

# Early Monthly Estimation of Mexico's Manufacturing Production Level using Electric Energy Consumption data

Daniel Alba Cuellar  
daniel.alba@inegi.org.mx

Hugo Hernandez Ramos  
hugo.hernandez@inegi.org.mx

## Abstract

Directly measuring the monthly Manufacturing Production Level in Mexico via national accounting methods is an elaborate process, yielding a preliminary figure approximately 40 days after the end of the reference month. A separate analysis conducted by INEGI (Mexico's National Statistical Office) showed that in Mexico's manufacturing sector, in principle there exists a significant linear relationship between electric energy consumption and Production Level. Currently, electric energy consumption data from the Federal Electricity Commission (CFE) are made available to INEGI approximately 15 days after the end of the reference month; this timeliness in the availability of CFE data, combined with the observed relationship, allowed INEGI to build an econometric model which produces early estimates for the Production Level Index just 20 days after the end of the reference period. In this paper we describe the initial analysis conducted by INEGI to determine the relationship between both variables; then we describe the characteristics and evolution of the "definitive" econometric model, and compare early estimates, computed in real time, against official Production Level figures, as an empirical means for evaluating early estimation accuracy. We observed that 93% of the time, the official figure is located inside the estimation limits, which were computed with a 95% confidence level; this means that, in this case, observed empirical accuracy approached the theoretical confidence level. Finally, in this paper we comment about INEGI's data sharing experience with CFE and talk about future steps to improve this nowcasting process.

**Keywords:** linear regression, nowcasting, macroeconomic indicators, electric energy consumption, manufacturing activity level index.

## 1. Introduction

The aggregate Production Level in Mexico's Manufacturing Sector is a key macroeconomic variable which gives policymakers important clues about the health status of the national economy, given Mexico's Manufacturing Sector significant contribution to the National Gross Domestic Product. The objective of this paper is to describe our methodology used to investigate the functional relationship between **electric energy consumption** and **production level** in Mexico's manufacturing sector; it is of great interest to know such functional relationship, since the process for measuring the Monthly Manufacturing Production Level Index (IMAI-31-33) by means of national accounting methods is rather elaborate and time-consuming, while the process for measuring the consumption of electric energy in Mexico's manufacturing sector is faster and more direct; one would hope that if the latter variable is known in a timely fashion, then it would be feasible to obtain, in an economical way, an early estimate of the manufacturing production level, assuming we know the functional relationship between both variables.

Analysis of Monthly data collected by the Federal Electricity Commission (CFE) and by INEGI (Mexico's National Statistical Office), show that there is a strong linear relationship between electric energy consumption and production level aggregated at the manufacturing sector level and at national level, so it seems appropriate to use linear regression methods to approximate the functional relationship between these variables. In practice, INEGI obtains, at establishment level, monthly observations for both the consumption of electric energy and the volume of production (industrial activity), although at different moments in time: on the one hand, electric energy consumption data for establishments are provided to INEGI by CFE, approximately 15 days after the end of the reference month, while on the other hand, volume of production data at establishment level are collected, analyzed, processed, and published by INEGI's System of National Accounts (SNA) in aggregate and preliminary form, approximately 40 days after the end of the reference month; the empirically observed linear relationship between the two variables at the aggregate national manufacturing sector level, coupled with the timeliness of the

data provided by CFE, allow INEGI to obtain good early estimates of the production level in Mexico's manufacturing sector.

Obtaining a regression model to generate good early estimates for the IMAI-31-33 indicator, based on electric energy consumption data, requires fairly long time series; for the production level in Mexico's manufacturing sector, INEGI has compiled monthly observations starting from January 1993; thus, as of February 2020, the IMAI-31-33 time series contains more than 27 years of monthly observations; on the other hand, INEGI, with data from CFE, has compiled monthly electric energy consumption observations starting from January 2013; from these CFE monthly observations, INEGI built an electric energy consumption indicator for Mexico's manufacturing sector, called ICEE (both ICEE and IMAI-31-33 are acronyms in Spanish). Thus, ICEE is a time series with more than 84 monthly observations (7 years of monthly observations). This of course means that we cannot use all the available IMAI-31-33 observations for the construction of a regression model if we want to include ICEE as an explanatory variable. Fortunately, now ICEE is long enough to build statistically valid regression models.

This paper is structured as follows: section 2 describes the origin and characteristics of the data used in our analysis; then, section 3 describes how we prepare our data for use in the construction of regression models. Section 4 describes the characteristics of some regression models fitted to the prepared data that were available to us back in 2015, when our objective was to arrive at a functional model for generating early estimations (nowcasts) for the IMAI-31-33 indicator as a function of electric energy consumption; we begin with a simple two-variable linear regression model, and then we progress towards models that overcome the deficiencies of their predecessors. Section 5 shows out-of-sample estimates obtained with the models from section 4, and their accuracy assessments by comparing them against officially published IMAI-31-33 values; the evidence shown in sections 4 and 5 will help us select a "definitive" working model to generate IMAI-31-33 nowcasts. Section 6 shows the complete sequence of IMAI-31-33 nowcasts generated by the "definitive" model selected in section 5, running from

August 2015 to February 2020; we compare this sequence of estimated values against officially published IMAI-31-33 values, and comment on the changes the “definitive” model has undergone across time. Finally, section 7 briefly describes INEGI’s data sharing experience with CFE, and discusses possible avenues for future work.

## 2. Origin and characteristics of the data

**CFE data.** As mentioned above, electric energy consumption data are provided to INEGI by CFE on a monthly basis, approximately 15 days after the end of the reference month; these data contain electric energy consumption values, at user (establishment) level, which start at January 2013. Table 1 shows the structure of CFE monthly data provided to INEGI.

Table 1

Field name	Description
rpu*	User's Permanent Record key (Establishment's ID key), assigned by CFE
num_cta	Service's Account Number associated to user, assigned by CFE
estatus**	Code for Status of Service (Active, Suspended, Closed down, etc.)
nombre*	User's name
direccion*	User's address (street name and building number)
colonia*	Name of residential subdivision for user's address
cve_edo*	Mexican State code for user's address
estado*	Mexican State name for user's address
cve_mpio*	Municipality code for user's address
municipio*	Municipality name for user's address
poblacion*	Name of Settlement or City for user's address
e_calles1*	Name of contiguous street # 1 for user's address
e_calles2*	Name of contiguous street # 2 for user's address
cp*	Postal code for user's address
Telefono	User's telephone number (censored by CFE)
rfc*	User ID as a Business, assigned by Mexico's Federal Tax Agency
fecha_alta	Service's starting date
Tarifa	Code for type of service tariff specified in contract between CFE and user
Hilos	Number of phases specified in contract between CFE and user
tipo_sumin	Code for supplying method specified in contract between CFE and user
carga_inst	Installed load specified in contract between CFE and user
dem_contra	Demand cap specified in contract between CFE and user
cve_empre	Centralized Collection Agency code
Empresa	Centralized Collection Agency name
giro_scian	Code of economic activity carried out in the establishment, according to CFE
x_lectura*	X coordinate for energy meter's geographic location, in UTM format
y_lectura*	Y coordinate for energy meter's geographic location, in UTM format
x_reparto*	X coordinate for invoice delivery place, in Latitude-Longitude format
y_reparto*	Y coordinate for invoice delivery place, in Latitude-Longitude format
medidor1	Energy meter #1 ID Code
medidor2	Energy meter #2 ID Code (if applicable)
medidor3	Energy meter #3 ID Code (if applicable)
medidor4	Energy meter #4 ID Code (if applicable)
medidor5	Energy meter #5 ID Code (if applicable)
tipo_fact**	Invoice issuing type (code) specified in contract
k_YYYYMM**	energy consumption (in kWh) recorded in year YYYY, month MM
i_YYYYMM**	cost of electric energy consumed in year YYYY, month MM, in current pesos

Structure of CFE Administrative Data provided to INEGI

In Table 1, fields with \* at the end of their names are used for record linkage activities, while fields with \*\* at the end of their names are used for determining which CFE user (establishment) records to include in the sample employed in the computation of the ICEE indicator. Of particular importance is field k\_YYYYMM, which in itself contains the electric energy consumption data, in KWh. As of February 2020, INEGI receives each month about five million CFE user records belonging to the industrial (including manufacturing), trade and services economic sectors. Fields k\_YYYYMM and i\_YYYYMM are available simultaneously in a monthly dataset for several months; CFE sends revised data for all months from the current year and from the previous year, so, for example, CFE data sent after the end of reference month February 2020, contains data for each month of 2019, and data for January 2020 and February 2020.

The record linkage activities (not described in this paper), and the analysis and experiments described below, were conducted by the area of Statistical Linkage of Economic Administrative Registers (DVERA, its acronym in Spanish), which is part of INEGI. The record linkage process allowed DVERA to retrieve the correct North American Industry Classification System (NAICS) code at establishment level from INEGI's Statistical Business Register (SBR); we'll see in section 3 why it is important to have the correct NAICS economic activity classification code associated to CFE user records. Although the CFE data already contain information on the class of economic activity (field "giro\_scian" in Table 1), it has been observed that NAICS classification codes from CFE are in many cases missing or incorrect, since this piece of information is filled in by CFE according to the user's own declaration when establishing a service contract with CFE. User's name, business ID from the Federal Tax Agency, and data from address and geographical location fields are used to link CFE records (uniquely identified by their "rpu" keys) to SBR establishment records by means of probabilistic record linkage techniques.

**Missing values in CFE data.** Some Missing values in fields k\_YYYYMM and i\_YYYYMM are usually observed for the most recent month (i.e., for the reference month). The sooner CFE delivers its monthly data to INEGI, the greater the number

of missing values at the reference month; this is probably due to unregistered monthly invoices at the time of data compilation by CFE; this hypothesis is later confirmed when revised CFE data arrives in subsequent months. To fill in missing values, DVERA uses a simple 2-step procedure:

1. For all CFE records  $r$  with complete electric energy consumption data in reference month  $M$  and previous month  $M - 1$ , belonging to a given manufacturing subsector  $S$  (according to the NAICS classification code from INEGI's SBR), compute the proportion  $p_S = \sum_{r \in S} c_{r,M} / \sum_{r \in S} c_{r,M-1}$ , where  $c_{r,M}$  is the electric energy consumption of establishment  $r$  at month  $M$ .  $p_S$  can be interpreted as an approximation to the monthly growth rate in electric energy consumption for manufacturing subsector  $S$  at reference month  $M$ .

2. For each record  $i$  in manufacturing subsector  $S$  with missing value  $c_{i,M}$ , assign  $c_{i,M} = p_S c_{i,M-1}$ .

NB: This 2-step procedure only takes into account records included in a sample of CFE records especially designed to build the ICEE indicator. Below we give some more details about the construction of this CFE sample.

