Forecasting Tourist Arrivals: The Use of Seasonal Unit Root Pre-testing to Improve Forecasting Accuracy.By Dillon Alleyne,Department of Economics,UWI,Mona.

I. Introduction

This paper focuses on forecasting tourist arrivals in a tourism dependent economy and argues that given the importance of the tourism sector to the Jamaican economy, accurate forecasts of tourist arrivals are of importance for planning by both the private and public sectors.
Tourism has emerged as one of the fastest growing industries in the Caribbean and it is the major foreign exchange earner in many countries in the region[1]. In Jamaica this is no exception with annual arrivals increasing from 532,864 visitors in 1978 to 2,116,035 in 2001, which is an annual growth rate of 5.8 percent. Over the same period, travel receipts rose from US$.148 M to US$1,332,597M and this represented an annual growth rate of 9.1 percent. The importance of tourism has been magnified by the stagnation in the rest of the economy and the increase in the service sector's contribution to GDP[2]. For example, in 1996 the annual percentage change in real GDP (at 1996 prices) was –1.0 percent and in 1998 it was –0.3 percent. In 1999 and 2000 the percentages were 0.7 and 1.7 percent respectively (Economic and Social Survey of Jamaica (ESSJ), 2001).

Because of the obvious importance of tourism in the Jamaica economy, the role of forecasting in tourism research is of considerable importance. Frechtling (2001) has pointed out that forecasting is useful in shaping demand in the short run and anticipating it in the long run to avoid unsold inventories and unfulfilled demand. In particular, because some investments are lumpy and require a long lead time, there is need for accurate forecasting to avoid the financial costs of

[1] The events of September 11th in the united States may have a temporary effect on world tourism growth.
[2] The contribution of the service sector to GDP in 1996 was 76.9 percent and in 2001 it was 84.2 percent.

excess capacity or the opportunity costs of unfulfilled demand (Frechtling, 2001). Secondly, because production and consumption may take place at the same time, forecasting can help anticipate and synchronise the process. In addition, since consumer satisfaction depends on complementary services, forecasting can help to anticipate the demand for such services.

Given the considerable public funds expended by the Jamaica Tourist Board to improve the image of the industry, and to expand the tourism market, accurate forecasting of demand can help to optimize the use of public funds. In addition, where revenue is collected from tourists on a per capita basis, and if these revenues are significant, accurate arrival forecasts can help to facilitate government expenditure policy. Gustavasson and *Nordström* (2000) have pointed out that due to many capacity- related decisions that must be made well in advance in the industry even small improvements in forecasting are worthwhile.

Forecasting is complicated however, by the strong seasonality of most tourism series. Franses (1998) has pointed out that a typical approach to the analysis of macroeconomic time series is to see seasonality as a form of data contamination, yielding little useful information and as a result seasonality is removed prior to data analysis. This was the approach often taken by many census and statistical departments (See Bjarne, Hylleberg et al. 2001). There are instances however, when seasonality itself is of interest and seasonally adjusted data is not very informative. In the case of tourism analysis seasonality is integral to the process and is crucial to the timing of the issuance of policy measures in addition to studying the long run trend. For this reason there are strong arguments for modelling seasonality rather than removing it (See Ghyles (1994), Hylleberg (1994) and Miron (1996)).

Typifying works which compare the forecasting performance of various econometric models are those Martin and Witt (1989), Witt and Witt (1995), Garcia-Ferrer and Queralt (1997) and *Nordström* (1999). Martin and Witt (1989) used least squares regression to model tourist arrivals as a function of a number of factors including price, income, airfare and special events (See Kulendran and King 1997)[3].

Among the other works that have used regression analysis to model the level of tourist arrivals have been Loeb (1982), Witt and Martin (1987) and Crouch et al. (1992), Carey (1991), Metzgen-Quemarez (1990), Worrell et al. (1997),Worrell (1995), Dharmaratne (1995) and Greenidge (2001).

This paper employs univariate forecasting techniques to forecast arrivals.This is a limited methodology relative to structural models which allow policy makers to determine how changes in particular variables can help to improve the industry. As Greeenidge (2001) has pointed out these models have no explanatory variables so the individual components are difficult to interpret. There is evidence however that the forecasting record of many uinivariate models have considerable forecasting accuracy and in many cases the price variables in structural models are difficult to predict Gustavasson and *Nordström* (2000).

While the literature on estimating and forecasting tourism demand is vast[4] there is an emerging literature on the impact of accounting for non-stationarity at the seasonal frequency (or unit root testing) on forecasting accuracy. Non technically it means that if a series has a unit root it is non-stationary and some degree of differencing is necessary to make this series stationary, in order to employ econometric techniques. For example, Witt et al. (1994) and Lim and McAleer (2000)

---

[3] Gustavasson and *Nordström* (2000) have argued that one of the reasons for relatively modest short-and medium-term forecasting performance of economic models is the difficulty of obtaining good forecasts of the price variable.

modeled a variety of tourist arrival data and the implications for seasonality. Later, Lim and McAleer (2001) employed univariate techniques to forecast quarterly tourist arrivals to Australia and to determine their forecasting accuracy using a variety of seasonal filters. Kulendran and King (1997) also employed a variety of models to rank the forecasting performance of various tourist arrival series using seasonal unit root testing. Other studies such as Gustavasson and *Nordström* (2000) utilised both Auto Regressive Integrated Moving Average (ARIMA) and Vector ARMA models to examine the impact of seasonal unit root pretesting on forecasting monthly tourism flows.

According to Clements and Hendry (1997,p, 341) the work of Hylleberg et al. (1990) [hereafter called HEGY] helped to popularize modeling economic time series as variables with possible seasonal unit roots as against the Box and Jenkins (1970) [5]methodology in which seasonal unit roots are implicitly assumed.

Central to understanding the nature of seasonality is to determine whether a series exhibit a stochastic or deterministic trend. In the former case, the series may require annual differencing due to a unit root at all frequencies, to be made stationary. In the latter case, a regression of the first difference of the series on seasonal dummies may be sufficient.

Tests for seasonal unit roots have also been developed by Hasza and Fuller (1982) and Osborn et al. (1988) and empirical studies employing such methods have been several (See Osborn and Birchenhall, et al.1999). Among the studies that have examined the impact of seasonal roots on forecasting accuracy are Franses (1991), Clements and Hendry (1997), Paap et al. (1997) and

[4] See Kulendran and King(1997) and Lim and McAleer (2001) for references
[5] This type of model is usually referred to as the airline model.

Osborn and Birchenhall et al. (1999). Many of the findings however, are not always consistent and vary by series, frequency and country.

The objective of this study is to determine the accuracy of forecasting various tourist arrival series based on seasonal unit root pretesting. The paper models the series using a variety of univariate forecasting techniques and compares the forecasts at various horizons. The series of interest are quarterly total visitor arrivals, long staying and short staying visitors and landed[6] visitors from the United States, Europe and Canada. The period of interest is from 1968:1 to 2001:3.

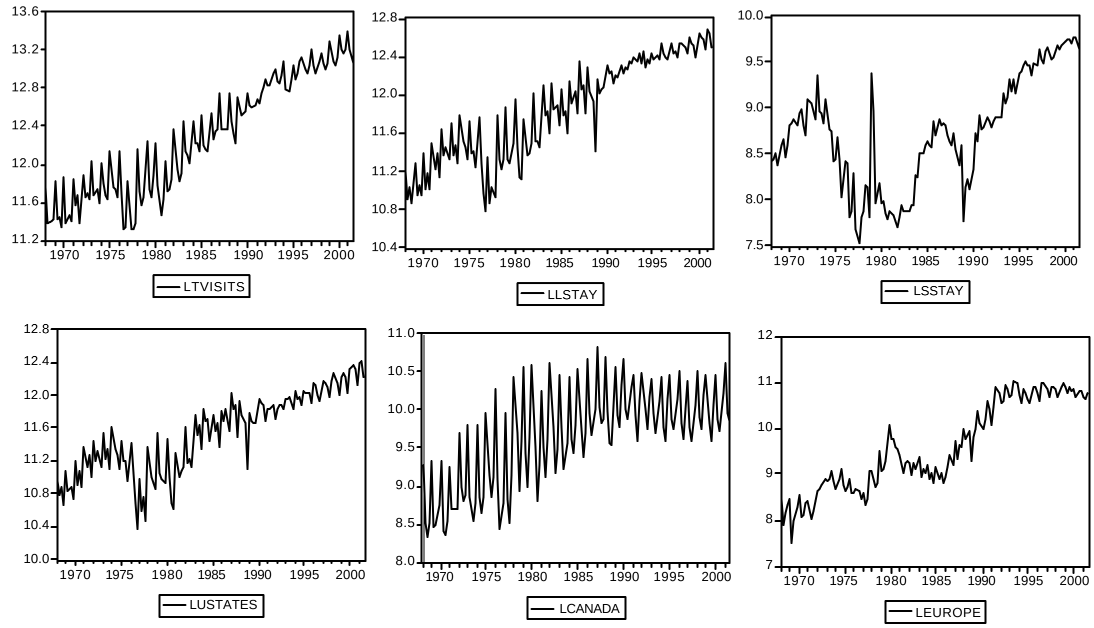
Section 2 examines the descriptive properties of the data. Section 3 looks at the models used for forecasting while section 4 outlines the HEGY procedure and tests for seasonal unit roots. Section 5 examines the forecasts and their accuracy in relation to the various models. Section 6 is the conclusion.

