Calculation of early estimates for the monthly manufacturing activity level index in Mexico as a function of electric energy consumption

COMISIÓN FEDERAL DE ELECTRICIDAD

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

11 September 2018

Use of Administrative Registers

Major challenge for National Statistical Offices:

Use of administrative registers for statistical purposes

Use of Administrative Registers

- INEGI has worked with national government agencies in Mexico to harmonize and link administrative registers to INEGI´s Statistical Business Register (SRB).
- INEGI's SBR is called RENEM: "Registro Estadístico de NEgocios en México"
- With administrative data linked to the SBR at establishment level, it is possible to build indicators which can be useful to explain the behavior of related existing economic variables
- The most significant advances in this respect, are with the electric utility company (CFE) and the tax administration agency (SAT). This presentation focuses on the work made jointly between CFE and INEGI

Objectives of using CFE´s administrative data

1. Link CFE´s data to a master sample (from the SBR), which contains Mexico's largest establishments in the industry, trade and services sectors. From this process, INEGI obtains an electric energy consumption index for the manufacturing sector (ICEE)

2. Use the ICEE in order to obtain an early estimate for the monthly manufacturing activity level index (IMAI 31-33) through an econometric model. This is possible to implement given the timeliness of CFE´s data, together with the high linear correlation observed between IMAI 31-33 and ICEE

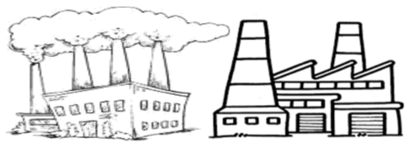
Master sample coverage in the manufacturing sector

Sector/ Subsector	Description	Coverage Percentage					
		Master Sample			Master Sample linked to CFE data		
		Establishments	Revenue	Employees	Establishments	Revenue	Employees
31-33	Manufacturing industries	4%	88%	68%	3%	79%	57%
311	Food products	2%	89%	47%	1%	77%	39%
312	Beverage and tobacco products	3%	93%	59%	2%	87%	48%
313	Textile mills	3%	95%	70%	2%	75%	53%
314	Textile product mills	1%	77%	30%	1%	65%	25%
315	Apparel manufacturing	4%	81%	66%	3%	64%	54%
316	Leather and allied products	7%	81%	60%	5%	75%	54%
321	Wood products	2%	66%	28%	1%	55%	22%
322	Paper products	12%	95%	82%	9%	72%	62%
323	Printing and related support activities	3%	73%	41%	2%	63%	36%
324	Petroleum and coal products	51%	81%	75%	42%	80%	74%
325	Chemical manufacturing	30%	72%	78%	21%	61%	65%
326	Plastics and rubber products	27%	92%	79%	21%	79%	67%
327	Nonmetallic mineral products	4%	93%	56%	2%	65%	38%
331	Primary metal manufacturing	44%	99%	91%	34%	89%	78%
332	Fabricated metal products	2%	84%	46%	1%	66%	38%
333	Machinery manufacturing	27%	94%	82%	21%	86%	73%
334	Computer and electronic products	60%	96%	93%	50%	89%	83%
335	Electrical equipment, appliance, components	44%	93%	86%	36%	85%	77%
336	Transportation equipment	48%	98%	94%	38%	91%	80%
337	Furniture and related products	2%	73%	43%	1%	59%	37%
339	Miscellaneous manufacturing	4%	89%	75%	3%	80%	67%

Joint Work CFE-INEGI

Comisión Federal de Electricidad

CFE provides INEGI with electric energy consumption data at contract level (establishment), approximately 4.8 million records per month



Industrial

Trade

Services

Agriculture

CFE - INEGI have worked on:

- Data harmonization
- Data linking
- Data Integration
- Data analysis

Objective: To produce monthly indicators on electric energy consumption

Results Obtained with linked CFE´s data

- Using records from the master sample linked to CFE data, INEGI builds the Electric Energy Consumption Index (ICEE) for the Manufacturing sector

- the Monthly Manufacturing Activity Level Index (IMAI 31-33) is published by the System of National Accounts, approximately 40 days after the end of the reference month

- Given the opportunity with which the ICEE index is built (approximately 12 to 15 days after the end of the reference month), and its high linear correlation with IMAI 31-33, it is feasible to obtain an early estimate for IMAI 31-33 through a linear regression model

Electric Energy Consumption Index (ICEE)

Construction of the ICEE index (X_t variable):

1. For month t, the electric energy consumption (in kWh) for each record (establishment) in the linked sample SBR-CFE is multiplied by a weight which depends on the manufacturing subsector the establishment belongs to; this weight also depends on month t, and is provided by the System of National Accounts. Note that electric energy consumption data comes from CFE, while economic activity information comes from the SBR

