Calculation of early estimates for the monthly manufacturing activity level index in Mexico as a function of electric energy consumption
Major challenge for National Statistical Offices:

Use of administrative registers for statistical purposes

- Electric energy consumption
- Taxes
- Social security
- Organizations
- Water consumption
- Associations
- Government offices
• INEGI has worked with national government agencies in Mexico to harmonize and link administrative registers to INEGI’s Statistical Business Register (SBR).

• 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
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

<table>
<thead>
<tr>
<th>Sector/Subsector</th>
<th>Description</th>
<th>Coverage Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Master Sample</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Establishments</td>
</tr>
<tr>
<td>31-33</td>
<td>Manufacturing industries</td>
<td>4%</td>
</tr>
<tr>
<td>311</td>
<td>Food products</td>
<td>2%</td>
</tr>
<tr>
<td>312</td>
<td>Beverage and tobacco products</td>
<td>3%</td>
</tr>
<tr>
<td>313</td>
<td>Textile mills</td>
<td>3%</td>
</tr>
<tr>
<td>314</td>
<td>Textile product mills</td>
<td>1%</td>
</tr>
<tr>
<td>315</td>
<td>Apparel manufacturing</td>
<td>4%</td>
</tr>
<tr>
<td>316</td>
<td>Leather and allied products</td>
<td>7%</td>
</tr>
<tr>
<td>321</td>
<td>Wood products</td>
<td>2%</td>
</tr>
<tr>
<td>322</td>
<td>Paper products</td>
<td>12%</td>
</tr>
<tr>
<td>323</td>
<td>Printing and related support activities</td>
<td>3%</td>
</tr>
<tr>
<td>324</td>
<td>Petroleum and coal products</td>
<td>51%</td>
</tr>
<tr>
<td>325</td>
<td>Chemical manufacturing</td>
<td>30%</td>
</tr>
<tr>
<td>326</td>
<td>Plastics and rubber products</td>
<td>27%</td>
</tr>
<tr>
<td>327</td>
<td>Nonmetallic mineral products</td>
<td>4%</td>
</tr>
<tr>
<td>332</td>
<td>Primary metal manufacturing</td>
<td>44%</td>
</tr>
<tr>
<td>332</td>
<td>Fabricated metal products</td>
<td>2%</td>
</tr>
<tr>
<td>333</td>
<td>Machinery manufacturing</td>
<td>27%</td>
</tr>
<tr>
<td>334</td>
<td>Computer and electronic products</td>
<td>60%</td>
</tr>
<tr>
<td>335</td>
<td>Electrical equipment, appliance, components</td>
<td>44%</td>
</tr>
<tr>
<td>336</td>
<td>Transportation equipment</td>
<td>48%</td>
</tr>
<tr>
<td>337</td>
<td>Furniture and related products</td>
<td>2%</td>
</tr>
<tr>
<td>339</td>
<td>Miscellaneous manufacturing</td>
<td>4%</td>
</tr>
</tbody>
</table>
CFE provides INEGI with electric energy consumption data at contract level (establishment), approximately 4.8 million records per month.

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
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

<table>
<thead>
<tr>
<th>Subsector</th>
<th>Description</th>
<th>May_2018</th>
<th>Jun_2018</th>
<th>Jul_2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>311</td>
<td>Food products</td>
<td>0.224</td>
<td>0.221</td>
<td>0.224</td>
</tr>
<tr>
<td>312</td>
<td>Beverage and tobacco products</td>
<td>0.060</td>
<td>0.059</td>
<td>0.060</td>
</tr>
<tr>
<td>313</td>
<td>Textile mills</td>
<td>0.009</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td>314</td>
<td>Textile product mills</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>315</td>
<td>Apparel manufacturing</td>
<td>0.020</td>
<td>0.020</td>
<td>0.020</td>
</tr>
<tr>
<td>316</td>
<td>Leather and allied products</td>
<td>0.008</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>321</td>
<td>Wood products</td>
<td>0.009</td>
<td>0.009</td>
<td>0.008</td>
</tr>
<tr>
<td>322</td>
<td>Paper products</td>
<td>0.018</td>
<td>0.018</td>
<td>0.018</td>
</tr>
<tr>
<td>323</td>
<td>Printing and related support activities</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>324</td>
<td>Petroleum and coal products</td>
<td>0.016</td>
<td>0.014</td>
<td>0.013</td>
</tr>
<tr>
<td>325</td>
<td>Chemical manufacturing</td>
<td>0.083</td>
<td>0.084</td>
<td>0.087</td>
</tr>
<tr>
<td>326</td>
<td>Plastics and rubber products</td>
<td>0.027</td>
<td>0.027</td>
<td>0.027</td>
</tr>
<tr>
<td>327</td>
<td>Nonmetallic mineral products</td>
<td>0.026</td>
<td>0.025</td>
<td>0.026</td>
</tr>
<tr>
<td>331</td>
<td>Primary metal manufacturing</td>
<td>0.064</td>
<td>0.065</td>
<td>0.068</td>
</tr>
<tr>
<td>332</td>
<td>Fabricated metal products</td>
<td>0.035</td>
<td>0.036</td>
<td>0.035</td>
</tr>
<tr>
<td>333</td>
<td>Machinery manufacturing</td>
<td>0.045</td>
<td>0.046</td>
<td>0.044</td>
</tr>
<tr>
<td>334</td>
<td>Computer and electronic products</td>
<td>0.081</td>
<td>0.082</td>
<td>0.082</td>
</tr>
<tr>
<td>335</td>
<td>Electrical equipment, appliance, components</td>
<td>0.029</td>
<td>0.033</td>
<td>0.031</td>
</tr>
<tr>
<td>336</td>
<td>Transportation equipment</td>
<td>0.203</td>
<td>0.201</td>
<td>0.194</td>
</tr>
<tr>
<td>337</td>
<td>Furniture and related products</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>339</td>
<td>Miscellaneous manufacturing</td>
<td>0.020</td>
<td>0.020</td>
<td>0.021</td>
</tr>
</tbody>
</table>
ICEE and IMAI 31-33 time series