2. Descriptive Properties of the data

The six data series on tourist (visitor) arrivals are all in logs and are shown in Figure I. These are total visitors (LTVISITS), long staying visitors (LLSTAY), short staying visitors (LSSTAY) and landed visitors from the United States (LUSTATES), Canada (LCANADA) and Europe (LEUROPE).

[6] Landed visitors exclude cruise passengers and armed forces. Visitors from the United States, Europe and Canada are referred to as landed visitors.

Figure I: Log of Various Quarterly Tourist(Visitor) Arrival Series:Total Arrivals(Tvisitors),Long Stay(LLstay) Short Stay(Lsstay), Visitors From United States(LUstates), Canada(Lcanada), Europe(Leurope)



Source: Statistical Digest.Bank of Jamica (various years).

The total number of visitors[7] and those long staying are highly correlated since a significant number of total arrivals are long staying guests. For example in 2001, long staying guests were 53 percent of total visitor arrivals.

The United States constitutes the largest market and events there dominate the total arrival series, thus annual landed visitors from the United States were 40.1 percent of total annual arrivals in 2001, while European guests were only 8 percent and Canadian arrivals were 4.8 percent of overall arrivals. Thus not surprisingly, there is considerable similarity among the series, LTVISITS, LLSTAY and LUSTATES, as each trend upwards with strong seasonality. The series

[7] The total number of visitors includes all visitors and armed forces personnel on shore leave. Foreign crews and other carrier personnel, foreign diplomats, technical assistance personnel and migrant workers are excluded. Short staying visitors are persons staying 1 to 2 nights.

also exhibit significant breaks in 1980 reflecting the period of violence leading up to general elections in that year, in 1988 due to Hurricane Gilbert, which ravaged the island, and in 1990 during the Gulf War.

The last quarter included the impact of the September 11 terrorist bombing in the United States and its impact on American travel to the country. The Canadian series shows a distinct seasonal pattern with what appears to be seasonal shift while the short stay and European series appear to be somewhat volatile and respond sharply to the events in the 1980's and 1990's.

Table I sets out the mean growth rates of each series, the sample variances and the $\overline{R}^2$ which is estimated from the regression of the change in the log of each series on four seasonal dummies.

$$\Delta y_t = \hat{d}_1 D_{1,t} + \hat{d}_2 D_{2,t} + \hat{d}_3 D_{3,t} + \hat{d}_4 D_{4,t} + e_t \qquad (1)$$

Where the $\hat{d}_i$ are the estimates of the quarterly growth rates in seasons i and $D_{i,t}$ are dummy variables which take the value of 1, for observations in season i and 0 otherwise. Franses (1996) and other have cautioned against the use of such an equation to make strong inferences, thus it is used only for illustration and is not definitive about the nature of seasonality. This is because a high $\overline{R}^2$ in this regression can arise from seasonal unit roots and need not imply any deterministic seasonal effect (Abeysinghe, 1994, Franses et al.1995, Miron, 1994). Formal tests are carried out later.[8]

[8] . Franses (1998), points out that stochastic seasonality means that the seasonal pattern changes over time. It is well known that in the case of a random walk the variance becomes large and the shocks have a permanent effect on the series. When the DGP is approximated by a random walk but seasonal dummies as in

$\Delta_1 y_t = \sum_{s=1}^{4} d_s D_{s,t} + e_t$ are used to reflect the process, the d_s will not be constant over time. Thus the

Table I: Summary Statistics of the First Difference of the Logs of the Series, from 1968:1 to 2001:3.The $\overline{R}^2$ is from equation (1)

	Mean	Variance	$\overline{R}^2$
LTVISITS	0.97	0.067	0.73
LLSTAY	0.98	0.080	0.70
LSSTAY	0.89	0.064	0.06
LUSTATES	0.93	0.080	0.72
LEUROPE	1.69	0.073	0.19
LCANADA	0.41	0.374	0.86

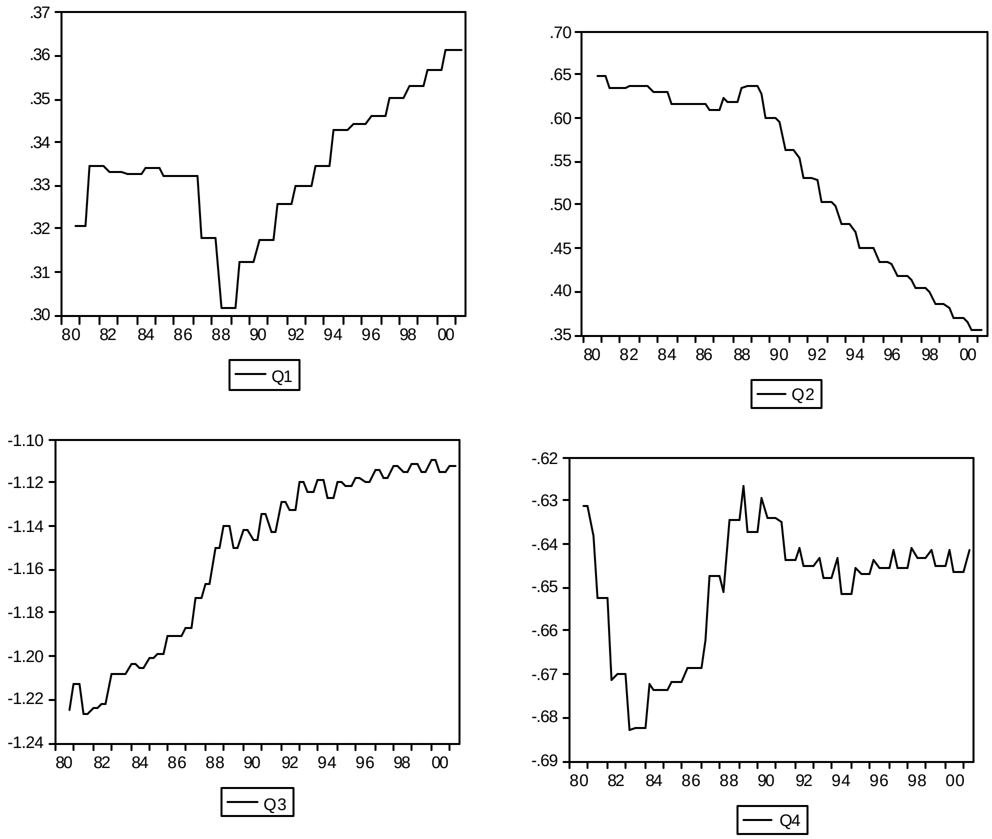
Source: Statistical Digest. Bank of Jamaica (various years).

The mean growth rates were strongest for the series LEUROPE, LLSTAY and LUSTATES which had significantly high variances while the Canadian series seem to have relatively low mean growth and high variance. This is not surprising since the Canadian series tended to show a declining growth rate in the 1990's. The initial results seem to suggest that the high $\overline{R}^2$ for LCANADA may reflect a constant seasonal pattern and this may be the case for total visitors, long staying visitors and the Unites States visitors.

The exceptions are LEUROPE and short staying visitors (LSSTAY), which have relatively low $\overline{R}^2$ values. When a regression was run on a subset of the series (1968:1-1979:4) and the rest of the observations were each added over time the coefficients for each of the series exhibited a slowly changing seasonal pattern which may reflect a stochastic rather than a deterministic trend. This also has important implications since it suggests that the seasonal pattern is changing Figure 2 reports the results for the Canadian series.

R^2 and the coefficients will reflect spurious results (Frances,1998:115).

Figure II:Estimates of Quarterly Growth rates of Arrivals from Canada by Seasons.



