19

Model Selection by Automated X-11 ARIMA Option and Related Problems: A Case Study of Thailand

Yosuke Noda

Introduction

In the study of business cycles, we need to process a great deal of data to remove trend and seasonal factors from the original time series. The X-11 softwear developed by the National Bureau of Economic Research (NBER) in the United States is a very well-known and commonly used program package available for the mainframe computer. However, the X-11 program can cause large distortions on both sides of the seasonal adjusted series because there is a lack of data. In order to avoid such distortions, the X-11 ARIMA softwear package was developed by Dogam of Time Series Research and Analysis Division Statistics Canada. The basic function of the X-11 ARIMA program performs three basic jobs: (1) models the original series, using the general multiplicative seasonal autoregressive integrated moving average (ARIMA) model by Box and Jenkins; (2) extrapolates one year at each end of the series using an ARIMA model, if any, that fits the series well; and (3) seasonally adjusts the extrapolated series, using various moving average processes in the X-11 program.

The ARIMA part of the X-11 ARIMA program plays a very important role in the extention of the original series with the extrapolated series and the estimation of seasonal factors. It is said that the reliability of making seasonal adjusted series obtained from the extended series by using the ARIMA model is greater than that of the original series, and the magnitude of the distortions at each end of the seasonally adjusted series is significantly reduced.

As we sometimes spend a lot of time trying to identify the ARIMA model, a function of selection to the ARIMA model is installed in the X-11 ARIMA so as not to waste time. This function is an automated option in which several types of ARIMA models are assumed. The automated option checks different criteria of the guideline to determine the model. A series we want to forecast does not always pass the criteria of the guideline, therefore an optimal ARIMA model can not necessarily be selected for the series by the automated option. It is recommended that a series be strongly affected by external factors and have extreme variations in it. The parts of the variations must be modified or removed from the series before using the automated option or identifying the ARIMA model by ourselves.

This paper covers three topics. The first is a brief explanation of a basic idea for an automated ARIMA option incoporated in the X-11 ARIMA program. The second is a trial application of the automated ARIMA option to monthly economic time series which have been used to compile DIs for Thailand, and to use this to discuss the ratio of success of the automated option. We also compare the models selected by the automated option with the models by AIC criterion. The third topic is the application of a univariate intervention model, also called a structural dummy variable model, to the series which the automated ARIMA model failed to fit. The intervention model is used to measure and evaluate the effects of changes in a time series, that is, to assess the impact of a discrete intervention which causes certain exceptional external events on the behavioral process.

General Multiplicative Seasonal ARIMA Model

The notation ∇ is used to denote the difference operator defined as $\nabla y_t = y_t - y_{t-1}$ for an economic series, and B to denote the backward shift operater $By_t = y_{t-1}$. We also define the Dth order seasonal difference with the length of the seasonal cycle of s periods known as span:

$$\nabla_s{}^D y_t = (1-B^s)^D y_t .$$

Suppose that a combination of the Dth order seasonal difference with span equal to s and the dth order consecutive difference is needed to make the series w_t stationary. Let w_t denote a nonstationary series with span s:

$$w_t = \nabla^d \nabla_s{}^D y_t .$$

Using the stationary series w_t, a general multiplicative ARIMA model is expressed as follows:

$$\phi(B)\Phi(B)w_t = \theta(B)\Theta(B)a_t ,$$

where random shock a_t, known as the white noise series, is assumed to be normally and independently distributed with mean 0 and variance $\sigma_a{}^2$ and independent of a_{t-1}. The expression $\phi(B)$ and $\theta(B)$ are polynomials of finite degree in B as a consecutive autoregressive and a moving average lag structures.

$$\phi(B) = 1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\,,$$
$$\theta(B) = 1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\,.$$

The expression $\Phi(B)$ and $\Theta(B)$ are polynomials of finite degree in B as a seasonal autoregressive and a moving average lag structures.

$$\Phi(B) = 1-\Phi_1 B^s-\Phi_2 B^{2s}-\cdots-\Phi_P B^{Ps}\,,$$
$$\Theta(B) = 1-\Theta_1 B^s-\Theta_2 B^{2s}-\cdots-\Theta_Q B^{Qs}\,.$$

In general, the following notations are used to express the seasonal ARIMA model: p=order of the consecutive autoregressive process; d= number of consecutive differences; q=order of the consecutive moving average process; P=order of the seasonal autoregressive process; D=number of seasonal differences; Q=order of the seasonal moving average process; and s=the span of the seasonality. We sometimes use s=12 for monthly time series. Alternatively, the general multiplicative seasonal ARIMA model can be summarized as

$$\text{ARIMA}(p, d, q)(P, D, Q)_s\,,$$

or simply $(p, d, q)(P, D, Q)_s \cdots$.

