Best Linear Unbiased Disaggregation of Annual GDP to Quarterly Figures: The Case of Malaysia

TILAK ABEYSINGHE1* AND CHRISTOPHER LEE2
1Department of Economics and Statistics, National University of Singapore
2Institute of South East Asian Studies, Singapore

ABSTRACT
Long series of quarterly GDP figures are still not available for many countries. This paper suggests an empirical procedure adapted from Chow and Lin (1971) to derive quarterly estimates from annual GDP figures and produces quarterly GDP by sectors for Malaysia from 1973Q1 onwards. A comparison of these estimates with some univariate interpolations using published quarterly figures for recent years show that the use of related series can produce substantially superior estimates of GDP compared to univariate methods. The data set is available from the authors. © 1998 John Wiley & Sons, Ltd.

KEY WORDS Chow–Lin procedure; univariate interpolation; seasonality; structural time-series approach

INTRODUCTION
With the expansion of multinational investments across developing countries many institutions such as the Asian Development Bank have begun to provide GDP growth forecasts and economic outlook for many developing countries on a regular basis. Many of these countries, however, record GDP figures only annually and because of a limited number of annual observations forecasting is often done more subjectively than objectively.

Long quarterly GDP series are often required for econometric modelling and forecasting. Although a number of univariate data interpolation† methods are readily available in some computer packages, with a little extra effort we can produce more reliable interpolations based on related series. The objective of this paper is to present a practical procedure adapted from Chow and Lin (1971) to estimate quarterly GDP from annual GDP for Malaysia. We generate

* Correspondence to: Tilak Abeysinghe, Department of Economics and Statistics, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, E-mail: ecstabey@leonis.nus.edu.sg or tilakabey@nus.edu.sg
† In this exercise we use the words ‘interpolation’ and ‘disaggregation’ interchangeably.
quarterly GDP by sectors from 1973Q1 onwards. Malaysia began publishing quarterly GDP by sectors only since 1987Q1. We use these series to compare our estimates with univariate interpolations.

The next section presents the basic tools of the Chow–Lin procedure that are necessary for disaggregating an annual variable to quarterly figures. The following two sections go through the empirical methodology we have implemented. In the concluding section we highlight the salient features of our approach. Data sources are given in an Appendix.

METHODOLOGY

There are basically two strands of time series disaggregation procedures found in the literature, those which use related series (see Guerrero, 1990) and those which use univariate methods (see Chan, 1993). As our main objective is to use related series to interpolate quarterly GDP, the following section provides an outline of the Chow–Lin procedure. As for univariate interpolations, we use three procedures available in SAS and avoid presenting details for brevity.

Chow–Lin procedure

Following the work of Chow and Lin (1971), a number of papers have appeared in the literature, which attempt to overcome some of the difficulties associated with the Chow–Lin procedure (see, for example, Fernandez, 1981; Litterman, 1983; Harvey and Pierse, 1984; Guerrero, 1990). In practice, depending on the data availability, one may have to resort to a combination of methods to obtain satisfactory interpolations.

The Chow–Lin procedure, which they presented to convert quarterly observations to monthly interpolations, can easily be adapted to convert annual aggregates to quarterly values. Assume that GDP figures are available annually over \( n \) years. Let \( y \) be a \( (4n \times 1) \) vector of quarterly GDP figures to be estimated. Let \( C \) be a \( (n \times 4n) \) matrix that converts \( 4n \) quarterly observations into \( n \) annual observations. This matrix is defined as:

\[
C = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & \cdots & \cdots & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & \cdots & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 1 & 1 & 1
\end{bmatrix}
\]  

(1)

Now suppose that quarterly \( y \) can be predicted using a multiple linear regression:

\[
y = X\beta + u
\]  