3. Forecasting Models and Seasonal Unit Root tests

This study considers the implications of pretesting for seasonal unit roots and the imposition of various seasonal filters on forecasting accuracy. Clements and Hendry (1997:p, 342) point out that there may be some tension between the number of roots suggested say by the HEGY procedure and the numbers implicit in the Box and Jenkins approach to seasonality. The methodology of Box and Jenkins (1970) assumes that all roots are present, and employs seasonal autoregressive moving average (SARMA) models which have had some success in forecasting. Their approach takes the form:

$$(1-L)(1-L^4)y_t = m + (1-q_1 L)(1-q_4 L^4)e_t \qquad (2)$$
$$= m + e_t + b_1 e_{t-1} + b_2 e_{t-4} + b_3 e_{t-5}$$

where $e_t \sim IN(0, s_e^2), |q_1| < 1, q_4 < 1 \quad and \quad L^K y_t = y_{t-k.}$

The seasonal filter $\Delta_4 = (1-L^4)$ captures the seasonal part of the model, which is the tendency of the series in each season to be correlated with values in the same season the year before. Finally the filter $\Delta_1 = (1-L)$ is the regular unit root and captures the stochastic trend proxied by a random walk (See Nelson and Plosser, 1982).[9] The Box-Jenkins filter $(1-L)(1-L^4) =$(1-L)(1-L)(1+L)(1-iL)(1+iL) can be disaggregated to show that the seasonal operator $\Delta_4 = (1-L^4)$ has four roots: +1, -1, +i and –i. The complex conjugate, +i and –i, are similar for quarterly data and are referred to as the annual cycle. The roots of the operator are +1, the zero cycle, -1 the biannual cycle or (1/2 cycle per quarter) and $\pm\ i$, the annual cycle or (1/4 cycle per quarter or 1 cycle per year). Thus the two complex roots can be written as[10] $(1-iL)(1+iL) = (1+L^2)$.

In the Box-Jenkins framework it is assumed that the series admits of two zero frequency roots and as well as roots at the biannual and semi-annual frequencies. Thus the regular plus seasonal roots can be written as integrated of order $I(1,1,1)$. Clements and Hendry (1997)[11] point out, that the HEGY procedure seldom finds roots at all frequencies, while the Box and Jenkins procedure may

---

[9] Clements and Hendry (1997) point out that other views exist (See Perron 1989,1990).

$$Note\ that\ (1-L^4) = (1-L^2)(1+L^2) = (1-L)(1+L)(1+L^2)$$
$$let\ i = \sqrt{-1}\ so\ that\ i^2 = -1. Thus\ -i^2 = 1$$

[10] $This\ yields\ ,(1-L^4) = (1-L)(1+l)(1-i^2L^2)$
$$= (1-L)(1+L)(1-iL)(1+iL)$$
$$= (1-L)(1+L+L^2+L^3)$$

imply over differencing which may convert level shifts in the mean to blips which may be taken to be outliers.

An alternative formulation might be an autoregressive moving average model in first difference with a possible constant and three seasonal dummies.

$$f_p(L)(1-L)y_t = a_0 + \sum_{i=1}^{3} a_i D_{it} + q_q(L)e_t \qquad (3)$$

In this case the D_{it} are dummies and $f_p(L)$ *and* $q_q(L)$ are polynomials in the backward shift operator L with the usual restrictions applying. This model assumes a deterministic trend. The idea is to pretest the series for unit roots and then compare the filtered results with results based on models (2) and (3). The filtered series for each variable is assumed to be the base model.

4. Testing For Seasonal Roots

In this section we employ the HEGY procedure to test for unit roots then to use the pre-tested results to transform the various series. The HEGY procedure is based on the following equation:

$$f(L)x_t = m_t + e_t \qquad (4)$$

where $f(L)$ is the seasonal differencing operator for the quarterly series x_t, m_t contains the deterministic component of a constant, three seasonal dummies and a time trend, and e_t is a white noise process. A root of $f(L)$ can be obtained from the following auxiliary regression equation:

[11] See Osborn 1990, Hylleberg et al.(1993) Franses 1996

$$\tilde{f}(L)y_{4t} = p_1 y_{1,t-1} + p_2 y_{2,t-1} + p_3 y_{3,t-2} + p_4 y_{3,t-1} + j_t y_{4,t-1} + m_t + e_t \qquad (5)$$

$$where \quad y_{4t} = (1-L^4)x_t; \quad y_{1,t-1} = (1+L+L^2+L^3)x_{t-1}; \quad y_{2,t-1} = -(1-L+L^2-L^3)x_{t-1}$$
$$y_{3,t-1} = -(1-L^2)x_{t-1}$$

This equation can be estimated by ordinary least squares (OLS) and the lag on y_{4t} is used to whiten the residuals. In this equation deterministic terms can be added without changing the distribution.

The three null and alternative hypotheses can be tested as follows.

$$(1) H_0 : p_1 = 0, \quad H_1 : p_1 < 0$$
$$(2) H_0 : p_2 = 0, \quad H_1 : p_2 < 0$$
$$(3) H_0 : p_3 = p_4 = 0, \quad H_1 : p_3 \neq 0 \quad and\ /\ or\ p_4 \neq 0$$

The t-test is used for the first two hypotheses and an F-test for the third. The distributions for the F and t-tests are tabulated by Hylleberg et al. (1990). In the first case, rejection of H_0 means that the variable is stationary at the 0 frequency. Rejection of the second means that there is stationarity at the semiannual frequency and rejection of the third means that there is stationarity at the annual frequency. Acceptance of all three hypotheses means that the variables are integrated at all frequencies and this can be written as $I(1, 1, 1)$.

An augmented Dickey-Fuller test, (Dickey and Fuller, 1979,1981), was first carried out on all the series and these were found to be nonstationary since they had unit roots at the zero frequency. Table 2 reports the seasonal unit root tests utilizing the HEGY procedure. [12]The results suggest that that all three hypotheses are accepted for the variables LVISITS, LUSTATES, LCANADA and LLSTAY at the 5% level of significance. These variables require the seasonal filter $(1-L^4)$

[12] A number of other testing procedures have been suggested in the literature.(See Franses and Taylor, 2000; Kawasaki and Franses1996; Arthur and Lopes 2001)

to be stationary. In the case of LEUROPE and LSSTAY, only the first hypothesis t_{Z1} is accepted thus the filter (1-L) is required.

Table II: HEGY Test for Seasonal Unit Roots, with Intercept, Seasonal Dummies and Trend (I, SD, Tr), for Period 1968:1 to 1998:4

	t_{p1}	t_{p2}	$F_{p3,p4}$	P	Lags
LVISITS	-2.122	-1.738	4.246	0.162	1,2,3,4
LUSTATES	-2.494	-1.904	6.112	0.121	1,2,3
LCANADA	-1.171	-1.580	1.180	0.506	1,2,3,4,5,6,7
LEUROPE	-2.196	-3.051*	10.057 *	0.712	1
LLSTAY	-2.213	-1.474	3.901	0.177	1,2,3,4
LSSTAY	-2.005	-5.647 *	37.637 *	0.134	1

P= LM test for serial correlation.

The test values for (hypothesis 1) t_{p1} at the 1%, 5% and 10% levels of significance are -4.15, -3.52 and -3.21.

For (hypothesis 2) t_{p2} the values are -3.57, -2.93 and -2.61 respectively and for (hypothesis 3) $F_{p3,p4}$ the values are 8.77, 6.62 and 5.55 respectively.

The table also reports the LM test for serial correlation and the number of lags required to reduce autocorrelation in the HEGY regression.[13] Figures 2 to 6 in Appendix I, graphs the variables after applying filters at the zero frequency (1-L), the Box and Jenkins filter $(1-L)(1-L^4)$, the seasonal filter $(1-L^4)$ and the residuals from equation (1) to remove possible deterministic seasonal effects.

[13] Gustavasson and *Nordström* point out that the addition of several lags may reduce the power of the test (See Hylleberg 1995).

Figure 2 reports the graphs for long staying visitors (LLSTAY). The first difference seems not to have removed much of the seasonality with wider swings before the 1990's while the Box-Jenkins filter may have over differenced the series. The seasonal filter seems to have removed most of the seasonal effects but has less volatility than the Box-Jenkins filter. The residual series (LLSTAYRES) seem also to have some seasonality left with greater seasonal impact in the 1990's. Very similar conclusions can be drawn for figures 3 and 4. In figures 5 and 6 the residual series, LSSTAYRES and LEUROPERES show that the seasonal dummies have little impact in relation to the original series which is not surprising given the low explanatory power of equation (1) in both cases.

5. Forecasting Results and Interpretation

The various models were run over a subset of the sample and the rest of the series was then used as the forecast horizon. The period 1968:1 to 1998:4 was the estimation period then multistep forecasts were done at time n for n+1 until n+h. In the present case h=8 and as the forecasting horizon increases the number of observations gets smaller. Thus when the forecast is at 1999:1 the number of observations is 11 and when it is 1999:2 it falls to 10 and so on.

Osborn and Birchenhall et al.(1999) have argued that although the same or different models should be used at different forecasting horizons, iterating a single model is usually the approach. Clements and Hendry (1996) support this by arguing that misspecification is necessary but not sufficient to justify different models at each horizon.

