ y varianza σ^2. Entonces, para $k > 0$ se cumple que

$$\Pr\left[\,\left|W - \mu\right| \geq k\sigma\right] \leq 1/k^2.$$

- Así, para $t = 2, \ldots, N$ y $k = 2$, se debe cumplir que

$$\Pr\left[\,\left|\widehat{Dy}_t - Dy_t\right| \geq 2\hat{\sigma}_\varepsilon\sqrt{(\hat{\alpha}^{-1} + \hat{\beta}^2)^{-1}}\right] \leq 1/4.$$

- Por lo tanto, una serie de datos de crecimiento oficial $\{\mathrm{Dz}_t\}$ se considera cercana a $\{\mathrm{Dy}_t\}$ (por ello es una **buena representación del verdadero crecimiento del PIB**) si a lo más ¼ de los datos de $\{\mathrm{Dz}_t\}$ están fuera de los intervalos de $\pm$ 2 errores estándar

$$\widehat{\mathrm{Dy}}_t \pm 2\hat{\sigma}_\varepsilon\sqrt{(\hat{\alpha}^{-1}+\hat{\beta}^2)^{-1}} \text{ para } t = 2, \ldots, \mathrm{N}.$$

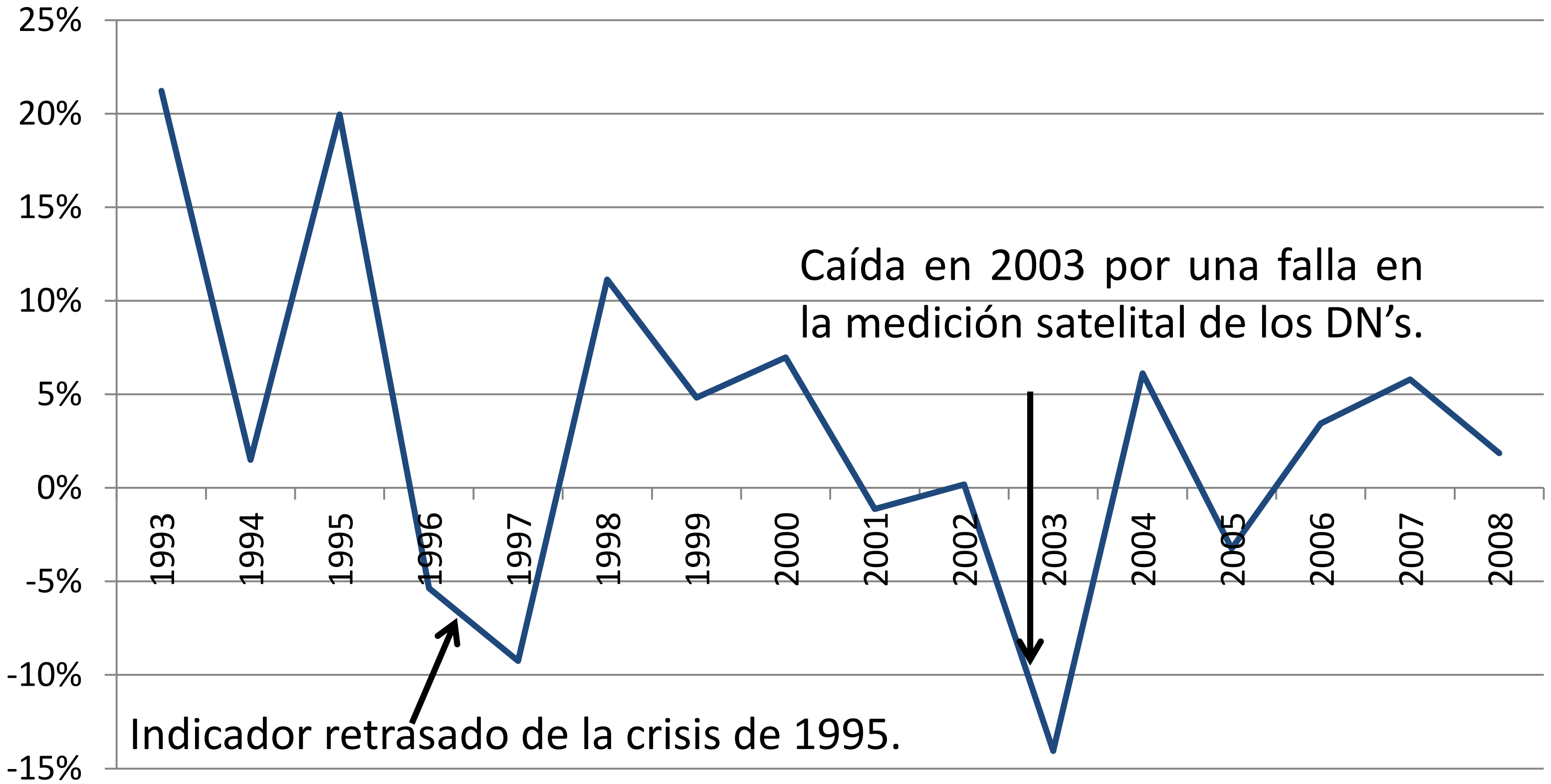
- Como no hay supuesto distribucional, no se puede hacer inferencia estadística propiamente, es decir, no se puede asignar nivel de confianza a los intervalos para los parámetros, ni probabilidad a los intervalos de predicción.