Basic Concept of Extrapolation by an Automated X-11 ARIMA

An automated option of the X-11 ARIMA program chooses an optimal ARIMA model from the five ARIMA models as follows:

1. Multiplicative model
   log $(0, 1, 1)(0, 1, 1)_{12}$, log $(0, 1, 2)(0, 1, 1)_{12}$, log $(2, 1, 1)(0, 1, 1)_{12}$, log $(0, 2, 2)(0, 1, 1)_{12}$, $(2, 1, 2)(0, 1, 1)_{12}$.
2. Additive model
   $(0, 1, 1)(0, 1, 1)_{12}$, $(0, 1, 2)(0, 1, 1)_{12}$, $(2, 1, 1)(0, 1, 1)_{12}$, $(0, 2, 2)(0, 1, 1)_{12}$, $(2, 1, 2)(0, 1, 1)_{12}$.
3. Log additive model
   log $(0, 1, 1)(0, 1, 1)_{12}$, log $(0, 1, 2)(0, 1, 1)_{12}$, log $(2, 1, 1)(0, 1, 1)_{12}$, log $(0, 2, 2)(0, 1, 1)_{12}$, log $(2, 1, 1)(0, 1, 1)_{12}$.

The notation log denotes log-transformed series as input series to the ARIMA model. After choosing the best model, the X-11 ARIMA program automatically adds one year of extrapolated series to the original series and makes extended series.

The model automatically selected is the first model to pass all of the following guideline criteria.

1. The mean absolute percentage error. Let y_t denote an original series and $\hat{y}_t$ a forecast one estimated by the automated option of X-11 ARIMA. The mean absolute percentage error of the forecast for one, two, and three years is represented as follows:

$$M = n^{-1}\sum_{t=1}^{n} | y_t - \hat{y}_t | / y_t .$$

   The percent error is less than or equal to 15 per cent where $n=$ 12, 24, and 36.

2. The Ljung and Box Q test. The Ljung and Box Q test is used to determine whether an ARIMA model fits the original series—in other words, whether residuals are white noise series. The check of the autocorrelation function of the residuals relies on the Ljung and Box Q statistics defined as

$$Q(K) = n(n+2)\sum_{j=1}^{K} \frac{1}{n-j} r_j^2(\hat{a}),$$

   where K is the number of terms in the summation. $Q(K)$ is very sensitive to the number K, so that $K=12$ and $K=24$ have proven to be useful for monthly economic time series, which tend to have twelve-month seasonal cycles. A sample autocorrelation of a_t at lag j, $r_j(a)$ can be calculated by the formula

$$r_j = c_j / c_0 .$$

   A sample autocovariance c_j is defined as

$$c_j = n^{-1}\sum_{t=1}^{n-j} \hat{a}_t \hat{a}_{t+j}, \qquad j \geqq 0,$$

   where a_t is the residual of the ARIMA model defined as

$$\hat{a}_t = [\hat{\theta}(B)\hat{\Theta}(B)]^{-1}\hat{\phi}(B)\hat{\Phi}(B)\nabla^d \nabla_s{}^D y_t .$$

   The Q test is also known as an application of the Portmanteau test on the residuals. If the residuals are white noise. Q is distributed approximately as χ^2 distribution, with the degrees of freedom $K(p+q+P+Q)$. The notations p, P, q, and Q are the numbers of parameters of the seasonal ARIMA model as mentioned above. A significant level for the Q test is 5 per cent. The X-11 ARIMA program does not mention the number K in the manual or in the printed output.