The root mean square error (RMSE) is used to measure the forecasting performance of the various models. Since the forecasts with respect to the future levels of each of the variables are converted to levels, the problems with the RMSE, as pointed out by Clements and Hendry (1993), are avoided.[14] We also report Theil's U[15] which together with the RMSE gives some idea of the magnitude of differences in forecasting performance of the different approaches.

The procedure employed to determine the model to represent each filter was to use a maximum lag structure then test down to determine the model that had significant coefficients and was most parsimonious. The best model was then chosen based on the Akaike information criterion and no significant serial correlation.

Tables 3-8 report the results for the variables LLSTAY, LUSTATES, LTVISITS, LCANADA, LEUROPE and LSSTAY. The HEGY procedure had identified the seasonal filter as most appropriate for the first four variables mentioned above and the models related to these are reported in tables 3-6. In each case the Box-Jenkins model performs the worst except for the early forecasting horizon in some cases. At longer horizons its forecasting accuracy deteriorates considerably. The seasonal filter outperforms the others for these four variables and improves considerably at longer horizons.

The exception is the variable LUSTATES, in which the seasonal filter does outperform the others but the forecasts deteriorate at longer horizons. The results casts doubt on the high $\overline{R}^2$ and significant coefficients of equation (1) and may reflect a stochastic seasonal pattern rather than a deterministic one. The results are in line with those from Franses (1998:114) where the Box-

[14] The generalized forecast error second moment(GFESM) measure developed by Clements and Hendry (1993) is not necessary in this case.

Jenkins model while performing well for the 1 step ahead forecasts performs badly for the multi-step ahead forecasts.

These results are also in accord with those of Franses (1991) for the forecasts at shorter horizons and those of Clements and Hendry (1997) for longer horizons. Franses (1991) who used monthly data, also points out that it is extremely important to determine the nature of the seasonality since this bears heavily on forecasting accuracy. Some of the other findings in the literature are also interesting and help explain our findings.

Clements and Hendry (1997) indicated that imposing roots at all frequencies lead to as good forecasts as imposing a small number of roots as is often suggested by the HEGY procedure. Given that the HEGY procedure, in this study, suggests seasonal roots at all frequencies for four out of six of the series, this may explain the superior forecasting results for the seasonal filter.

Paap et al. (1997) compared the forecast performance of autoregressive seasonal unit root models with seasonal mean shift models and found that the mean shift models outperformed the models that incorrectly imposed too many roots. On the other hand, if there were more than one seasonal unit roots present, the HEGY procedure outperformed the mean shift model. The results however, are not in agreement with Taylor (1997) who found little difference between models with different sets of unit roots imposed.

Gustavasson and *Nordström* (2000) using monthly series have suggested that the reason for the better performance of the seasonal filter may be due to shifts in deterministic factors over the forecasting period.

---

[15] The models were run in RATS Version 5 and the RMSE and Theil's U are as defined in the reference

The last two series, LEUROPE and LSSTAY are more volatile than the rest and the HEGY procedure suggested that they were best handled by the 0 frequency filter. Tables 7 and 8 report the results for these two series. For the LEUROPE series, the regular filter with seasonal dummies performs best, relative to other model with a regular filter without dummies, however from horizon 6-8 the seasonal filter is better. In the case of short staying visitors, the seasonal filter performs better generally but more so at the shorter horizon. The regular filter with dummies is the next best model in terms of forecasts but it has the best diagnostic features. Again, the Box-Jenkins approach performs reasonable well at the short horizon but deteriorates at longer horizons.

The results give no guidance whether there is a greater penalty to misspecify a stochastic trend relative to a deterministic trend, but it seems that erring on the side of differencing the series produces good forecasts at longer horizons. This is certainly an area for more intense research.

6. Conclusion

The results are not out of line with findings from several other studies and suggest that seasonal unit root pretesting is important for model specification (see Lim and McAleer, 2000). They also suggest that it may be better to err on the side of imposing seasonal unit roots rather than seasonal dummies and that seasonal unit root pretesting can help improve forecasting accuracy for the tourism series. Franses (1991) has shown that misspecification in general can affect forecasting

---

manual.

accuracy, thus emphasis has to be placed on accurate identification of the series. These results are for quarterly series and may not be generalisable at other frequencies.

References

Arthur, C. B, and da Silva Lopes. 2001. "The order of integration for quarterly macroeconomic time series: A simple testing strategy". The Royal Economic Society Annual Conference 2002.Paper No.55.

Abeysinghe, T. 1994. Deterministic seasonal models and spurious regressions. *Journal of Econometric* 61: 259–72.

Bank of Jamaica. *Statistical Digest* (Various years).

Bjarne, B., S. Hylleberg, M. Nielsen, L. Skipper, and S. Stentoft. 2001. Seasonality in economic models". Working Paper No.2001–16. Centre for Dynamic Modelling in Economics. University of Aarhus, Denmark.

Box, G. E. P., and G. M. Jenkins. 1970. *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

Carey, Kathleen. 1991. "Estimation of tourism demand". *Atlantic Economic Journal* 19, no. 3 (September):32-40.

Clemetns, P., and D. Hendry. 1993. "On the limitations of comparing mean squared forecast errors". *Journal of Forecasting* 12: 617–37.

Clements, P., and D. Hendry. 1996. "Muti-step estimation for forecasting". *Oxford Bulletin of Economics and Statistics* 58: 653–80.

Clements. P., and D. Hendry. 1997. "An empirical study of seasonal unit roots in forecasting". *International Journal of Forecasting* 13: 341–55.

Crouch, G. I., L. Schultz , and P. Valerio. 1992. "Marketing international tourism to Australia: A regression analysis". *Tourism Management* 13: 196–208.

Dharmaratne, G. S.1995. "Forecasting tourist arrivals in Barbados". *Annals of Tourism Research* V.22, no. 4: 804–18.

Dickey, D.A and Fuller,W .A1979."Distribution of the Estimators for Autoregressive Time series With a Unit Root".*Journal of American Statistical Association*, 74: 427-431.

Dickey, D.A and Fuller,W .A.1981."Likelihood Ratio Statistics for Autoregressive Time series With a Unit Root".*Econometrics*, 49:1057-1022.

Economic and Social Survey of Jamaica (ESSJ), *Planning Institute of Jamaica*, 2001.

Franses, P. H. 1991. "Seasonality, non-stationarity and the forecasting of monthly time series". *International Journal of Forecasting* 7: 199–208.

Franses, P. H. 1996. *Periodicity and Stochastic trends in Economic time series*. Oxford: Oxford University Press.

Franses, H. P. 1998. *Time Series Models for Business and Economic Forecasting*. Cambridge: Cambridge University Press.

Franses, P. H., S. Hylleberg, and H. S. Lee. 1995. "Spurious deterministic seasonality". *Economics Letters* 48: 249–56.

Franses H. P., and A. M. Taylor. 2000. "Determining the order of differencing in seasonal time series processes". *Econometrics Journal* 3: 250–64.

Frechtling, D. C. 2001. *Forecasting Tourism Demand: Methods and Strategies*. Butterworth, Heinemann. Linacre House, Jordan Hill, Oxford, OX2 8DP.Great Britain.

Garcia-Ferrier, A., and A. Queralt. 1997. "A note on forecasting tourist demand in Spain". *International Journal of Forecasting* 13: 539–49.

Ghyles, E. 1994. "On the Economics and Econometrics of Seasonality". In *Advances in Econometrics, 6th World Congress of the Econometrics Society*, edited by C. A Sims. Cambridge: Cambridge University Press.

Greenidge, K. 2001. "Forecasting tourism demand. An STM approach". *Annals of Tourism Research* 28, no.1 :98–112.

Gustavasson. P., and J. Nordström. 2000. "The impact of seasonal unit roots and Vector ARMA modeling on forecasting monthly tourism flows". Working Paper 150. Trade Union Institute for Economic Research and Stockholm School of Economics, Sweden.

Hasza, D. P., and W. A. Fuller. 1982. "Testing for nonstationary parameter specification in seasonal time series models". *Annals of Statistics* 10: 1209–16.

Hylleberg, S. 1994. "Modelling seasonal variation". In *Nonstationaary Time Series Analysis and Cointegration*, edited by C. P. Hargreaves. Cambridge: Oxford University Press.

Hylleberg, S. 1995. "Testing for seasonal unit roots: General to specific or specific to general?" *Journal of Econometrics* 69: 5–25.

Hylleberg, S., R. F. Engle, C. W. J. Granger and B. S. Yoo. 1990. "Seasonal integration and cointegration". *Journal of Econometrics* 44: 215–38.

Kawasaki, Y., and H. P. Franses. 1996. "A model selection approach to detect seasonal unit roots". Timbergen Institute Rotterdam – Papers, 96-180.Netherlands.pp:1-20.

Kulendran, N., and M. L. King. 1997. "Forecasting international quarterly tourist flows using error correction and time–series models". *International Journal of Froecasting*13: 319–27.

