# Temporal Disaggregation and Forecasts of an UnobservableTimeSeries. The Case of Mexico's Monthly GDP 

Víctor M. Guerrero ${ }^{1)}$<br>${ }^{1)}$ Instituto Tecnológico Autónomo de México (ITAM), Department of Statistics<br>(guerrero@itam.mx)


#### Abstract

The temporal disaggregation problem consists of deriving high frequency data from less frequent observations of a time series. This problem usually occurs when carrying out analysis of the economic situation. Several analysts have proposed methodologies to obtain high frequency (say monthly) data from less frequent (say quarterly) observations of such an important economic variable as Gross Domestic Product (GDP).

This work proposes a model-based solution to this problem. In fact, a method is proposed to disaggregate historical values of the unobserved time series in one step. Another method is also suggested to predict its future values. The procedures involved are derived from a statistical model that links the unobserved data with a preliminarily estimated series and with another series of temporally aggregated values. It is assumed that the preliminary series can be estimated by the use of related variables through a Linear Regression Model. This procedure produces a preliminary series that does not necessarily satisfy the accounting restrictions that the unobserved one is expected to fulfill (e.g. the monthly GDP values must average to the quarterly, observable, figure). Thus the accounting restrictions are taken into consideration by means of an already known theoretical result that produces the Minimum Mean Square Error Linear Estimator of the unobserved series, given the preliminary series.

To operationalize the previous result, a time series model for the differences between the preliminary and the unobserved series is required. Since that model cannot be obtained from observed data, it is suggested to apply a result that leads us to estimate an Auto-Regressive Moving Average (ARMA) model from the aggregated differences and then disaggregate this model to get the required model for the unobserved differences. Another model is needed to forecast the preliminary series from its own past. Once this model as well as that for the differences, have been estimated, it is possible to obtain forecasts of the unobserved series.The already established results employed are known to be optimal in a statistical sense and allow the analyst to make inferences about the unobserved series.

Mexico's monthly GDP is employed as an illustrative example. In Mexico, GDP is measured only on a quarterly basis. However, the basic need of analyzing the economic situation many times requires more frequent data. Thus the suggested procedures are applied to derive the monthly figures of GDP.


## 1. Introduction

Several analysts have proposed different methodologies to obtain high frequency data (say monthly) from less frequent observations (say quarterly) of such an important economic variable as Gross Domestic Product (GDP). Friedman (1962) is one of the pioneers in this area and suggested using related variables to estimate the unobserved one from observations on the others. His proposal was incomplete since the method does not produce an estimated series satisfying the accounting restrictions that the unobserved variable has to fulfill. Some other works did pay attention to the accounting restrictions, but did not employ related variables. Such was the case of Cohen, Müller and Padberg (1971). Nowadays the methods proposed by Chow and Lin (1971) and Denton (1971) are probably the most frequently used in practice, because they take into account both the information provided by related variables and the temporal restrictions on the unobserved series. These methods fail to consider the fact that the most important feature of a time series variable is its autocorrelation structure. In contrast, in the solutions suggested by Guerrero (1990) and Wei and Stram (1990) the main focus was placed on the use of the appropriate
autocorrelation structure. Unfortunately, the latter methods are not completely advisable in practice for reasons to be mentioned later. Some other works dealing with the temporal disaggregation problem, are those of Hillmer and Trabelsi (1987), Chen, Cholette and Dagum (1997) and Nieto (1998).

This paper presents a practical method that shares some of the most desirable features of the previous ones. In fact, (a) it employs related variables to obtain a preliminary series, (b) it includes the appropriate autocorrelation structure, estimated from observed data and (c) disaggregates the aggregated series in a statistical optimal way.

## 2. A Statistical Model

Let $\left\{Z_{t}\right\}$, for $t=1, \ldots, m n$, be an unobserved series, where $n \geq 1$ denotes the number of whole periods (say quarters) and $m \geq 2$ is the intraperiod frequency (say months, in which case $m=3$ ). Let us suppose that $\left\{W_{t}\right\}$ is a preliminary series of estimates of the unobserved data. Given this series we postulate that

$$
\begin{equation*}
\mathrm{Z}_{\mathrm{t}}=\mathrm{W}_{\mathrm{t}}+\mathrm{S}_{\mathrm{t}} \text {, with }\left\{\mathrm{S}_{\mathrm{t}}\right\} \text { an unobserved stationary process with mean zero. } \tag{1}
\end{equation*}
$$

Assumption 1. An Autoregressive and Moving Average (ARMA) model captures the structure of $\left\{\mathrm{S}_{\mathrm{t}}\right\}$, that is

$$
\begin{equation*}
\phi_{S}(B) S_{t}=\theta_{S}(B) e_{t} \tag{2}
\end{equation*}
$$

where $\phi_{S}(B)=1-\phi_{S, 1} B-\ldots-\phi_{S, p} B^{p} \quad$ and $\quad \theta_{S}(B)=1+\theta_{S, 1} B+\ldots+\theta_{S, q} B^{q}$ are polynomials in the operator $B$ such that $\mathrm{BX} \mathrm{X}_{\mathrm{t}}=\mathrm{X}_{\mathrm{t}-1}$ for every variable X and t . Those polynomials are prime with the roots of $\phi_{\mathrm{S}}(\mathrm{x})=0$ and $\theta_{\mathrm{S}}(\mathrm{x})=0$ outside the unit circle, in such a way that they correspond to a stationary and invertible process. Besides, $\left\{e_{t}\right\}$ is a Gaussian white noise process with mean zero and variance $\sigma_{e}^{2}$.