(2)

where \( X \) is a \( (4n \times k) \) matrix of \( k \) predictor variables which are observed quarterly and \( u \) is a \( (4n \times l) \) random vector with zero mean and covariance matrix \( V (4n \times 4n) \). Using subscript \( a \) to denote annual figures, equation (2) can be converted to a regression of annual aggregates:

\[
y_a = Cy = CX\beta + Cu = X_a\beta + u_a
\]  

(3)

© 1998 John Wiley & Sons, Ltd.  
To implement the Chow–Lin procedure first apply the GLS method to equation (3) to obtain $\hat{\beta} = (X_a'V_a^{-1}X_a)^{-1}(X_a'V_a^{-1}y_a)$ where $V_a = (CVC)$ and obtain $\hat{u}_a = y_a - X_a\hat{\beta}$. Finally the Chow–Lin best linear unbiased predictor (disaggregate) of $y$ is derived from:

$$\hat{y} = X\hat{\beta} + VC'(CVC)^{-1}\hat{u}_a \quad (4)$$

The first term on the RHS of equation (4) gives the predicted quarterly $y$ based on observed quarterly $X$ and estimated $\beta$ from annual totals. What the second term does is to allocate annual residuals $\hat{u}_a$ to the four quarters of the year such that the annual sum of the interpolated values equal the observed value $y_a$.

A major drawback of the Chow–Lin procedure is that $V$ is unknown.\(^1\) Subsequent developments in the literature (see the citations above) deal primarily with this issue. Two assumptions Chow and Lin (1971) have considered are (a) $V = \sigma^2I_n$ and (b) $u_t = \rho u_{t-1} + e_t$, where $e_t$ is white noise and $|\rho| < 1$. In cases where a more general structure is required a method suggested by Guerrero (1990) seems appropriate. Our experience shows that by spending time on constructing good related series we can easily avoid the use of complicated models and restrict ourselves to assumptions (a) and (b) above.

Under assumption (a) $\hat{\beta}$ reduces to the OLS estimator $(X_a'X_a)^{-1}(X_a'y_a)$, and the second term on the RHS of equation (4) amounts to allocating one quarter of the annual residual to each quarter of the year. Under assumption (b) $V$ takes the form

$$V = \sigma^2_e \left[\begin{array}{cccc}
1 & \rho & \rho^2 & \ldots & \rho^{4n-1} \\
\rho & 1 & \rho & \ldots & \rho^{4n-2} \\
\rho^{4n-1} & \ldots & \ldots & \ldots & 1
\end{array}\right]$$

and the procedure needs an estimate of $\rho$. Unfortunately there may not be a convenient way of estimating $\rho$. If a sufficient length of quarterly data is available then one may estimate $\rho$ from the OLS residuals of equation (2). In this exercise we use a procedure suggested by Chow and Lin (1971). To estimate $\rho$ we have to solve the polynomial

$$\rho^7 + 2\rho^6 + 3\rho^5 + 4\rho^4 + (3 - 2\hat{\rho}_a)\rho^3 + (2 - 4\hat{\rho}_a)\rho^2 + (1 - 6\hat{\rho}_a)\rho - 4\hat{\rho}_a = 0 \quad (5)$$

where $\hat{\rho}$ is the estimated first order autocorrelation coefficient from the OLS residuals of the annual-data regression (3). We derived equation (5) using the relationship that $\rho_a$ is equal to the ratio of the off-diagonal element to the diagonal element of the matrix $V_a = (CVC)$. Under the assumption that both $\rho_a$ and $\rho$ are positive equation (5) has only one positive real root which is an estimate of $\rho$.

### DISAGGREGATION OF GDP

A successful implementation of regression (2) requires $u_t$ to be a stationary series. This means that (2) must form a cointegrating regression if GDP and the predictor variables are integrated series.

\(^1\)Another point to bear in mind is that the Chow–Lin theory assumes that the predictor series ($X$) are fixed. As in many econometric applications this is unlikely in practice. Therefore, we may have to resort to the assumption that $X$ is independent or uncorrelated with $u$, in which case the optimal properties hold only asymptotically.