2. All weighted electric energy consumption values are added, obtaining S_t

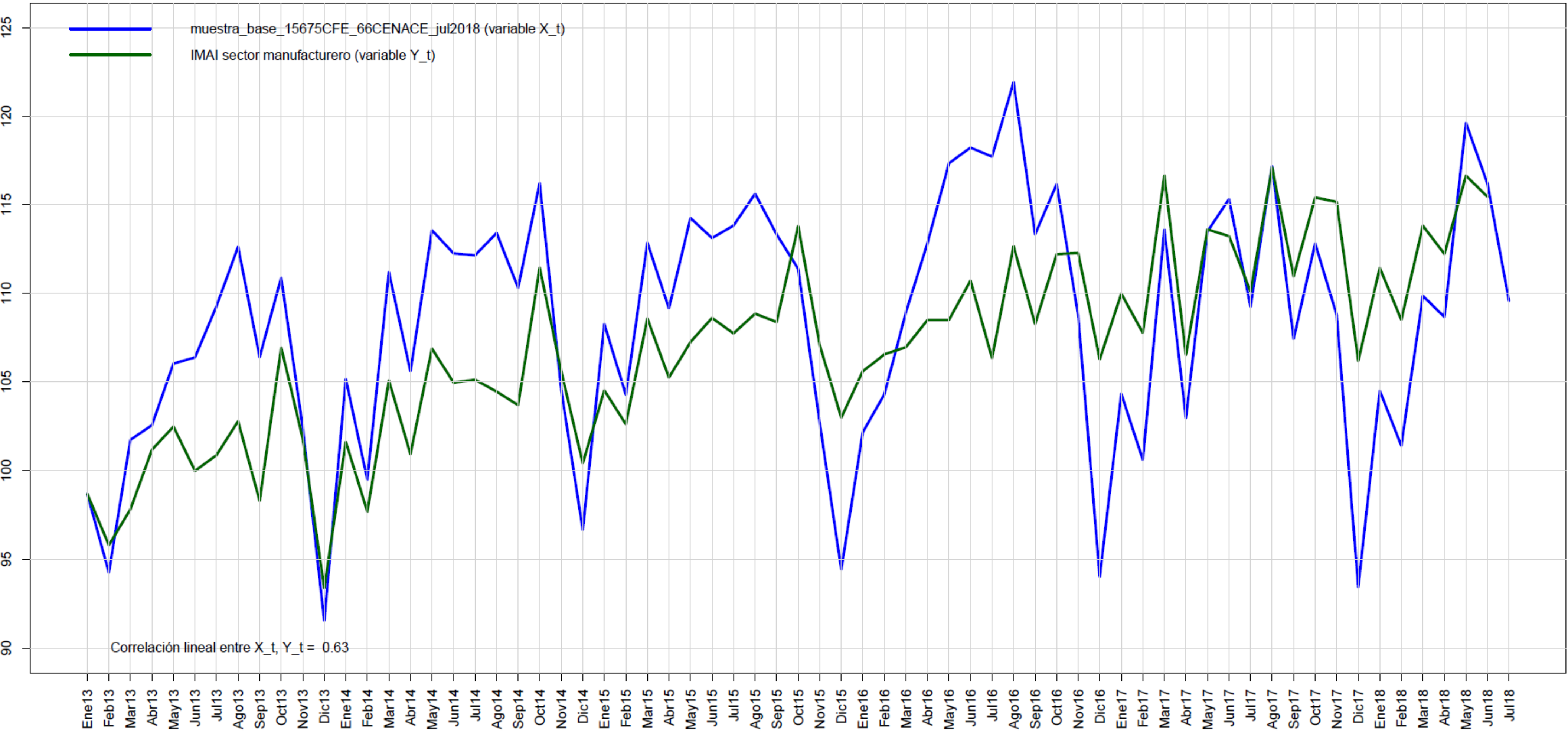
3. Finally, S_t scale is changed to coincide with IMAI 31-33 (variable Y_t) on a base month (January 2013)

Electric Energy Consumption Index (ICEE)

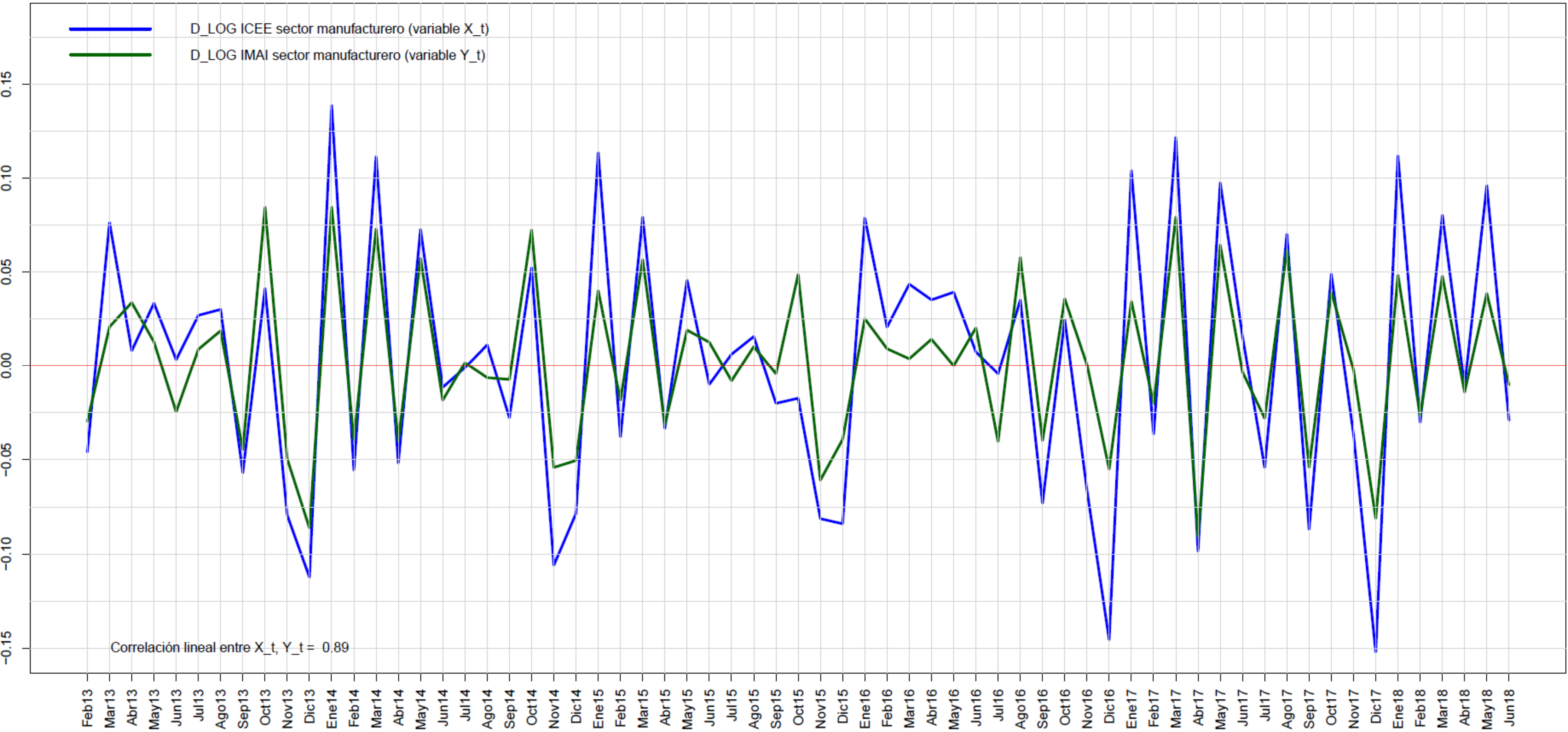
Weights for the manufacturing subsectors, provided by the System of National Accounts SNA in Mexico