3. No evidence of over-difference. It is assumed that when the sum of the consecutive MA parameters or the seasonal MA parameter is greater than 0.9, an over-difference occurs. Suppose that $\theta(B)$ is over-differenced.

$$\begin{aligned}\theta(B) &= 1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\\ &= (1-B)(1-\phi_1 B-\cdots-\phi_{q-1}B^{q-1})\\ &= 1-(1-\phi_1)B-(\phi_2-\phi_1)B^2-\cdots-(\phi_{q-1}-\phi_{q-2})B^{q-1}\\ &\quad -(0-\phi_{q-1})B^q\,.\end{aligned}$$

The sum of the MA parameters is

$$\sum_{j=1}^{q}\theta_j = (1-\phi_1)-(\phi_2-\phi_1)-\cdots-(\phi_{q-1}-\phi_{q-2})-(0-\phi_{q-1}) = 1\,.$$

It is said that a model generated by a over-differenced series will be more complicated than a model generated by a stationary series obtained with the minimum amount of difference.

If the first model fulfills the guideline criteria, then the X-11 ARIMA does not try the others. If the first model fails, then the X-11 ARIMA tries the second model, and so on. When none of the five models is found acceptable, a message which indicates that extrapolated values have not been incoporated into the unadjusted series is given, and the best model should be resubmitted using the option corresponding to the user's ARIMA model identification.

Selection by the Automated ARIMA Option

We applied the automated ARIMA option to monthly economic time series which have been used as the components of the DIs for Thailand in our SEPIA project. The twenty-one economic time series and periods of series used are shown in Table 19–1. As mentioned above when a model fulfills the criteria of the guideline for the series, then X-11 ARIMA give us the information about the model.

The results of the model selection by the automated option for the series are shown in Table 19–2. For the seven series CPINON, DSALE, LCBDOM, M1, NUTOUR, PETROL, and STOCKP, all of the guideline criteria fulfilled were by the automated ARIMA option with a type of multiplicative model assigned as a default option. The ratio of success of the

Table 19–1
Economic Time Series

Variable	Description	Start year	End year	Size
BEER	Production of beer	Jan. 70	May 89	233
COMVIECL	Production of commercial vehicles	Jan. 70	May 89	233
CONAR	Construction area permited in Metropolitan Bangkok	Jan. 70	May 89	233
CPINON	CPI-nonfood	Jan. 79	June 89	126
DSALE	Value of sales for department stores in Bangkok	Jan. 70	May 89	233
ELECTRIC	Electricity used by large users of Metropolitan Electricity Agency	Jan. 70	May 89	233
EXPINDEX	Export index (unit value)	Jan. 70	March89	231
GOVREVEN	Total government actual revenue	Jan. 70	May 89	233
GYPSUM	Production of gypsum	Jan. 70	April 89	232
IMP	Import of capital goods	Jan. 70	April 89	232
IMPINDEX	Import index (unit value)	Jan. 70	May 89	233
LCBDOM	Domestic bills of commercial banks	Jan. 70	May 89	233
LIGNIT	Production of lignite	Jan. 70	April 89	232
MOTCYCLE	Production of motorcycles	Jan. 72	May 89	209
M_1	Money supply	Jan. 70	July 89	235
NOCHEQUE	Number of checks	Jan. 71	June 89	222
NUTOUR	Number of tourists	Jan. 76	June 89	162
PETROL	Production of petroleum products	Jan. 70	May 89	233
STOCKP	Stock price index	Jan. 76	July 89	163
TINPLATE	Production of tinplate	May 79	May 89	121
TOTCLEAR	Total of clearing	Jan. 71	June 89	222

automated option occupies 33 per cent. The six series CPINON, DSALE, LCBDOM, M1, NUTOUR, and PETROL, corresponding to approximately a 28 per cent success ratio, are modeled by Log $(0, 1, 1)(0, 1, 1)_{12}$, and the one series STOCKP, corresponding to approximately a 5 per cent success ratio, by Log $(0, 1, 2)(0, 1, 1)_{12}$.