**CFE records included in the calculation of the ICEE indicator.** Depending on the type of contract that an establishment makes with CFE, its electric energy consumption invoice can be issued once every month, or once every other month; this is indicated by the “tipo\_fact” field from Table 1. Large establishments generally receive their CFE invoices (and their consumptions recorded into CFE database) once every month. These large establishments have a high probability of being included in the sample employed for calculating the ICEE indicator, since the main inclusion criterion is to consider those CFE establishments who are statistically linked to a special subset of INEGI's SBR, called *INEGI master sample*, which contains the records of the largest establishments in the national economy (the size of an SBR establishment is measured jointly by number of employees and income), and serves as a basis for constructing samples used by INEGI's national economic surveys. As of February 2020, our sample for building the ICEE indicator contains

17,085 CFE records from the manufacturing sector with monthly invoicing frequency; it is worth mentioning that this sample has been growing steadily since we started with this project back in the second half of 2015. When the first early IMAI-31-33 estimate was generated, corresponding to reference month August 2015, our sample for building the ICEE indicator contained 7,744 CFE records. Periodically, CFE data are analyzed by DVERA to update the CFE sample for computing the ICEE indicator by detecting closed down records (which does not necessarily imply that the corresponding establishment has closed down) or new records; these new records sometimes correspond to new establishments and sometimes correspond to existing establishments who updated their contract with CFE. From here, we can see that CFE data can also be used to periodically update INEGI's SBR.

***Complementary electric energy consumption data.*** In Mexico, there are a few manufacturing businesses who also generate electric energy for self-consumption. Mexico's Federal Government has an administrative register, at establishment level, for these self-sufficient businesses, which is managed by an agency, very closely related to CFE, called the *National Center of Energy Control* (CENACE for its acronym in Spanish). Starting from the calculation of the early IMAI-31-33 estimate corresponding to reference month August 2016, we incorporated 43 CENACE records to our CFE sample for computing the ICEE indicator. As of February 2020, our sample includes 66 CENACE records, who represent 1.7% of the total electric energy consumption recorded in our sample across all months. Of course, these 66 CENACE records were also linked to INEGI's SBR, in order to retrieve their correct NAICS economic activity codes.

***Data from the System of National Accounts.*** DVERA obtains production level data for manufacturing industries in Mexico directly from the Manufacturing Production Level Index, IMAI-31-33. This indicator, aggregated at the economic sector level, is published on a monthly basis by the System of National Accounts of Mexico (SNA) at the INEGI website. Since the construction process for the IMAI-31-33 indicator by the SNA depends on survey-collected data, it is published approximately 40 days after the end of the reference month. For the manufacturing

sector in Mexico, we initially observed that the linear correlation coefficient between ICEE and IMAI-31-33 is high, close to positive 1 (see figure 1); from here, we see that it is feasible to have a good notion, several days in advance, about the magnitude of the next monthly “true” published IMAI-31-33 value, by using CFE electric energy consumption data, along with a properly built regression model.

**NB:** In this paper we work only with non-seasonally adjusted time series; this choice, in our experience, produces better results, as we have seen after several exercises with linear regression models involving both non-seasonally adjusted and seasonally adjusted time series. If we want to obtain a seasonally-adjusted nowcast for IMAI-31-33 as a function of electric energy consumption, we can simply append our non-seasonally adjusted nowcast to the non-seasonally adjusted IMAI-31-33 time series itself, and then use a seasonally adjusting software, such as x-13arima-seats.

### 3. Data preparation to conduct a regression analysis

With the data provided by CFE and SNA, monthly data on electric energy consumption and on volume of production, respectively, are available for the main manufacturing industries in Mexico. From these data, it is possible to build two monthly frequency time series:

- Time series  $X_t$ : electric energy consumption index for Mexico's manufacturing sector (ICEE)
- Time series  $Y_t$ : production level for Mexico's manufacturing sector

Here, subscript  $t$  indicates the month to which a measurement for any of these two variables corresponds.

The  $Y_t$  variable is simply a subset of the IMAI-31-33 time series mentioned above, which can be located within the IMAI file: Original series - Physical volume indices, base 2013 = 100, namely in the row labeled as “31-33 - Industrias manufactureras”. This file can be directly downloaded on the following link (in Spanish): <https://www.inegi.org.mx/temas/imai/default.html#Tabulados>.

NOTE: For the construction of IMAI-31-33, SNA uses information mainly from INEGI's Monthly Survey on Manufacturing Industries (EMIM for its acronym in Spanish); this national-level economic survey is also based on INEGI's SBR master sample. See [1] for more information.

To generate variable  $X_t$  from CFE data, the following procedure was defined:

1. For each month  $t$ , sum the *weighted* electric energy consumption values corresponding to all CFE records in our sample:

$$s_t = \sum_r w_{s(r),t} c_{r,t};$$

$c_{r,t}$  is the electric energy consumption (in kWh) of establishment  $r$  within our sample in month  $t$ , while  $w_{s(r),t}$  is a weight that depends both on month  $t$ , and on the manufacturing subsector of economic activity  $s(r)$  to which establishment  $r$  belongs to. Weights  $w_{s,t}$  satisfy the convexity property; i.e., the sum of all weights for a fixed month  $t$  equals 1. Table 2 shows typical values for weights  $w_{s,t}$ .

2. Re-scale time series  $s_t$  so that it coincides with  $Y_t$  at a base period (say  $t = 1$ ):

$$X_t = \frac{Y_1}{s_1} s_t.$$

Table 2

Manufacturing subsector $s$	Description	Weights $w_{s,t}$ (t corresponds to May 2019)
311	Food manufacturing	0.2295
312	Beverage and tobacco product manufacturing	0.0598
313	Textile mills	0.0086
314	Textile product mills	0.0045
315	Apparel manufacturing	0.0192
316	Leather and allied product manufacturing	0.0076
321	Wood product manufacturing	0.0089
322	Paper manufacturing	0.0176
323	Printing and related support activities	0.0058
324	Petroleum and coal products manufacturing	0.0136
325	Chemical manufacturing	0.0806
326	Plastics and rubber products manufacturing	0.0252
327	Nonmetallic mineral product manufacturing	0.0244
331	Primary metal manufacturing	0.0620
332	Fabricated metal product manufacturing	0.0315
333	Machinery manufacturing	0.0449
334	Computer and electronic product manufacturing	0.0873
335	Electrical equipment, appliance, and component manufacturing	0.0287
336	Transportation equipment manufacturing	0.2102
337	Furniture and related product manufacturing	0.0095
339	Miscellaneous manufacturing	0.0206

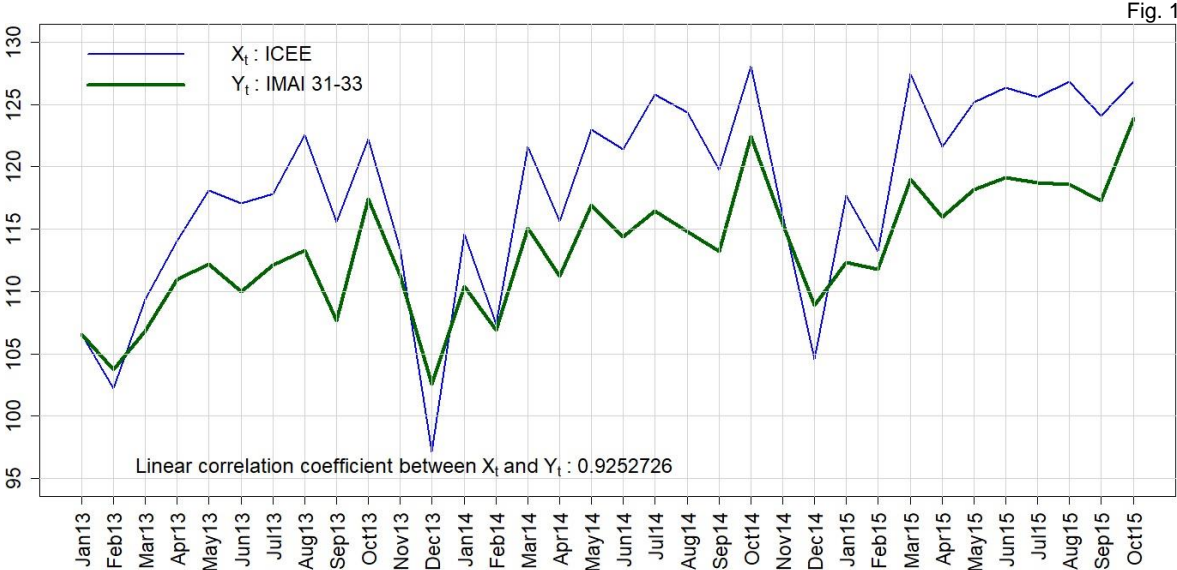
Manufacturing subsector weights for month May 2019. These weights are provided by SNA and vary slightly from month to month

Manufacturing subsector weights  $w_{s,t}$ , just like the IMAI-31-33 indicator, are also provided by SNA; furthermore, their availability coincides with that of the IMAI 31-33 indicator; that is, weights  $w_{s,t}$  are available approximately 40 days after the end of the reference month. In order to estimate the manufacturing subsector weights for the reference month (say, month  $M$ ) for which electric energy consumption data are already available, but not yet for the IMAI 31-33 indicator, take the most recent weight  $w_{s,M-1}$  for manufacturing subsector  $s$ , and multiply it by the factor  $w_{s,M-12}/w_{s,M-13}$ , which represents the monthly growth rate that weight  $w_{s,t}$  experienced a year before reference month  $M$ ; in this way, we have

$$\widehat{w}_{s,M} = (w_{s,M-12}/w_{s,M-13})w_{s,M-1}, \text{ where } \widehat{w}_{s,M} \text{ is an estimate for weight } w_{s,M}.$$

The data  $(X_t, Y_t)$  used in the analysis presented in sections 4 and 5 of this paper correspond to the data available to us back in the second half of 2015, when our objective was to build a “definitive” working model useful in explaining the relationship between  $X_t$  and  $Y_t$  and at the same time useful in generating “nowcasts”

for  $Y_t$ . Thus, we have 34 monthly observations available for our present analysis:  $t = 1$  corresponds to January 2013,  $t = 2$  corresponds to February 2013, ...,  $t = 34$  corresponds to October 2015. Figure 1 shows the time series corresponding to variables  $X_t$  and  $Y_t$ . Additionally, the values of these two variables are listed in appendix 1 of this document, for the sake of experiment reproducibility.



Time series for IMAI-31-33 ( $Y_t$  variable, bold line) and for the ICEE indicator ( $X_t$  variable, thin line), shown across time interval January 2013 – October 2015. Note that monthly movements in both time series are similar to each other, which translates to a strong linear positive relationship (sample Pearson correlation coefficient equal to 0.93); this suggests the feasibility of estimating  $Y_t$  values as a function of  $X_t$  values, via a linear regression model, since the most recent value for  $X_t$  is available almost a month before the official publication of the corresponding  $Y_t$  value

**4. Models fitted to the analysis variables**

Once the analysis variables  $X_t, Y_t$  are prepared, the objective now is to generate “nowcasts” of  $Y_t$  as a function of  $X_t$  for reference month  $t = M$ . To reach this objective, we chose to use linear regression models, commonly used in econometrics. This section describes the main models that we progressively fitted to the analysis variables  $X_t$  and  $Y_t$ , indicating in each case their characteristics. When we conducted this model construction exercise, back in 2015, our approach was to progressively find a regression model such that subsequent models improved upon

the deficiencies found in its predecessors. This model construction exercise was done with help of the R Statistical Computing Software [2].