- Para resumir el procedimento, conviene enunciar los **pasos que deben seguirse** para aplicarlo.

1. Estimar β con $\hat{\beta}_1^{-1}$ (en función de la mediana de Dx).
2. Elegir un valor de λ entre 0 y 1.
3. Calcular $\hat{\alpha}$, $\hat{\sigma}_\varepsilon^2$ y $\widehat{D\mathbf{y}}$ con los valores $\hat{\beta}$ y λ.
4. Calcular intervalos de $\pm 2\hat{\sigma}_\varepsilon$ para el crecimiento del PIB y contar los elementos de $\{Dz_t\}$ fuera de dichos intervalos.
5. Repetir los pasos 2, 3 y 4 para otros valores de λ. El menor de estos valores, que deje ¼ de los elementos de $\{Dz_t\}$ fuera de las cotas de Tchebysheff, se considera el estimador $\hat{\lambda}$.

4. Aplicación empírica para México

- La siguiente figura muestra el crecimiento de las luces en México, de 1993 a 2008, medido como el cambio anual del DN promedio.



Luces nocturnas en México

1992

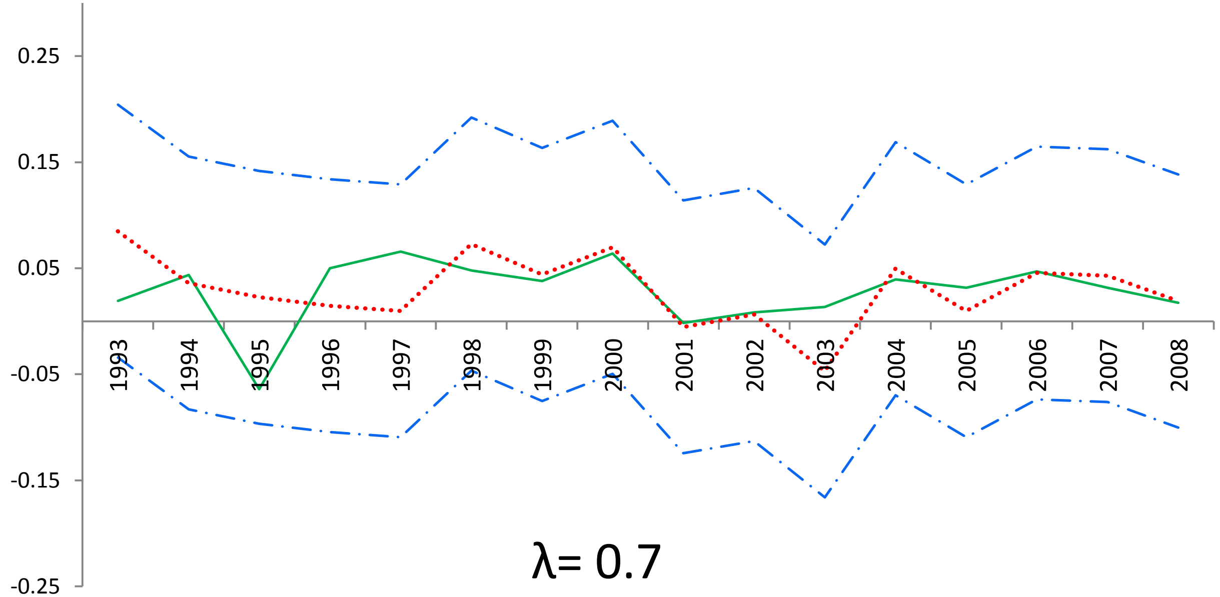
2008