Subsector	Description	May_2018	Jun_2018	Jul_2018
311	Food products	0.224	0.221	0.224
312	Beverage and tobacco products	0.060	0.059	0.060
313	Textile mills	0.009	0.010	0.009
314	Textile product mills	0.004	0.005	0.005
315	Apparel manufacturing	0.020	0.020	0.020
316	Leather and allied products	0.008	0.008	0.008
321	Wood products	0.009	0.009	0.008
322	Paper products	0.018	0.018	0.018
323	Printing and related support activities	0.007	0.007	0.007
324	Petroleum and coal products	0.016	0.014	0.013
325	Chemical manufacturing	0.083	0.084	0.087
326	Plastics and rubber products	0.027	0.027	0.027
327	Nonmetallic mineral products	0.026	0.025	0.026
331	Primary metal manufacturing	0.064	0.065	0.068
332	Fabricated metal products	0.035	0.036	0.035
333	Machinery manufacturing	0.045	0.046	0.044
334	Computer and electronic products	0.081	0.082	0.082
335	Electrical equipment, appliance, components	0.029	0.033	0.031
336	Transportation equipment	0.203	0.201	0.194
337	Furniture and related products	0.010	0.010	0.010
339	Miscellaneous manufacturing	0.020	0.020	0.021

ICEE and IMAI 31-33 time series

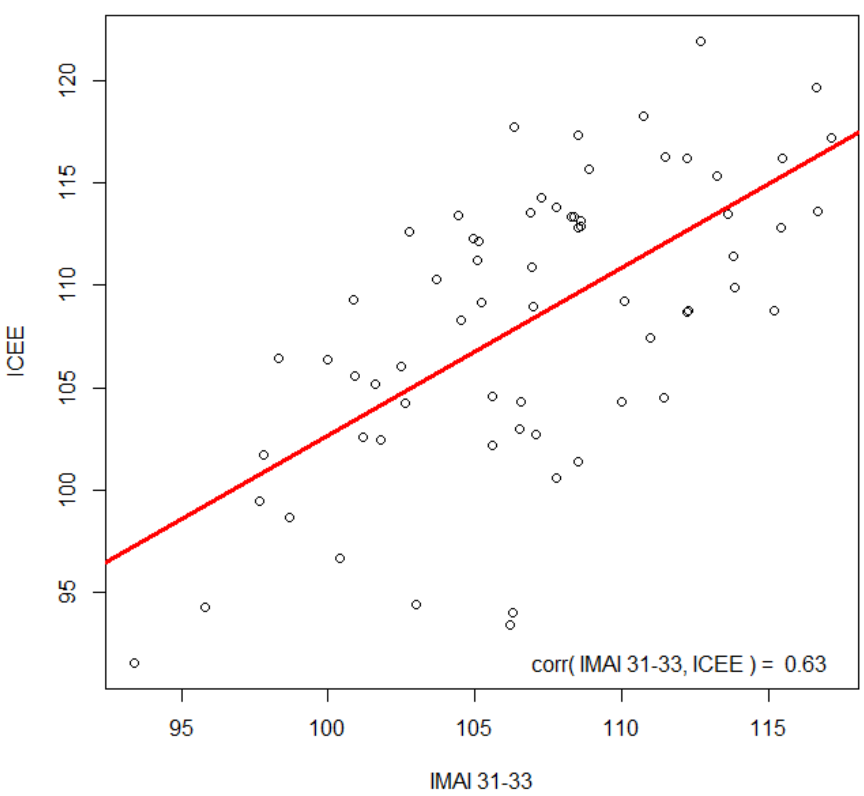


Logarithmic differences: ICEE and IMAI 31-33

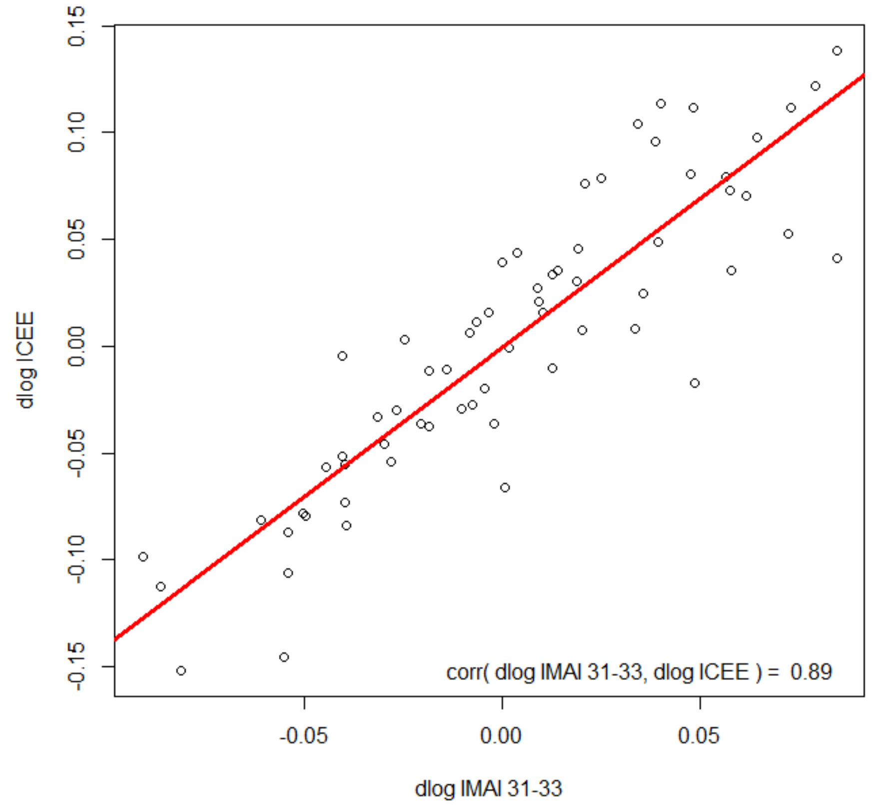


Scatter plots for ICEE and IMAI 31-33

Original Variables



Transformed Variables



Regression Model for Estimating IMAI 31-33

Logarithmic differences as approximations to monthly and annual variations

Let X_t be a time series, where sub index $t = 1,2,3, \dots$ distinguishes months. Let $vm_t \coloneqq \frac{X_t}{X_{t-1}} - 1$ and $va_t \coloneqq \frac{X_t}{X_{t-12}} - 1$ be the monthly and annual variations, respectively, for series X at month t. X_t logarithmic difference for two consecutive months $t-1, t$ is defined as:

$$\nabla_m \ln(X_t) := \ln(X_t) - \ln(X_{t-1})$$

Using properties of logarithms,

$$\nabla_m \ln(X_t) = \ln\left(\frac{X_t}{X_{t-1}}\right) = \ln\left[\frac{(1+vm_t)X_{t-1}}{X_{t-1}}\right] = \ln(1 + vm_t) \approx vm_t$$

In other words, $\widehat{vm}_t := \nabla_m \ln(X_t)$ is a good approximation for X_t monthly variation.