#### 4.1 Simple linear regression model

The first model fitted to our data was a simple linear regression model of the form

$$Y_t = \alpha + \beta X_t + \varepsilon_t. \quad (1.0)$$

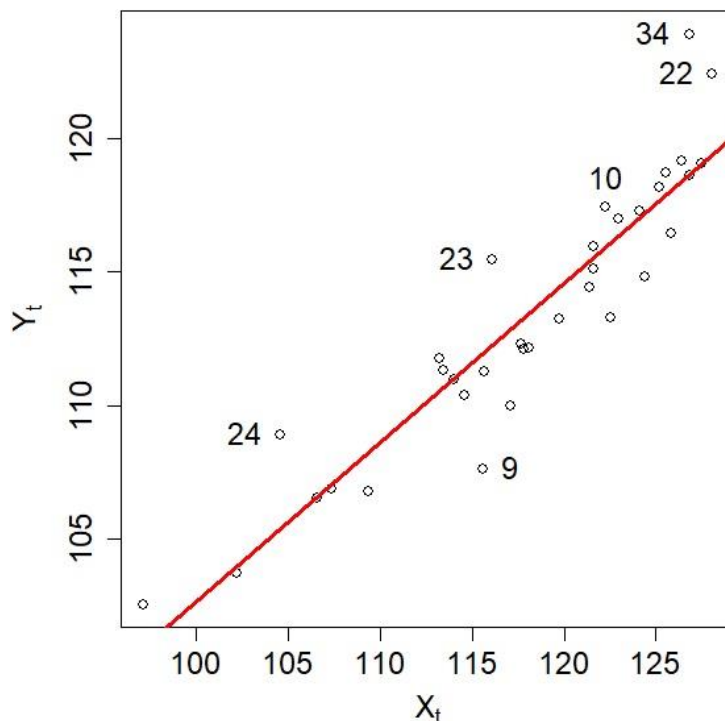
For a model of the form (1.0) to be considered as an adequate fit to the data  $(X_t, Y_t)$ , in theory the disturbances  $\varepsilon_t$  must follow a white noise process; i.e.,  $\varepsilon_t$  variates must be independent and identically distributed, each with normal distribution with mean zero and constant variance  $\sigma^2$ ; to abbreviate, we denote this requirement as  $\varepsilon_t \sim \text{iid } N(0, \sigma^2)$ . By using the data in appendix 1 to fit a model of the form (1.0) via Ordinary Least Squares (OLS), we obtain the estimation equation

$$\hat{Y}_t = 43.14049 + 0.59528 X_t;$$

thus, the estimated parameters are intercept  $\hat{\alpha} = 43.14049$  (5.103), slope  $\hat{\beta} = 0.59528$  (0.043); figures in parentheses indicate standard errors. From here, it is possible to see that both coefficients are statistically significant. Additionally, adjusted  $R^2 = 0.852$ , and the corresponding F-test of overall significance has a p-value of  $5.1 \times 10^{-15}$ . Model residuals  $r_t = Y_t - \hat{Y}_t$  (sample estimates of the  $\varepsilon_t$  disturbances) appear to be approximately centered around 0 (min = -4.3, median = -0.04, max = 5.3). All these model hypothesis tests indicate a good fit so far.

The corresponding fitted regression line is shown in Figure 2, along with data points  $(X_t, Y_t)$  used to build model (1.0).

Fig. 2



Regression line  $\hat{Y}_t = 43.14049 + 0.59528X_t$ , together with data points  $(X_t, Y_t)$ ,  $t = 1, 2, \dots, 34$ . Data points farthest from the regression line, vertically-wise (outliers), are labeled with the number of the month  $t$  to which they correspond; in this way, we see that outliers are present in sep-2013 (9), oct-2014 (22), nov-2014 (23), dec-2014 (24), and oct-2015 (34). We also labeled oct-2013 (10), which is at a fair vertical distance from the regression line, without being an outlier.

Before accepting this model as the definitive one, additional tests are necessary to verify that the residuals  $r_t$  come from a white noise process. The Durbin-Watson (DW) test for the fitted model of the form (1.0) gives a test statistic equal to 1.11, with a p-value close to zero; this strongly suggests that we must reject the null hypothesis which states that the model residuals are serially uncorrelated; from here, we conclude that there is strong evidence that the model residuals  $r_t$  follow a 1<sup>st</sup> order autoregressive process. To correct the problem of auto-correlated residuals in a linear regression model, there are several alternatives: one is to apply the Cochrane-Orcutt correcting procedure to model the disturbances using a first-order autoregressive model of the form  $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ , with  $v_t \sim \text{iid } N(0, \sigma^2)$ ; a second alternative is to use generalized least squares together with the autoregressive structure of the  $\varepsilon_t$  disturbances (Cochrane-Orcutt uses OLS); a third option is to incorporate explanatory lagged variables  $X_{t-1}, Y_{t-1}$  into the regression model; this

third approach is known as auto-regressive modeling with distributed lags. For more information about these approaches, see [3].

An additional problem with this simple linear regression model (1.0) fitted to our data has to do with the outliers shown in Figure 2. As can be seen, outliers manifest themselves at different months, and for almost 3 years of monthly data values, there are outliers at October (outliers 22 and 34; 10 is not an outlier, but it is not either one of the closest points to the regression line). This suggests the need to add to subsequent linear regression models an indicator variable for the month of October as an explanatory variable; this indicator variable could be roughly defined as

$$I_{oct} = \begin{cases} 1 & \text{if } t \text{ corresponds to an october month} \\ 0 & \text{otherwise} \end{cases} .$$

#### 4.2 Autoregressive model with distributed lags

If the disturbances from regression model (1.0) follow a first-order autoregressive process  $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ , with  $v_t \sim \text{iid } N(0, \sigma^2)$ , which appears to be the case when fitting model (1.0) to our data, then it is possible to correct this autoregressive effect if we instead use a model of the form

$$Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta(X_t - \rho X_{t-1}) + v_t, \quad (2.0)$$

which is obtained by subtracting  $\rho Y_{t-1} = \alpha\rho + \beta\rho X_{t-1} + \rho\varepsilon_{t-1}$  from model (1.0). Of course,  $\rho$  corresponds to the autocorrelation coefficient which characterizes the 1<sup>st</sup> order auto-correlated disturbances from model (1.0). The Cochrane-Orcutt procedure obtains an estimate of  $\rho$  using an iterative algorithm, although it is also possible to re-parameterize model (2.0) in order to obtain an autoregressive model with distributed lags of the form

$$Y_t = \alpha^* + \rho Y_{t-1} + \beta X_t + \gamma^* X_{t-1} + v_t, \quad (2.1)$$

where  $\alpha^*$  represents the quantity  $\alpha(1 - \rho)$ , and  $\gamma^*$  represents the quantity  $-\rho\beta$ . The parameters from model (2.1) can be estimated via OLS.

By fitting a model of the form (2.1) to the data shown in appendix 1 via OLS, together with an indicator variable  $I_{oct}$ , we obtained the following estimation equation

(note that, in this case, for variables  $X_t$ ,  $Y_t$  and  $I_{oct}$ , the initial value for  $t$  corresponds to February 2013):

$$\hat{Y}_t = 12.54 + 0.69Y_{t-1} + 0.52X_t - 0.33X_{t-1} + 5.21 I_{oct}.$$

It is possible to see that all regression coefficients in fitted model (2.1) are statistically significant; adjusted  $R^2 = 0.948$ , and model residuals appear to be centered at zero (min =  $-2.6$ , median =  $0.3$ , max =  $2.0$ ).

The Durbin-Watson test for fitted model (2.1) yields a test statistic equal to  $2.44$ , with an associated p-value close to  $0.3$ . From here, we see that fitted model (2.1), which includes an indicator variable for October, is free of residual first-order autocorrelation, and all its explanatory variables are statistically significant.

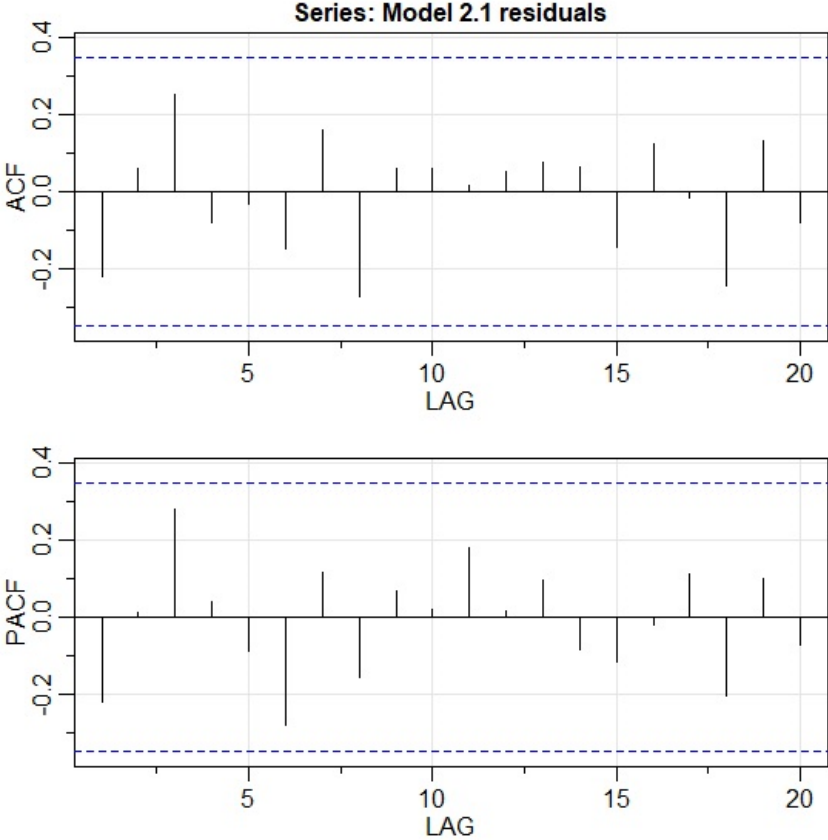
Additionally, a Breusch-Pagan test for heteroscedasticity (non-constant variance) was conducted on the model residuals, producing a test statistic equal to  $0.716$  with an associated p-value close to  $0.4$ ; The Breusch-Pagan test's null hypothesis states that there is no heteroscedasticity in the residuals; The p-value in this case tells us that there is not enough evidence to reject the null hypothesis, so it is concluded that fitted model (2.1) successfully passes the heteroscedasticity test.