- La med(Dx) para el periodo 1992-2008 se obtuvo como promedio de los crecimientos de DN's para 2006 y 2008, de forma que $\hat{\beta}_1 = 1.238$; y la elasticidad estimada de las luces respecto al ingreso es $\hat{\beta} = 0.808$.
- El supuesto de no-correlación serial se verificó al calcular el coeficiente autorregresivo de orden 1, -0.24, cuyo error estándar 0.29 condujo a un p-valor de 0.42 que no-rechaza el supuesto.
- El cuadro que sigue muestra valores de $\hat{\alpha}$ y $\hat{\sigma}_\varepsilon^2$ para diferentes $\hat{\lambda}$'s.

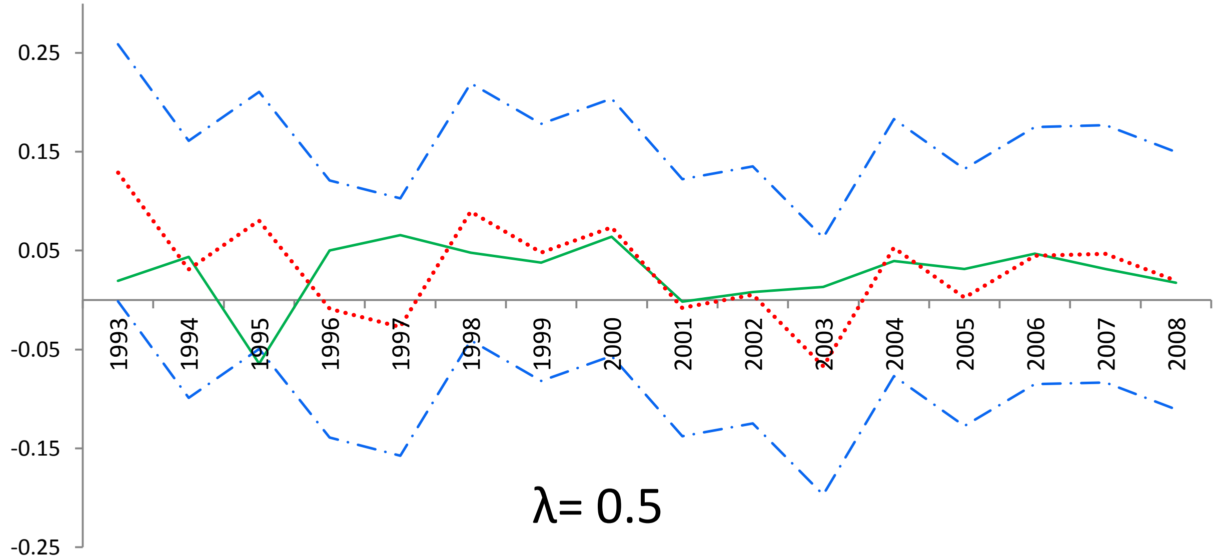
Parámetros estimados correspondientes a $\boldsymbol{\hat{\beta} = 0.808}$

$\hat{\lambda}$	**0.9**	**0.8**	**0.7**	**0.6**	**0.5**	**0.4**	**0.3**	**0.2**	**0.1**
$\hat{\alpha}$	0.170	0.383	0.656	1.021	1.532	2.298	3.574	6.127	13.786
$\hat{\sigma}_\varepsilon^2$	0.010	0.009	0.008	0.007	0.006	0.004	0.003	0.002	0.001

Crecimientos oficial (línea sólida) y estimado (punteada), con intervalos de $\pm 2\hat{\sigma}_\varepsilon$

