Temporal Dissaggregation Method for Estimating Indonesia's Monthly Gross Domestic Product

M Irsyad Ilham[1]

[1] BPS- Statistics Indonesia
Indonesia, email: irsyad.ilham@bps.go.id

Abstract:

Estimation of higher frequency data (say monthly) from less frequent observation of such an important economic variable as Gross Domestic Product (GDP) is a useful monitoring tool for policy-makers. It is also a valuable input in the analysis and development of models in the fields of both macroeconomics and finance. However, the estimate of GDP conducted BPS-Statistics Indonesia, as many other National Statistics Offices, is measured only on a quarterly basis. If monthly data are available, the basic need of analyzing the economic situation could be satisfied more reasonably. Hence, producing the estimate of monthly GDP become important in order to observe the economic progress alongside provide more timely data for the users to take crucial decisions. Therefore, this paper aims to estimate and forecast the monthly Gross Domestic Product in order to obtain the estimated value of the economic activity and monthly-series forecast. In order to derive high frequency data from less frequency observation, this article proposed a direct solution to disaggregate the historical values of quarterly GDP. The novelties of this paper are : (1) Conducting the disaggregation by economic sector for manufacturing sector and (2) Calculating the compatibility test either in each estimates in order to validate the direct disaggregation method. The data used in this research are the real Indonesia's quarterly GDP (including manufacturing sector) and the monthly production index of large and medium manufacturing for the period of 2000 until 2018. The procedures are derived from the statistical model that links the unobserved data with a preliminary estimated series and with the series of aggregated values. Once the monthly GDP series have been estimated, the model therefore to be incorporated into the forecast process. The results showed that there was a strong relationship between GDP and the monthly production index of large and medium manufacturing. In addition, the direct disaggregation for each estimates of monthly GDP series was validated by compatibility test and showed that no reason to doubt the compatibility between preliminary and disaggregated series. Therefore, this production index was considered as a good predictor and was suitable to be used as coincident indicator to describe monthly GDP. According to the findings, the estimates series for those monthly GDP (also include manufacturing) have the similar upward trend each other and pointing to the positive impulses and expectations of a continuous increase by 2019 and 2020.
Keywords: temporal disaggregation; time series; regression model; forecasting; monthly GDP

1. Introduction:

Not having a time series at the desired frequency is common problem for researchers and analyst (Sax and Steiner, 2013). For example, instead of the interpolation or distribution of economic time series data's case, the delayed release of quarterly national account data by national statistics agency is an impediment to the early understanding of the current economic situation.This condition implies that macroeconomic policy making in real time faces the perennial problem of uncovering what is actually happening to the economy (Mitcell, *et.al*. 2005). However, in order to get a picture of the evolution of the current economic activity comprehensively, the methodology to get high frequencies value of economic data needed to be observed.

One of the important economic variable, which released by some National Statistics Office like BPS-Statistics Indonesia, is Gross Domestic Product (GDP). However, the estimates is measured only on a quarterly basis. If monthly data are available, the basic need of analyzing the economic situation could be satisfied more reasonably. Several analyst have proposed methodologies to obtain high frequency data from less frequent observation of such an important economic variable as Gross Domestic Product (GDP). Friedman (1962), one of the pioneers in this area, suggested that we should use related variable to estimate the unobserved one from observations of the others. Chow and Lin (1971) and Denton

(1971) had proposed the method which take into account both the information provided by related variables and the temporal restrictions on the unobserved series. Nevertheless, those methods consider the autocorrelation structure of the time series variable in a subjective way. In contrast, the solution of Guerrero (1990) and Wei and Stram (1990) focused mainly on the use of the appropriate autocorrelation structure.

This empirical study apply the proposed model to Indonesia's quarterly real GDP data for the period of 2000: I to 2018: IV with 76 observations. Transforming quarterly real GDP figures into monthly GDP series use a disaggregation method which applied by Guererro (2003). This study present method that: (a) uses related variables to obtain a preliminary series, (b) includes the appropriate autocorrelation structure deduced from observed data and (c) disaggregate the series in a statistically optimal way. In addition, The novelties of this paper conducted the disaggregation by economic sector for manufacturing sector and thus, calculating the compatibility test either in each estimates in order to validate the direct disaggregation method.