Lim C., and M. McAleer. 2000. "A seasonal analysis of Asian tourist arrivals to Australia". *Applied Economics* 32: 499–509.

Lim. C., and M.McAleer. 2001. "Time series forecasts of international tourism demand for Australia". Discussion paper No.533. The Institute of Social and Economic Research, Osake University, Japan.

Loeb, P. D. 1982. "International travel to the United States: An econometric evaluation". *Annals of Tourism Research* 9: 7–20.

Martin, C. A., and F. S. Witt. 1989. "Accuracy of forecasting: An assessment of future directions". *International Journal of Forecasting* 2: 15–39.

Metzgen-Quemarez, Ydahlia. 1990. "Estimating the demand for international tourist services". Woodrow Wilson School, Princeton University.

Miron, J. A. 1994. "The economics of seasonal cycles". In *Advances in Econometrics, 6th World Congress of the Econometrics Society*, edited by C. A Sims. Cambridge: Cambridge University Press.

Miron, J. A. 1996. *The Economics of Seasonal Cycles*. Cambridge, MA: MIT Press.

Nelson, C. R., and C. I. Plosser. 1982. "Trends and random walks in macroeconomic time series". *Journal of Monetary Economics* 10: 139–62.

Nordström J. 1999. "Unobserved components in international tourist demand".Swedish Working papers in Economics. Umeå Economic Studies, No. 502, Umeå University.

Osborn, D. R., H. Saeed, and C. R. Birchenhall. 1999. "Seasonal unit roots and forecasts of the two-digit European industrial production". *International Journal of Forecasting*15 : 27–47.

Osborn, D. R, A. P. L. Smith, and C. R. Birchenhall. 1988. "Seasonality and the order of integration for consumption". *Oxford Bulletin of Economics and Statistics* 50: 361–77.

Paap, R., P. H. Franses, and H. Hoek. 1997. "Mean shifts, unit roots and forecasting seasonal time series". *International Journal of Forecasting* 13: 357–68.

Perron, P. 1989. "The great crash, the oil price shock and the unit root hypothesis". *Econometrica* 57: 1361–1401.

Perron, P. 1990. "Testing for a unit root in a time series with changing mean". *Journal of Business and Economic Statistics* 8: 501–15.

Taylor, A. M. R. 1997. "On the practical problem of comparing unit roots tests". *International Journal of Tourism* 13: 307–318.

Witt, F. S, and C. A. Martin. 1987. "Econometric models for forecasting international international tourist demand". *Journal of Travel Research* 3: 23–30.

Witt, S.F , C.A.Witt.1995."Forecasting Tourism demand:A review of empirical research".*International Journal of Forecasting* 11: 447-475.

Witt, A. C, S. F. Witt, and N. Wilson. 1994. "Forecasting international flows". *Annals of Tourism Research* 21, no. 3: 612–28.

Worrell, DeLisle.1995. "Forecasting tourism demand in Barbados". Central Bank of Barbados, Working Papers. . Bridgetown, Barbados:Central Bank of Barbados

Worrell, D, K. Greenidge and K. Dalrymple.1977. “Forecasting Tourist demand in Barbados”. In Central Bank of Barbados Working Papers, Vol.II. Bridgetown, Barbados:Central Bank of Barbados.

Appendix I

Figure 2 : Log of Quarterly Long Staying Tourist,1968:01-2001:03:The results of Applying Various Filters

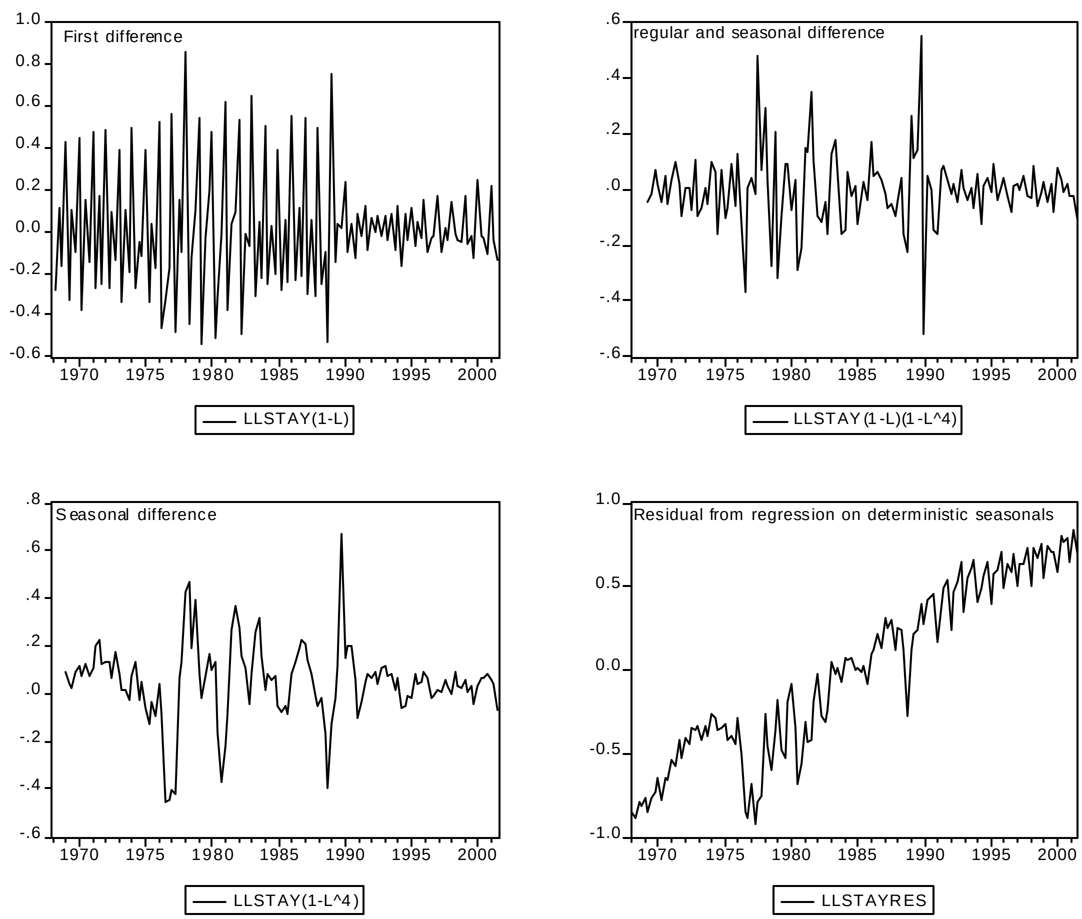


Figure 3: Log of Quarterly Total Visitor Arrivals, 1968:01-2001:03:The results of Applying Various Filters

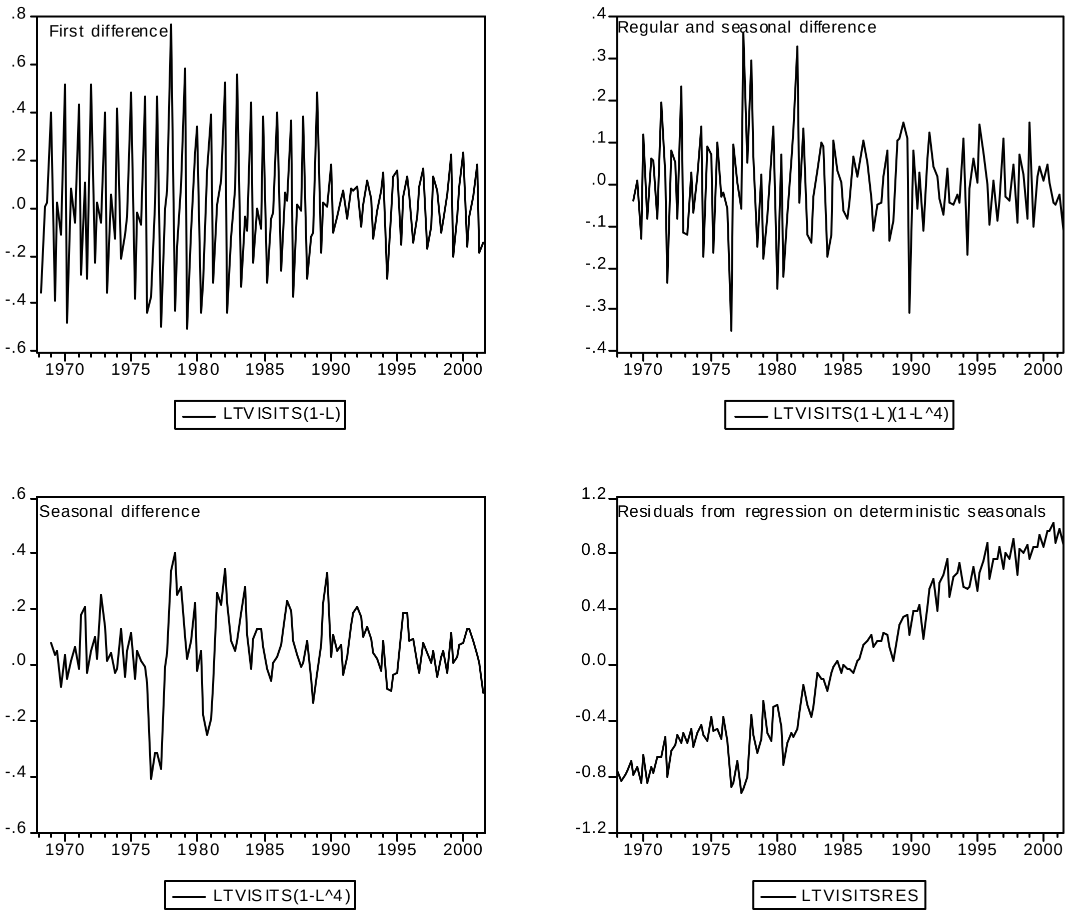


Figure 4: Log of Quarterly US Arrivals 1968:01-2001:03:The Results of Applying various Filters