When specifying an additive model, the success ratio is reduced and only three series, CPINON, DSALE, and PETROL, occupying around 14 per cent, fulfilled the criteria because LCBDOM, M1, NUTOUR, and STOCKP, which passed in the multiplicative model, failed in the additive model type. All of the three series are modeled by $(0, 1, 1)(0, 1, 1)_{12}$ without log transformation. Suppose that a seasonal model is $(0, 1, 1)(0, 1, 1)_{12}$ and the variance of white noise a_t is σ^2.

$$\nabla\nabla_{12} w_t = (1-\theta B)(1-\Theta B^{12})a_t.$$

The autocovariance function and autocorrelation function of $\nabla\nabla_{12} w_t$ at lag j, γ_j and ρ_j are shown as follows:

Table 19–2
Models Selected by the Automated Option

Variable	M-type	A-type	M-type (AIC)	A-type (AIC)
BEER	×	×		
COMVIECL	—	—		
CONAR	×	×		
CPINON	Log $(0,1,1)\ (0,1,1)_{12}$	$(0,1,1)\ (0,1,1)_{12}$	Log $(0,1,2)\ (0,1,1)_{12}$	$(0,1,2)\ (0,1,1)_{12}$
DSALE	Log $(0,1,1)\ (0,1,1)_{12}$	$(0,1,1)\ (0,1,1)_{12}$	Log $(2,1,1)\ (0,1,1)_{12}$	$(2,1,1)\ (0,1,1)_{12}$
ELECTRIC	—	—		
EXPINDEX	—	—		
GOVREVEN	—	—		
GYPSUM	×	×		
IMP	×	×		
IMPINDEX	—	—		
LCBDOM	Log $(0,1,1)\ (0,1,1)_{12}$	×	Log $(0,1,1)\ (0,1,1)_{12}$	
LIGNIT	×	×		
MOTCYCLE	—	—		
M_1	Log $(0,1,1)\ (0,1,1)_{12}$	×	Log $(0,1,1)\ (0,1,1)_{12}$	
NOCHEQUE	—	—		
NUTOUR	Log $(0,1,1)\ (0,1,1)_{12}$	×	Log $(0,1,1)\ (0,1,1)_{12}$	
PETROL	Log $(0,1,1)\ (0,1,1)_{12}$	$(0,1,1)\ (0,1,1)_{12}$	Log $(0,1,1)\ (0,1,1)_{12}$	$(0,1,1)\ (0,1,1)_{12}$
STOCKP	Log $(0,1,2)\ (0,1,1)_{12}$	×	Log $(2,1,1)\ (0,1,1)_{12}$	
TINPLATE	—	—		
TOTCLEAR	—	—		

Note: ×=All models failed. —=no data available. M-type means multiplicative model type, and A-type means additive model type. AIC=Akaike Information Criterion.

$$
\begin{array}{ll}
\gamma_0 = (1+\theta^2)(1+\Theta^2)\sigma^2 & \rho_0 = 1\,, \\
\gamma_1 = -\theta(1+\Theta^2)\sigma^2 & \rho_1 = -\theta/(1+\theta^2)\,, \\
\gamma_2 = \cdots = \gamma_{10} = 0 & \rho_2 = \cdots = \gamma_{10} = 0\,, \\
\gamma_{11} = \theta\sigma^2 & \rho_{11} = \theta/(1+\theta^2)(1+\Theta^2)\,, \\
\gamma_{12} = -\Theta(1+\theta^2)\sigma^2 & \rho_{12} = -\Theta/(1+\Theta^2)\,, \\
\gamma_{13} = \theta\sigma^2 & \sigma_{13} = \theta/(1+\theta^2)(1+\Theta^2)\,. \\
\cdots & \cdots
\end{array}
$$

According to the results of the autocorrelation function for the multiplicative model, M1 shows the typical type of $(0, 1, 1)(0, 1, 1)_{12}$ model which has significant spikes at lag 0, 1, 11, 12, and 13. Ignoring the spike at lag 0, DSALE and PETROL have spikes at 1, 12, and 13, NUTOUR has two spikes at lags 1 and 12, and CPINON, LCBDOM, and STOCKP have only one spike at lag 12.

One of the disadvantages of the Box-Jenkins approach is that it requires a lot of time and skill at the identification stage. An alternative approach is to use AIC (Akaike Information Criterion), which is now more comprehensively accepted as one of the most reliable methods for order determination.

$$\mathrm{AIC} = -2 \log (\text{maximum likelihood}) + 2\,(\text{number of free parameters})\,.$$

When n is the number of effective observations to which the ARMA(p, q) model is fitted, $n=N-P$, N is the number of observations, and $\hat{\sigma}^2$ is the maximum likelihood estimator of the residual variance, the maximum log likelihood for the model is

$$\hat{L} = -n/2 \log \hat{\sigma}^2 - n/2\,.$$

Ignoring the second term because $-n/2$ is constant, the AIC is defined by the formula for the ARMA(p, q) model.

$$\mathrm{AIC} = n \log \hat{\sigma}^2 + 2(p+q)\,.$$

The model which has a minimum AIC statistic calculated is assumed to be the best model.