2. Methodology:

There were several methods to disaggregate low frequency series, such as: Denton and Denton-Cholette, and Chow-Lin, Fernandez, and Litterman. For the first method, It use a single indicator as their preliminary series:$p = X$, where X is a $n \; x \; 1$ matrix. Particularly, it may use a constant and then it can be considered as simple mathematical method. In fact, there is no regression in Denton-like methods. For the second, this method use generalized least square regression on low frequency data of $k \geq 1$ variables on a model expressed.

Both two methods above are probably the most frequently used in practice nowadays because they take into account both the information provided by related variables and the temporal restrictions on the unobserved series. Furthermore, In order to derive high frequency data from less frequency observation, this article proposed a direct solution to disaggregate the historical values of quarterly GDP

Let $\{Z_t\}$, for t = 1,..., mn, be an unobserved series, where n > 1 denotes the number of whole periods (say quarters) and m ≥ 2 is the intra-period frequency (say months, in which case m = 3). Let us suppose that $\{W_t\}$ is a possibly non-stationary series of preliminary estimates of unobserved data. Then, given $\{W_t\}$, the following relation holds:

$Z_t = W_t + S_t$, with $\{S_t\}$ a zero-mean unobserved stationary process ……………… (1)

Based on Guerero (2003), the disaggregation model is complemented by following two assumptions.

Assumption 1. An auto-regressive and Moving Average (ARMA) model captures the dynamic structure of $\{S_t\}$, that is

$$\phi_S(B)S_t = \theta_S(B)e_t \; \ldots\ldots\ldots (2)$$

Where $\phi_S(B) = 1 - \phi_{S,1}B - \cdots - \phi_{S,p}B^p$ and $\theta_S(B) = 1 + \theta_{S,1}B + \cdots + \theta_{S,q}B^q$ are polynomials in the backshift operator B such that $BX_t = X_{t-1}$ for every variable X and every index t.

Assumption 2. The series $\{W_t\}$ can be represented by the following Autoregressive Integrated and Moving Average (ARIMA) model:

$$\phi_W(B)d(B)W_t = \theta_W(B)a_t \ldots\ldots(3)$$

Where d(B) is a differencing operator that renders $\{d(B)W_t\}$ stationary, whereas $\phi_W(B)$ and $\theta_W(B)$ are the corresponding autoregressive (AR) and moving average (MA) polynomials whose roots are outside the unit circle. The process $\{a_t\}$ is zero-mean Gaussian White Noise with variance σ_a^2 and is uncorrelated with $\{e_t\}$.Model (2) can be written equivalently as

$$S_t = \psi_S(B)e_t \ldots\ldots\ldots\ldots\ldots\ldots(4)$$

With $\psi_S(B) = 1 + \psi_{S,1}B + \psi_{S,2}B^2 + \cdots$ the pure MA polynomial, obtained from the relation $\psi_S(B)\phi_S(B) = \theta_S(B)$ by equating coefficients of powers of B. Expression (4) leads us to

$$\boldsymbol{S} = \boldsymbol{\Psi_S} \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

With $\boldsymbol{S} = (S_1 \dots S_{mn})'$ and $\boldsymbol{e} = (e_1 \dots e_{mn})'$, where the prime sign denotes transposition, and $\boldsymbol{\Psi_S}$ is mn X mn lower triangular matrix with 1's on its main diagonal, $\psi_{S,1}$ on its first sub-diagonal, $\psi_{S,2}$ on its second sub-diagonal and so on. On the other hand, the aggregated series $\{Y_1 \dots Y_n\}$ can be written as $Y_t = \sum_{j=1}^{m} c_j Z_{m(i-1)+j}$ for $i = 1, \dots, n$ Where the c_j's are known constants defined by the type of aggregation. For instance, disaggregation means distribution if the Y_i are flow values, in which case $\boldsymbol{c}' = (c_1 \dots c_{mn})' = (1, \dots, 1)$ whereby the whole set of restrictions can written as

$$\boldsymbol{Y} = C\boldsymbol{Z} \quad \dots\dots\dots\dots\dots (6)$$

The first stage of disaggregation is regression. Regression of Y_i with additional indicator variable produce the discrepancy between Y_i and aggregate W_t, D_i which has form $D_i = Y_i - WAGR_t$.