Another test that can be done on the residuals from the fitted model with form (2.1) is that of Cramér-von Mises, to verify the normality of the residuals. In this case, we obtained a test statistic equal to  $0.0502$ , with associated p-value close to  $0.5$ . In the Cramér-von Mises test, the null hypothesis states that the residuals are normally distributed; the p-value indicates that we cannot reject the null hypothesis, so it is concluded that the residuals from model (2.1) are normally distributed.

When working with linear regression models which contain two or more explanatory variables, as is the case with model (2.1), we must make sure that there is no high linear correlation among explanatory variables; for this, we applied a multicollinearity test (for more details, see [4]), obtaining *Variance Inflation Factors* (VIFs) equal to  $7.6$ ,  $1.2$  and  $8.2$ ; if any VIF is greater than  $10$ , then we have multicollinearity problems, which does not occur in this case. Therefore, fitted model (2.1) does not have multicollinearity problems.

Finally, the graphs for the sample autocorrelation function (ACF) and for the sample partial autocorrelation function (PACF), shown in Figure 3, confirm that the residuals from fitted model (2.1) are uncorrelated over time:

Fig. 3



Sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) for residuals of fitted model (2.1). The lagged correlations do not extend beyond horizontal dotted lines, which indicates that there is no evidence of residual autocorrelation.

From here, we see that fitted model (2.1) is adequate, although it could be argued that it has several estimable parameters, and a couple of VIF’s are rather large. Next we’ll investigate a more parsimonious alternative.

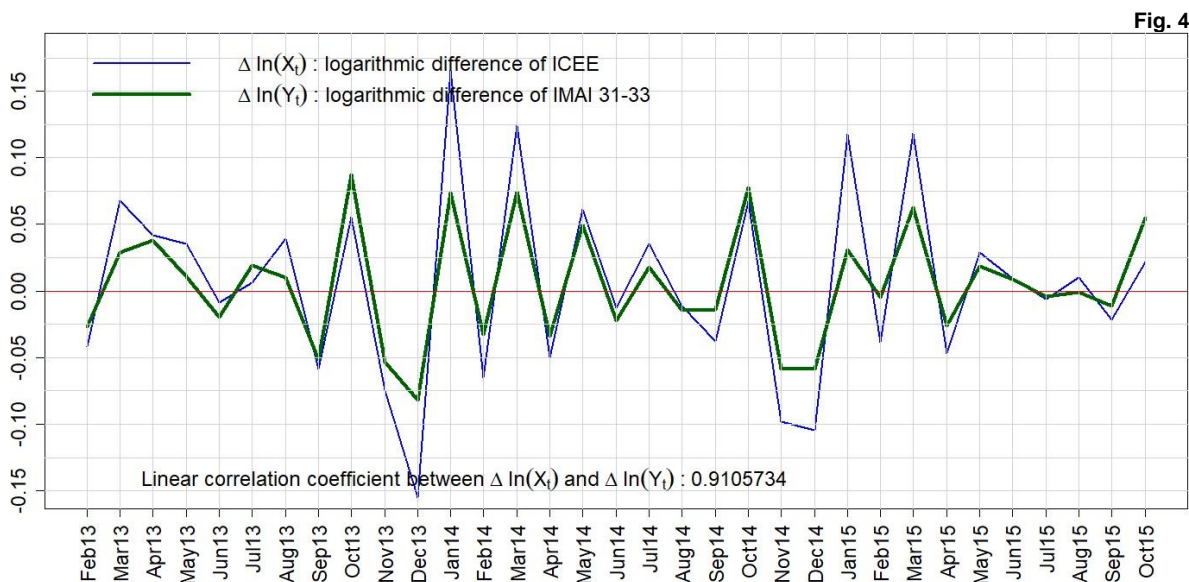
**4.3 Logarithmic differences model**

A proposed modeling alternative to that presented in subsection 4.2 is the *logarithmic differences model*, which is built by means of the following expression:

$$\Delta \ln(Y_t) = \beta \Delta \ln(X_t) + \varepsilon_t, \tag{3.0}$$

where  $\Delta \ln(Y_t) := \ln(Y_t) - \ln(Y_{t-1}) = \ln(Y_t/Y_{t-1})$  is the logarithmic difference of  $Y_t$ ,  $\Delta \ln(X_t)$  is the logarithmic difference of  $X_t$ ,  $\beta$  is the parameter to be estimated (via OLS), and  $\varepsilon_t$  are the usual stochastic white-noise disturbances in a linear regression model. The advantage of this model lies in its parsimony; that is, it is not necessary to estimate many parameters, unlike model (2.1). Note that this model does not directly estimate values for  $Y_t$ ; instead, it estimates the *monthly growth rate* of  $Y_t$ . To see this, first note that  $\ln(1 \pm r) \approx \pm r$  if  $r$  is a reasonably small value; for example,  $\ln(1 + 0.05) = 0.04879 \approx 0.05$ , while  $\ln(1 - 0.05) = -0.05129 \approx -0.05$ . Now, the logarithmic difference  $\Delta \ln(Y_t) = \ln(Y_t/Y_{t-1})$  is in fact the natural logarithm of the variation of  $Y_t$  with respect to its previous monthly value  $Y_{t-1}$ ; if this monthly growth is reasonably small, then  $Y_t/Y_{t-1}$  will be of the form  $1 + r$ ; for example, if  $Y_T = 118.1$ ,  $Y_{T-1} = 116.0$  for  $t = T$ , then  $Y_T/Y_{T-1} = 1.018103 = 1 + 0.018103$ , which means that when  $t = T$ ,  $Y_t$  grows, or varies 1.8%, with respect to its previous monthly value  $Y_{t-1}$ ; note that in this example,  $\Delta \ln(Y_T) = \ln(Y_T/Y_{T-1}) = 0.01794153$  is reasonably close to the true  $Y_t$  monthly growth rate for  $t = T$ .

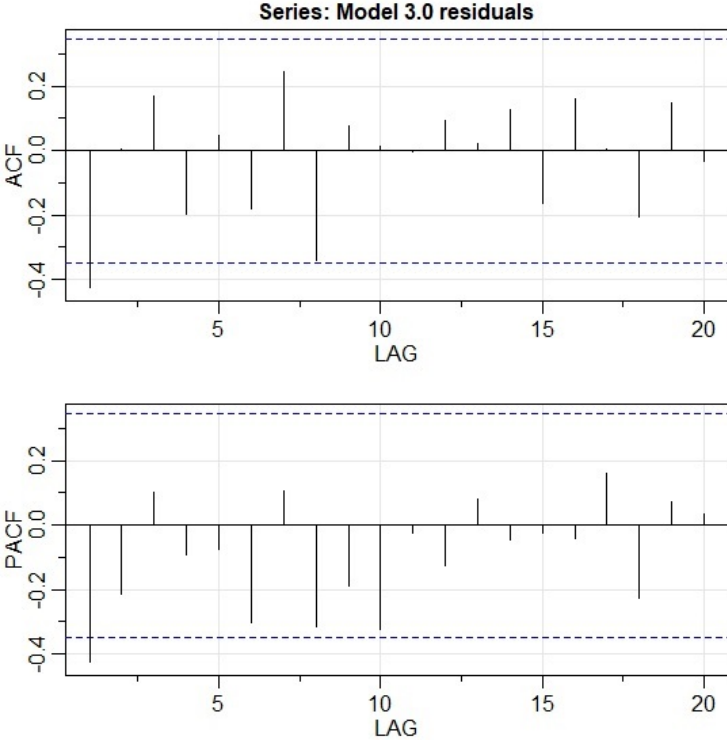
Figure 4 shows a graph of the transformed variables  $\Delta \ln(X_t)$  and  $\Delta \ln(Y_t)$ , obtained directly from the data in appendix 1.



Variables  $\Delta \ln(X_t)$  (thin line),  $\Delta \ln(Y_t)$  (bold line) used in the construction of model (3.0). Notice that the initial value for  $t$  corresponds to February 2013; in this case, we cannot obtain  $\Delta \ln(X_t)$  for January 2013. Sample Pearson correlation coefficient between variables is 0.91.

By fitting, via OLS, a model of the form (3.0) to the data represented graphically in Figure 4, together with an indicator variable for the months of October, we obtained the estimation equation  $\Delta \widehat{\ln(Y_t)} = 0.523002 \Delta \ln(X_t) + 0.047991 I_{\text{Oct}}$ . We also applied the same battery of tests from subsection 4.2 to fitted model (3.0), sans the multicollinearity test. We found out that fitted model (3.0) successfully passes all tests, except for the Durbin-Watson test, as can also be seen from Figure 5.

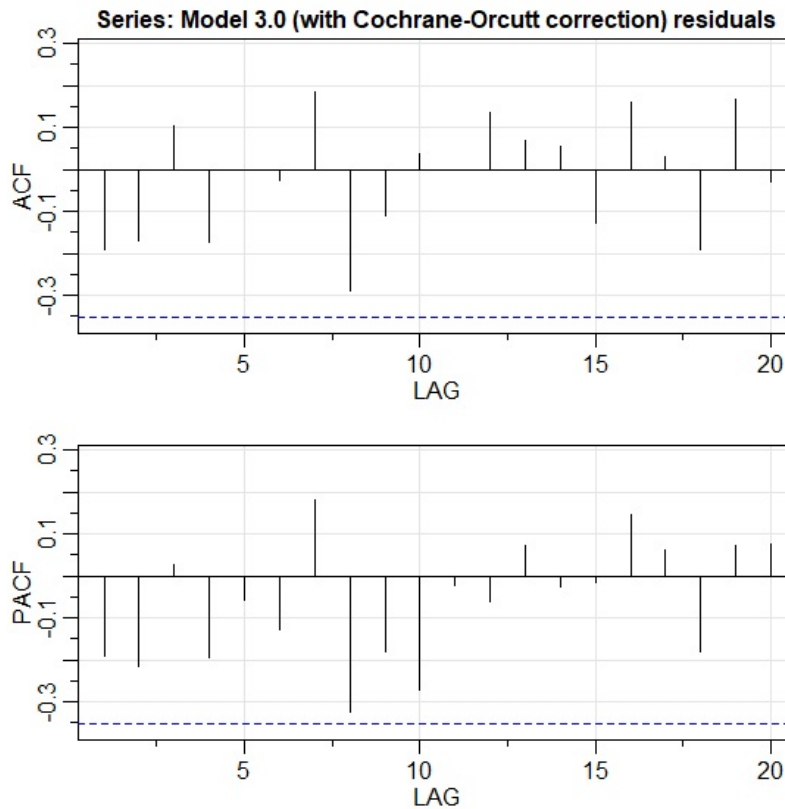
Fig. 5



Sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) for residuals of fitted model (3.0). 1<sup>st</sup> lag correlation extends beyond the lower horizontal dotted line, which indicates possible presence of 1<sup>st</sup> order negative autocorrelation in the model residuals.