We show the AIC value for five models of the multiplicative model in Table 19–3. The minimum AIC value occurs with $(0, 1, 1)(0, 1, 1)_{12}$ for LCBDOM, M1, NUTOUR, and PETROL, which were selected also by the automated option of X-11 ARIMA. However, the minimum AIC value occurs with $(0, 1, 2)(0, 1, 1)_{12}$ for CPINON against $(0, 1, 1)(0, 1, 1)_{12}$ by the automated option. The model $(2, 1, 1)(0, 1, 1)_{12}$ has the minimum AIC value for DSALE against $(0, 1, 1)(0, 1, 1)_{12}$. The model $(2, 1, 1)(0, 1, 1)_{12}$ has the minimum AIC value for STOCKP against $(0, 1, 2)(0, 1, 1)_{12}$.

Table 19–3
AIC of Multiplicative Model Type

Variable	Log (0,1,1) (0,1,1)$_{12}$	Log (0,1,2) (0,1,1)$_{12}$	Log (2,1,1) (0,1,1)$_{12}$	Log (0,2,2) (0,1,1)$_{12}$
CPINON	−513.39	−517.41	*	−494.26
DSALE	−476.78	−477.43	−480.12	−425.84
LCBDOM	−711.18	−709.39	−708.03	−694.84
M_1	−888.32	−887.99	−887.11	−870.16
NUTOUR	−260.23	−258.24	−256.24	−221.61
PETROL	−188.84	−187.01	−186.48	−150.65
STOCKP	−313.29	−317.30	−321.25	−309.72

Note: Symbol * means that this model does not converge.

Table 19–4
AIC of Additive Model Type

Variable	(0,1,1) (0,1,1)$_{12}$	(0,1,2) (0,1,1)$_{12}$	(2,1,1) (0,1,1)$_{12}$	(0,2,2) (0,1,1)$_{12}$	(2,1,2) (0,1,1)$_{12}$
CPINON	264.50	260.33	263.82	262.81	273.94
DSALE	1333.12	1333.10	1330.13	1336.03	1368.73
PETROL	4735.62	4735.82	4737.55	4737.59	*

For the additive model type in Table 19–4, the minimum AIC value occurs with $(0, 1, 1)(0, 1, 1)_{12}$ for PETROL, and $(0, 1, 2)(0, 1, 1)_{12}$ and $(2, 1, 1)(0, 1, 1)_{12}$ for CPINON and DSALE, respectively.

For the series that the built-in five models using the automated option failed, we have to resubmit to get an appropriate model by using the option corresponding to the user's ARIMA model identification or other techniques.

Univariate Intervention Model

Exogenous time series ξ_{jt} $j=1\cdots r$ are dummy variables or indicator variables with 1 and 0 denoting the presence or nonpresence of unique events (also called interventions) at time t. The indicator variable is denoted by S_t, with T referring to the time period at which the event starts.

$$\xi_{jt} = S_t^{(T)} = \begin{bmatrix} 0 & t<T \\ 1 & t\geq T. \end{bmatrix}$$

An univariate intervention model belonging to a certain class of a transfer function model is represented as the combination of intervention and noise N_t which may be modeled by the seasonal ARIMA process.

$$y_t = \sum_{j=1}^{r} \frac{\omega_j(B)}{\delta_j(B)} \xi_{jt} + N_t ,$$

$$\nabla^d \nabla_s^D \phi(B)\Phi(B)N_t = \theta(B)\Theta(B)a_t ,$$

where the notations given in the seasonal ARIMA model are explained above, r is the number of interventions, and $\omega_j(B)$ and $\delta_j(B)$ are polynomials of finite degree in B.

$$\omega(B) = \omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_r B^r ,$$

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_s B^s .$$

The autocorrelation function and partial autocorrelation function of y_t do not necessarily show the ARIMA process, because the effects of the intervention distort the pattern of the autocorrelation function and partial autocorrelation function. Therefore we need the following three steps to determine the intervention model.