Similarly, $\nabla_a \ln(X_t) := \ln(X_t) - \ln(X_{t-12})$ approximates X_t annual variation.

Note that $\nabla_a \ln(X_t) = [\ln(X_t) - \ln(X_{t-1})] + [\ln(X_{t-1}) - \ln(X_{t-2})] + \cdots + [\ln(X_{t-11}) - \ln(X_{t-12})]$

i.e., $\widehat{va}_t = \widehat{vm}_t + vm_{t-1} + \cdots + vm_{t-11}$

Regression Model for Estimating IMAI 31-33

Logarithmic difference model for estimating IMAI 31-33
as a function of the electric energy consumption index (ICEE)

$$\nabla_m \ln Y_t = \beta_1 \nabla_m \ln X_t + \beta_2 i_{oct} + \beta_3 i_{nov} + \beta_4 i_{dic} + \beta_5 i_{ene} + \varepsilon_t,$$
$$\varepsilon_t = \rho \varepsilon_{t-1} + \upsilon_t$$

Y_t is IMAI 31-33 for month t; X_t is the ICEE index for month t

i_{oct} = 1 for October 2013, 2014, 2015, and 2016; 0 otherwise
i_{nov} =1 for November 2016 and 2017; 0 otherwise
i_{dic} = 1 for December 2016 and 2017; 0 otherwise
i_{ene} =1 for January 2015, 2016, 2017 and 2018; 0 otherwise

Note : ε_t possesses an AR(1) structure.
Cochrane-Orcutt method is used for estimating the model's parameters.

Regression Model for Estimating IMAI 31-33

Obtaining IMAI 31-33 estimations from the logarithmic difference model

The logarithmic difference model generates direct estimations $\widehat{vm}_t$ for IMAI 31-33 monthly variations. From these values $\widehat{vm}_t$, it is also possible to obtain estimations $\hat{Y}_t$ for the IMAI 31-33 index itself:

from $\widehat{vm}_t = \ln \frac{\hat{Y}_t}{Y_{t-1}}$, one obtains

$$\begin{aligned} \ln \hat{Y}_t - \ln Y_{t-1} &= \widehat{vm}_t \\ \ln \hat{Y}_t &= \widehat{vm}_t + \ln Y_{t-1} \\ \boldsymbol{\hat{Y}_t} &= \mathbf{exp}(\boldsymbol{\widehat{vm}_t})\, \boldsymbol{Y_{t-1}} \end{aligned}$$

To estimate IMAI 31-33 for month t, we multiply the true IMAI 31-33 value for month $t-1$ by the exponential function (natural logarithm inverse) evaluated at the monthly variation estimation for month t.

Regression Model for Estimating IMAI 31-33

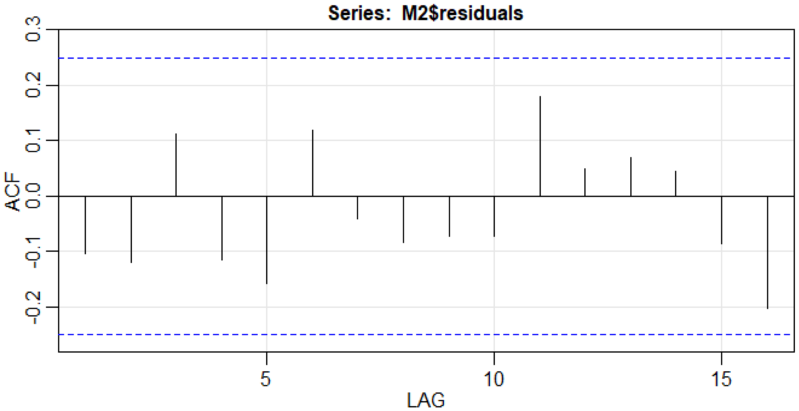
Coefficients after fitting a Logarithmic Difference model to the variables:

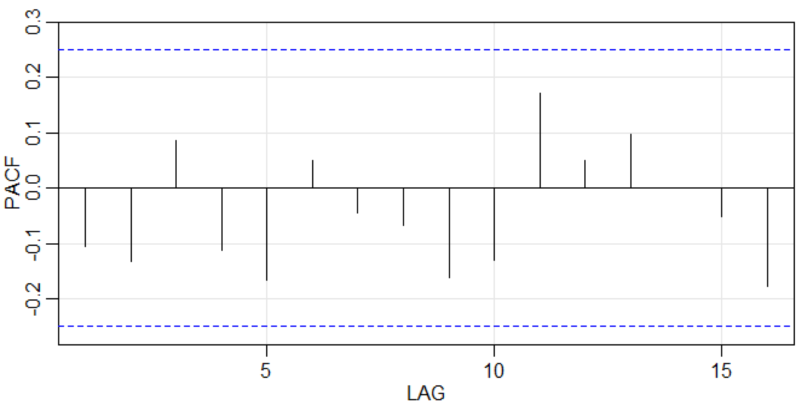
```
Call:
lm(formula = COdlog_Y ~ COdlog_X + COOCT + CONOV + CODIC + COENE - 1)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.031464 -0.009098 -0.002061  0.005431  0.022554 

Coefficients:
          Estimate Std. Error t value Pr(>|t|)    
COdlog_X  0.577010   0.033256  17.351  < 2e-16 ***
COOCT     0.046529   0.006402   7.268 9.53e-10 ***
CONOV     0.025890   0.009738   2.659  0.01008 *  
CODIC     0.019346   0.010982   1.762  0.08330 .  
COENE    -0.022526   0.007072  -3.185  0.00231 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.01363 on 59 degrees of freedom
Multiple R-squared:  0.8676,    Adjusted R-squared:  0.8563 
F-statistic:  77.3 on 5 and 59 DF,  p-value: < 2.2e-16

CO Rho Coefficient:
       Estimate Std. Error t value Pr(>|t|)    
e1[-n]  -0.4006     0.1156  -3.465 0.000959 ***
---
```