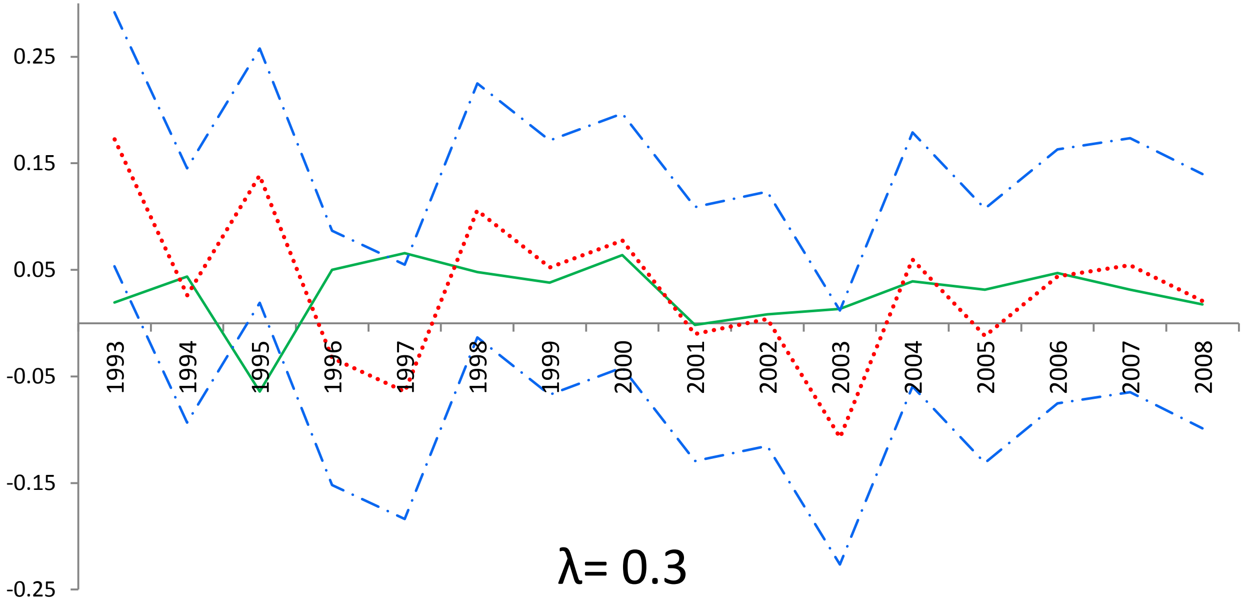


No hay observación que salga de su intervalo. (0/16 < 1/4)

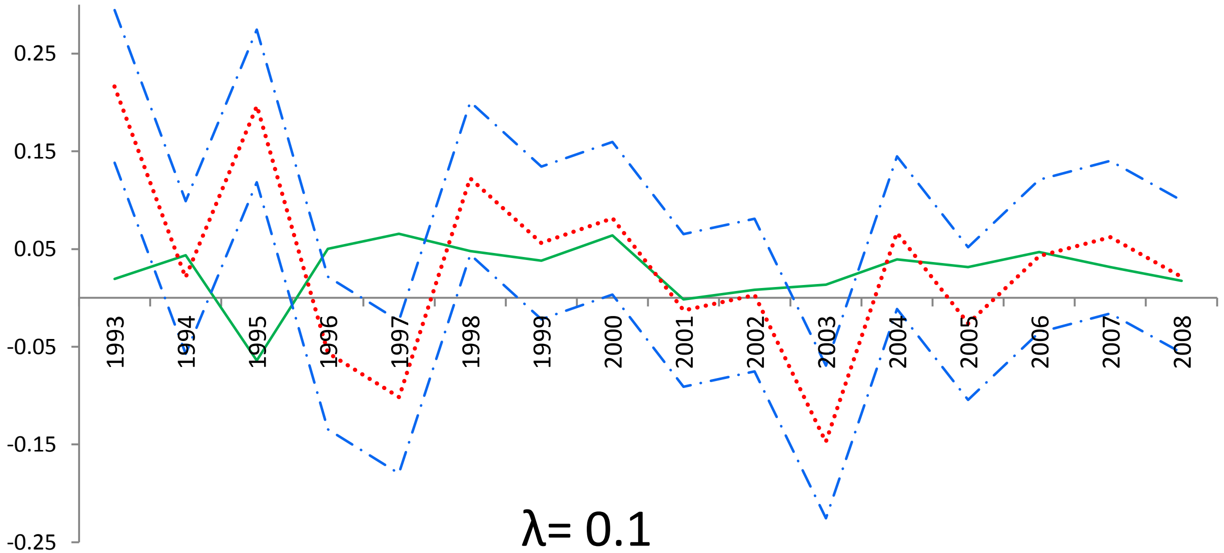


Sólo **una observación** sale de su intervalo. (1/16 < 1/4)

Crecimientos oficial (línea sólida) y estimado (punteada), con intervalos de $\pm 2\hat{\sigma}_{\varepsilon}$