1. If the data before or after the intervention is sufficiently long, the ARIMA model of noise N_t is tentatively identified from the data. After the identification, a tentative intervention model is checked by means of using the data.
2. From the possible effects of the intervention, we specify the form of $\omega(B)$ and $\delta(B)$. The residuals can be calculated as mentioned before.

$$\hat{N}_t = y_t - \sum_{j=1}^{r} \hat{\delta}_j(B)^{-1} \hat{\omega}_j(B) \xi_{jt} .$$

   The ARIMA model of N_t is identified by using the pattern of autocorrelation and partial autocorrelation.
3. Using all the data, all of the parameters in the model are jointly estimated in this step.

Application of the Intervention Model

We applied the intervention model to the variable BEER which failed to pass the X-11 ARIMA criteria guideline. Figure 19–1 shows an original series and a log-transformed series for the variable BEER made from January 1970 to May 1989, a total of 233 monthly observations. Because there is not only a remarkable upward trend in the original series of BEER, but there also appears to be an increase in the variation of the data along the trend, (in other words, the amplitude increases as time passes), the use of the logarithmic transformation seems appropriate in order to get a constant amplitude for the series. To obtain a more exact transformation, the Box and Cox transformation might also have been needed.

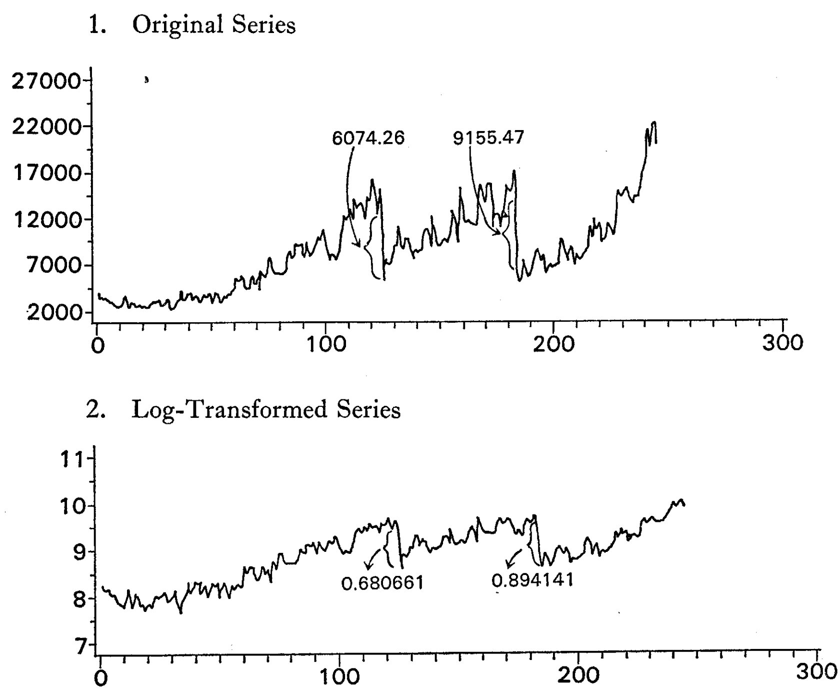

Fig. 19–1
Time Plot of BEER

It seems that there are two sudden drops in Figure 19–1 caused by the twin effects of a rise in sales tax inducing a reduction in the production of beer. The sales tax on beer in Thailand rose by 40 per cent in May 1980 and again by 28 per cent in March 1985. Therefore we specify two interventions $\xi_{1t}=S_t^{(T1)}$ and $\xi_{2t}=S_t^{(T2)}$ with T_1=June 1980 and T_2=April 1985.

$$\xi_{1t} = S_t^{(T1)} = \begin{bmatrix} 0 & t: & \text{from January 1970 to May 1980} \\ 1 & t: & \text{from June 1980 to May 1989,} \end{bmatrix}$$

$$\xi_{2t} = S_t^{(T2)} = \begin{bmatrix} 0 & t: & \text{from January 1970 to March 1985} \\ 1 & t: & \text{from April 1985 to May 1989.} \end{bmatrix}$$

Since we have comparatively long periods of data before the first intervention, the first 124 observations from January 1970 to May 1980 are used to identify a tentative ARIMA model for N_t. To identify the ARIMA model from the autocorrelation function and partial autocorrelation function, using a difference is recommended in order to make the series N_t

stationary. At the same time, since the series has a seasonality of order 12, a seasonal difference ∇_{12} may be needed.