© 1998 John Wiley & Sons, Ltd.  
Quite often it is possible to interpolate GDP from the production indexes of three broad categories, agriculture, industry, and service, depending on the data availability. In the case of Malaysia an index of industrial production is available. An index of agricultural production can be constructed using published quarterly growth rates (see Appendix). Usually an index of service production is not available for many countries.

A regression of GDP on agricultural and industrial production indexes cannot form a co-integrating regression as the omitted variable, service production, is unlikely to be stationary. To overcome this problem the basic methodology we use is to estimate sectoral (agriculture, industry, and service) output separately and then aggregate them to derive the GDP. The output of each sector and the corresponding production index are expected to be cointegrated because the latter is an estimate of the former with deviations representing stationary measurement errors. This approach easily accounts for seasonal variation as well. As seasonal variation of many GDP series appear to be evolving (Abeysinghe, 1994) any direct interpolation of GDP may have to be done with seasonally adjusted data to avoid distorting the seasonal component.

Preliminary investigations show that AR(1) errors for industry and service sectors and white-noise errors for agriculture are appropriate assumptions to make. Estimation was carried out using the entire data span from 1973 to 1993 in order to compare our estimates with published quarterly figures which are available from 1987 onwards. The fitted models pass through a number of standard diagnostic tests which are not reported for brevity. The estimation procedure used for each sector is briefly described below.

**Industry**

The estimation of industrial output is the least problematic because an industrial production index is readily available. The GLS estimates based on annual data is given below:

\[
\hat{Output}_t = -655.43 + 89.61 \text{Index}_t + 878.32D_t - 24.43D_t \text{Index}_t
\]

\[
(364.19) \quad (7.81) \quad (489.73) \quad (8.47)
\]

\[
R^2 = 0.99 \quad \hat{\rho}_a = 0.447 \quad \hat{\rho} = 0.724
\]

where the parenthesized values are standard errors and \( D \) is a dummy variable such that \( D = 0 \) if year \( \leq 1983 \) and \( D = 1 \) if year \( \geq 1984 \). This dummy is introduced to account for the shift in the index in 1984 due to a rebasing of the index (see Appendix).

Estimation of equation (6) was carried out using the SAS AUTOREG procedure which provides an estimate of \( \rho_a \), the annual autocorrelation coefficient. Then we used equation (5) to estimate \( \rho \), the quarterly autocorrelation coefficient, which is required to work out the second component on the RHS of equation (4). The coefficient estimates of equation (6) and the quarterly observations generate the first component on the RHS of equation (4). Adding the two components provided interpolated quarterly industrial output.

**Agriculture**

The Appendix describes how a preliminary estimate of agricultural output was obtained using quarterly growth rates published by the Bank Negara Malaysia. The second step is to reconcile these estimates with published annual figures. The difference between the published figures and our estimates (annual) appears to be white noise. Therefore one quarter of this difference was
allocated to the four quarters of the corresponding year so that the four quarters add up to the annual figures.

**Service**

A direct measure of the service sector quarterly output is not available. We therefore first created a quarterly index by predicting the service sector output based on total import and export trade and total commercial bank loans. Our choice of related series was limited by the availability of quarterly data since 1973.

The model we used for this purpose is a structural time-series model:

\[
\begin{align*}
S_t &= \mu_t + \gamma_t + \alpha_1 \text{Loans}_t + \alpha_2 \text{Trade}_t + \epsilon_t \\
\mu_t &= \mu_{t-1} + \beta + \eta_t \\
S(L)\gamma_t &= \omega_t
\end{align*}
\]

where \(S_t\) = service output, \(\mu_t\) and \(\gamma_t\) are trend and seasonal respectively, \(S(L) = 1 + L + L^2 + L^3\), and \(L\) is the lag operator. The three disturbance terms \((\epsilon_t, \eta_t, \omega_t)\) are assumed to be mutually uncorrelated white-noise processes with zero means and constant variances \((\sigma^2_\epsilon, \sigma^2_\eta, \sigma^2_\omega)\). The first equation is a measurement equation with constant \(\alpha_1\) and \(\alpha_2\) and the other two are transition equations (see Harvey, 1989, for details).