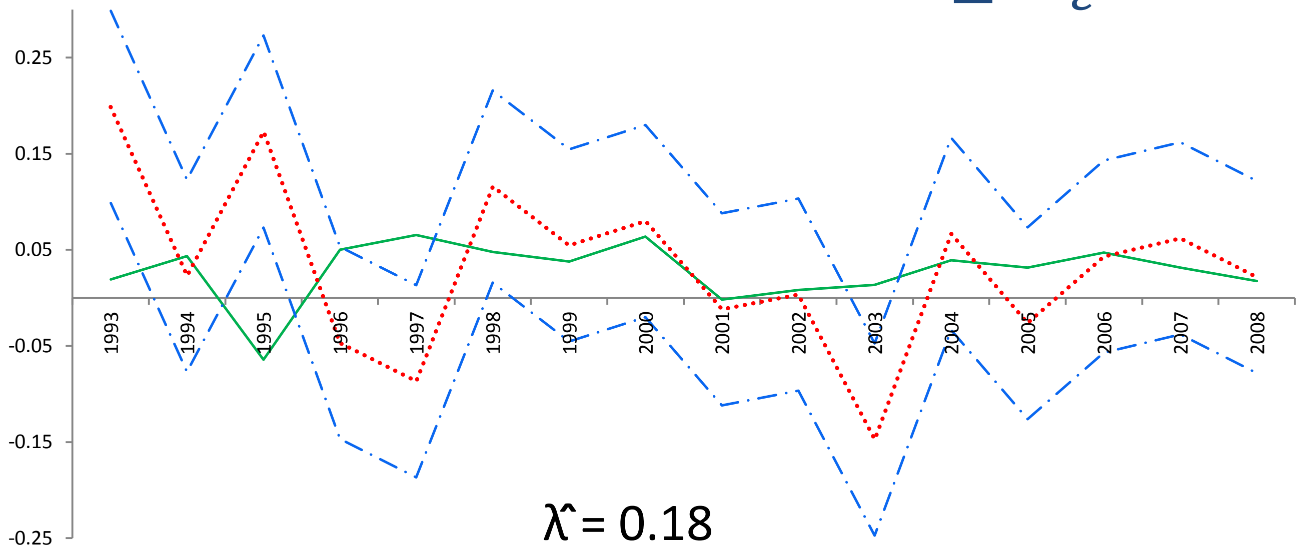


Cuatro observaciones están fuera de los intervalos. (4/16 = 1/4)



Cinco observaciones se salen de sus intervalos. (5/16 > 1/4)

Crecimientos oficial (línea sólida) y estimado (punteada), con intervalos de $\pm 2\hat{\sigma}_{\varepsilon}$



Cuatro observaciones caen fuera de sus intervalos. (4/16 = 1/4)

- Nótese que $\hat{\boldsymbol{\lambda}}$ **= 0.18 es el menor valor** (entre los que se usaron) que cumple con el requisito del Teorema de Tchebysheff.
- La estimación del verdadero PIB real conduce a un crecimiento promedio para 1992-2008 de 3.27%, mientras que el crecimiento ofcial promedio es de 2.82%.
  - Por lo tanto, el crecimiento real se subestimó 0.45% por año.
  - HSW dedujeron una subestimación anual de 0.43%.

5. Consideraciones finales

- El modelo de HSW permite mejorar el cálculo oficial de crecimiento del PIB; aquí se propone una simplificación del procedimiento de cálculo, que emplea la intensidad de las luces en forma eficiente.

- El método propuesto estima los parámetros de cada país **en forma individual** y **robusta**, con un método simple y no-parámetrico.

- No usa calificaciones para los datos, ni la misma ponderación del crecimiento de la intensidad de las luces, para distintos países

- La **falta de credibilidad** en los datos oficiales se mitiga al usar datos de luces nocturnas, para medir el crecimiento en forma complementaria.

¡Gracias por su atención!

Referencias

Aruoba BS, Diebold FX, Nalewaik J, Shorfheide F, Song D (2013). Improving GDP Measurements: A Measurement-Error Perspective. **National Bureau of Economic Research**, Working paper 18954: 1-34.

Chen X, Nordhaus W (2011). Using luminosity data as a proxy for economic statistics. **The Proceedings of National Academy of Sciences** 108(21): 8589-8594.

Henderson JV, Storeygard A, Weil DN (2012). Measuring Economic Growth from Outer Space. **American Economic Review** 102(2): 994-1028.

Nordhaus W, Chen X (2014). A sharper image? Estimates of the precision of nighttime lights as a proxy for economic statistics. **Journal of Economic Geography** (Advanced Access, May 7, 2014): 1-30.