From the result of the crosscorrelation function between differenced and interventions, seen in Figure 19–2, only one spike appears significantly at lag 0 for the series $\nabla\nabla_{12}y_t$. We then determine that the lag structure of the interventions are $\omega_j(B)=\omega_{j1}$ and $\delta_j(B)=1$, $j=1, 2$. Therefore the univariate intervention model is expressed as follows:

$$
\begin{aligned}
y_t &= \underset{(-8.64)}{-0.680661}\xi_{1t} \underset{(-11.48)}{-0.894141}\xi_{2t}+N_t,\\
\nabla\nabla_{12}(1\underset{(2.06)}{-0.139301}B^3\underset{(-3.19)}{+0.212645}B^7)N_t &= (1\underset{(8.98)}{-0.533624}B)(1\underset{(15.53)}{-0.828525}B^{12})a_t,
\end{aligned}
$$

where the values in the parentheses are t-value. The standard error, AIC criterion, and number of residuals of the model are 0.0985518, −347.049, and 208, respectively. The Ljung and Box Q test with $K=12$ and degrees of freedom 8 is 2.53, and the p-value 0.961. Also the $Q(24)$ with degrees of freedom 20 is 22.14 and the p-value 0.333. The mean absolute average percentage errors of the extrapolated data for the past one, two, and three years and their estimated parameters of the transfer function models are shown in Tables 19–5 and 19–6 respectively. The percentages error, 6.3 per cent and 5.4 per cent, are less than 15 per cent, which is the maximum percentage point of the criteria. As the percentage error for the past three years is almost equal to 15 per cent, we accept that the transfer function model passes the percentage error. Therefore the transfer function model, instead of the ARIMA model, can be used to extrapolate one year at each end of the series. Because we applied the model to log-transformed series, it was difficult to interpret the meanings of the model. Therefore instead of using log-transformed series, we tried the original series and got the results of the intervention model, expressed as follows.

$$
\begin{aligned}
y_t &= \underset{(-9.10)}{-6047.26}\xi_{1t} \underset{(-14.29)}{-9155.47}\xi_{2t}+N_t,\\
\nabla\nabla_{12}(1\underset{(2.57)}{-0.175409}B^3\underset{(-2.48)}{+0.160959}B^7)N_t &= (1\underset{(9.66)}{-0.563659}B)(1\underset{(13.77)}{-0.731230}B^{12})a_t.
\end{aligned}
$$

The standard error, AIC criterion, and number of residuals of the model are 839.48, 3608.56, and 220, respectively. From the Ljung and Box Q test we can recognize that the residuals are white noise. Thus the model still works for the original series.

Since the unit of the production of beer (BEER) is 1,000 liters, the interpretation of this intervention model is that the monthly average production of beer dropped by about 6047.26 liters after the first intervention of the

1. $\nabla\nabla_{12}y_t$, $\nabla\nabla\xi_{1t}$

LAG	CORRELATION	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
-15	0.06893	|*
-14	-0.06422	*|
-13	0.18334	|****
-12	0.18405	|****
-11	-0.07814	**|
-10	0.03241	|*
-9	-0.05109	*|
-8	-0.00353	|
-7	-0.03872	*|
-6	-0.12665	***|
-5	0.09865	|**
-4	0.00383	|
-3	-0.14961	***|
-2	0.12478	|**
-1	-0.13859	***|
0	-0.36044	*******|
1	0.13764	|***
2	-0.05524	*|
3	0.05865	|*
4	0.04844	|*
5	-0.01801	|
6	0.07812	|**
7	-0.10615	**|
8	0.00348	|
9	0.04349	|*
10	-0.01138	|
11	0.02894	|*
12	0.14807	|***
13	-0.03762	*|
14	-0.00025	|
15	0.00486	|