Model (7) has the advantage that it allows for a stochastic trend and seasonal with a limiting case of a deterministic trend and seasonal which occurs when \(\sigma^2_\eta = \sigma^2_\omega = 0\). Moreover, the regression need not be a cointegrating relation. In fact, this procedure can be used directly to estimate GDP or any other series using related series. The main requirement is that a sufficient length of quarterly data must be available to estimate the model.

Model (1) was estimated from quarterly data over 1987Q1 to 1993Q4 using the STAMP package. Quarterly predictions of \(S_t\) since 1973 can easily be generated by putting the data series in reverse order and carrying out estimation and forecasting as is normally done. The resulting estimates of model (7), with standard errors in parentheses, are \(\hat{\sigma}^2_\epsilon = 7979.0 (2471.9), \hat{\sigma}^2_\eta = 9.46 (27.01), \hat{\sigma}^2_\omega = 0.00 (1.00), \hat{z}_1 = 0.068 (0.005), \hat{z}_2 = 0.033 (0.009)\). The fitted model passes through the diagnostics provided in STAMP.

The next step of interpolating the service sector output is the same as that for the industrial sector. The GLS estimates of annual (1973–93) service output on the service index derived from model (7) are given below

\[
\begin{align*}
\text{Output}_t &= -870.27 + 40.32 \text{Index}_t \\
&\quad (417.60) \quad (2.79) \\
R^2 &= 0.99 \quad \hat{\rho}_a = 0.71 \quad \hat{\rho} = 0.88
\end{align*}
\]

It should be noted that if a sufficient length of quarterly data of service output is not available then the approach in model (7) cannot be used. Instead, we have to estimate a model of the form (8) with appropriate predictor variables such as \(\text{Loans}\) and \(\text{Trade}\). However, one may have to use a large number of predictor variables, at the expense of degrees of freedom, in order to get white noise or AR(1) residuals.

RESULTS

Figures 1–4 show the estimated series over 1973Q1 to 1993Q4 and published series over 1987Q1 to 1993Q4. For GDP, unpublished estimates provided by the Department of Statistics, Malaysia, run from 1980Q1. The estimated series coincide very closely with the published series. Some departure from the published series can be observed in industrial output over five consecutive quarters (Figure 3) and service output over four consecutive quarters (Figure 4). These effects, however, cancel out in the aggregation process as observed in GDP in Figure 1.

In order to assess the quality of our estimates against the published figures Table I compares the root mean square errors (RMSE%) expressed as a percentage of the mean values of output over 1987–93. In the table we have included three univariate interpolations (spine, join, step) available in SAS. As mentioned earlier, such univariate interpolations are readily available in some computer packages (see Chan, 1993, for details).

Table I shows that, with the exception of the service sector, the univariate interpolations are inferior to our regression estimates. We have to bear in mind that univariate interpolations do not
allow for seasonal variation. Usually the estimation errors are larger when the estimates have to account for seasonal variation. It appears, however, that the selected related series for the service sector do not capture the variation of service output very well. Despite the presence of seasonal variation our estimates of GDP and agricultural output are far superior to univariate estimates.
As mentioned earlier, the estimation errors of sectoral output cancel out in the aggregation process of GDP.

**Further disaggregation of GDP**

Quite often we need quarterly sectoral output figures at a more disaggregated level than the three broad categories above. This can be accomplished by interpolating sectoral shares of GDP by a univariate method and then multiplying the share estimates by the interpolated quarterly GDP figures. This approach is effectively a combination of univariate and related-series methods, the related series being the GDP. In this respect, our earlier effort of interpolating GDP can be considered as an intermediate step in deriving reliable estimates of GDP.