Regression Model for Estimating IMAI 31-33

Adjusted Model´s diagnostics with Cochrane-Orcutt correction:

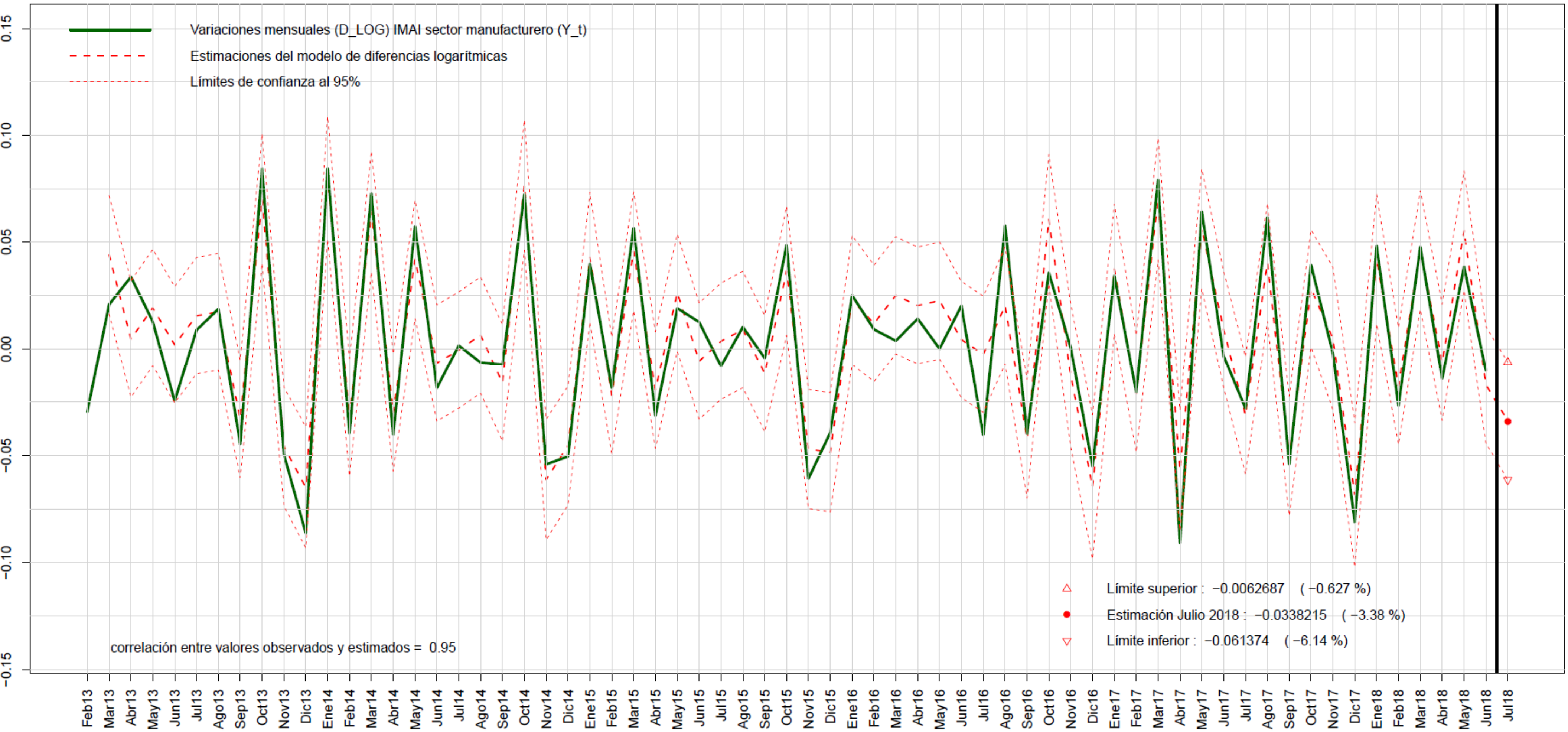
```
> # durbin-watson test
> dwt(M2)
 lag Autocorrelation D-W Statistic p-value
   1     -0.06334221      2.072043   0.604
 Alternative hypothesis: rho != 0

> # Breusch-Pagan test
> ncvTest(M2)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 4.651002e-06    Df = 1     p = 0.9982793

> # normality of residuals test
> shapiro.test(M2$residuals)

        Shapiro-Wilk normality test

data:  M2$residuals
W = 0.9767, p-value = 0.2666
```