To correct this 1<sup>st</sup> order autocorrelation problem, we applied the Cochrane-Orcutt correction procedure, originally defined in [5], to fitted model (3.0); in this way, we obtained the estimation equation  $\Delta \widehat{\ln(Y_t)} = 0.547802 \Delta \ln(X_t) + 0.043357 I_{\text{Oct}}$ , with residual term  $r_t = \hat{\rho}r_{t-1}$ , where  $\hat{\rho} = -0.4277$  ( $\hat{\rho}$  is used to correct model coefficients). This time, all tests on this corrected model (3.0) are successfully passed, getting an adjusted  $R^2$  equal to 0.93 and no autocorrelation problems on model residuals, as Figure 6 shows.

Fig. 6



Sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) for residuals of fitted model (3.0) with Cochrane-Orcutt correction. Lagged correlations do not extend beyond the horizontal dotted lines, which indicates that there is no evidence of residual autocorrelation.

We conclude that fitted model (3.0) with Cochrane-Orcutt correction passes all our criteria for obtaining an adequate linear regression model. In the next section we'll test empirically the out-of-sample forecasting accuracy of the models obtained here. The results from this section and the next section will help us decide on a "definitive" model.

## 5. Comparisons of out-of-sample estimates among regression models

In this section we compute out-of-sample estimates using the models built in section 4 and then comparing such estimates against officially published values. As we previously saw, the data used to build our models span the time interval from January 2013 to October 2015; therefore, circumscribing ourselves to the data from

appendix 1, we have the November 2015  $X_t$  value as the only element for producing out-of-sample estimates for  $Y_t$  values. Thus, in order to “nowcast” the November 2015 IMAI 31-33 value, we incorporate the November 2015 ICEE value (namely, 112.61) into the estimation equations obtained in section 4; it is worth mentioning that the models described in section 4 were originally built on December 9, 2015, and their predictions for the month of November 2015 were compared against the “observed” November 2015 value for the IMAI 31-33 indicator, published on January 11, 2016 as a preliminary figure (namely, 117.5659). Additionally, with the published IMAI 31-33 figures, we obtained the observed (or “true”) annual and monthly growth rates: the observed annual growth rate of the IMAI 31-33 indicator for November 2016 is  $100(Y_{\text{nov2015}}/Y_{\text{nov2014}} - 1) = 100(117.5659/115.4457 - 1) = 1.84$  (see appendix 1 for  $Y_{\text{nov2014}}$  value). Likewise, the observed monthly growth rate of the IMAI 31-33 indicator for November 2016 is  $100(Y_{\text{nov2015}}/Y_{\text{oct2015}} - 1) = 100(117.5659/124.2666 - 1) = -5.39$ . Note that in the calculation of the observed monthly growth rate, we are using the revised figure for the October 2015 IMAI 31-33 indicator (124.2666) instead of the corresponding  $Y_{\text{oct2015}}$  value found in appendix 1 (123.8842).

We’ll use these observed values as part of an additional criterion for assessing the models built in section 4. Table 3 shows out-of-sample estimations from models (1.0), (2.1) and (3.0) along with their corresponding absolute errors with respect to observed values.

Table 3.

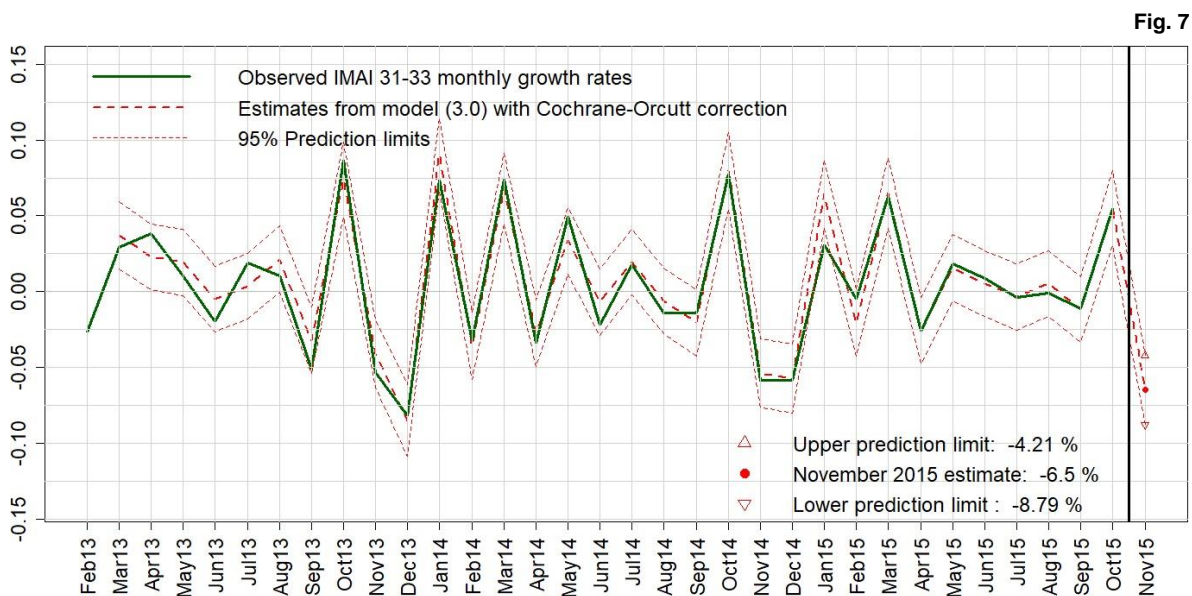
	Observed value	Estimate Model (1.0)	Absolute error Model (1.0)	Estimate Model (2.1)	Absolute error Model (2.1)	Estimate Model (3.0)	Absolute error Model (3.0)
<b>IMAI 31-33</b>	117.57	110.17	7.40	114.56	3.01	116.09	1.48
<b>Annual growth rate</b>	1.84	-4.57	6.41	-0.77	2.61	0.56	1.28
<b>Monthly growth rate</b>	-5.39	-11.07	5.68	-7.53	2.14	-6.50	1.11

November 2015 out-of-sample estimations and absolute estimation errors for models (1.0), (2.0) and (3.0) with Cochrane-Orcutt correction

Appendix 3 explains how the estimated values in table 3 were computed.

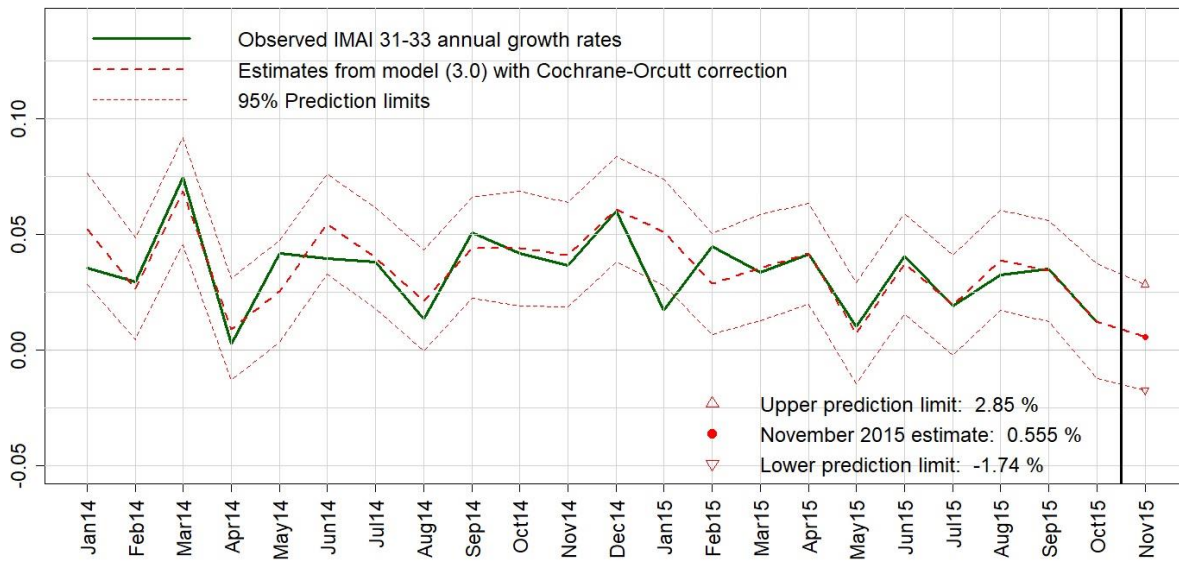
Estimation errors shown in table 3 indicate us that the logarithmic differences model (3.0) with Cochrane-Orcutt correction produces the most accurate November 2015 IMAI 31-33 nowcasts. From the model diagnostics and hypothesis testing procedures described in section 4, we see that both the distributed lags model (2.1) and the logarithmic differences model (3.0) with Cochrane-Orcutt correction have desirable statistical properties, although model (2.1) has more estimable parameters than model (3.0) and could potentially be more unstable as new monthly observations become available. We then conclude that, from the models built in section 4, the logarithmic differences model (3.0) with Cochrane-Orcutt correction is the most parsimonious, robust and is the one which produces the most accurate forecasts; thus we select model (3.0) as our “definitive” model.

Figures 7, 8 and 9 show graphs for in-sample and out-of-sample estimations produced with our definitive model, using the data from appendix 1. The in-sample estimations for IMAI 31-33 values and for annual growth rates, were obtained by applying the same expressions derived in appendix 3.