Expressions (6) and (1), written in vector notation as $\mathbf{Z} = \mathbf{W} + \mathbf{S}$ with $W = (W_1, \dots, W_{mn})'$ allow us to use the Basic Combination Rule derived by Guerrero and Pena (2000). Moreover (5) implies that $\Sigma_S = \sigma_e^2 \Psi_S \Psi_S'$. In addition, $\mathbf{W}$ is the minimum Mean Squared Error Linear Estimator (MMSELE) of $\mathbf{Z}$ based on $\mathbf{W}$. The MMSELE of $\mathbf{Z}$ given $\mathbf{W}$ and $\mathbf{Y}$ given by

$$\hat{\mathbf{Z}} = \mathbf{W} + \hat{A}(\mathbf{Y} - \mathbf{CW}) \quad \dots\dots\dots\dots\dots\dots\dots\dots (7)$$

Where

$$\hat{A} = \Psi_S \Psi_S' C'(C\Psi_S \Psi_S' C')^{-1} \quad \dots\dots\dots\dots\dots\dots\dots (8)$$

An estimate of $\boldsymbol{\Psi_S}$ can be obtained from an estimated model for the aggregated differences

$$D = CS = CZ - CW = Y - CW$$

We assume $\{D_i\}$ that admits the ARMA model

$\phi_D(L)D_i = \theta_D(L)\varepsilon_i, \qquad \text{for } i = 1, \dots, n$

With $\phi_D(L) = 1 - \phi_{D1}L - \dots - \phi_{DP}L^P$ and $\theta_D(L) = 1 + \theta_{D1}L + \dots + \theta_{DQ}L^Q$ polynomials in the backhift operator L acting on the aggregate variable, in this case the use of an ARMA mode due to the process $\{S_t\}$ produces another ARMA process.

Now, we use Wei and Stram's (1990) method to disaggregate the model. Firstly, choose the seasonal AR and MA polynomials with the same parameter values as those of the model for the aggregated series. Secondly, deseasonalize both the aggregated and disaggregated series by means of the filters

$$FD_i = \boldsymbol{\Phi}_D(L^{E/m})\boldsymbol{\Theta}_D(L^{E/m})^{-1} D_i \quad \text{and} \quad FS_t = \boldsymbol{\Phi}_S(B^E)\boldsymbol{\Theta}_S(B^E)^{-1} S_t$$

Then apply the procedure for nonseasonal series to obtain the model

$\phi_S(B)FS_t = \theta_S(B)e_t$

Thus, the complete model for the disaggregated series of difference becomes

$\phi_S(B)\boldsymbol{\Phi}_S(B^E)S_t = \boldsymbol{\Theta}_S(B^E)\theta_S(B)e_t$

The assumption of $\{W_t\}$ is a true series of preliminary estimates of $\{Z_t\}$ can be validated empirically via a compatibility test. Since $\{Z_t\}$ is unobservable, the test should be carried out with the observed values of $\{\mathbf{CW}\}$ and $\{\mathbf{CZ} = \mathbf{Y}\}$. That is, given $\{W_t\}$

$\mathbf{Y} - \mathbf{CW} = C\Psi_S e \sim N(0, \sigma_e^2 C\Psi_S \Psi_S' C')$

Then a test statistic for the null hypothesis that the two sets of observed data are compatible, i.e., for $H_0\!: \mathrm{E}(\mathbf{Y}|\mathbf{W}) = \mathbf{CW}$, becomes K_{calc}

$K = (\mathbf{Y} - C\mathbf{W})'(C\hat{\Psi}_S \hat{\Psi}_S' C')^{-1}(\mathbf{Y} - C\mathbf{W})/\hat{\sigma}_e^2$

Whose asymptotic distribution, against which K should be compared, is Chi-square with n degrees of freedom.