Correlación lineal entre $X_t, Y_t = 0.63$
Logarithmic differences: ICEE and IMAI 31-33

D_LOG ICEE sector manufacturero (variable X_t)
D_LOG IMAI sector manufacturero (variable Y_t)

Correlación lineal entre X_t, Y_t = 0.89
Scatter plots for ICEE and IMAI 31-33

Original Variables

Transformed Variables

\[ \text{corr( IMAI 31-33, ICEE )} = 0.63 \]

\[ \text{corr( dlog IMAI 31-33, dlog ICEE )} = 0.89 \]
Logarithmic differences as approximations to monthly and annual variations

Let $X_t$ be a time series, where sub index $t = 1, 2, 3, \ldots$ distinguishes months. Let $vm_t := \frac{X_t}{X_{t-1}} - 1$ and $va_t := \frac{X_t}{X_{t-12}} - 1$ 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:

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

Using properties of logarithms,

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

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

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

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

i.e., $\hat{va}_t = \hat{vm}_t + vm_{t-1} + \cdots + vm_{t-11}$
Logarithmic difference model for estimating IMAI 31-33 as a function of the electric energy consumption index (ICEE)

\[ \nabla_m \ln Y_t = \beta_1 \nabla_m \ln X_t + \beta_2 i_{\text{oct}} + \beta_3 i_{\text{nov}} + \beta_4 i_{\text{dic}} + \beta_5 i_{\text{ene}} + \varepsilon_t, \]

\[ \varepsilon_t = \rho \varepsilon_{t-1} + \nu_t \]

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

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

**Note**: \( \varepsilon_t \) possesses an AR(1) structure.
Cochrane-Orcutt method is used for estimating the model’s parameters.
Obtaining IMAI 31-33 estimations from the logarithmic difference model

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

$$\ln \hat{Y}_t - \ln Y_{t-1} = \hat{v}_t$$

$$\ln \hat{Y}_t = \hat{v}_t + \ln Y_{t-1}$$

$$\hat{Y}_t = \exp(\hat{v}_t) Y_{t-1}$$

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 $t$. 

15
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

Variaciones mensuales (D_LOG) IMAI sector manufacturer (Y_1)
Estimaciones del modelo de diferencias logarítmicas
Límites de confianza al 95%

correlación entre valores observados y estimados = 0.95

Límite superior: -0.0062987 (≈ -0.627 %)
Estimación Julio 2018: -0.0338215 (≈ -3.38 %)
Límite inferior: -0.051374 (≈ -6.14 %)
Estimates for annual variations up to July 2018

Variaciones anuales (d_log) IMAI sector manufacturas
Var. anuales estimadas a través del modelo de diferencias logarítmicas
Limites de confianza al 95%

corrrelación entre valores observados y estimados = 0.8

Limite superior : 0.0412  (4.12 %)
Estimación Julio 2018 : 0.0136  (1.36 %)
Limite inferior : -0.014  (-1.39 %)
Estimates for the IMAI 31-33 index up to July 2018

Correlation between observed and estimated values = 0.96

Upper limit: 114.73
Estimation July 2018: 111.62
Lower limit: 108.58
Historical assessments for model’s estimates in real time
Assessments for IMAI 31-33 early estimates

Vertical green line indicates change in base year from 2008 to 2013

Correlation between observed and estimated values = 0.92
Assessments for estimated annual variations

corrrelación entre valores observados y estimados = 0.68
Assessments on estimated monthly variations

Variaciones mensuales observadas
Estimaciones generadas a través del modelo
Límites de confianza al 95%

correlación entre valores observados y estimados = 0.92
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