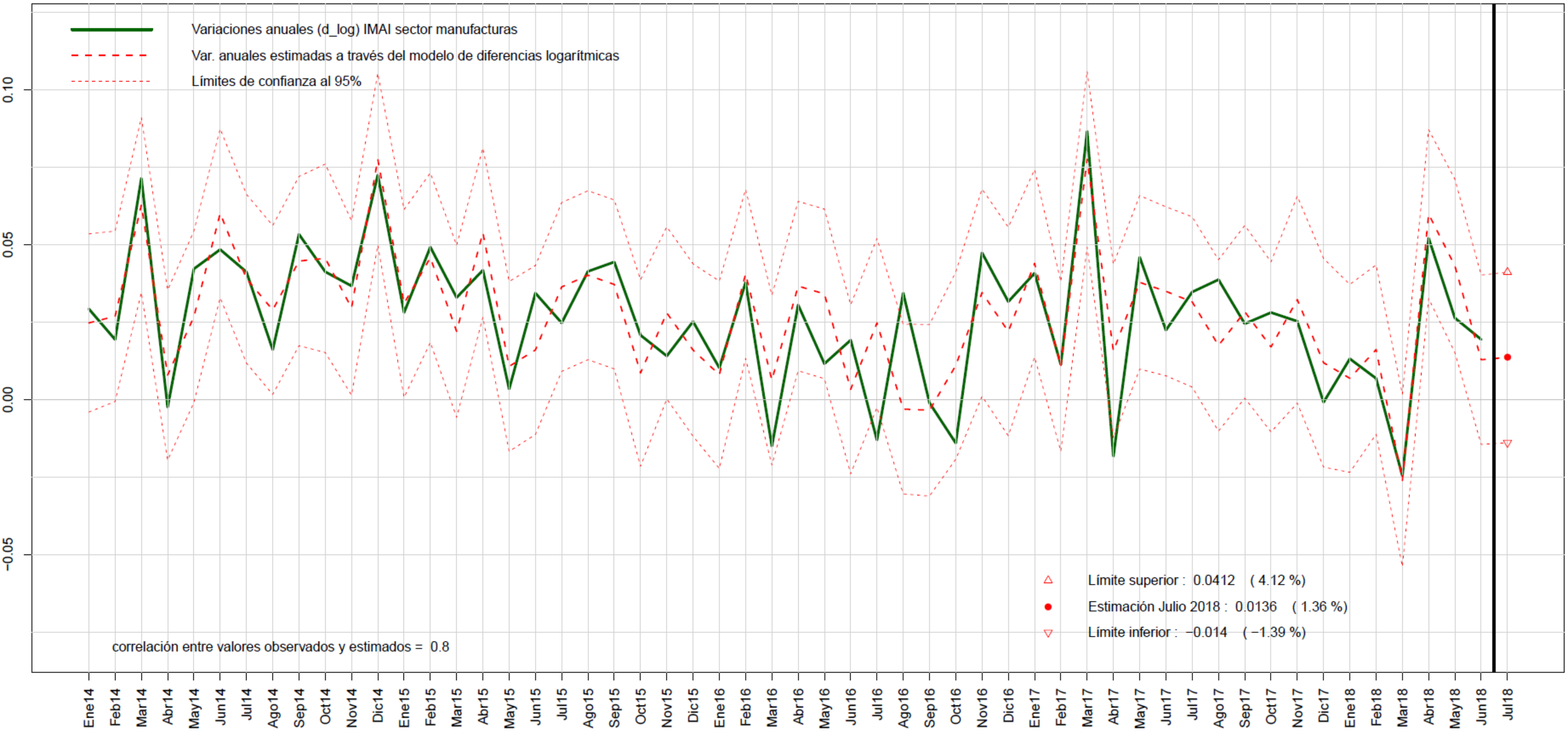


Estimates obtained with the model

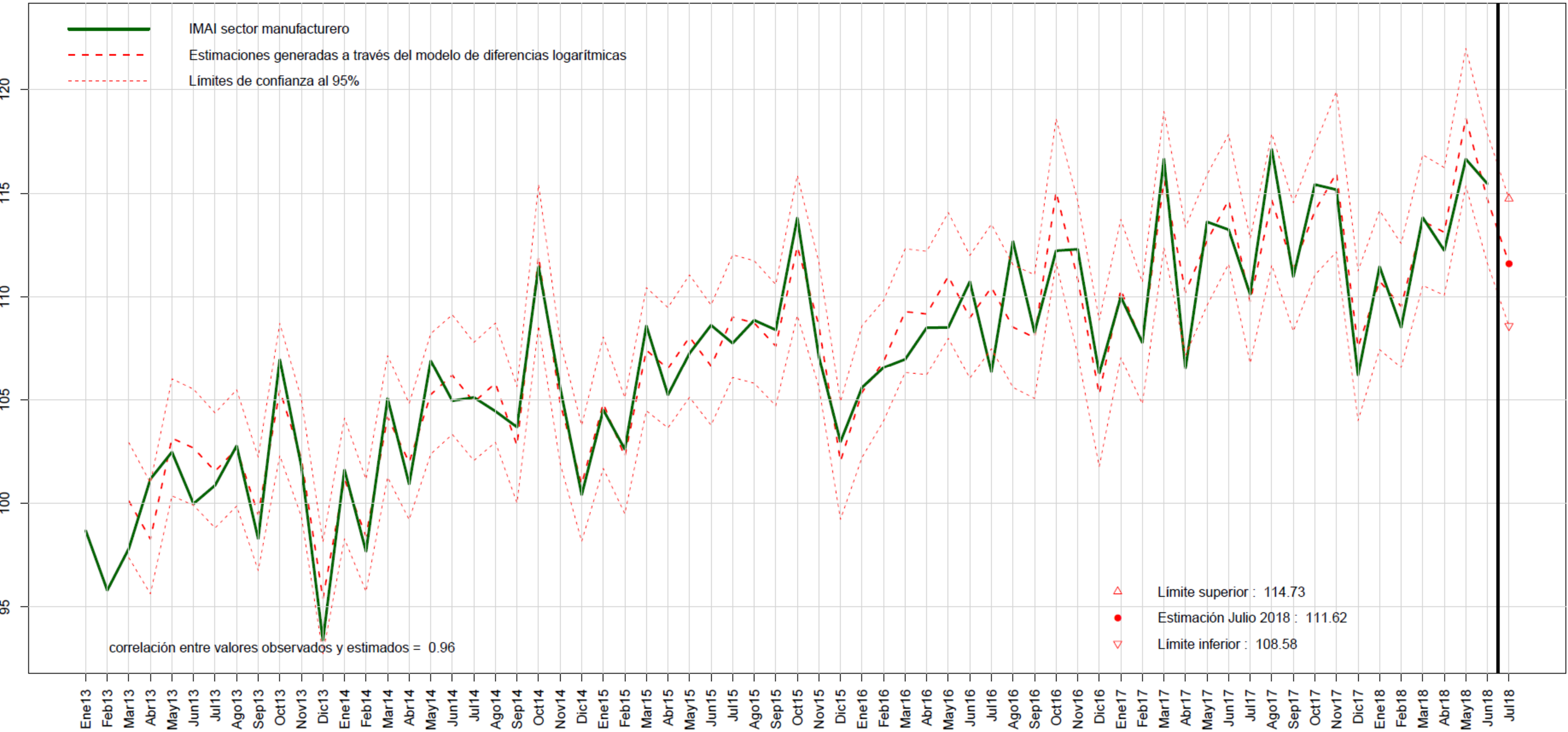
Estimates for monthly variations up to July 2018