3. Results

In Indonesia, as in many other countries, the real GDP is measured only on a quarterly basis. BPS-Statistics Indonesia also produce the figures of monthly Large Medium Manufacturing production Index (IP). A quarterly indicator, IPAGR, was built by averaging the monthly figures of IP and a linear regression model was fitted to the aggregated data, yielding the following results for quarters $i = 1, \dots, 76$ (that is, 2000: I to 2018: IV) with standard errors in parentheses

$$GDP_i = -751695 + 23865{,}03 IPAGR_i, \quad R^2 = 0{,}9711$$
$$(50052{,}57) \quad (478{,}56)$$

Then the monthly preliminary data were obtained for $t = 1, \dots, 228$ (January 2000 to December 2018) with the equation

$$W_t = -751695 + 23865{,}03 IP_t$$

The autocorrelation for the series $\{D_i\}$ were calculated from the 72 data points of the series and they allowed us to identify a seasonal ARMA model for $\{D_i\}$. The estimation results of such a model are

$$(1 - 0{,}5943L^4)D_i = \widehat{\varepsilon_i} \quad \text{with } \hat{\sigma}_\varepsilon = 65770{,}49$$

In order to get a disaggregated model, the following seasonal AR polynomial was defined:

$$\widehat{\Phi}(B) = 1 - 0{,}5934B^4$$

And a de-seasonalized series was obtained from $\{D_i\}$ by applying the filter

$$FD_i = D_i - 0{,}5934 D_{i-4} \text{ for } i = 5, \dots, 76$$

Thus the non-seasonal AR and MA polynomials of the disaggregated series of differences were identified by analyzing the sample autocorrelation of $\{FD_i\}$. None of these autocorrelations differed significantly from zero. Hence the polynomial orders were chosen as $p = 0$ and $q = p + 1$, following Wei and Stram's (1990) method.Since the aggregation involved in the present case is of the form

$$FD_i = \frac{1}{3}(1 + B + B^2)FS_{3i}$$

Then the auto-covariances of the aggregated and disaggregated series are

$$\begin{pmatrix} \hat{\gamma}_{FS}(0) \\ \hat{\gamma}_{FS}(1) \end{pmatrix} = \begin{pmatrix} 3 & -12 \\ 0 & 9 \end{pmatrix} \begin{pmatrix} 4320392593 \\ 1736797823 \end{pmatrix} = \begin{pmatrix} -7880396090 \\ 15631180403 \end{pmatrix}$$

This result is inadmissible because it leads one to estimate the first autocorrelation of $\{FS_t\}$ as $\hat{\rho}_{FS}(1) = \hat{\gamma}_{FS}(1)/\hat{\gamma}_{FS}(0) = -1{,}98$. A possible explanation of the aforementioned result is that some hidden periodicity of order $m = 3$ is present in $\{FS_t\}$. Thus the MA polynomial was assumed of order $q = 3$. The corresponding system of equations became

$$\begin{pmatrix} \hat{\gamma}_{FS}(0) \\ \hat{\gamma}_{FS}(3) \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 9 \end{pmatrix} \begin{pmatrix} 4320392593 \\ 1736797823 \end{pmatrix} = \begin{pmatrix} 12961177780 \\ 5210393468 \end{pmatrix}$$