2. $\nabla\nabla_{12}y_t$, $\nabla\nabla_{12}\xi_{2t}$

LAG	CORRELATION	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
-15	0.00371	|
-14	0.03100	|*
-13	-0.09564	**|
-12	0.26423	|*****
-11	0.08016	|**
-10	-0.14077	***|
-9	-0.03414	*|
-8	0.03336	|*
-7	-0.03594	*|
-6	0.07657	|**
-5	0.00118	|
-4	-0.05356	*|
-3	-0.03193	*|
-2	0.04432	|*
-1	0.05852	|*
0	-0.46158	*********|
1	-0.09255	**|
2	0.07502	|**
3	0.11125	|**
4	-0.04880	*|
5	-0.04709	*|
6	0.01143	|
7	0.01252	|
8	-0.02093	|
9	0.05449	|*
10	0.00504	|
11	-0.11735	**|
12	0.25699	|*****
13	0.00813	|
14	-0.01150	|
15	-0.04611	*|

Fig. 19-2
Cross Correlation

Table 19-5
Mean Absolute Average Percentage Error of the Transfer Function Model for BEER

	Past 1 Year	Past 2 Years	Past 3 Years
Percentage error	6.3	5.4	15.3

Table 19-6
Estimated Parameters of the Transfer Function Models

Parameter	Past 1 Year	Past 2 Years	Past 3 Years
θ_1	0.523899 (0.059966)	0.533631 (0.059375)	0.524955 (0.061873)
Θ_1	0.898443 (0.081789)	0.828386 (0.061270)	0.924601 (0.118923)
ϕ_3	0.138157 (0.067941)	0.139286 (0.067416)	0.128439 (0.070500)
ϕ_7	−0.121990 (0.067635)	−0.212645 (0.068081)	−0.225470 (0.069924)
ω_{11}	−0.683052 (0.077693)	−0.680653 (0.078307)	−0.687316 (0.078649)
ω_{21}	−0.896594 (0.076907)	−0.894146 (0.075875)	−0.902747 (0.077858)

Note: Figures in parentheses are standard deviations.
Parameters of the transfer function model:

$$\begin{cases} y_t = \omega_{11}\xi_{1t} + \omega_{21}\xi_{2t} + N_t \\ (1-\phi_3 B^3 - \phi_7 B^7)\nabla\nabla_{12} N_t = (1-\theta_1 B)(1-\Theta_1 B^{12})a_t . \end{cases}$$

40 per cent rise in sales tax in March 1984 and by about 9155.47 liters after the second intervention of the 28 per cent rise in sales tax in April 1985.

Conclusion

We introduced the basic idea of the automated ARIMA option and the intervention model. Using economic time series for Thailand, the ratio of success of the automated ARIMA option with a multiplicative model occupies 33 per cent, while with an additive model it was around 14 per cent. When specifying a multiplicative model, the six series were modeled by Log $(0, 0, 1)(0, 0, 1)_{12}$ and one series by Log $(0, 1, 2)(0, 1, 1)_{12}$. When specifying an additive model, only three series are modeled by $(0, 1, 1)(0, 1, 1)_{12}$ without log transformation. However, when we use the AIC to determine the best model from the five built-in models for each series, the different model was selected in comparison with the model selected by the automated X-11 ARIMA option. Only four series were selected equally by both the AIC and the automated option for the multiplicative model, and only one series for the additive model.

Finally, for the series which failed to pass the automated option with the five built-in models, we had to resubmit to get an appropriate model using the option corresponding to the user's ARIMA model identification or other techniques. One technique was to apply the intervention model to a series given by the effects of changes and extend the unadjusted series with one year of extrapolated data, even though this is very difficult to fit to an ordinary ARIMA model.

Bibliography

Box, G.E.P., and Tiao, G. C. 1975. "Intervention Analysis with Applications to Economic and Environmental Problems." *Journal of the American Statistical Association* 70.

Dagum, Estela Bee. 1988. *The X11ARIMA/88 Seasonal Adjustment Method: Foundation and User's Manual.* Ottawa: Time Series Research and Analysis Division Statistics Canada.

Lassarre, S., and Tan, S. H. 1982. "Evaluation of Safety Measures on the Frequency and Gravity of Traffic Accidents in France by Means of Intervention Analysis." In *Time Series Analysis: Theory and Practices 1.* New York: North Holland Publishing Co.

SAS/ETS User's Guide, Version 5 edition. 1984. Cary, N. C.: SAS Institute.

Vandaele, Walter. 1983. *Applied Time Series and Box-Jenkins Model.* New York: Academic Press.