Assumption 2. The following Autoregressive Integrated and Moving Average (ARIMA) model is valid

$$
\begin{equation*}
\phi_{\mathrm{w}}(\mathrm{~B}) \mathrm{d}(\mathrm{~B}) \mathrm{W}_{\mathrm{t}}=\theta_{\mathrm{w}}(\mathrm{~B}) \mathrm{a}_{\mathrm{t}} \tag{3}
\end{equation*}
$$

where $d(B)$ is a differencing operator that renders $\left\{d(B) W_{t}\right\}$ stationary. Whereas $\phi_{W}(B)$ and $\theta_{W}(B)$ are the autoregressive (AR) and moving average (MA) polynomials whose roots are outside the unit circle. The process $\left\{\mathrm{a}_{\mathrm{t}}\right\}$ is a zero-mean Gaussian white noise with variance $\sigma_{a}^{2}$ and is uncorrelated with $\left\{\mathrm{e}_{\mathrm{t}}\right\}$.

Model (2) can be written equivalently as

$$
\begin{equation*}
S_{t}=\psi_{S}(B) e_{t} \tag{4}
\end{equation*}
$$

with $\psi_{S}(B)=1+\psi_{\mathrm{S}, 1} \mathrm{~B}+\psi_{\mathrm{S}, 2} \mathrm{~B}^{2}+\ldots$ the pure MA polynomial, obtained from the relation $\psi_{\mathrm{S}}(\mathrm{B}) \phi_{\mathrm{S}}(\mathrm{B})=\theta_{\mathrm{S}}(\mathrm{B})$, by equating coefficients of powers of B. Expression (4) enables us to write

$$
\begin{equation*}
\mathbf{S}=\Psi_{\mathrm{S}} \mathbf{e} \tag{5}
\end{equation*}
$$

with $\mathbf{S}=\left(\mathrm{S}_{1}, \ldots, \mathbf{S}_{\mathrm{mn}}\right)^{\prime}$ and $\mathbf{e}=\left(\mathrm{e}_{1}, \ldots, \mathrm{e}_{\mathrm{m}}\right)^{\prime}$, where the prime sign denotes transposition, and $\Psi_{\mathrm{S}}$ is an $\mathrm{mn} \times \mathrm{mn}$ lower triangular
matrix with 1 's on the main diagonal, $\psi_{\mathrm{S}, 1}$ on its first subdiagonal, $\psi_{\mathrm{S}, 2}$ on its second subdiagonal and so on. For (5) to be completely equivalent to (4), for $\mathrm{t}=1, \ldots, \mathrm{mn}$, we require that $\mathrm{e}_{\mathrm{t}}=0$ for $\mathrm{t} \quad 0$.

On the other hand, the aggregated data of the unobserved series can be written as $\left\{Y_{1}, \ldots, Y_{n}\right\}$ with

$$
\begin{equation*}
Y_{i}={ }_{j=1}^{m} c_{j} Z_{m(i-1)+j} \text { for } i=1, \ldots, n \tag{6}
\end{equation*}
$$

where the g's are known constants, defined by the type of aggregation under consideration. Let us now define the matrix $\mathbf{C}=\otimes \mathbf{c}^{\prime}$ with $\otimes$ denoting Kronecker product and $\mathbf{c}^{\prime}=\left(\mathrm{c}_{1}, \ldots, \mathrm{c}_{\mathrm{n}}\right)$. If we let $\mathbf{Y}=\left(\mathrm{Y}_{1}, \ldots, \mathrm{Y}_{\mathrm{n}}\right)^{\prime}$ and $\mathbf{Z}=\left(\mathrm{Z}_{1}, \ldots, \mathrm{Z}_{\mathrm{m}}\right)^{\prime}$ then the whole set of restrictions can be written as

$$
\begin{equation*}
\mathbf{Y}=\mathbf{C Z} \tag{7}
\end{equation*}
$$

## 3. Optimal Disaggregation

From (7) and (1) written as $\mathbf{Z}=\mathbf{W}+\mathbf{S}$, where $\mathbf{W}=\left(\mathrm{W}_{1}, \ldots, \mathrm{~W}_{\mathrm{m}}\right)^{\prime}$, will allow us to use the Basic Combination Rule of Guerrero and Peña (2000). First note that $E(\mathbf{Z} \mid \mathbf{W})=\mathbf{W}$, so that $\mathbf{W}$ is the Best Linear Estimator (BLE) of $\mathbf{Z}$ based on $\mathbf{W}$. By BLE it will be understood a linear estimator with minimum Mean Square Error (MSE). It should be noted that rather than estimator we could have used the term predictor, but we reserve the term predictor for the forecasting situation. Moreover, (5) implies that $\Sigma_{S}=\sigma_{\mathrm{e}}^{2} \Psi_{\mathrm{S}} \Psi^{\prime}{ }_{\mathrm{S}}$. Hence we get the following theoretical result.

Proposition. The BLE of $\mathbf{Z}$, given $\mathbf{W}$ and $\mathbf{Y}$, is given by

$$
\begin{equation*}
\mathbf{Z}=\mathbf{W}+\mathrm{A}(\mathbf{Y}-\mathbf{C W}) \tag{8}
\end{equation*}
$$

with MSE matrix

$$
\begin{equation*}
\operatorname{MSE}(\mathbf{Z})=\sigma_{\mathrm{e}}^{2}\left(\mathrm{I}_{\mathrm{mn}}-\mathrm{AC}\right) \Psi_{\mathrm{S}} \Psi_{\mathrm{S}}^{\prime} \tag{9}
\end{equation*}
$$

where

$$
\begin{equation*}
\mathrm{A}=\Psi_{\mathrm{S}} \Psi_{\mathrm{S}}^{\prime} \mathrm{C}^{\prime}\left(\mathrm{C}_{\mathrm{S}} \Psi_{\mathrm{S}}^{\prime} \mathrm{C}^{\prime}\right)^{-1} \tag{10}
\end{equation*}
$$