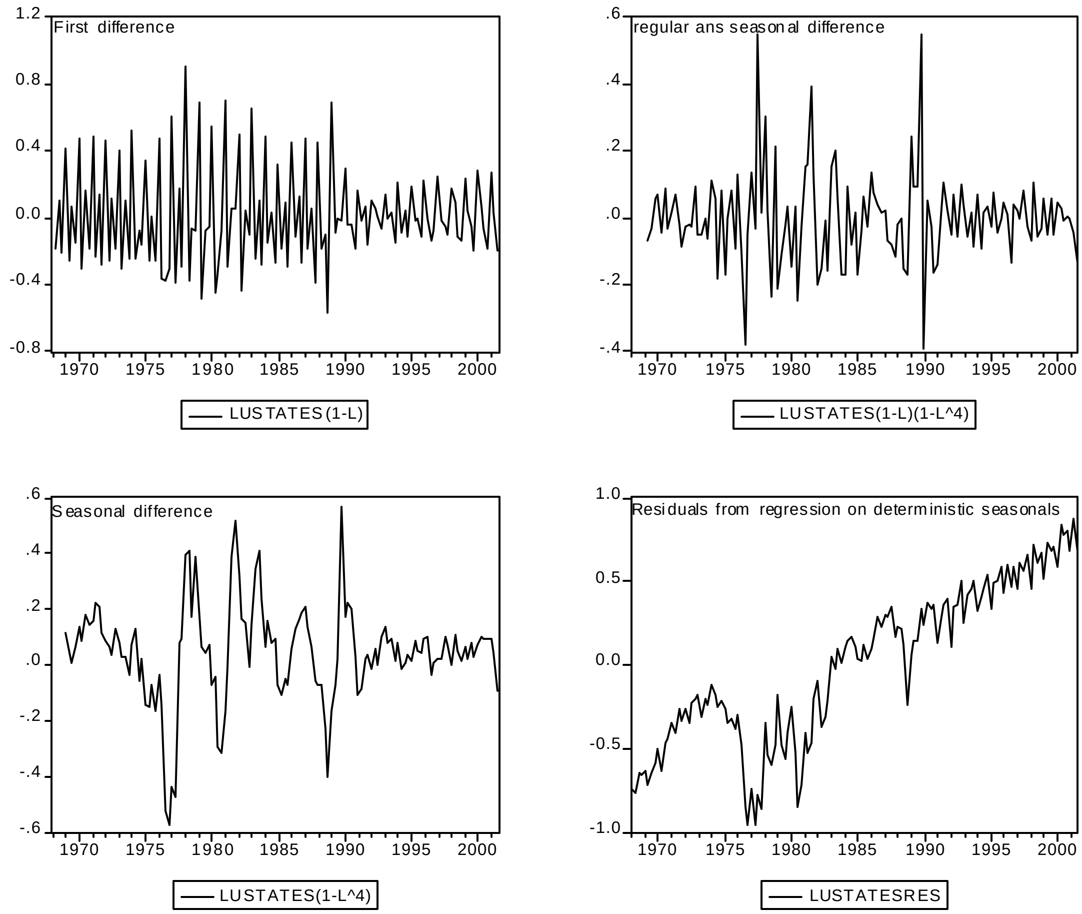


Figure 5: Log of Quarterly Arrivals From Canada:The Results of Applying Various Filters

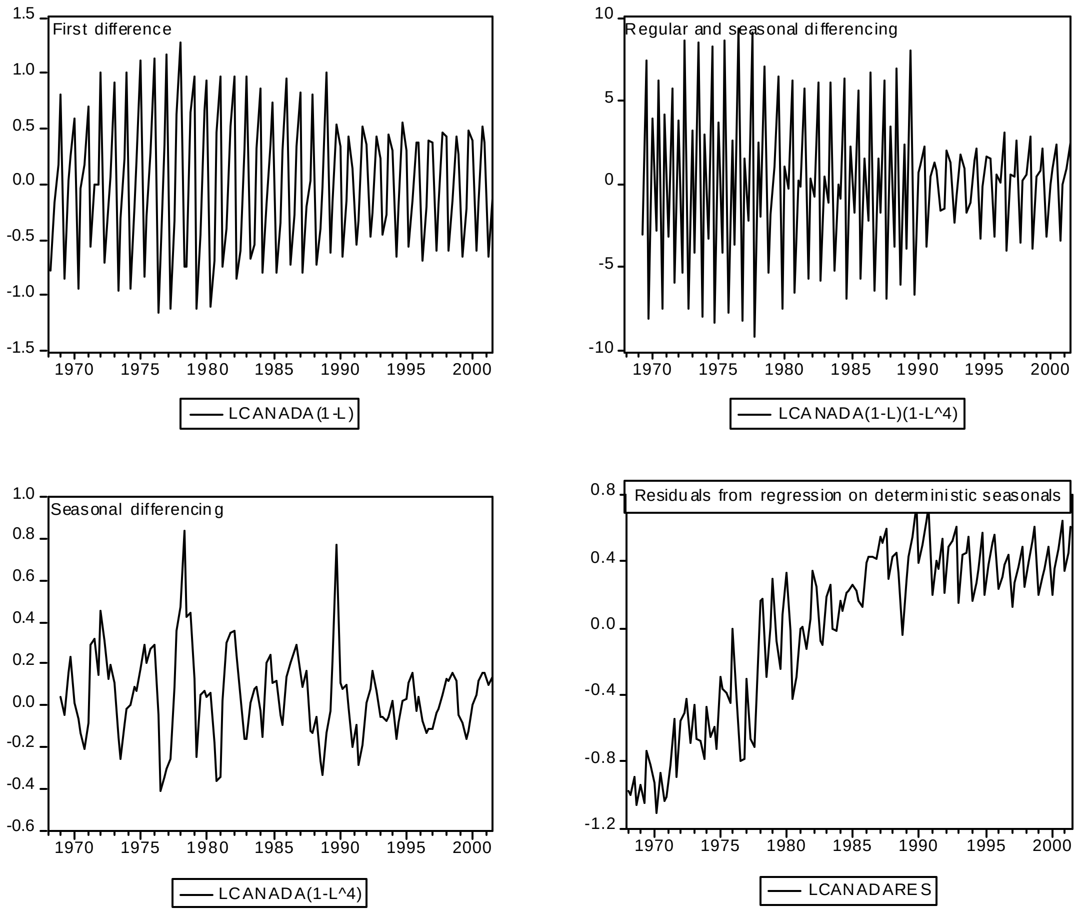


Figure 6 : Log of Quarterly Short Staying Tourists 1968:01-2001:03:The Results of Applying Various Filters

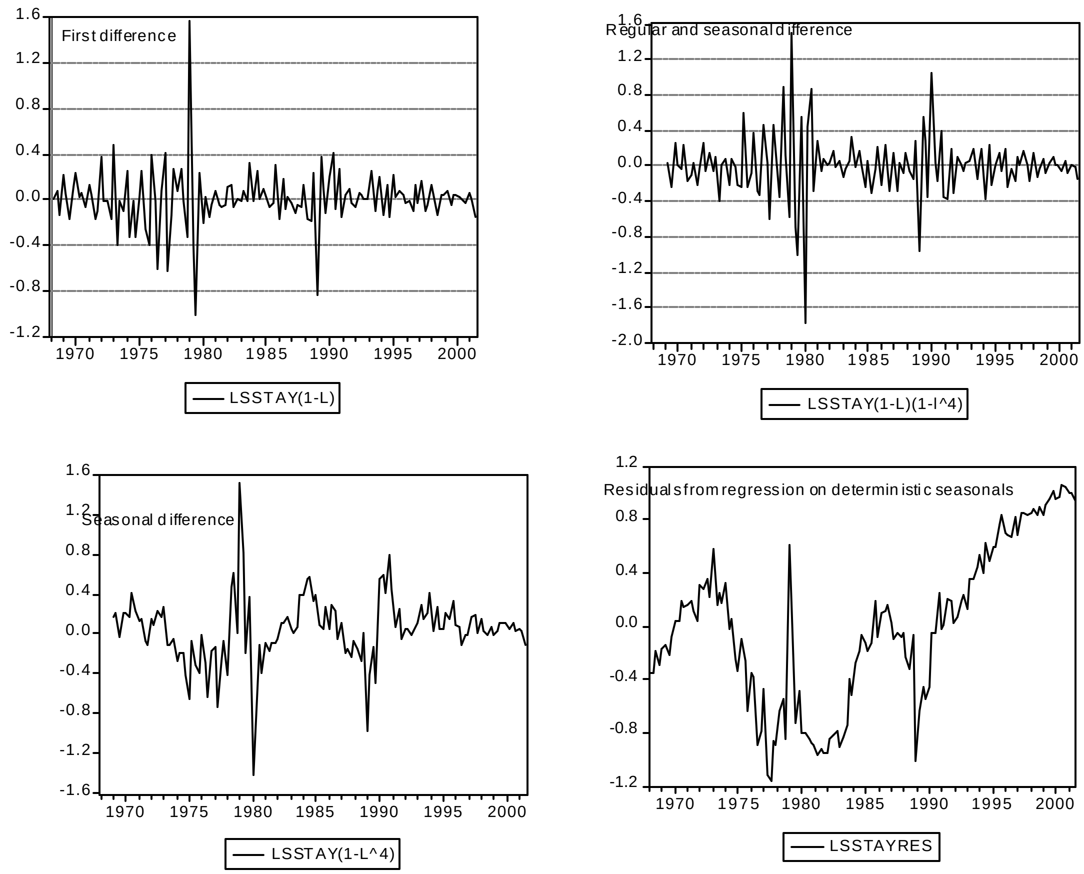
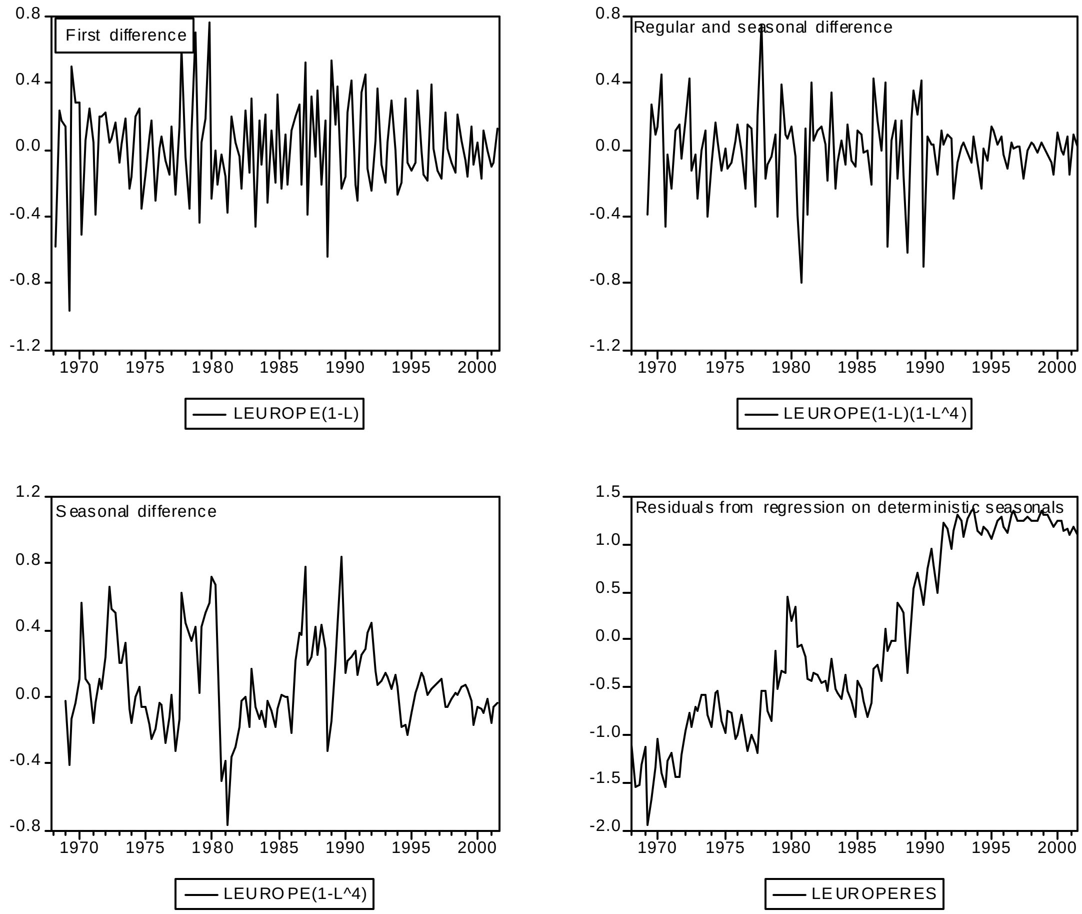


Figure 7: Log of Quarterly Tourist Arrivals from Europe 1968:01-2001:03:The Results of Applying Various Filters



Appendix II

Table 3:Forecasting Results for Total Visitors (LTVISITS)								
	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value
Variables	(1-L)		(1-L)		$(1-L^4)$		$(1-L)(1-L^4)$	
Constant			0.01	3.39	0.05	3.45		
AR(1)	-0.93	-9.68	-0.30	-4.30	0.68	9.87		
AR(2)	-0.61	-6.15	-0.30	-4.33				
AR(3)	-0.63	-8.33	-0.29	-4.21				
AR(4)			0.66	9.29				
MA(1)	0.60	5.56					-0.21	-2.29
SMA(1)			-0.46	-5.23	-0.46	-5.57	-0.52	-6.58
DUM(1)	0.36	8.21						
DUM(2)	-0.30	-6.93						
AIC,SBC	32.5,49.2		18.0,34.6		15.3,23.6		31.7,40.1	
Q(8)	12.8		3.7		3.7		9.17	
Forecasting Horizon								
	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U
1	0.051	0.32	0.056	0.35	0.052	0.33	0.055	0.35
2	0.052	0.27	0.056	0.29	0.053	0.27	0.067	0.35
3	0.055	0.36	0.060	0.39	0.059	0.38	0.092	0.59
4	0.057	0.65	0.062	0.70	0.061	0.69	0.114	1.30
5	0.057	0.30	0.083	0.44	0.075	0.40	0.168	0.90
6	0.073	0.34	0.069	0.32	0.058	0.27	0.227	1.06
7	0.088	0.50	0.062	0.35	0.055	0.31	0.336	1.92
8	0.075	0.63	0.036	0.30	0.035	0.29	0.427	3.61

AR(1) to AR(4) =Autoregressive terms of orders 1 to 4 respectively. MA(1) = Moving average of order one. SMA(1) = Seasonal moving average of order one. DUM(1) to DUM(3)= Dummy variables for quarters 1 to 3. AIC and SBC are Akaike and Schwartz Bayesian criterion respectively. Q(8) is the Q -statistic for serial correlation.

Table 4:Forecasting Results for US Arrivals (LUATATES)								
	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value
Variables	(1-L)		(1-L)		$(1-L^4)$		$(1-L)(1-L^4)$	
Constant	-0.17	-1.91						
Ar(1)			-0.12	-1.30	-0.19	-2.22		
AR(2)					-0.17	-2.03		
AR(3)					-0.18	-2.13		
AR(4)					0.40	4.25		
SAR(1)	0.84	9.52	0.96	41.90	0.96	74.57		
MA(1)							-0.14	-1.53
SMA(1)	-0.38	-2.78	-0.51	-5.95	-0.94	-17.5	-0.53	-6.83
DUM(1)	0.55	4.21						
DUM(2)	0.01	0.15						
DUM(3)	0.13	1.04						
AIC,SBC	76.6,93.3		71.4,79.6		68.8,85.3			75.6,84.0
Q(8)	7.38		6.64		8.18			6.68
Forecasting Horizon								
	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U
1	0.054	.31	0.059	0.34	0.065	.37	0.053	0.30
2	0.062	.29	0.068	0.32	0.073	.34	0.065	0.30
3	0.063	.38	0.063	0.37	0.062	.37	0.066	0.39
4	0.068	.85	0.069	0.85	0.064	.80	0.079	0.98
5	0.101	.49	0.113	0.55	0.143	.69	0.120	0.58
6	0.108	.44	0.124	0.50	0.189	.77	0.149	0.61
7	0.089	.44	0.117	0.58	0.161	.81	0.181	0.90
8	0.110	.88	0.110	0.88	0.104	.97	0.195	1.56

AR(1) to AR(4)= Autoregressive terms of orders 1 to 4 respectively. MA(1) = Moving average of order one. SMA(1) =Seasonal moving average of order 1. DUM(1) to DUM(3)= Dummy variables for quarters 1 to 3. AIC and SBC are Akaike and Schwartz Bayesian criterion respectively. Q(8) is the Q-statistic for serial correlation.