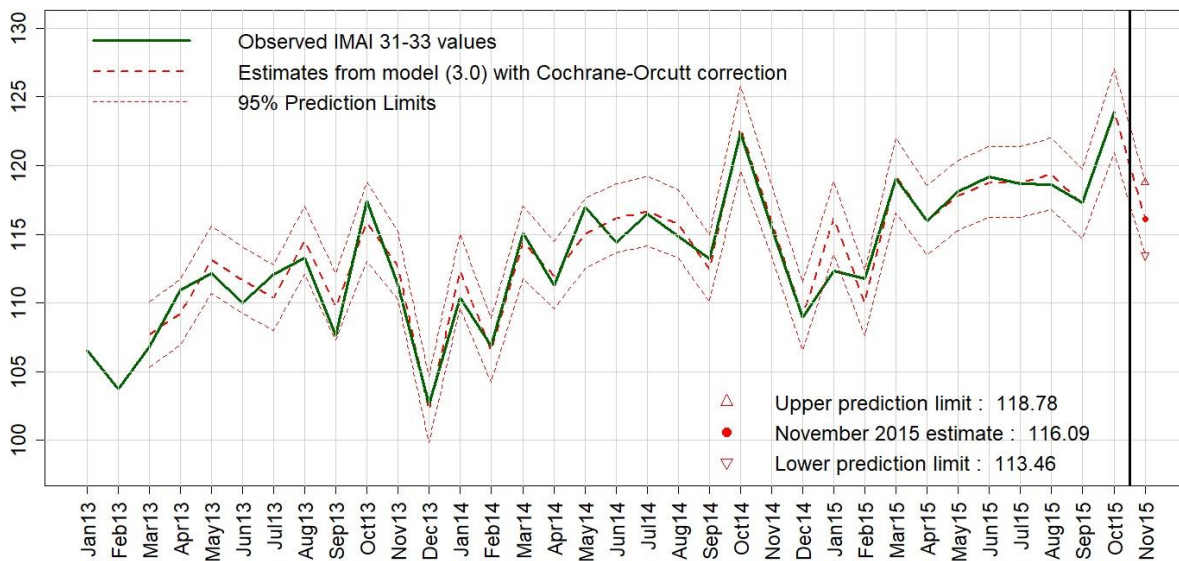
In-sample (from February 2013 to October 2015) and out-of-sample (November 2015) estimates of IMAI 31-33 monthly growth rates (dotted line), generated from the model (3.0) with Cochrane-Orcutt correction, via its estimation equation. Solid line represents observed IMAI 31-33 monthly growth rates. Correlation between observed and estimated values is 0.97

Fig. 8



In-sample and out-of-sample estimates of IMAI 31-33 annual growth rates (dotted line), generated from model (3.0) with Cochrane-Orcutt correction, by using the expression  $\Delta \ln(\widehat{Y}_t) + \Delta \ln(Y_{t-1}) + \dots + \Delta \ln(Y_{t-11})$ . Solid line represents observed IMAI 31-33 annual growth rates. Correlation between observed and estimated values is 0.79

Fig. 9

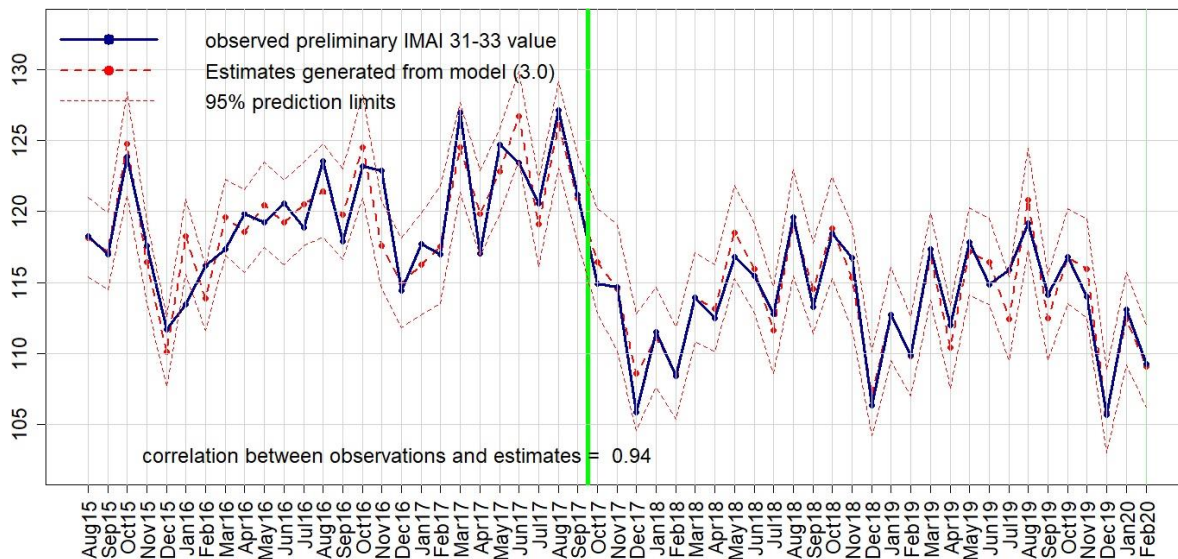


In-sample and out-of-sample estimates of IMAI 31-33 values (dotted line), generated from model (3.0) with Cochrane-Orcutt correction, by using the expression  $\widehat{Y}_t = Y_{t-1} e^{\Delta \ln(\widehat{Y}_t)}$ . Solid line represents observed IMAI 31-33 values. Correlation between observed and estimated values is 0.96

## 6. Evaluating logarithmic differences model in real time

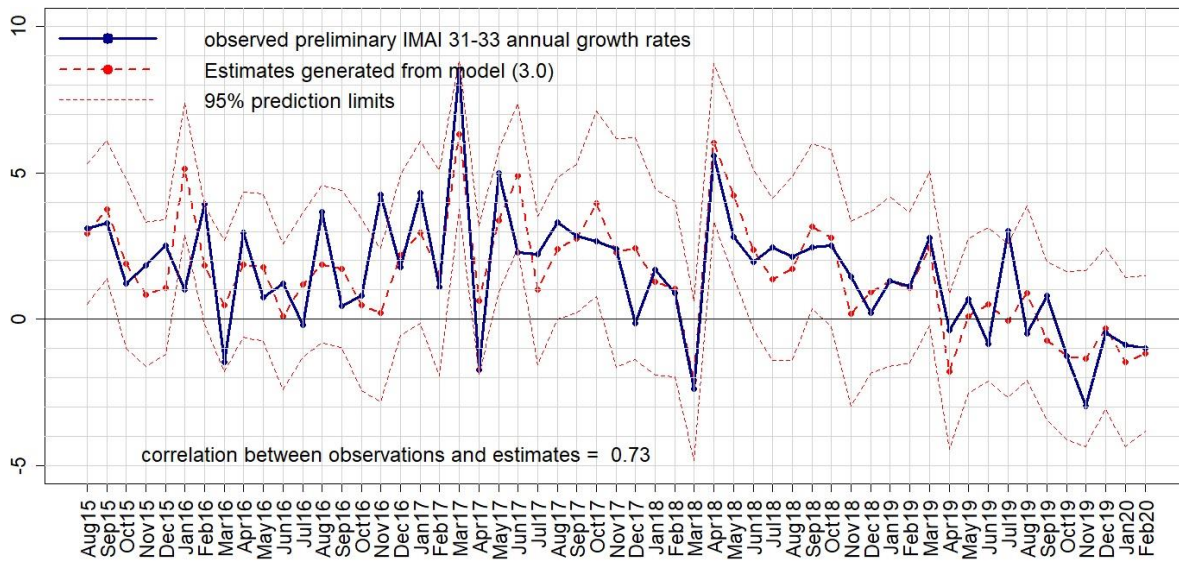
Having decided on what model to use in order to generate subsequent nowcasts for the IMAI 31-33 indicator, DVERA has fitted, on a monthly basis, linear regression models of the form (3.0) to produce such nowcasts, beginning at reference month August 2015, and verifying each time that the fitted logarithmic differences model passes all the statistical tests for model adequacy. Figures 10, 11, and 12 show the comparisons of IMAI 31-33 nowcasts against published IMAI 31-33 values. It is worth mentioning that these comparison graphs are updated each month, appending each new time an additional estimate and corresponding observed value, published as a preliminary figure (no revised figures for officially published IMAI 31-33 values are displayed in Figures 10, 11 and 12). Appendix 2 shows a table with dates for reception of CFE data, calculation of nowcasts, and official publication of IMAI 31-33 values.

Fig. 10.



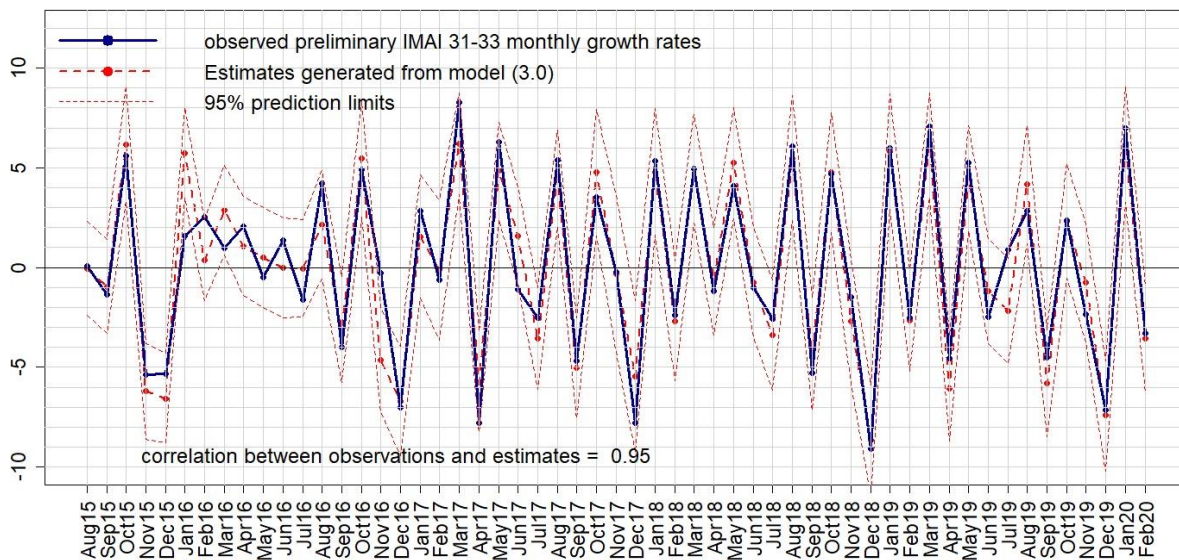
Dashed lines represent the point and interval nowcasts for the IMAI 31-33 indicator, calculated from a logarithmic differences model, while solid line represents IMAI 31-33 official values, initially published as preliminary figures. The solid vertical line in this graph indicates a change of base year from 2008 to 2013 in officially published INEGI's SNA figures, which includes the IMAI 31-33 indicator. Empirical correlation between observed and estimated values equals 0.94

Fig. 11



Dashed lines are the point and interval nowcasts for the IMAI 31-33 annual growth rates, calculated from a logarithmic differences model, while solid line represents official values for IMAI 31-33 annual growth rates, initially published as preliminary figures. Empirical correlation between observed and estimated values equals 0.73

Fig. 12



Dashed lines are the point and interval nowcasts for the IMAI 31-33 monthly growth rates, calculated from a logarithmic differences model, while solid line represents official values for IMAI 31-33 monthly growth rates, initially published as preliminary figures. Empirical correlation between observed and estimated values equals 0.95

Table 4 shows the distribution of absolute errors between observed and estimated historical values. From Table 4 and from Figures 10, 11 and 12, we can see that for any month in the time interval August 2015 – February 2020, magnitudes

of absolute estimation errors are similar across the 3 estimated quantities: index values, annual and monthly growth rates. For most months, the officially published value has “landed” inside the 95% prediction intervals, which have a mean width of between 5 and 6 percentage points for any of the 3 estimated quantities (this means that, with 95% confidence, our point estimates have a maximum associated error of  $\pm 3\%$ ). Table 4 shows that, empirically, absolute errors tend to group around 1; a very few times, the absolute error has been close to 0, and a very few times, it has been large. In fact, only 4 times the observed IMAI 31-33 value has landed outside the 95% prediction interval: in January 2016, in November 2016, in June 2017 and in July 2019. This means, that, empirically, officially published values have landed inside the prediction intervals 92.7% of the time, which is close to the theoretical (or nominal) 95% confidence level incorporated into prediction intervals.