An estimate of $\Psi_{\mathrm{S}}$ can be obtained from the estimated model for the aggregated differences

$$
\begin{equation*}
\mathbf{D}=\mathbf{C S}=\mathbf{C Z}-\mathbf{C W}=\mathbf{Y}-\mathbf{C W} . \tag{11}
\end{equation*}
$$

That is, we assume that $\left\{D_{i}\right\}$ admits the ARMA model

$$
\begin{equation*}
\phi_{\mathrm{D}}(\mathrm{~L}) \mathrm{D}_{\mathrm{i}}=\theta_{\mathrm{D}}(\mathrm{~L}) \varepsilon_{\mathrm{i}}, \text { for } \mathrm{i}=1, \ldots, \mathrm{n} \tag{12}
\end{equation*}
$$

with $\phi_{\mathrm{D}}(\mathrm{L})=1-\phi_{\mathrm{D} 1} \mathrm{~L}-\ldots-\phi_{\mathrm{DP}} \mathrm{L}^{\mathrm{P}} \quad$ and $\quad \theta_{\mathrm{D}}(\mathrm{L})=1+\theta_{\mathrm{D} 1} \mathrm{~L}+\ldots+\theta_{\mathrm{DQ}} \mathrm{L}^{\mathrm{Q}}$ the polynomials in the backshift operator L acting on the aggregated variable. We can use here an ARMA model because the temporal aggregation of an ARMA process, in this case the process $\left\{\mathrm{S}_{\mathrm{t}}\right\}$, produces another ARMA process with different orders for its polynomials. Since $\left\{D_{i}\right\}$ is obtained from the series $\left\{Y_{i}\right\}$ and $\left\{W_{t}\right\}$, model (12) can be built by applying standard time series techniques.

Once that model is built we can use Wei and Stram's (1990) method to disaggregate it. This method produces a model for $\left\{S_{t}\right\}$ from (12) when the series $\left\{D_{i}\right\}$ has no hidden periodicity of order $m$, then the model for the disaggregated series becomes

$$
\begin{equation*}
\phi_{S}(B) \Phi_{S}\left(B^{E}\right) S_{t}=\Theta_{S}\left(B^{E}\right) \theta_{S}(B) e_{t} \tag{13}
\end{equation*}
$$

where E is the seasonality length. Finally, the weights $\psi_{\mathrm{S}, 1}, \psi_{\mathrm{S}, 2}, \ldots$ used by (5) are obtained by equating the coefficients of powers of $B \operatorname{in} \psi_{S}(B) \phi_{S}(B) \Phi_{S}\left(B^{E}\right)=\Theta_{S}\left(B^{E}\right) \theta_{S}(B)$. It should also be noticed that the matrix (9) will produce different variances for the disaggregated values. Since this might be due in part to the initial conditions $\mathrm{e}_{\mathrm{f}}=0$ for $\mathrm{t} \leq 0$, an adjustment to correct for this nonstationarity problem consists in equating all the diagonal elements to the theoretical variance, e.g. if $\left(1-\Phi B^{E}\right) S_{t}=\left(1+\theta_{1}+\ldots+\theta_{q} B^{q}\right) e_{t}$ with $0 \leq q \leq E$, then $\operatorname{Var}\left(S_{t}\right)=\left(1+\theta_{1}^{2}+\ldots+\theta_{q}^{2}\right) \sigma_{e}^{2} /\left(1-\Phi^{2}\right)$.

## 4. Forecasting Future Disaggregated Values

The problem to solve now is that of forecasting the vector $\mathbf{Z}_{\mathrm{F}}=\left(\mathrm{Z}_{\mathrm{mN}+1}, \ldots, \mathrm{Z}_{\mathrm{mN}+\mathrm{H}}\right)^{\prime}$, with $\mathrm{mN} \geq \mathrm{mn}$ the number of previously disaggregated values and $H \geq 1$ the forecast horizon. We assume there are no aggregated values $\left\{Y_{i}\right\}$ available for $i>N$, but we count on the estimated vectors $\mathbf{W}$ and $\mathbf{S}$. As in Nieto (1998), we consider two different situations about the the series $\left\{\mathrm{W}_{\mathrm{t}}\right\}$ during the forecast horizon. (1) There are no preliminary observations for $t>m N$ and (2) observations $W_{m N+1}, \ldots, W_{m N+\eta}$, with $1 \leq \eta \leq H$, are available. In the first case the forecast will be defined as $\mathbf{Z}_{\mathrm{F}}^{(1)}=\left(Z_{\mathrm{mN}+1}^{(1)}, \ldots, Z_{\mathrm{mN}+\mathrm{H}}^{(1)}\right)^{\prime}$ with

$$
\begin{equation*}
\mathrm{Z}_{\mathrm{mN}+\mathrm{h}}^{(1)}=\mathrm{W}_{\mathrm{mN}+\mathrm{h}}+\mathrm{S}_{\mathrm{mN}+\mathrm{h}} \text { for } \mathrm{h}=1, \ldots, \mathrm{H} \tag{14}
\end{equation*}
$$

where $\mathrm{W}_{\mathrm{mN}+\mathrm{h}}$ and $\mathrm{S}_{\mathrm{mN}+\mathrm{h}}$ are obtained from their respective models. Thus the forecasts satisfy

$$
\begin{equation*}
\phi_{\mathrm{w}}(\mathrm{~B}) \mathrm{d}(\mathrm{~B}) \mathrm{W}_{\mathrm{mN}+\mathrm{h}}=\theta_{\mathrm{w}}(\mathrm{~B}) \mathrm{a}_{\mathrm{mN}+\mathrm{h}} \quad \text { and } \phi_{\mathrm{S}}(\mathrm{~B}) \mathrm{S}_{\mathrm{mN}+\mathrm{h}}=\theta_{\mathrm{S}}(\mathrm{~B}) \mathrm{e}_{\mathrm{mN}+\mathrm{h}} \tag{15}
\end{equation*}
$$

with $\mathrm{W}_{\mathrm{mN}+\mathrm{h}-\mathrm{j}}=\mathrm{W}_{\mathrm{mN}+\mathrm{h}-\mathrm{j}}$ and $\mathrm{S}_{\mathrm{mN}+\mathrm{h}-\mathrm{j}}=\mathrm{S}_{\mathrm{mN}+\mathrm{h}-\mathrm{j}}$ if $\mathrm{j} \quad \mathrm{h}$, and $\mathrm{a}_{\mathrm{mN}+\mathrm{h}}=\mathrm{e}_{\mathrm{mN}+\mathrm{h}}=0$ for $\mathrm{h} \quad 1$. Now, since

$$
\begin{equation*}
\mathrm{W}_{\mathrm{mN}+\mathrm{h}}-\mathrm{W}_{\mathrm{mN}+\mathrm{h}}=\psi_{\mathrm{j}=0}^{\mathrm{h}-1} \psi_{\mathrm{W}, \mathrm{j}} \mathrm{a}_{\mathrm{mN}+\mathrm{h}-\mathrm{j}} \quad \text { and } \quad \mathrm{S}_{\mathrm{mN}+\mathrm{h}}-\mathrm{S}_{\mathrm{mN}+\mathrm{h}}={ }_{\mathrm{j}=0}^{\mathrm{h}-1} \psi_{\mathrm{S}, \mathrm{j}} \mathrm{e}_{\mathrm{mN}+\mathrm{h}-\mathrm{j}} \tag{16}
\end{equation*}
$$

with the weights $\psi_{\mathrm{W}}$ coming from $\psi_{\mathrm{W}}(\mathrm{B}) \phi_{\mathrm{W}}(\mathrm{B}) \mathrm{d}(\mathrm{B})=\theta_{\mathrm{W}}(\mathrm{B})$ and the $\psi_{\mathrm{s}}$ 's from model (5).
Expression (16) can be written as

$$
\begin{equation*}
\mathbf{W}_{\mathrm{F}}-\mathbf{W}_{\mathrm{F}}=\Psi_{\mathrm{W}}^{(\mathrm{H})} \mathbf{a}_{\mathrm{F}} \quad \text { and } \quad \mathbf{S}_{\mathrm{F}}-\mathbf{S}_{\mathrm{F}}=\Psi_{\mathrm{S}}^{(\mathrm{H})} \mathbf{e}_{\mathrm{F}} \tag{17}
\end{equation*}
$$

where the vectors with subindex F are defined in analogy with $\mathbf{Z}_{\mathrm{F}}$, and $\Psi_{\mathrm{W}}{ }^{(\mathrm{H})}$ is the lower triangular matrix with elements 1, $\psi_{\mathrm{W}, 1}, \ldots, \psi_{\mathrm{WH}-1}$ on its first column, with $0,1, \psi_{\mathrm{W}, 1}, \ldots, \psi_{\mathrm{WH}-2}$ on its second column and so on, while $\Psi_{\mathrm{S}}{ }^{(\mathrm{H})}$ is defined in a similar fashion as $\Psi_{\mathrm{w}}{ }^{(\mathrm{H})}$. From (17) it follows that

$$
\begin{equation*}
\mathbf{Z}_{\mathrm{F}}-\mathbf{Z}_{\mathrm{F}}^{(1)}=\mathbf{W}_{\mathrm{F}}+\mathbf{S}_{\mathrm{F}}-\mathbf{W}_{\mathrm{F}}-\mathbf{S}_{\mathrm{F}}=\Psi_{\mathrm{W}}^{(\mathrm{H})} \mathbf{a}_{\mathrm{F}}+\Psi_{\mathrm{S}}^{(\mathrm{H})} \mathbf{e}_{\mathrm{F}} \tag{18}
\end{equation*}
$$

in such a way that

In the second case we have

$$
\mathrm{Z}_{\mathrm{mN}+\mathrm{h}}^{(2)}=\begin{array}{lll}
\mathrm{W}_{\mathrm{mN}+\mathrm{h}}+\mathrm{S}_{\mathrm{mN}+\mathrm{h}} & \text { if } \quad \mathrm{h}=1, \ldots, \eta  \tag{20}\\
\mathrm{~W}_{\mathrm{mN}+\mathrm{h}}+\mathrm{S}_{\mathrm{mN+h}} & \text { if } & \mathrm{h}=\eta+1, \ldots, \mathrm{H}
\end{array}
$$

with $W_{m N+h-j}=W_{m N+h-j}$ if $j \quad h-\eta$ and $S_{m N+h-j}=S_{m N+h-j}$ if $j \quad h$. Therefore

$$
\begin{align*}
\mathrm{Z}_{\mathrm{mN}+\mathrm{h}}-\mathrm{Z}_{\mathrm{mN}+\mathrm{h}}^{(2)}= & \mathrm{S}_{\mathrm{mN}+\mathrm{h}}-\mathrm{S}_{\mathrm{mN}+\mathrm{h}} & \text { if } \mathrm{h}=1, \ldots, \eta  \tag{21}\\
& \left(\mathrm{~W}_{\mathrm{mN}+\mathrm{h}}-\mathrm{W}_{\mathrm{mN}+\mathrm{h}}\right)+\left(\mathrm{S}_{\mathrm{mN}+\mathrm{h}}-\mathrm{S}_{\mathrm{mN}+\mathrm{h}}\right) & \text { if } \mathrm{h}=\eta+1, \ldots, \mathrm{~N}
\end{align*}
$$

so that

$$
\begin{aligned}
& \text { H- }-1 \quad \cdots{ }_{\text {H-1 }} \\
& \underset{j=0}{ } \Psi_{\mathrm{W}, \mathrm{j}} \mathrm{a}_{\mathrm{mN}+\mathrm{H}-\mathrm{j}}+\Psi_{\mathrm{j}=0} \Psi_{\mathrm{S}, \mathrm{j}} \mathrm{e}_{\mathrm{mN}+\mathrm{H}-\mathrm{j}}
\end{aligned}
$$

with $0_{\eta}$ the $\eta \times \eta$ zero matrix. Hence, the MSE matrix of the forecast vector is

$$
\begin{equation*}
\operatorname{MSE}\left(\mathrm{Z}_{\mathrm{F}}^{(2)}\right)=\sigma_{\mathrm{e}}^{2} \Psi_{\mathrm{S}}^{(\mathrm{H})} \Psi_{\mathrm{S}}^{(\mathrm{H})}{ }^{\prime}+\sigma_{\mathrm{a}}^{2} 0_{\eta}^{0_{\eta}} \quad \Psi_{\mathrm{W}}^{(\mathrm{H}-\eta)} \Psi_{\mathrm{W}}^{(\mathrm{H}-\eta)^{\prime}} \tag{23}
\end{equation*}
$$

In summary, the forecast of $\mathbf{Z}_{\mathrm{F}}$ is the sum of the forecasts of $\mathbf{W}_{\mathrm{F}}$ and $\mathbf{S}_{\mathrm{F}}$ obtained separately. The MSE matrix of this forecast is the sum of the corresponding matrices for the individual forecast vectors.

## 5. Monthly disaggregation of Mexico's GDP

In Mexico, as in many other countries, GDP is measured only on a quarterly basis (see Table 1). If monthly data were available, the basic need of analysis of GDP in the recent past and the near future could be analyzed. A step in this direction was given by INEGI (Instituto Nacional de Estadística, Geografía e Informática, MEXICO) when it started to produce figures of the monthly indicator of the global economic activity (IMGAE, Base 1993=100) since January 1993. We thus have access to $\mathrm{Y}=\mathrm{GDP}$ (quarterly) and $\mathrm{W}=\mathrm{IMGAE}$ (monthly). IMGAE is a general indicator calculated with the same methodology as GDP, although with less coverage. It takes into account only the Industrial Sector and the Services Sector of the Mexican Economy, but its high intraperiod correlation with GDP will be evident below. When this work was carried out, the available data on IMGAE ran from January 1993 to December 1999.

Table 1. Mexico's Real GDP (in millions of pesos at 1993 prices), aggregated preliminary series and estimated series of differences

|  | Quarter | GDP | INDIAGR | GDPAGR | D |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1993 | I | $1,248,725.34$ | 99.37 | $1,248,463.50$ | 261.84 |
|  | II | $1,260,351.97$ | 100.60 | $1,263,707.25$ | $-3,355.27$ |
|  | III | $1,211,579.72$ | 96.73 | $1,215,916.04$ | $-4,336.32$ |
|  | IV | $1,304,126.86$ | 103.27 | $1,296,666.70$ | $7,460.15$ |
|  |  |  |  |  |  |
|  | I | $1,277,838.03$ | 102.37 | $1,285,542.89$ | $-7,704.85$ |
|  | II | $1,331,435.05$ | 105.93 | $1,329,626.16$ | $1,808.89$ |
|  | III | $1,267,386.31$ | 102.30 | $1,284,718.90$ | $-17,332.59$ |
|  | IV | $1,372,142.33$ | 109.00 | $1,367,529.53$ | $4,612.80$ |
|  |  |  |  |  |  |
| 1995 | I | $1,272.241 .55$ | 101.62 | $1,276,321.16$ | $-4,079.61$ |
|  | II | $1,209,052.70$ | 96.51 | $1,213,147.06$ | $-4,094.36$ |
|  | III | $1,165,580.18$ | 92.62 | $1,165,076.38$ | 503.80 |
|  | IV | $1,275,557.48$ | 100.24 | $1,259,282.00$ | $16,275.49$ |
|  | I | $1,273,078.05$ | 100.81 | $1,266,356.12$ | $6,721.92$ |
|  | II | $1,287,401.28$ | 102.22 | $1,283,766.66$ | $3,634.61$ |
|  | III | $1,248,655.10$ | 99.53 | $1,250,431.75$ | $-1,766.65$ |
|  | IV | $1,366,292.01$ | 107.71 | $1,351,527.45$ | $14,764.56$ |
|  | I | $1,331,526.94$ | 105.64 | $1,325,992.75$ | $5,534.19$ |
|  | II | $1,395,247.46$ | 110.66 | $1,388,018.61$ | $7,228.85$ |
|  | III | $1,342,047.95$ | 108.06 | $1,355,948.87$ | $-13,900.92$ |
|  | IV | $1,457,278.33$ | 115.98 | $1,453,831.52$ | $3,446.81$ |
|  |  |  |  |  |  |
| 1998 | I | $1,430,820.67$ | 114.01 | $1,429,452.38$ | $1,368.29$ |
|  | II | $1,454,490.59$ | 115.82 | $1,451,850.95$ | $2,639.63$ |
|  | III | $1,411,536.62$ | 113.81 | $1,426,933.67$ | $-15,397.06$ |
|  | IV | $1,495,691.40$ | 119.11 | $1,492,464.42$ | $3,226.98$ |
|  | I | $1,457,161.35$ | 115.95 | $1,453,470.11$ | $3,691.23$ |
|  | II | $1,500,167.45$ | 120.05 | $1,504,081.03$ | $-3,913.58$ |
|  | III | $1,472,607.44$ | 118.39 | $1,483,604.18$ | $-10,996.74$ |
|  | IV | $1,574,096.55$ | 125.41 | $1,570,398.65$ | $3,697.90$ |

A quarterly indicator, called INDIAGR, was first built by averaging the monthly figures of IMGAE, then a Linear Regression Model was fitted to the aggregated data yielding the following results for quarters $i=1, \ldots, 28$ (that is, 1993:I to 1999:IV) with standard errors in parentheses.

$$
\begin{align*}
\mathrm{GDP}_{\mathrm{i}} & =20311.79+12359.80 \text { INDIAGR }_{\mathrm{i}}, \quad \overline{\mathrm{R}}^{2}=0.9938, \mathrm{DW}=2.23  \tag{24}\\
& (20231.38)(188.04)
\end{align*}
$$

The monthly preliminary data was obtained for $t=1, \ldots, 84$ (January 1993 to December 1999) with the equation

$$
\begin{equation*}
\mathrm{W}_{\mathrm{t}}=20311.79+12359.80 \mathrm{IMGAE}_{\mathrm{t}} \tag{25}
\end{equation*}
$$

These figures were aggregated to the quarter to get the values of $\operatorname{GDPAGR}_{\mathrm{i}}$ for $\mathrm{i}=1, \ldots, 28$, as well as the differences $\mathrm{D}_{\mathrm{i}}=\mathrm{GDP}_{\mathrm{i}}-\mathrm{GDPAGR}_{\mathrm{i}}$ and they are shown in Table 1. Equation (24) shows a strong linear relationship between GDP and INDIAGR and the Durbin-Watson statistic does not show evidence of inadequacy. Since the slope coefficient is more than 65 times its standard error INDIAGR was considered a good predictor of GDP. Even though the intercept is not significantly
different from zero, it was included in equation (25) to avoid possible biases when predicting $\left\{\mathrm{W}_{\mathrm{t}}\right\}$.
The autocorrelations for series $\left\{D_{i}\right\}$ were calculated from the 28 data points of the series and they allowed us to identify a seasonal ARMA model. The estimation results of such a model are

$$
\begin{align*}
& \left(1-0.6001 \mathrm{~L}^{4}\right) \mathrm{D}_{\mathrm{i}}=\varepsilon_{\mathrm{i}} \text { with } \sigma_{\varepsilon}=6905.45  \tag{26}\\
& \quad(0.1730) \tag{0.1730}
\end{align*}
$$

and Ljung-Box statistic $\mathrm{Q}^{\prime}(5)=4.95$. When comparing this value against a $\chi^{2}$ distribution with 2 degrees of freedom, there is no reason to doubt of the model adequacy. In order to disaggregate this model, the following seasonal AR polynomial was defined

$$
\begin{equation*}
\Phi(\mathrm{B})=1-0.6001 \mathrm{~B}^{12} \tag{27}
\end{equation*}
$$

and a deseasonalized series was obtained from $\left\{D_{i}\right\}$ by applying the filter

$$
\begin{equation*}
\mathrm{FD}_{\mathrm{i}}=\mathrm{D}_{\mathrm{i}}-0.6001 \mathrm{D}_{\mathrm{i}-4} \text { for } \mathrm{i}=5, \ldots, 28 . \tag{28}
\end{equation*}
$$

The nonseasonal AR and MA polynomials of the disaggregated series of differences were identified by analyzing the sample autocorrelation function of $\left\{\mathrm{FD}_{\mathrm{i}}\right\}$. None of these autocorrelations was significantly different from zero. Hence, according to Wei and Stram's (1990) method, the polynomial orders must be $\mathrm{p}=0$ and $\mathrm{q}=\mathrm{p}+1=1$.

Since the aggregation in the present case is of the form

$$
\begin{equation*}
\mathrm{FD}_{\mathrm{i}}=\frac{1}{3}\left(1+\mathrm{B}+\mathrm{B}^{2}\right) \mathrm{FS}_{3 \mathrm{i}} \tag{29}
\end{equation*}
$$

then the variance and autocovariances of the aggregated and disaggregated series become

$$
\begin{gather*}
\gamma_{\mathrm{FD}}(0)=\frac{1}{9}\left(1+\mathrm{B}+\mathrm{B}^{2}\right)^{2} \gamma_{\mathrm{FS}}(2)=\frac{1}{9}\left[\gamma_{\mathrm{FS}}(-2)+2 \gamma_{\mathrm{FS}}(-1)+3 \gamma_{\mathrm{FS}}(0)+2 \gamma_{\mathrm{FS}}(1)+\gamma_{\mathrm{FS}}(2)\right]  \tag{30}\\
\gamma_{\mathrm{FD}}(1)=\frac{1}{9}\left(1+\mathrm{B}+\mathrm{B}^{2}\right)^{2} \gamma_{\mathrm{FS}}(5)=\frac{1}{9}\left[\gamma_{\mathrm{FS}}(1)+2 \gamma_{\mathrm{FS}}(2)+3 \gamma_{\mathrm{FS}}(3)+2 \gamma_{\mathrm{FS}}(4)+\gamma_{\mathrm{FS}}(5)\right] \tag{31}
\end{gather*}
$$

where $\gamma_{\mathrm{FD}}(\mathrm{k})=0=\gamma_{\mathrm{FS}}(\mathrm{k})$ for $\mathrm{k} \quad 0, \quad 1$. The following system of equations was then obtained

$$
\begin{align*}
& \gamma_{\mathrm{FD}}(0)  \tag{32}\\
& \gamma_{\mathrm{FD}}(1)
\end{align*}=\begin{array}{ccc}
3 / 9 & 4 / 9 & \gamma_{\mathrm{FS}}(0) \\
0 & 1 / 9 & \gamma_{\mathrm{FS}}(1)
\end{array}
$$

which, given that $\gamma_{\mathrm{FD}}(0)=47647902.75$ and $\gamma_{\mathrm{FD}}(1)=\gamma_{\mathrm{FD}}(0) \rho_{\mathrm{FD}}(1)=8187991.91$, yields

$$
\begin{align*}
& \gamma_{\mathrm{FS}}(0)  \tag{33}\\
& \gamma_{\mathrm{FS}}(1)
\end{align*}=\begin{array}{ccc}
3 & -12 & 47647902.75 \\
0 & 9 & 8187991.91
\end{array}=\begin{gathered}
44687805.38 \\
73691927.91
\end{gathered}
$$

This result is inadmissible because it leads to estimate the first autocorrelation of series $\left\{\mathrm{FS}_{\mathrm{t}}\right\}$ as $\rho_{\mathrm{FS}}(1)=\gamma_{\mathrm{FS}}(1) / \gamma_{\mathrm{FS}}(0)=1.6490$, which does not make any sense since for an $\mathrm{MA}(1)$ model the first autocorrelation must satisfy $\mathrm{p}_{\mathrm{FS}}(1) \mid \leq 0.5$. A possible explanation of that result is that some hidden periodicity of order $\mathrm{m}=3$ exists in $\left\{\mathrm{FS}_{\mathrm{t}}\right\}$. Thus the MA polynomial was assumed of order $\mathrm{q}=3$, so that $\gamma_{\mathrm{FD}}(\mathrm{k})=0$ for $\mathrm{k} \neq 0, \pm 1$ and $\gamma_{\mathrm{FS}}(\mathrm{k})=0$ for $\mathrm{k} \neq 0, \pm 3$. The corresponding system of equations became

$$
\begin{gather*}
\gamma_{\mathrm{FD}}(0)  \tag{34}\\
\gamma_{\mathrm{FD}}(1)
\end{gather*}=\begin{array}{ccc}
3 / 9 & 0 & \gamma_{\mathrm{FS}}(0) \\
0 & 3 / 9 & \gamma_{\mathrm{FS}}(3)
\end{array}
$$

with solution

$$
\begin{align*}
& \gamma_{\mathrm{FS}}(0)  \tag{35}\\
& \gamma_{\mathrm{FS}}(3)
\end{align*}=\begin{array}{lll}
3 & 0 & 47647902.75 \\
0 & 3 & 8187991.91
\end{array}=\begin{array}{r}
142943708.30 \\
24563975.72
\end{array}
$$

These autocovariances enabled us to estimate the MA(3) parameter of the model for $\left\{\mathrm{FS}_{\mathrm{t}}\right\}$. Since the theoretical autocovariances for that model are $\gamma_{\mathrm{FS}}(0)=\left(1+\theta_{3}^{2}\right) \sigma_{\mathrm{e}}^{2}$ and $\gamma_{\mathrm{FS}}(3)=\theta_{3} \sigma_{\mathrm{e}}^{2}$, the estimator $\theta_{3}$ comes out by solving

$$
\begin{equation*}
\gamma_{\mathrm{FS}}(3)-\gamma_{\mathrm{FS}}(0) \theta_{3}+\gamma_{\mathrm{FS}}(3) \theta_{3}^{2}=0 \tag{36}
\end{equation*}
$$

Thus we obtained $\theta_{3}=0.1772$ or $\theta_{3}=5.6420$ so that $\theta_{3}$ is chosen to be the former value, in order to ensure invertibility of the model. Hence the estimated model for $\left\{\mathrm{S}_{\mathrm{t}}\right\}$ is given by

$$
\begin{equation*}
\left(1-0.60001 B^{12}\right) S_{t}=\left(1+0.1772 B^{3}\right) e_{t} \tag{37}
\end{equation*}
$$

with $\sigma_{\mathrm{e}}^{2}=\gamma_{\mathrm{FS}}(3) / \theta_{3}=138589937.5$. This model is of type $\left(1-\Phi \mathrm{B}^{\mathrm{E}}\right) \mathrm{S}_{\mathrm{t}}=\left(1+\theta \mathrm{B}^{3}\right) \mathrm{e}_{\mathrm{t}}$ so that the following weights for the pure MA representation are obtained $\psi_{3+12(j-1)}=\Phi^{j-1} \theta$ for $j=1,2, \ldots, \psi_{12 j}=\Phi^{j}$ for $j=0,1, \ldots$ and $\psi_{j}=0$ otherwise. Finally, a correction for nonconstant variance is obtained by modifying the diagonal elements of $\Psi_{S} \Psi_{\mathrm{S}}$ ' so that they take on the values $\operatorname{Var}\left(\mathrm{S}_{\mathrm{t}}\right) / \sigma_{\mathrm{e}}^{2}=\left(1+\theta^{2}\right) /\left(1-\Phi^{2}\right)$.