Estimates for annual variations up to July 2018

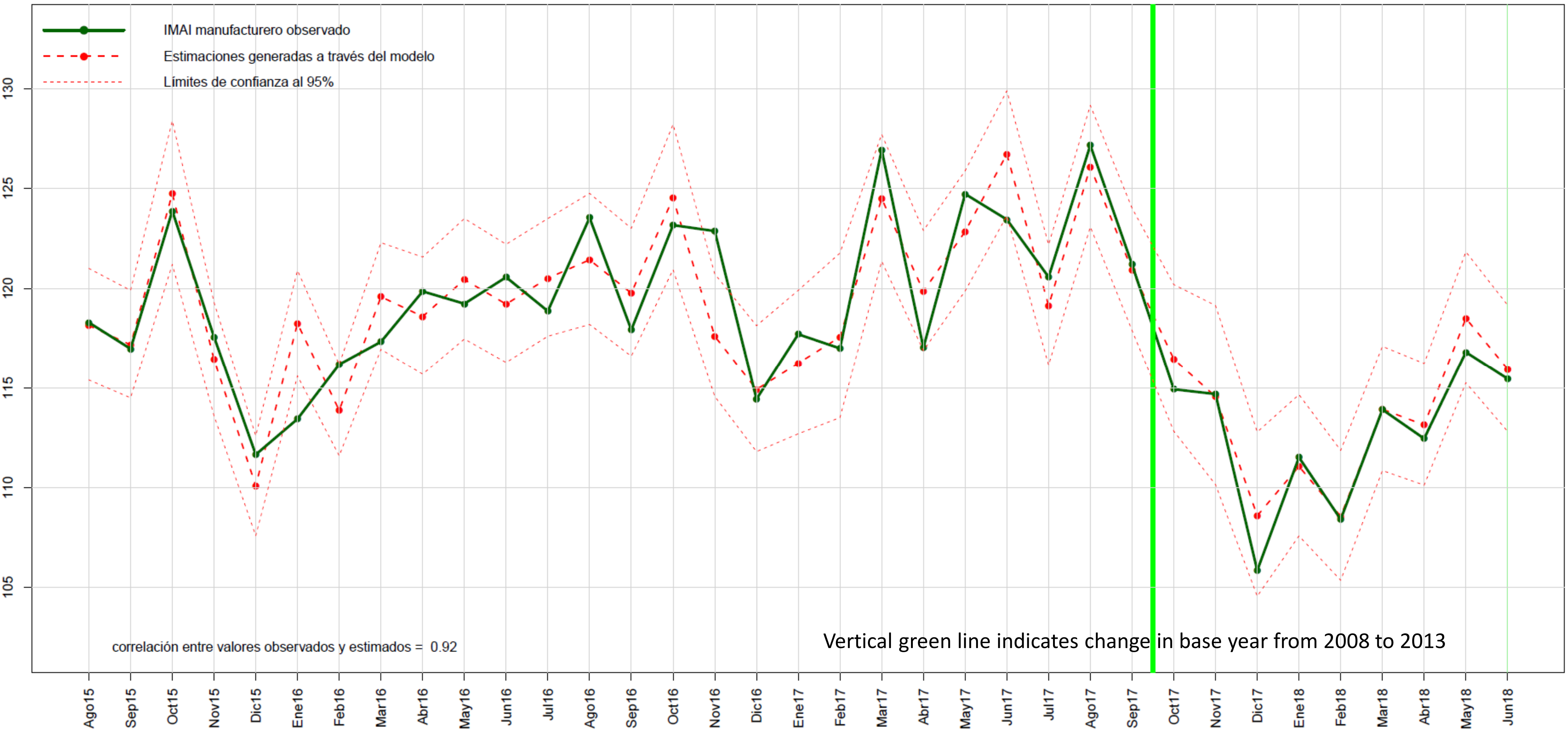


Estimates for the IMAI 31-33 index up to July 2018

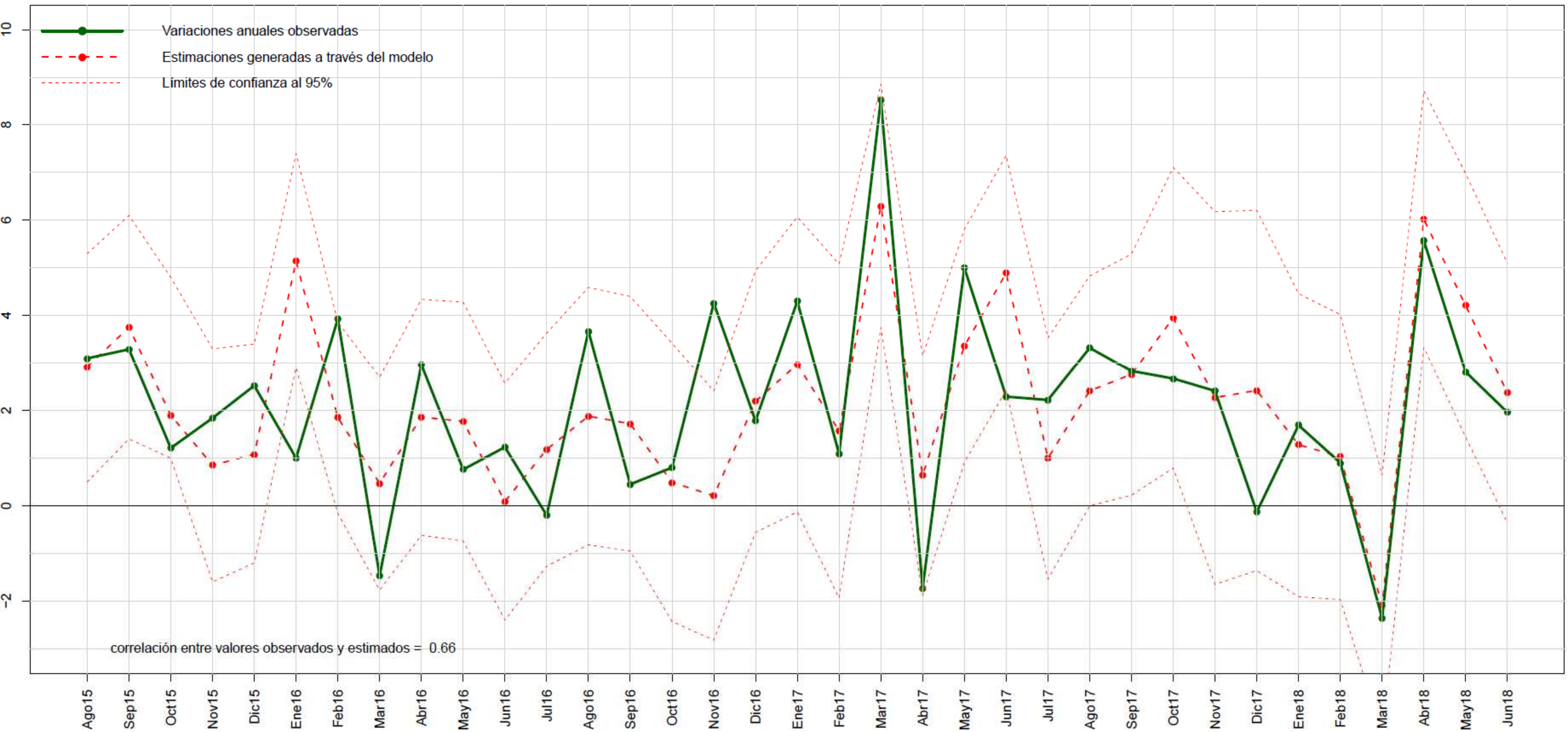


Historical assessments for model's estimates in real time

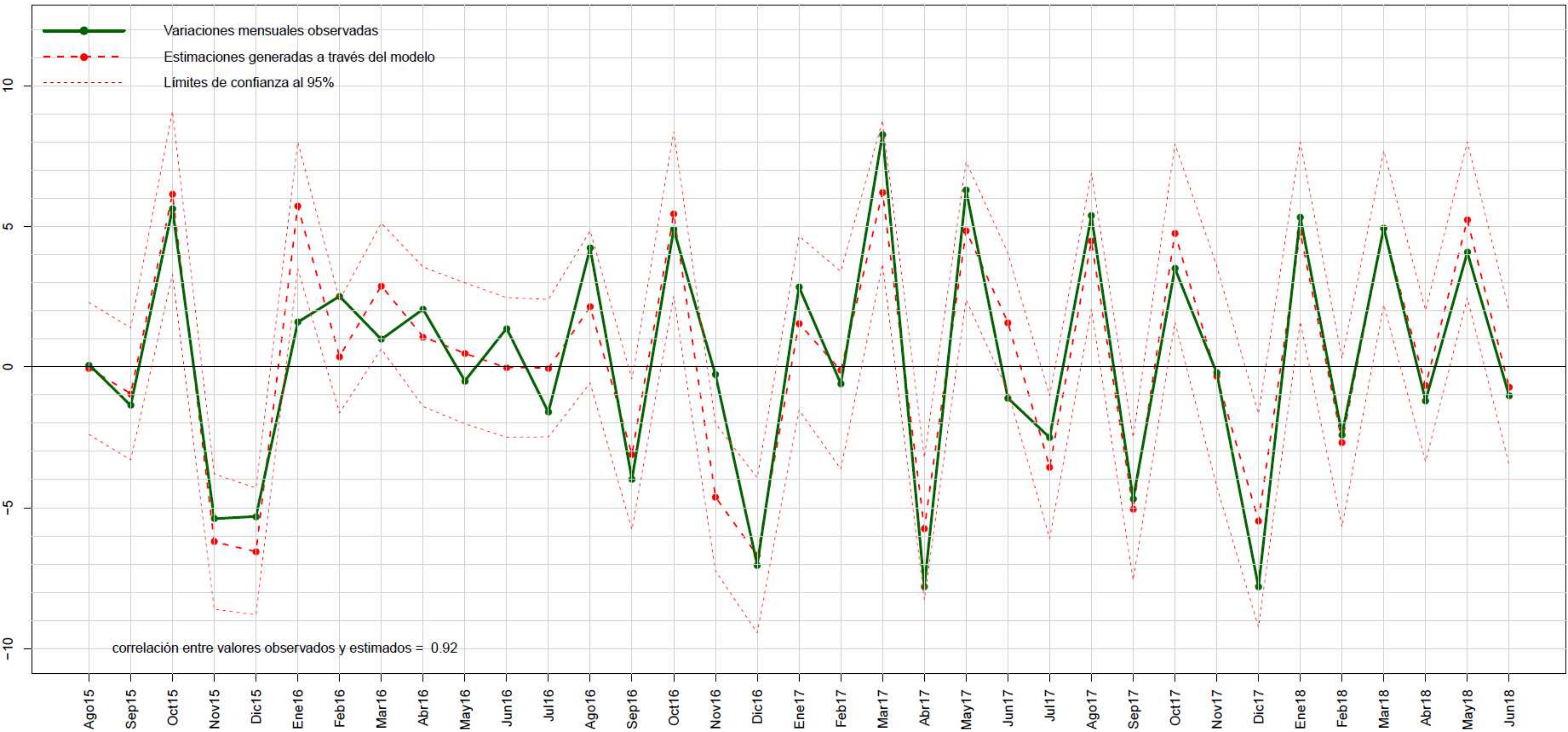
Assessments for IMAI 31-33 early estimates



Assessments for estimated annual variations



Assessments on estimated monthly variations



Current, future work and perspectives

- Currently, these results are communicated to officers from some Mexican government agencies, like the Central Bank, the Secretariat of Economy and the Tax office. The documents related to these communications clearly state that these results are still in an experimental phase

- INEGI is currently working in order to publish these results in its web page, as experimental Statistics

- Although the electric energy consumption data involved in this project can't be classified as big data in the strict sense, the involved data linkage process certainly involves a great data volume

- If an electric utility company could provide electric energy consumption data by connecting its electric meters to a data network, then more precise, opportune and detailed economic statistics could be produced; this could be viewed as an extension of the project presented here

Thank you for your attention

Questions?

Conociendo México

01 800 111 46 34

www.inegi.org.mx

atencion.usuarios@inegi.org.mx

@inegi_informa

INEGI Informa

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA