A Cross-Section Method for Modelling Seasonal Time Series

Tilin Yi and Charlie Presland

School of Information Technology, Charles Sturt University, Bathurst NSW 2795, Australia

Abstract A new method of modelling seasonal time series , a cross-section method, is proposed. Compared with conventional methods of analysing seasonal time series, the cross-section method is effective and easy to understand. Basically, seasonality is removed by cross-sectioning the time series, ie. by modelling a cross-section of data at each season of the time series to obtain a distinct fitted model for each season of the year. The method is illustrated on a simulated time series and on two practical time series, one of which is well-known in the literature.

Key Words: seasonal time series, seasonality, ARIMA model, Box-Jenkins, cross-section modelling, forecasting

1 Introduction

Many modelling methods are available for the analysis of seasonal time series. For example, decomposition isolates estimates of the seasonal, trend and cyclical components of the time series, projects these components into the future, and synthesises the projections into forecasts of future values of the time series in accordance with some assumed model (eg., additive, multiplicative, mixed). Smoothing methods, such as Winter's three-parameter model, obtain seasonally adjusted forecasts which in the final step are re-seasonalised to yield the forecast. In Box-Jenkins [1976] ARIMA modelling, seasonal differencing may be employed to obtain a transformed time series which is stationary in its seasonal component. Then analysis of the correlations at the seasonal lags enables the appropriate seasonal component of the model to be identified and fitted in the usual way.

Many time series met with in practice (especially economic time series) contain seasonal effects including increasing seasonal variation. Bowerman et al [1990] discuss the approaches to modelling and forecasting these kinds of series. They describe four type of models:

Data Transformation: This is the traditional approach used in ARIMA modelling, and includes seasonal differencing and the usual logarithmic or power transformations.

Double Seasonal Difference Model: For an ARIMA model requiring second order seasonal differences at the identification stage, two terms are required in the fitted equation to obtain the forecast of season j that is r years beyond time origin t.

Seasonal Intervention Model: This model is a modification of multiple regression which uses seasonal differences and dummy variables as shown in the equation below to model the time series as partly deterministic and partly stochastic.

$$(1-B^s)Y_t = \beta_0 + \sum_j \beta_j d_{jt} + e_t$$

where, B is backward operator, β_0, β_j are unknown parameters, s is length of seasonality , d_{jt} is dummy variables which is 1 when observation Y_t is in season j and zero otherwise for s-1 out of s seasons. The stochastic components { e_t } follow an ARIMA model of the form

$$(1-B)^d \phi(B)\Theta(B^s)e_t = \phi(B)\Theta(B^s)a_t,$$

where a_t's are independently and identically distributed with mean zero and fixed variance, ie, $a_t \sim iid(0,\sigma^2)$.

Seasonal Interaction Model: In place of the seasonal differencing of the seasonal intervention model, this model includes deterministic and seasonal trend estimates as shown.

$$Y_t = \beta_0 + \beta_0^* t + \sum_j \beta_j d_{jt} + \sum_j \beta_j^* d_{jt} t + e_t$$

where e_t, and d_{jt} as defined before, and $\beta_0, \beta_0^*, \beta_j, \beta_j^*$ are unknown parameters. (also see Bowerman et al [1990])

2 Option Selection Principles

When faced with the choice of selecting which model to use for a particular time series, the following principles are suggested as a guide [Bowerman et al 1990].

- "The use of a data transformation implies that the time series does not exhibit linear increasing seasonal variation in the original metric." The question of whether linear increasing seasonal variation exists in the original series is often difficult to answer.

- In following the usual identification process for ARIMA modelling, significant autocorrelations would be observed slowly dying down through seasonal lags 1s, 2s, 3s, ... if the second seasonal difference is to be justified as part of a model. In fact, for many time series requiring double seasonal differencing, this does not occur. For example, Wichern [1973] examines a non-stationary, non-seasonal time series $(1-B)Y_t = (1-\theta B)a_t$ and shows that as $\theta \rightarrow 1$, the sample autocorrelations approach zero at all lags. For a seasonal model $(1-B^s)Y_t = (1-\theta_s B^s)a_t$, Wichern's discussion could be applied to argue that as $\theta \rightarrow 1$, the sample autocorrelations approach zero at the seasonal lags 1s, 2s, 3s, ... even though the series is non-stationary in its seasonality.

- Seasonal Intervention Model: The seasonal intervention model may be appropriate, as an alternative to the usual ARIMA, when the latter's identification stage fails to identify a double seasonal difference model while the plot of the time series shows increasing seasonal variation.

- When, in fitting a seasonal intervention model, a unit root is found in a seasonal moving average operator in the ARIMA model for the e_t,, a seasonal interaction model could be tried. Compared with other double seasonal differencing related models, interaction model has the most deterministic features.

In general, nonlinear increasing seasonal variation is more appropriately modelled using the transformed data model, ie an ARIMA model of the transformed time series. When the seasonal variation is increasing linearly, the seasonal intervention model could be more appropriate. Sometimes, however, it is uncertain which model is best.

Many time series displaying increasing seasonal variation have a strong deterministic component. This is associated with the failure of correlation function analysis to identify true double seasonal difference models, and in the estimation stage with finding unit roots in moving-average part of the model.

3 Cross-Section Modelling Method

Suppose a seasonal time series with increasing variation is $\{Y_t\}$ t=1,2,...n. and the seasonality length is s. Obviously, for quarterly data, s=4; for monthly data, s=12.

All the models introduced above try to fit a unique model to the given time series with increasing seasonal variation. We note that the seasonal impacts could be analysed more accurately if we concentrate on the analysis of the sub-time series { SY_{it} } , i=1,2,...s consisting of all the i^{th} season data only.

Further, we assume n=K*s + N, 0 < N < s-1. Then we can re-write original time series as follows.

Seasons	1	2	...	N	...	s-1	s
1	Y_1	Y_2	...	Y_N	...	Y_{s-1}	Y_s
2	Y_{s+1}	Y_{s+2}	...	Y_{s+N}		Y_{2s-1}	Y_{2s}

...	...	...	...	...	...	...	...
K	$Y_{(K-1)s+1}$	$Y_{(K-1)s+2}$	...	$Y_{(K-1)s+N}$	...	Y_{Ks-1}	Y_{Ks}
K+1	Y_{Ks+1}	Y_{Ks+2}	...	Y_{Ks+N}			
	SY_1	SY_2		SY_N		SY_{s-1}	SY_s

Suppose, we cut time series cross seasons vertically and get s sub-time series , { SY_{it} }, t=1,2,...,K+1 for i=1,2,...N; t=1,2,...,K for i=N+1, N+2,...,s.

The most important feature of these sub-time series is that each of them consists of data from the same season. It is reasonable to expect that { SY_{it} } i=1,2...S, t=1,2,...,K+1 for i=1,2,...N; t=1,2,...,K for i=N+1, N+2,...,S reflects the most important characteristics for the i^{th} season. For different seasons, different models are to be built to catch up these characteristics.

Our forecasting procedure is:

- Divide original time series Y_t across the different seasons, getting s different sub-time series, { $SY_{1,t}$ }, { $SY_{2,t}$ },...., { $SY_{s,t}$ }
- For s sub-time series { $SY_{i,t}$ }, i=1,2,...s, fit different models, M_1 , M_2 , ...M_s .
- Use the s fitted models to obtain different seasonal forecasts.

4 Examples

4.1 Example One

In order to show the effectiveness of cross-section method, we consider function

$$Y_t = \exp^{\frac{t}{100}} \cos t, t > 0 \qquad (1)$$

Clearly, Y_t is a cyclical function with varying amplitudes and period 2π. For simplicity, we consider only 4 points in each period over 10 periods, like a series of ten years of quarterly data.

Period	$\frac{\pi}{5}$	$\frac{7\pi}{10}$	$\frac{12\pi}{10}$	$\frac{17\pi}{10}$
1	0.8141	-0.6009	-0.8401	0.62
2	0.8669	-0.6398	-0.8946	0.6602
3	0.9231	-0.6813	-0.9526	0.7031
4	0.983	-0.7255	-1.0144	0.7486
5	1.0467	-0.7725	-1.0801	0.7972
6	1.1146	-0.8226	-1.1502	0.8489
7	1.1869	-0.876	-1.2248	0.9039
8	1.2639	-0.9328	-1.3042	0.9626
9	1.3458	-0.9933	-1.3888	1.025
10	1.4331	-1.0577	-1.4788	1.0914

We have four sub-time series { $SY_{i,t}$ },i=1,2,3,4; t=1,2,..,10. Although original time series varies significantly, one common feature of each sub-time series is that it is either monotone increasing or decreasing. In other words, we can use four simple models to fit each sub-time series individually rather than finding a complicated model to fit the entire time series.

In this case, the model $SY_{i,t}$=1.06485$SY_{i,t-1}$,i=1,2,3,4, has been fitted. It perfectly fits all of the sub time series. R^2 =100 %

When we consider fitting a unique model to the entire time series, as in the conventional time series analysis modelling process, it is difficult to envisage using a nonlinear model (1) to fit historical data. But, by separating the different seasons of data and fitting a simple model for each season, we are able to incorporate all the seasonal characteristics of the original time series and obtain a prefect fit and forecast in this deterministic example.

4.2 Example Two

The feasibility of cross-section modelling has been seen in the above example. Due to the specificity of that example, we now analyse a well-known practical time series here to illuminate the effectiveness of cross-section modelling with real data.

4.2.1 Data

The data are monthly totals of international airline passengers (in thousands) from January

1949 to December 1960 [Box and Jenkins 1976]. This time series comprises 144 observations. It is obvious that the series is seasonal with increasing seasonal variation [see Figure 1]. Box and Jenkins [1976] use a seasonal model of the form ARIMA$(0,1,1)(1,1,0)^{12}$

$$(1-B)(1-B^{12})Y_t = (1-\theta B)(1-\Theta B^{12})e_t$$

to fit the logged airline data, where e_t ~ idd $N(0, \sigma^2)$. They obtain $\hat{\theta}=0.4$, $\hat{\Theta}=0.6$, and $\hat{\sigma}^2 = 1.34 * 10^{-3}$. But,as indicated by Chatfield and Prothero [1973], it doesn't follow that the model which fits the transformed data best will also fit the original data the best. We will compare Box-Jenkins model performance with cross-section model later.

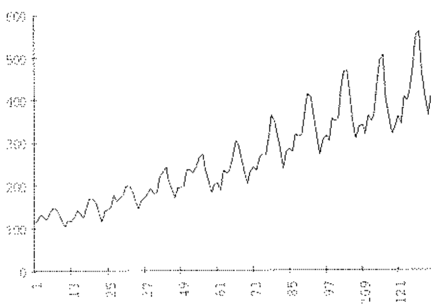

Figure 1: International Airline Passengers (in thousands)

4.2.2 Cross-Section Modelling

Since this series is made up of monthly data, we construct s=12 models to fit individual seasons of data. Twelve sub time series, $\{SY_{1,t}\}$, $\{SY_{2,t}\}$,..... $\{SY_{s,t}\}$ have been obtained for Jan, Feb,...., Dec data respectively. We consider that in the season k at time t, $\{SY_{k,t}\}$is affected by data from:

- the previous two seasons of the previous year, ie $SY_{k-2,t-1}$ and $SY_{k-1,t-1}$
- the same season of the previous year, ie $SY_{k,t-1}$
- the next two seasons of the previous year, ie $SY_{k+1,t-1}$ and $SY_{k+2,t-1}$; and
- the previous two seasons of the current year, $SY_{k-1,t}$, $SY_{k-2,t}$

So the model takes the general form

$$SY_{k,t} = f(SY_{k-2,t-1}, SY_{k-1,t-1}, SY_{k,t-1}, SY_{k+1,t-1}, SY_{k+2,t-1}, SY_{k-1,t}, SY_{k-2,t})$$

We use data from January 1949 to December 1959 to estimate the models and retain the last year's data for evaluation. Using stepwise regression with the above-mentioned factors, we obtain the fitted models:

- January: $SY_{1,t}=-6.167+1.056\ SY_{12,t-1}$, $R^2=99.6\ \%$
- February: $SY_{2,t}=24.207+1.147SY_{1,t}-0.277\ SY_{4,t-1}$, $R^2=99.7\ \%$
- March: $SY_{3,t}=20.024+1.048\ SY_{1,t}$, $R=99.3\ \%$
- April:$SY_{4,t}=-6.806+0.997\ SY_{3,t}$, $R^2=99.6\ \%$
- May: $SY_{5,t}=-23.685+1.080SY_{3,t}$, $R^2=99.8\ \%$
- June: $SY_{6,t}=-0.033+1.108SY_{7,t-1}+0.581\ SY_{4,t}-0.598\ SY_{8,t-1}$, $R^2=99.8\ \%$
- July: $SY_{7,t}=-8.598+1.150\ SY_{6,t}$, $R^2=99.6\ \%$
- August: $SY_{8,t}=16.660+1.081\ SY_{7,t-1}$, $R^2=99.6\ \%$
- September: $SY_{9,t}=24.051+0.796SY_{7,t}$, $R^2=99.7\%$
- October: $SY_{10,t}=10.403+0.549\ SY_{9,t}+0.253SY_{8,t}$, $R^2=99.8\ \%$
- November: $SY_{11,t}=-2.269+1.075\ SY_{10,t}-0.242\ SY_{11,t-1}$, $R^2=99.9\ \%$
- December: $SY_{12,t}=12.788+1.07\ SY_{11,t}$ $R^2=99.9\%$

The forecasting results of the cross-section method and those of Box-Jenkins ARIMA$(0\ 1\ 1)(0\ 1\ 1)^{12}$ are compared in Figure 2.

Obs	Cro-Sec	A.E.	P.E.	ARIMA	A.E.	P.E.
417	421.51	4.51	1.08%	418.93	1.93	0.46%
391	397.99	6.99	1.79%	399.81	8.81	2.25%
419	461.77	42.77	10.21%	467.15	48.15	11.49%
461	453.58	7.42	1.61%	455.08	5.92	1.28%
472	475.03	3.03	0.64%	472.1	0.1	0.02%
535	536.4	1.4	0.26%	545.97	10.97	2.05%
622	608.26	13.74	2.21%	619.11	2.89	0.47%
606	609.05	3.05	0.50%	625.51	19.51	3.22%
508	508.23	0.23	0.04%	525.98	17.98	3.54%
461	443.51	17.49	3.79%	461.45	0.45	0.10%
390	386.9	3.1	0.79%	405.61	15.61	4.0
432	427.16	4.84	1.12%	451.96	19.96	4.60%

MSE	MAPE		MSE	MAPE
208.48	2.0%		325.8	2.86%

Figure 2 Comparison of Cross-Section Model and ARIMA (A.E. denotes Absolute Error. P.E. denotes Percentage Error)

From figure 2, it can be seen that cross-section modelling is better than ARIMA(0 1 1)(0 1 1)^{12} in term of MSE and MAPE. The residual analysis shows no correlations among the residuals.

Lags	Cross- Section		ARI MA	
	ACF	PACF	ACF	PACF
1	-0.112	-0.112	-0.163	-0.163
2	-0.262	-0.278	-0.449	-0.488
3	-0.02	-0.097	0.106	-0.111
4	0.13	0.043	-0.14	-0.488
5	-0.124	-0.14	0.172	0.019
6	-0.166	-0.182	0.152	-0.149
7	0.24	0.15	-0.323	-0.261
8	-0.113	-0.193	0.023	-0.221
9	-0.111	-0.073	0.191	-0.139
10	0.023	-0.042	-0.03	-0.115

Figure 3:ACF and PACF of residuals

It seems that cross-section modelling method looks like more complicated than ARIMA model. But in fact, it is not true. Most practical forecasting users with basic statistical knowledge can understand regression analysis but maybe not familiar with Box-Jenkins' methodology. Cross-section method provides a new modelling approach for practical users.

4.3 Example Three

In order to further demonstrate the effectiveness of cross-section method, we consider a further example which is contained in Bowerman and O'Connell [1993]. The available historical data consists of the number of monthly average occupied rooms for the 15 years from 1977 to 1991 of four hotels in a city. The data are plotted in Figure 4.

Bowerman and O'Connell [1993] use an ARIMA(5,0,0)(0,1,1)^{12} model for the transformed data $Y_t^* = Y_t^{0.25}$, where Y_t is raw data. Using cross-section method, we can get 12 sub-time series each consisting of the same season's data. It is apparent that each sub-time series has approximately linear relationship with time. This suggests that we try a simple linear regression using time as the independent

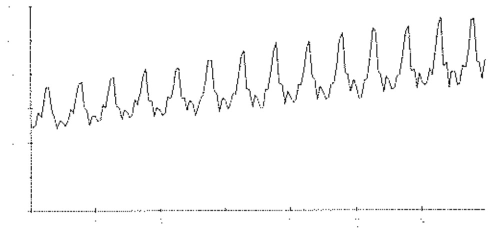

Figure 4: Monthly Average Occupied Rooms

We use the first 14 years' data to estimate parameters and the last year's data for evaluation purposes. The fitted models are:

January: $SY_{1,t}$=484+21.3*t, R^2 =98.5 %
(26.61)

February: $SY_{2,t}$=455+19.8*t, R^2 =97.4 %
(20.19)

March: $SY_{3,t}$=484+18.3*t, R^2 =95.7 %
(15.58)

April: $SY_{4,t}$=561+19.0*t, R^2 =97.9 %
(22.73)

May: $SY_{5,t}$=536+21.0*t, R^2 =98.4 %
(26.32)