Under this approach, however, we can interpolate the data series only in the seasonally adjusted form. If interpolated sectoral shares are multiplied by the seasonally unadjusted GDP then we are forcing the seasonal variation of each disaggregated series to be the same as that of the GDP series.

In the case of Malaysia, annual output figures are available by ten sectors. We use the share-based procedure suggested above and a direct univariate disaggregation to estimate quarterly sectoral output and then compare the results over the period 1987Q1–1993Q4. Both univariate interpolations (sectoral GDP shares and totals) are carried out using the spline interpolation given in SAS. Interpolated sectoral shares are then multiplied by the seasonally adjusted GDP series which comprises our estimates for 1973Q1–1979Q4 and Malaysian Department of Statistics estimates for the rest of the period.

Table II shows the RMSEs of interpolations against published figures as in Table I. In the table we have excluded two sectors (finance & business services and other) because the quarterly and annual classifications are incompatible. Our final estimates for these two sectors follow the annual classification.
Table II shows that, except for the transport and communication sector, the share-based estimates are more reliable than the estimates based on direct univariate disaggregation. Even for the transport and communication sector the share-based estimates may turn out to be more reliable when we consider the entire estimation period. A comparison of agricultural sector in Tables I and II shows that our original estimates of this sector are better than the share-based estimates. Therefore, our final estimates by sector comprise the seasonally adjusted original agriculture series and share-based estimates for the other series. In the absence of good related series, the above share-based approach can also be used to interpolate the quarterly expenditure components of GDP.

**CONCLUSION**

The procedure we have used for time-series disaggregation of Malaysian GDP is technically simple but labour intensive. The main feature of the procedure is to work out quarterly production indexes and use them as predictor series of the corresponding sectoral output. The interpolated GDP is simply the sum of the interpolated sectoral outputs. As we have seen, this approach, the ‘related-series approach’ in general, can produce far superior estimates of GDP compared to univariate interpolation methods commonly used.

Our procedure entails two advantages. One is that the basic regression (2), where the sector’s output is regressed on the corresponding production index, can be assumed to form a cointegrating regression because the disturbance term represents mainly the (stationary) measurement errors. There may be circumstances, however, where working with sectors is not possible due to a lack of quarterly data. Nevertheless it is still possible to use the related-series approach by regressing GDP on a series such as exports which is recorded monthly or quarterly by many developing countries. The cost of this approach is likely to be a non-cointegrating regression. One solution to this problem is to follow the approach advocated by Fernandez (1981) and extended by Litterman (1983). The latter proposed to assume that \( u_t \) in equation (2) is a unit root process and the differenced series is an AR(1) process \( (\Delta u_t = \rho \Delta u_{t-1} + e_t, |\rho| < 1) \). He then proposed a procedure to estimate \( \rho \) based on Chow and Lin (1971).

The second advantage of our approach is that it takes care of seasonality reasonably well. A regression of GDP on exports, for example, apart from lacking in cointegration, will force the interpolated GDP to contain the same seasonal as that of exports. Since this is rather unrealistic, the disturbance term in equation (2) will not be free from seasonal variation and the assumptions of white noise or AR(1) errors are unlikely to hold. In our approach, given that the disturbance term represents mainly the measurement errors, if a seasonal autocorrelation is present in the residuals this is likely to be small in magnitude and may be ignored in practice. An alternative solution to this problem is to use seasonally adjusted data.