Table 4

	<b>Index value</b>	<b>Annual growth rate</b>	<b>Monthly growth rate</b>
minimum	0.00	0.00	0.00
1st quartile	0.46	0.40	0.44
median	1.26	1.01	0.91
mean	1.32	1.11	1.11
3rd quartile	1.70	1.44	1.42
maximum	5.27	4.14	4.34

Distribution of absolute estimation errors between observed and estimated historical values for IMAI 31-33 index values, annual and monthly growth rates. Estimated values were obtained with the logarithmic differences model across time interval August 2015 – February 2020

As can be seen from Figure 10, there was a change in the base year for the IMAI 31-33 indicator between September 2017 and October 2017; before that change, the IMAI 31-33 indicator average for the year 2008 was 100.00; after the change, the IMAI 31-33 indicator average for the year 2013 is now 100.00. Basically, this only translates as a rescaling of the IMAI 31-33 indicator; our procedure for estimating monthly growth rates via a logarithmic differences model is unaffected by this change in SNA methodology for constructing the IMAI 31-33 indicator. Note that  $Y_t$  values from appendix 1 correspond to published IMAI 31-33 values under the 2008 base year.

**Evolution of the logarithmic differences model (3.0).** The model chosen for producing successive IMAI 31-33 nowcasts has basically retained its initial form across time interval August 2015 – February 2020; the only modifications we have done to this “definitive” model consist in gradually adding (as needed) indicator variables for modeling significant seasonal effects, similar to the  $I_{oct}$  indicator variable. The logarithmic differences linear regression model for producing the February 2020 IMAI 31-33 estimates has the following functional form:

$$\Delta \log Y_t = \beta_1 \Delta \log X_t + \beta_2 I_{aug} + \beta_3 I_{oct} + \beta_4 I_{nov} + \beta_5 I_{jan} + \varepsilon_t,$$

where  $\Delta \log Y_t$  and  $\Delta \log X_t$  are the respective logarithmic differences for response and explanatory variables  $Y_t$  and  $X_t$  at month  $t$ , as explained in section 4;  $\varepsilon_t$  follows a first order autoregressive process which is corrected via the Cochrane-Orcutt procedure. As of reference month February 2020, the estimation equation is

$$\Delta \log Y_t = 0.5467 \Delta \log X_t + 0.0184 I_{aug} + 0.0420 I_{oct} + 0.0265 I_{nov} - 0.0149 I_{jan};$$

note that the effect of individual seasonal indicator variables on response variable  $\Delta \log Y_t$  is an order of magnitude less than the corresponding effect of  $\Delta \log X_t$ . We compared statistical tests outputs among all fitted logarithmic differences models of the form (3.0) across months and found out that the common estimated coefficients (for variables  $\Delta \log X_t$  and  $I_{oct}$ , and for coefficient  $\rho$ ) have small variance. This is empirical evidence that the relationship between variables  $X_t, Y_t$  is structurally stable across time, independent of business cycles.

## 7. INEGI’s data sharing experience with CFE, conclusions and future work

The monthly data on electric energy consumption that CFE transmits to INEGI are packed into 16 files, each corresponding to one of the 16 regions in which CFE divides the Mexican territory. Transmission of CFE data to INEGI is usually made two weeks after the end of the reference month; sometimes, however, due to technical or administrative difficulties, CFE data have been transmitted to INEGI a few days later than usual. Overall, there have been no months during the realization of this project in which no data has been received from CFE; this has enabled the

successful realization of an empirical evaluation, in real time, of early IMAI 31-33 estimates. In appendix 2, we can see the dates on which INEGI has received data on electric energy consumption from CFE, and the comparison between early IMAI 31-33 estimates and published IMAI 31-33 values. It is important to emphasize that the agreement between INEGI and CFE to share information is rather informal, since a memorandum of understanding (MoU) has not been signed yet. As of this date, INEGI and CFE relationship continues to be cordial, and both institutions are working in the elaboration of a proper MoU.

The IMAI 31-33 nowcasts, generated from August 2015 to February 2020 with the help of a logarithmic differences model relating manufacturing production level to electric energy consumption at national level, have been communicated to some Federal Government Agencies in Mexico, such as the Central Bank of Mexico (Banxico), the Mexican Social Security Institute (IMSS), The Office of the Treasury and Public Credit (SHCP), the Tax Administration Service (SAT), CFE itself, and some other key users within INEGI. Each month, right after generating nowcasts for the IMAI 31-33 indicator, DVERA prepares an official letter with a technical annex, which is sent to the INEGI Presidency, responsible for disseminating the results to other Federal Government Agencies. This official letter clearly states that the results obtained are of an experimental nature. As of the writing of this article, INEGI is studying the possibility of publishing these results on its website, under the category of experimental statistics.

**Some concluding remarks and possible lines for future work.** By processing monthly electric energy consumption data from the majority of large manufacturing establishments in Mexico, we have produced an electric energy consumption index that has a significant linear relationship to the Monthly Manufacturing Production Level in Mexico. This enables INEGI to produce nowcasts for Mexico's Manufacturing Production Level, given the timeliness in the availability of electric energy consumption data. The monitoring in real time of such nowcasts for the last four years has provided empirical evidence in favor of the structural stability of the relationship between electric energy consumption and production level in Mexico's

manufacturing sector. In order to improve the quality of the IMAI 31-33 nowcasts, DVERA is working continually to keep an updated sample of large manufacturing establishments, and is monitoring the evolution of the variables involved, in order to update the “definitive” model if the need arises.

As future work, INEGI is contemplating the possibility of adding an additional explanatory variable to the “definitive” model; specifically, the explanatory variable considered is the monthly production of vehicles in Mexico’s automotive subsector; these data are collected jointly by the Mexican Association of the Automotive Industry (AMIA) and by INEGI. It has been observed that this monthly variable, which is updated only 10 days after the end of the reference month, has a high linear correlation with the IMAI 31-33 indicator. Preliminary exercises to build regression models with IMAI 31-33 as the response variable, and ICEE and the production of vehicles as explanatory variables, have already been carried out.

### **Acknowledgements**

The authors wish to express their deepest gratitude to former INEGI President Eduardo Sojo Garza-Aldape and current INEGI president Julio Alfonso Santaella Castell, for their valuable suggestions in designing the core strategy for building and testing regression models. Special thanks go to Enrique De-Alba-Guerra, INEGI Vice-President, for coming up with the idea of incorporating logarithmic differences into the modeling phase. The authors also wish to thank Arturo Blancas, Head of the Economic Statistics Directorate, Susana Perez, Director of Economic and Agriculture Censuses, and Gerardo Durand, Director of Economic Administrative Registers, for their valuable advice in the realization of this project. The effort from several INEGI colleagues in collecting and organizing the data used in this project is also appreciated. Last but not least, INEGI is deeply indebted to the CFE people, who has provided Electric Energy Consumption data without failing a single month.

## References

- [1] Sistema de Cuentas Nacionales de México. Cuentas de Corto Plazo y Regionales: Fuentes Metodológicas. 20 de Agosto de 2013. INEGI. Link (in Spanish, accessed on August 14, 2019): [https://www.inegi.org.mx/contenidos/programas/imai/2008/metodologias/SCNM\\_Metodologia\\_06.pdf](https://www.inegi.org.mx/contenidos/programas/imai/2008/metodologias/SCNM_Metodologia_06.pdf)
- [2] The R project for statistical computing: <https://www.r-project.org/>
- [3] Davidson, R., & MacKinnon, J. G. (1993). Estimation and inference in econometrics. OUP Catalogue. Oxford University Press, number 9780195060119
- [4] Kutner, M. H.; Nachtsheim, C. J.; Neter, J. (2004). Applied Linear Regression Models (4th ed.). McGraw-Hill Irwin
- [5] D. Cochrane & G.H. Orcutt (1949). Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms. Journal of the American Statistical Association. Volume 44, Issue 245: pages 32–61.

## Appendix 1. Data used in the construction of regression models

$X_t$ variable (Electric Energy Consumption Indicator for Mexico's Manufacturing Sector)					
jan-13	106.568152673383	jan-14	114.625237758366	jan-15	117.661685560915
feb-13	102.206667645052	feb-14	107.385758694238	feb-15	113.239376473580
mar-13	109.354644360592	mar-14	121.593568885955	mar-15	127.468887386019
apr-13	114.018480260521	apr-14	115.651850551774	apr-15	121.600394586046
may-13	118.093431044136	may-14	122.944929934959	may-15	125.156131097925
jun-13	117.073664454848	jun-14	121.397629555821	jun-15	126.368183422728
jul-13	117.837745378730	jul-14	125.798065936502	jul-15	125.567145531943
aug-13	122.554400895002	aug-14	124.362758669831	aug-15	126.822346966188
sep-13	115.584758323962	sep-14	119.778461889876	sep-15	124.084937798500
oct-13	122.215669912183	oct-14	128.065613726125	oct-15	126.795582223334
nov-13	113.448650003214	nov-14	116.099148775570	nov-15	112.609678608192
dec-13	97.149123345972	dec-14	104.573452963140		

$Y_t$ variable (Manufacturing Production Level Index in Mexico)							
jan-12	104.310303360161	jan-13	106.568152673383	jan-14	110.407782662540	jan-15	112.318007211117
feb-12	104.843887511884	feb-13	103.749250762350	feb-14	106.874856884563	feb-15	111.783049452478
mar-12	112.600605883904	mar-13	106.803072271226	mar-14	115.093379457459	mar-15	119.026895895035
apr-12	104.464240014746	apr-13	110.967188200699	apr-14	111.261263734624	apr-15	115.968640523484
may-12	111.753828772396	may-13	112.156581682476	may-14	116.957215562526	may-15	118.141639147214
jun-12	111.546679656662	jun-13	109.984403182279	jun-14	114.408078441256	jun-15	119.163876391430
jul-12	108.111150984771	jul-13	112.106373141296	jul-14	116.461208364193	jul-15	118.710597945713
aug-12	110.795476356651	aug-13	113.277152880165	aug-14	114.818286476764	aug-15	118.600066105228
sep-12	106.790160592184	sep-13	107.627008738172	sep-14	113.241801242496	sep-15	117.291984751955
oct-12	113.413349269598	oct-13	117.404211419188	oct-14	122.404319453146	oct-15	123.884236478199
nov-12	110.500150510510	nov-13	111.312710282635	nov-14	115.445734213437		
dec-12	100.813657763214	dec-13	102.574854163090	dec-14	108.913118770310		

## Appendix 2. Dates on reception of CFE data and production of early estimates; evaluation of early estimates for IMAI 31-33 values