These auto-covariances enabled us to estimate the MA(3) parameter of $FS_t = (1 + \theta_3 B^3)e_t$.Since the theoretical auto-covariances for that model are $\gamma_{FS}(0) = (1 + \theta_3 B^2)\,\sigma_\varepsilon^2$ and $\gamma_{FS}(3) = \theta_3 \sigma_\varepsilon^2$ the estimator $\hat{\theta}_3$ came out by solving the equation

$$\hat{\gamma}_{FS}(3) - \hat{\gamma}_{FS}(0)\hat{\theta}_3 + \hat{\gamma}_{FS}(3)\hat{\theta}_3^2$$

That is

$$\hat{\theta}_3 = \frac{\hat{\gamma}_{FS}(0) \pm \sqrt{\gamma_{FS}^2(0) - 4\gamma_{FS}^2(3)}}{2\hat{\gamma}_{FS}(3)}$$

By plugging the estimated values in this equation we obtained $\hat{\theta}_{31} = 0{,}5042$ and $\hat{\theta}_{32} = 1{,}9834$. Then $\hat{\theta}_3$ was chosen as the former value, in order to ensure invertibility of the model. Hence the estimated model for series $\{S_t\}$ is given by

$$(1 - 0{,}5943L^{12})S_t = (1 + 0{,}5042B^3)\widehat{\varepsilon_i} \quad \text{with } \hat{\sigma}_\varepsilon = \hat{\gamma}_{FS}(3)/\hat{\theta}_3 = 10334136901$$

Once the estimated model for series $\{S_t\}$ is available, Preposition 1 can be applied to disaggregate the GDP series directly. The Figure 1 shows that the preliminary and the direct disaggregated series follow each other closely. Besides the direct disaggregation was validated by the compatibility statistic whose value $K_{calc} = 94{,}11$(which 76 degrees of freedom) yielded a significant level of 5 percent. This result lent empirical support to the assumption of compatibility between preliminary and disaggregated series. Thus, an ARIMA model for the preliminary series $\{W_t\}$ was built, yielding the following results:

$$(1-B)(1-B^{12})W_t = (1 - \underset{(0,0133)}{0,8981}B^{10})(1 - \underset{(0,0173)}{0,8739}B^{12})\hat{a}_t$$

Then we forecasted the unobservable monthly GDP for the next two years (2019 and 2020) by calculating $\{Z_t\}$ based on the result of estimated preliminary series $\{W_t\}$ and the series $\{S_t\}$. The results as well as follows,

Figure 1. Monthly Disagregation of Indonesia's Gross Domestic Product, and forecast result with 95% limit

4. Discussion, Conclusion and Recommendations:

As a brief, the results showed that there was a strong relationship between GDP and the monthly production index of large and medium manufacturing. In addition, the direct disaggregation for each estimates of monthly GDP series was validated by compatibility test and showed that no reason to doubt the compatibility between preliminary and disaggregated series. Therefore, this production index was considered as a good predictor and was suitable to be used as coincident indicator to describe monthly GDP.

To this paper, we also conducted to estimate the monthly GDP for industry sector. The GDP selected as the availability of the data rather than other sector. As well as the result for GDP itself, for that sector there was a strong relationship between the GDP of Industry and the monthly monthly production index of large and medium manufacturing. , the direct disaggregation for each estimates of monthly Industrial GDP sector series was validated by compatibility test and showed that no reason to doubt the compatibility between preliminary and disaggregated series. The Figure 2 showed the similarity behaviour of the unobservable series (monthly GDP of Industry) with GDP.