Once the estimated model for series $\left\{\mathrm{S}_{\mathrm{t}}\right\}$ is known, the Proposition can be applied to disaggregate the GDP directly. The corresponding results of such an application are shown in Figure 1. This figure shows in particular that the preliminary and the direct disaggregated series follow each other very closely. Next we can forecast Mexico's GDP, and to that end we require an ARIMA model for the preliminary series $\left\{\mathrm{W}_{\mathrm{t}}\right\}$. An estimated model for that series became

$$
\begin{equation*}
(1-B)\left(1-B^{12}\right) W_{t}=\left(1-0.3438 B^{10}\right)\left(1-0.8684 B^{12}\right) a_{t}, \quad \sigma_{a}=23462.34 \tag{38}
\end{equation*}
$$

(0.1241)
(0.0895)

This model is empirically supported since the Ljung - Box statistic $Q^{\prime}=19.93$ with 15 degrees of freedom, when compared against a Chi-square distribution produced a significance level of 0.17 . Such a model was then employed to forecast the preliminary series. Table 2 contains the GDP forecasts together with their standard errors, prediction limits and annual rate of growth, when no preliminary observations are yet available for the forecast horizon (January, 2000 to December, 2000).

Figure 1. Monthly disaggregation of Mexico's Real GDP (in millions of pesos at 1993 prices). Preliminary series and disaggregated series, with $95 \%$ limits.


Table 2. Forecasting results for Mexico's Real GDP (with 0 preliminary observations)

| Year | $90 \%$ <br> Limit | Forecasts | $90 \%$ <br> Limit | Standard <br> error |
| :--- | :--- | :--- | :--- | :--- |
| 2000 | $1,482,411.56$ | $1,525,593.09$ | $1,568,774.62$ | $26,250.17$ |
|  | $1,458,401.34$ | $1,516,317.33$ | $1,574,233.32$ | $35,207.29$ |
|  | $1,521,916.28$ | $1,591,514.26$ | $1,661,112.24$ | $42,308.80$ |
|  | $1,447,215.58$ | $1,526,872.83$ | $1,606,530.08$ | $48,423.86$ |
|  | $1,459,792.00$ | $1,548,306.93$ | $1,636,821.87$ | $53,808.47$ |
|  | $1,473,325.45$ | $1,569,888.96$ | $1,666,452.47$ | $58,701.22$ |
|  | $1,427,837.11$ | $1,531,828.11$ | $1,635,819.11$ | $63,216.41$ |
|  | $1,402,266.54$ | $1,513,188.78$ | $1,624,111.03$ | $67,429.94$ |
|  | $1,375,613.78$ | $1,493,058.92$ | $1,610,504.07$ | $71,395.22$ |
|  | $1,476,694.97$ | $1,600,319.31$ | $1,723,943.66$ | $75,151.58$ |
|  | $1,468,618.97$ | $1,594,810.54$ | $1,721,002.12$ | $76,712.20$ |
|  | $1,485,695.86$ | $1,614,403.47$ | $1,743,111.08$ | $78,241.71$ |
| Average |  | $1,552,175.21$ |  |  |

## 6. Conclusions

The proposed procedures are supported by several intermediate and already known results that are optimal to solve a specific part of the problem of temporal disaggregation and forecasting of an unobservable time series. Each of those results is derived on the basis of assumptions that must be empirically validated to maintain its optimality. The most important assumptions were mentioned when deriving the theoretical results here employed. However, another assumption that must be taken into account is that the models for the series of differences is unaffected by structural breaks. Even if no abrupt structural changes occur it is appropriate to check for gradual changes in the series and adapt the models accordingly. Besides, whenever possible, the related variables should be improved to cover more sectors and geographic regions. Finally, the application here detailed for disaggregating Mexico's Real GDP could also be carried out with Mexico's Current GDP. Such an application would basically imply working first with the GDP Implicit Deflator to disaggregate it and then deriving the disaggregated Current GDP from the disaggregated figures of Real GDP and its Implicit Deflator. Many other applications can be devised for
this methodology when analyzing National Accounts, including multiple time series (simultaneous) disaggregation.

## Acknowledgments

The author thankfully acknowledges the financial support of INEGI and Asociación Mexicana de Cultura, A.C. to carry out this work. He is also grateful to Blanca Sainz for many helpful discussions on the topic of this paper.

## References

[1] Friedman, M.; The interpolation of time series by related series, Journal of the American Statistical Association, Vo 1. 57, pp729-757, 1962.
[2] Cohen, K.J., Müller, W. and Padberg, M.W.; Autoregressive approaches to disaggregation of time series data, Applied Statistics, Vol. 20, pp119-29, 1971.
[3] Chow, G.C. and Lin, A.; Best linear interpolation, distribution and extrapolation of time series by related series, Review of Economics and Statistics, Vol. 53, pp372-375, 1971.
[4] Denton, F.T.; Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization, Journal of the American Statistical Association, Vol. 66, pp99-102, 1971.
[5] Guerrero, V.M., Temporal disaggregation of time series. An ARIMA-based approach, International Statistical Review, Vol. 58, pp29-46, 1990.
[6] Wei, W.W.S. and Stram, D.D.; Disaggregation of time series models, Journal of the Royal Statistical Society, Vol. B-52, pp453-467, 1990.
[7] Hillmer, S.O. and Trabelsi, A.; Benchmarking of economic time series, Journal of the American Statistical Association, Vol. 82, pp1064-1071, 1987.
[8] Chen, Z.G., Cholette, P.A. and Dagum, E.B.; A nonparametric method for benchmarking survey data via signal extraction, Journal of the American Statistical Association, Vol. 92, pp1563-1571, 1997.
[9] Nieto, F.; Ex-post and ex-ante prediction of unobserved economic time series: A case study, Journal of Forecasting, Vol. 17, pp35-58, 1998.
[10] Guerrero, V.M. and Peña, D.; Linear combination of restrictions and forecasts in time series analysis, Journal of Forecasting, Vol. 19, pp103-122, 2000.