Month of Operation	CFE data reception date	Date of early estimate publication	IMAI 31-33 early estimates						SNA official values
			Reference month	Lower limit	Upper limit	Point estimate	Observed value	Absolute error	Date of publication
dec-2015	08-dec-2015	11-dec-2015	nov-2015	113.6	119.3	116.4	117.6	1.2	11-jan-2016
jan-2016	07-jan-2016	12-jan-2016	dec-2015	107.3	113.0	110.1	111.7	1.6	11-feb-2016
feb-2016	10-feb-2016	12-feb-2016	jan-2016	115.6	120.9	118.2	113.4	4.8	11-mar-2016
mar-2016	08-mar-2016	14-mar-2016	feb-2016	111.6	116.2	113.9	116.2	2.3	11-apr-2016
apr-2016	07-apr-2016	13-apr-2016	mar-2016	116.9	122.3	119.6	117.3	2.3	12-may-2016
may-2016	11-may-2016	13-may-2016	apr-2016	115.7	121.6	118.6	119.9	1.3	10-jun-2016
jun-2016	10-jun-2016	14-jun-2016	may-2016	117.5	123.5	120.4	119.2	1.2	12-jul-2016
jul-2016	12-jul-2016	18-jul-2016	jun-2016	116.3	122.2	119.2	120.6	1.4	11-aug-2016
aug-2016	10-aug-2016	12-aug-2016	jul-2016	117.6	123.5	120.5	118.9	1.6	09-sep-2016
sep-2016	12-sep-2016	14-sep-2016	aug-2016	118.2	124.8	121.4	123.5	2.1	12-oct-2016
oct-2016	12-oct-2016	14-oct-2016	sep-2016	116.6	123.0	119.8	117.9	1.9	11-nov-2016
nov-2016	14-nov-2016	15-nov-2016	oct-2016	120.9	128.2	124.5	123.2	1.3	12-dec-2016
dec-2016	13-dec-2016	14-dec-2016	nov-2016	114.6	120.7	117.6	122.9	5.3	11-jan-2017
jan-2017	13-jan-2017	19-jan-2017	dec-2016	111.8	118.1	114.9	114.4	0.5	10-feb-2017
feb-2017	17-feb-2017	17-feb-2017	jan-2017	112.7	119.9	116.2	117.7	1.5	14-mar-2017
mar-2017	16-mar-2017	17-mar-2017	feb-2017	113.5	121.8	117.6	117.0	0.6	11-apr-2017
apr-2017	21-apr-2017	24-apr-2017	mar-2017	121.4	127.7	124.5	126.9	2.4	12-may-2017
may-2017	16-may-2017	18-may-2017	apr-2017	116.9	122.9	119.9	117.0	2.8	09-jun-2017
jun-2017	13-jun-2017	15-jun-2017	may-2017	119.9	125.9	122.8	124.7	1.9	12-jul-2017
jul-2017	12-jul-2017	13-jul-2017	jun-2017	123.6	129.9	126.7	123.4	3.3	11-aug-2017
aug-2017	17-aug-2017	18-aug-2017	jul-2017	116.2	122.2	119.1	120.6	1.4	11-sep-2017
sep-2017	12-sep-2017	13-sep-2017	aug-2017	123.1	129.2	126.1	127.2	1.1	12-oct-2017
oct-2017	16-oct-2017	17-oct-2017	sep-2017	117.9	124.0	120.9	121.2*	0.3	10-nov-2017
nov-2017	14-nov-2017	17-nov-2017	oct-2017	112.8	120.2	116.5	114.9	1.6	12-dec-2017
dec-2017	11-dec-2017	14-dec-2017	nov-2017	110.2	119.1	114.6	114.7	0.1	11-jan-2018
jan-2018	30-jan-2018	01-feb-2018	dec-2017	104.6	112.8	108.6	105.9	2.7	09-feb-2018
feb-2018	22-feb-2018	27-feb-2018	jan-2018	107.6	114.7	111.1	111.5	0.4	13-mar-2018
mar-2018	16-mar-2018	23-mar-2018	feb-2018	105.4	111.9	108.6	108.4	0.2	11-apr-2018
apr-2018	19-apr-2018	04-may-2018	mar-2018	110.9	117.1	113.9	113.9	0.0	11-may-2018
may-2018	22-may-2018	04-jun-2018	apr-2018	110.1	116.2	113.1	112.5	0.6	11-jun-2018
jun-2018	20-jun-2018	05-jul-2018	may-2018	115.3	121.8	118.5	116.8	1.7	12-jul-2018
jul-2018	10-jul-2018	01-aug-2018	jun-2018	112.8	119.2	116.0	115.5	0.5	10-aug-2018
aug-2018	17-aug-2018	03-sep-2018	jul-2018	108.6	114.7	111.6	112.8	1.2	11-sep-2018
sep-2018	19-sep-2018	02-oct-2018	aug-2018	115.5	123.0	119.2	119.6	0.4	12-oct-2018
oct-2018	19-oct-2018	30-oct-2018	sep-2018	111.4	117.8	114.5	113.3	1.2	09-nov-2018
nov-2018	27-nov-2018	30-nov-2018	oct-2018	115.3	122.4	118.8	118.5	0.3	12-dec-2018
dec-2018	19-dec-2018	08-jan-2019	nov-2018	111.7	119.0	115.3	116.8	1.5	11-jan-2019
jan-2019	10-jan-2019	01-feb-2019	dec-2018	104.2	110.1	107.1	106.3	0.8	11-feb-2019
feb-2019	22-feb-2019	01-mar-2019	jan-2019	109.5	116.0	112.7	112.7	0.0	13-mar-2019
mar-2019	26-mar-2019	03-apr-2019	feb-2019	107.0	112.6	109.8	109.8	0.0	11-apr-2019
apr-2019	12-apr-2019	23-apr-2019	mar-2019	113.8	119.9	116.8	117.3	0.5	10-may-2019
may-2019	29-may-2019	31-may-2019	apr-2019	107.5	113.4	110.4	112.0	1.6	11-jun-2019
jun-2019	11-jun-2019	27-jun-2019	may-2019	114.1	120.3	117.2	117.8	0.6	12-jul-2019
jul-2019	10-jul-2019	22-jul-2019	jun-2019	113.4	119.5	116.4	114.9	1.5	9-aug-2019
aug-2019	16-aug-2019	27-aug-2019	jul-2019	109.5	115.4	112.4	115.9	3.1	11-sep-2019
sep-2019	20-sep-2019	27-sep-2019	aug-2019	117.3	124.5	120.8	119.2	1.6	11-oct-2019
oct-2019	23-oct-2019	31-oct-2019	sep-2019	109.5	115.5	112.5	114.1	1.6	11-nov-2019
nov-2019	26-nov-2019	05-dec-2019	oct-2019	113.5	120.2	116.8	116.8	0.0	12-dec-2019
dec-2019	13-dec-2019	19-dec-2019	nov-2019	112.5	119.5	115.9	114.0	1.9	10-jan-2020
jan-2020	27-jan-2020	05-feb-2020	dic-2019	103.0	108.8	105.9	105.7	0.2	11-feb-2020
feb-2020	26-feb-2020	06-mar-2020	ene-2020	109.2	115.7	112.4	113.1	0.7	13-mar-2020
mar-2020	13-mar-2020	24-mar-2020	feb-2020	106.2	112.0	109.1	109.3	0.2	8-apr-2020

### Appendix 3. Formulas for computing nowcasts using the models described in section 4

**model (1.0):** the estimation for the November 2015 IMAI 31-33 value,  $\hat{Y}_t$ , was obtained by simply using the corresponding estimation equation, together with the  $X_t$  value for November 2015 from appendix 1; the estimation for the corresponding annual variation was computed by using the formula  $100(\hat{Y}_t/Y_{t-12} - 1)$ , while the corresponding monthly growth rate was computed with the formula  $100(\hat{Y}_t/Y_{t-1} - 1)$ ; note that we must use the values from appendix 1 to compute the estimates for annual and monthly growth rates, since they are computed in real time; i.e., before the published revised figures  $Y_{t-12}$  and  $Y_{t-1}$  are available.

**model (2.1):** the estimation for the November 2015 IMAI 31-33 value,  $\hat{Y}_t$ , as in the case of model (1.0), was obtained by using the corresponding estimation equation, together with the needed values  $Y_{t-1}$ ,  $X_t$  and  $X_{t-1}$  from appendix 1. The corresponding estimations for the annual and monthly growth rates were obtained using the same procedure as in the case of model (1.0).

**model (3.0) with Cochrane-Orcutt correction:** unlike models (1.0) and (2.1), in this case we are directly estimating the November 2015 IMAI 31-33 monthly growth rate from the corresponding estimation equation; for this we need to plug in the values  $X_{nov2015}$  and  $X_{oct2015}$  from appendix 1. Now, to estimate the corresponding annual growth rate, we first note that  $\log(Y_t/Y_{t-12})$  acts as an approximation to the annual growth rate of  $Y_t$ , in the same way  $\log(Y_t/Y_{t-1})$  acts as an approximation to the monthly growth rate of  $Y_t$ ; we also observe that

$$\log Y_t - \log Y_{t-12} = \log Y_t - \log Y_{t-1} + \log Y_{t-1} - \log Y_{t-2} + \dots + \log Y_{t-11} - \log Y_{t-12},$$

or equivalently,

$$\log(Y_t/Y_{t-12}) = \Delta \ln(Y_t) + \Delta \ln(Y_{t-1}) + \dots + \Delta \ln(Y_{t-11}).$$

From this last expression, it seems reasonable to approximate the annual growth rate of  $Y_{nov2015}$  by using the sum  $\Delta \widehat{\ln}(Y_t) + \Delta \ln(Y_{t-1}) + \dots + \Delta \ln(Y_{t-11})$ , where  $\Delta \widehat{\ln}(Y_t)$  is the output from the estimation equation, and the quantities

$\Delta \ln(Y_{t-1}), \Delta \ln(Y_{t-2}), \dots, \Delta \ln(Y_{t-11})$  can be computed from the data in appendix 1. This is precisely the procedure we used to estimate the annual growth rate from model (3.0) shown in table 3.

Finally, it is also possible to compute an estimation for the IMAI 31-33 indicator from the quantity  $\Delta \widehat{\ln(Y_t)}$ ; just solve for  $Y_t$  in expression  $\ln(Y_t) - \ln(Y_{t-1}) = \Delta \ln(Y_t)$ :

$$\ln(Y_t) = \Delta \ln(Y_t) + \ln(Y_{t-1})$$

$$Y_t = e^{\Delta \ln(Y_t) + \ln(Y_{t-1})}$$

$$Y_t = Y_{t-1} e^{\Delta \ln(Y_t)}$$

From this last expression, it seems reasonable to approximate  $Y_t$  by using

$$\hat{Y}_t = Y_{t-1} e^{\Delta \widehat{\ln(Y_t)}}.$$

We used this last expression to estimate the IMAI 31-33 value from model (3.0) shown in table 3.