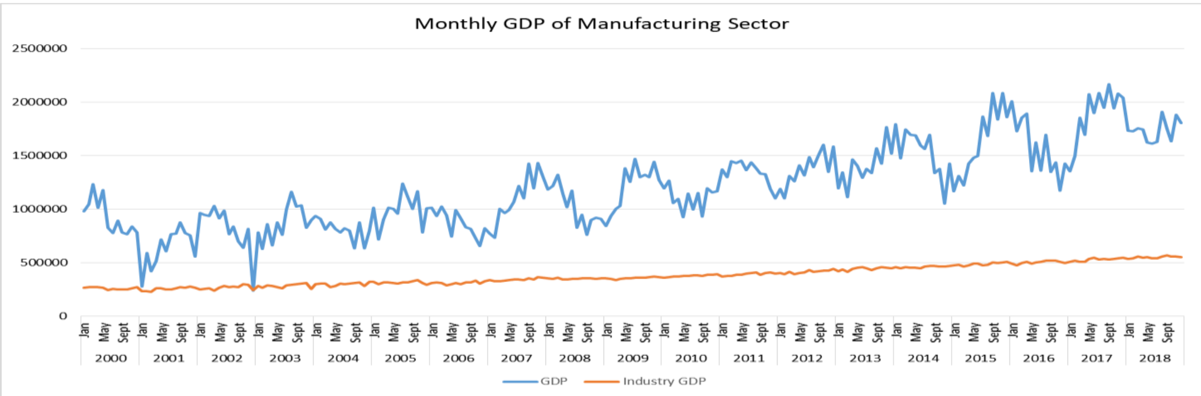

Figure 2. Monthly GDP of Indonesia's Manufacturing Sector

The temporal disaggregation method are supported by several intermediate and already known results that are optimal when it comes to solving a specific part not only for the problem of temporal disaggregation, but also the problem of forecasting of an unobservable time series. Each of those results is derived on the basis of assumptions that must be empirically validated in order to maintain its optimality. As mentioned above, there was two assumption underlying the method of temporal discussion that we conducted here. First, the dynamic series {St} followed an ARMA besides preliminary series {Wt} followed ARIMA process.

However, another assumption that must be taken into account is that the series of differences {St} is unaffected by structural breaks. In our model, we use ARIMA or ARMA model which often called The Box Jenkins Methodology. This methodology had one key assumption that the structure of data-generating process does not change (Enders, 2014). However, in some circumstances, there may be reasons to suspect in the data generating process. For example, the oil price shock, the tragedy of September 2011, financial crisis 2008, or even the recent COVID19 pandemic which had any significant impact on the coefficients. Besides, it should be clear that the ARMA model for the aggregated differences {Di} is obtained with possibly a very small amount of data since only n whole periods are available.

Some recommendation for future research should make use of new and reliable data to continue evaluating mathematical and statistical methods, examine additional factors that may contribute to their ability to estimate the temporal dynamics of target series, and ultimately further the improvement of temporal disaggregation and bench marking methods. Whenever possible it could be improved to cover all macroeconomic sectors (according to the System National Account) and geographic regions. Many other applications can be devised for this methodology when analyzing national accounts, including simultaneous disaggregation of multiple time series.

References:

1. Chow, G.C. and Lin, A. (1971). Best Linear Interpolation, Distribution and Extrapolation of Time Series by Related Series. Review of Economics and Statistics, 33, 372–373.
2. Denton, F.T. (1971). Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization. Journal of the American Statistical Association, 66, 99–102.
3. Enders, Walter. (2014). Applied Econometric Time Series 4th Edition. John Wiley and Sons
4. Friedman, M. (1962). The Interpolation of Time Series by Related Series. Journal of the American Statistical Association, 37, 729–737.
5. Guerrero, V.M. (1990). Temporal Disaggregation of Time Series. An ARIMA-based Approach. International Statistical Review, 38, 29–46.
6. Guerrero, V.M. and Pen˜a, D. (2000). Linear Combination of Restrictions and Forecasts in Time Series Analysis. Journal of Forecasting, 19, 103–122.
7. Guerrero, V.M. (2003). Monthly Disaggregation of a quarterly Time Series and Forecasts Its Unobservable Monthly Values, Journal of Official Statistics. Vol. 19,No.03, 2003, pp. 215-235
8. Mitchell et al. (2005). An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth. The Economic Journal. 115 (February), F108–F129.
9. Sax, Christoph & Steiner Peter. (2013). Temporal Disaggregation of Time. The R Journal Vol. 5/2
10. Wei, W.W.S. and Stram, D.D. (1990). Disaggregation of Time Series Models. Journal of the Royal Statistical Society, Series B, 32, 433–467.