Table 5:Forecasting Results for Long-staying Visitors (LLSTAY)								
	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value
Variables	(1-L)		(1-L)		$(1-L^4)$		$(1-L)(1-L^4)$	
Constant			0.01	2.44	0.04	2.50		
Ar(1)	-0.13	-1.4	-0.29	-4.10	0.69	10.25		
AR(2)			-0.28	-4.15				
AR(3)			-0.29	-4.23				
AR(4)			0.66	9.24				
SAR(1)	0.88	13.2						
MA(1)							-0.17	-1.84
SMA(1)	-0.42	-3.4	-0.45	-4.93	-0.46	-5.49	-0.50	-6.25
DUM(1)	0.32	2.6						
DUM(2)	-0.18	-1.5						
DUM(3)	-0.03	-0.26						
AIC,SBC	74.2,90.8		60.7,77.3		57.88,66.2		74.9,83.3	
Q(8)	7.77		2.72				7.51	
Forecasting Horizon								
	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U
1	0.050	0.38	0.051	0.39	0.048	0.36	0.051	0.38
2	0.058	0.41	0.057	0.40	0.054	0.38	0.062	0.44
3	0.058	0.46	0.056	0.44	0.054	0.43	0.064	0.51
4	0.063	1.06	0.057	0.97	0.057	0.95	0.073	1.23
5	0.083	0.58	0.073	0.51	0.068	0.47	0.122	0.85
6	0.093	0.60	0.080	0.51	0.074	0.47	0.171	1.09
7	0.068	0.50	0.070	0.51	0.067	0.49	0.201	1.49
8	0.094	1.32	0.094	1.32	0.094	1.32	0.070	0.98

AR(1) to AR(4) Autoregressive terms of orders 1 to 4 respectively. MA(1) = Moving average of order one. SAR(1)= seasonal autoregressive model of order one. SMA(1) –=Seasonal moving average of order one. DUM(1) to DUM(3)= Dummy variables for quarters 1 to 3. AIC and SBC are Akaike and Schwartz Bayesian criterion respectively. Q(8) is the Q-statistic for serial correlation.

Table 6:Forecasting Results for CANADIAN Arrivals (LCANADA)								
	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value
Variables	(1-L)		(1-L)		$(1-L^4)$		$(1-L)(1-L^4)$	
Constant					0.04	2.51		
Ar(1)	-0.34	-4.64	-0.30	-4.34	0.72	8.02		
AR(2)	-0.38	-4.98	-0.32	-4.52	-0.17	-1.56		
AR(3)	-0.34	-4.56	-0.29	-4.23	0.14	1.31		
AR(4)	0.60	7.85	0.67	9.44	-0.33	-3.12		
SMA(1)	-0.42	-4.63	-0.40	-0.40	-0.15	-1.21	-0.11	-1.29
DUM(1)	-0.27	-1.15					-0.44	-5.40
DUM(2)	0.31	1.35						
AIC,SBC	125,144		125,139		109,126		141,146	
Q(8)	14.5		14.5		4.60		16.7	
Forecasting Horizon								
	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U
1	0.061	0.14	0.061	0.13	0.054	0.12	0.062	0.14
2	0.083	0.13	0.088	0.14	0.065	0.10	0.097	0.15
3	0.094	0.20	0.106	0.22	0.069	0.15	0.139	0.30
4	0.078	0.68	0.110	0.95	0.054	0.46	0.171	1.48
5	0.112	0.25	0.141	0.31	0.075	0.17	0.254	0.57
6	0.135	0.22	0.166	0.27	0.079	0.12	0.358	0.58
7	0.153	0.32	0.183	0.39	0.076	0.16	0.488	1.03
8	0.118	0.71	0.173	1.04	0.051	0.30	0.569	3.40

AR(1) to AR(4)= Autoregressive terms of orders 1 to 4 respectively. MA(1) = Moving average of order one. SMA(1) = Seasonal moving average of order one. DUM(1) to DUM(3)= Dummy variables for quarters 1 to 3. AIC and SBC are Akaike and Schwartz Bayesian criterion respectively. Q(8) is the Q-statistic for serial correlation.

Table 7:Forecasting Results for European Arrivals (LEUROPE)								
	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value
Variables	(1-L)		(1-L)		$(1-L^4)$		$(1-L)(1-L^4)$	
Constant								
Ar(1)	-0.21	-2.23	-0.25	-3.10	-0.26	-3.06		
AR(2)	-0.17	-1.88	-0.22	-2.70	-0.23	-2.65		
AR(3)			-0.23	-2.77	-0.23	-2.69		
AR(4)			0.44	5.15	0.40	3.11		
SAR(1)	0.51	6.53						
MA(1)							-0.28	-3.22
SMA(1)			0.14	1.45	0.07	0.56	-0.31	-3.51
SMA(2)					0.19	1.74		
DUM(1)	-0.006	-0.06						
DUM(2)	0.26	3.12						
DUMMY	-0.072	-1.64						
AIC,SBC	179.0,195.5		190.5,204.4		192.2,208.9		212.3,217.9	
Q(8)	8.29		0.76		1.18		4.84	
Forecasting Horizon								
	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U
1	0.060	0.56	0.064	0.59	0.065	0.60	0.072	0.66
2	0.066	0.59	0.068	0.61	0.067	0.60	0.100	0.90
3	0.073	0.68	0.078	0.73	0.077	0.71	0.124	1.15
4	0.085	0.88	0.093	0.96	0.092	0.95	0.164	1.70
5	0.094	0.65	0.106	0.73	0.104	0.72	0.266	1.84
6	0.130	0.76	0.133	0.78	0.128	0.75	0.455	2.66
7	0.146	0.94	0.150	0.97	0.140	0.90	0.709	4.59
8	0.162	0.97	0.174	1.04	0.159	0.95	1.081	6.47

AR(1) to AR(4)= Autoregressive terms of orders 1 to 4 respectively. MA(1) = Moving average of order one. SMA(1) to SMA(2) = Seasonal moving average of order two. DUM(1) to DUM(3)= Dummy variables for quarters 1 to 3. Dummy is an indicator variable with a mean shift at 1981:1. AIC and SBC are Akaike and Schwartz Bayesian criterion respectively. Q(8) is the Q-statistic for serial correlation.

Table 8:Forecasting Results for European Arrivals (LEUROPE)								
	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value	LTVISITS	t-value
Variables	(1-L)		(1-L)		$(1-L^4)$		$(1-L)(1-L^4)$	
Constant								
Ar(1)	-0.43	-4.96	-0.40	-4.33				
AR(2)	-0.35	-3.98	-0.36	-3.47				
AR(3)			0.08	0.80				
MA(1)					0.64	7.91	-0.38	-4.48
MA(2)					0.42	4.44		
MA(3)					0.60	6.79		
SMA(1)			0.18	1.70	-0.41	-3.92	-0.87	-18.93
DUM(1)	0.14	3.31						
DUM(2)	-0.09	-2.25						
PULSE	-0.50	-2.42	-0.53	-2.25				
DUMMY					0.10	2.38		
AIC,SBC	227.5,241.5		232.7,246.6		235,249		237,242	
Q(8)	8.59		8.69		11.8		16.17	
Forecasting Horizon								
	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U	RMSE	Theil's U
1	0.062	0.99	0.059	0.93	0.059	0.94	0.048	0.76
2	0.070	0.91	0.070	0.91	0.060	0.78	0.064	0.84
3	0.072	1.12	0.069	1.06	0.063	0.97	0.080	1.24
4	0.066	0.75	0.093	1.07	0.067	0.76	0.113	1.29
5	0.071	0.65	0.108	1.00	0.086	0.80	0.114	1.06
6	0.084	0.68	0.109	0.88	0.067	0.54	0.196	1.59
7	0.087	0.71	0.106	0.87	0.074	0.60	0.182	1.50
8	0.050	0.41	0.105	0.87	0.080	0.67	0.219	1.82

AR(1) to AR(4)= Autoregressive terms of orders 1 to 4 respectively. MA(1) to MA(3) = Moving average of order three. SMA(1) = Seasonal moving average of order one. DUM(1) to DUM(3)= Dummy variables for quarters 1 to 3. Dummy is an indicator variable with a mean shift at 1981:1. PULSE is a dummy variable with 1 in at 1981:1 and zero elsewhere. AIC and SBC are Akaike and Schwartz Bayesian criterion respectively. Q(8) is the Q-statistic for serial correlation.