Table II. A comparison of sectoral estimates with published figures: RMSE% over 1987Q1–1993Q4 mean output (spline interpolation, seasonally adjusted)

<table>
<thead>
<tr>
<th>Agriculture</th>
<th>Mining</th>
<th>Manufacturing</th>
<th>Utilities</th>
<th>Construction</th>
<th>Commerce</th>
<th>Transport &amp; Communication</th>
<th>Public administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share-based</td>
<td>1.60</td>
<td>2.67</td>
<td>1.21</td>
<td>1.37</td>
<td>2.29</td>
<td>1.43</td>
<td>1.70</td>
</tr>
<tr>
<td>Direct</td>
<td>2.18</td>
<td>3.01</td>
<td>1.99</td>
<td>1.39</td>
<td>2.53</td>
<td>1.90</td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.77</td>
</tr>
</tbody>
</table>
As for future work it is worth exploring the structural time-series approach utilized here because of its simplicity in implementing without having to worry about non-stationarity and seasonality of disturbances. A major drawback of the method is that it may distort the trend and seasonal components when extrapolations are carried out too far away from the estimation period. Without an extensive Monte Carlo and empirical evaluation it is difficult to judge the merits of this method at this stage.

The two data sets we generated (seasonally unadjusted series by three broad sectors, and seasonally adjusted series by ten sectors) can be obtained from us as an Excel file through e-mail by writing to ecstabey@leonis.nus.edu.sg (first author) or leechris@merlion.iseas.edu.sg (second author).

APPENDIX: DATA SOURCES

Annual data on GDP by sectors can be obtained from a number of sources including *Malaysia Yearbook of Statistics*. For the years 1973–77 the data series was adjusted to the 1978 base using sectoral shares.

Quarterly GDP by sectors over 1987Q1–1993Q4 were made available to us by the Bank Negara Malaysia. The Malaysian Department of Statistics supplied us with unpublished quarterly GDP figures since 1980Q1.

Quarterly industrial production index (1988 = 100) was obtained from the IMF, International Financial Statistics (CM-ROM). This series shows a discontinuity, easily observed in growth rates, in 1984 due to a rebasing exercise.

Quarterly growth rates of agriculture and forestry are from the *Quarterly Bulletin of Bank Negara Malaysia*. The bank has published growth rates based on an index constructed by it since 1970. We were unable to obtain this index. A consistent series of these growth rates is also not available. Since 1975 the bank has not published the fourth quarter (annual) growth rate. Instead, it has published the first quarter growth against the preceding quarter. These figures are, however, based on provisional estimates. As we had no other choice we used these growth rates and 1987 agriculture and forestry output levels to work out the output series backward.

Total commercial bank loans and total import and export trade figures, which we used to construct a service output index, were obtained from the *Quarterly Bulletin of Bank Negara Malaysia*.

ACKNOWLEDGEMENTS

We would like to thank the Department of Statistics, Malaysia, and the Bank Negara Malaysia for supplying GDP data series. We also like to thank Garry Phillips, Robert Shumway and an anonymous referee for their valuable comments and a large number of participants at the Fourth Malaysian Econometric Conference (October, 1996) for their useful suggestions.

REFERENCES


Author’s biographies:

Tilak Abeysinghe is a senior lecturer in Econometrics in the Department of Economics and Statistics of the National University of Singapore. He holds an MSc in Statistics from the University of Colombo, Sri Lanka, an MA in Economics from Thammasat University, Thailand and a PhD in Economics from the University of Manitoba, Canada. His research interests include seasonality in regression, dynamic modelling with variables observed at different frequencies and macroeconometric modelling for the Singapore economy. He is a joint editor of the *Singapore Economic Review*.

Christopher Lee is a researcher at the Institute of South East Asian Studies. He obtained his BA (Hons) degree from the University of New South Wales, Australia. He is currently pursuing a Masters degree at the National University of Singapore. His research interests include macroeconomic modelling and total factor productivity measurements.

Authors’ addresses:

Tilak Abeysinghe, Department of Economics and Statistics, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, E-mail: ecestabay@leonis.nus.edu.sg or tilakabey@nus.edu.sg

Christopher Lee, Institute of South East Asian Studies, Heng Mui Keng Terrace, Pasir Panjary, Singapore 119596. E-mail: leechris@merlion.iseas.edu.sg