June: $SY_{6,t}$=605+24.7*t, R^2 =99.2 %
(36.46)

July: $SY_{7,t}$=680+31.3*t, R^2 =98.0 %
(23.32)

August: $SY_{8,t}$=690+33.3*t, R^2 =98.9 %
(31.38)

September: $SY_{9,t}$=555+23.0*t, R^2 =97.7 %
(21.43)

October: $SY_{10,t}$=532+25.0*t, R^2 =99.1 %
(35.71)

November: $SY_{11,t}$=464+20.5*t, R^2 =98.2 %
(24.28)

December: $SY_{12,t}$= 513+24.0*t. R^2 =99.4 %
(44.47)

Note: The numbers in parentheses are associated estimate standard errors

The forecasting results are shown in Figure 5.

Obs	Cro-Sec	A.E.	P.E.	ARIMA	A.E.	P.E.
811	782.5	28.5	3.5%	764.87	46.13	5.7%
732	731.81	0.19	0.026%	735.45	3.45	0.47%
745	739.23	5.77	0.77%	745.7	0.7	0.094%
844	826.38	17.62	2.09%	832.97	11.03	1.3%
833	829.81	3.19	0.38%	816.44	16.56	1.98%
935	951	16.0	1.7%	941.44	6.44	0.69%
1110	1118.19	8.19	0.74%	1094.33	15.67	1.41%
1124	1156.27	32.27	2.87%	1128.25	4.25	0.378%
868	877	9.0	1.04%	860.75	7.25	0.835%
860	882.15	22.15	2.58%	865.7	5.70	0.663%

762	751.19	10.81	1.42%	738.58	23.42	3.07%
877	848.27	28.73	3.28%	830.92	46.08	5.25%
		MSE	MAPE		MSE	MAPE
		337.05	1.70%		466.52	1.82%

Figure 5: Comparison of Cross-Section Model and ARIMA Model

The results show that there is no significant difference between the forecasting accuracy of simple linear regression models and that of ARIMA model. The former one is much easier to understand and explain compared with the latter which is quite complicated for the practical users.

Residual analysis results also show that no significant correlations among the residuals.

It is pointless to use a complex model for forecasting such a data set to get worse forecasts while a series of simpler models can be used to get better forecasts.

5 Summary and Conclusion

In this paper, we propose a cross-section method for seasonal time series with increasing variation. Examples with both simulated and practical time series show that this method is effective. It is well-known that a unique "best forecasting method" does not exist. The " best method" is subject to many considerations such as the nature of the time series, the criteria used to evaluate the fit.

From our case studies the forecasting errors produced by the cross-section method are smaller than those of the ARIMA model and no correlations exist among the residuals.

One of the most significant features of the cross-section modelling method is that it is very straightforward and is much easier for the practical users to understand and to use. In fact, cross-section modelling is a modelling strategy which could simplifies a single complex modelling construction procedure to a several simple modelling constructions and achieves the same, even better, forecasts.

Another significant feature is that unlike most forecasting models which use a unique model (eg seasonal ARIMA model) or same type models (eg seasonal intervention model uses different dummy variables to represent different seasons) for all seasons, cross-section method may have constructed totally different models for different seasons to reflect the most important seasonal effects.

Obviously, like any forecasting model, we can not guarantee cross-section method always outperforms than other models. But, what we can say is that for some types of time series, as series discussed in this paper, cross-section is easier and forecast accuracy is better in terms of MSE and MAPE.

We conclude that among the many methods for seasonal time series analysis, the cross-section method could be considered as an alternative.

6 Acknowledgment

The first author would like to thank Charles Sturt University for some Time Release Assistance. This research work is supported by the Department of Mathematics of Charles Sturt University.

7 References

Box G E P and Jenkins G M, Time series analysis, forecasting and control, 2nd Ed., San Francisco: Holden-Day, 1976

Bowerman B L, Koehler A B and Pack D J, Forecasting time series with increasing seasonal variation, *Journal of Forecasting*, No 5, Vol 9, 419-436, 1990

Bowerman B L and R O'Connell, Forecasting and time series: An applied approach, 3 ed, Duxbury Press, 1993

Chatfield C and Prothero D L, Box-Jenkins seasonal forecasting: problems in a case study, *Journal of the Royal States Soc A*, Part 3, 136, 295-315, 1973

Wichern D W, The behaviour of the sample autocorrelation function for an integrated moving average process, *Biometrika*, 60, No 2, 235-239, 1973