Model Selection, Forecasting and Monthly Seasonality of Hotel Nights in Denmark

Nils Karl Sørensen

University of Southern Denmark, Grundtvigs allé 150, DK-6400 Sønderborg, Denmark. **F** *(+45) 65 50 12 28,* **J** *nks@sam.sdu.dk*

Abstract

Foreigners' demand for hotel nights in Denmark by nationality are examined using monthly time series covering 30 years, and divided into 11 nationalities. Special attention is given to the role of seasonality.

Three univariate seasonal presentations of non-stationary data with different characteristics are considered, a stochastic, deterministic, and an error correction mechanism (ECM) approach taking into account economic as well as climatic variables.

Based on a presentation of different measures to evaluate the forecasting performance a model selection is undertaken. It is found that the single variable presentations in most cases are superior to the ECM. On the other hand the ECM presentation provides a more detailed description of the evolution of inbound tourism. In many cases it is found that the climatic indicators have significant influence on tourism. With regard to the single variable models it is found that seasonality in general is of stochastic nature, but the deterministic presentation is in many cases superior in forecasting performance.

Theme: *Tourism, regional econometric modelling*
Key words: *Tourism, seasonality, forecasting, climatic variables.*
JEL Classification: *C32.*

1. Introduction

The growth of the world-wide tourism industry over the past few decades has generated great interest in tourism demand modelling and forecasting. Contributions by Witt and Witt (1992,1995) and Song and Witt (2000) have shown that no single forecasting method performs consistently best across different situations.

This result is in accordance with econometric methodology. It is advocated that the »classical textbook« method of econometrics, i.e. the direct estimation of a theoretical model, and then diagnostic checking its validity by testing for autocorrelation, heteroskedasticity etc. is fundamentally wrong, and give no description of what is going on in the real world, see e.g. Clements and Hendry (1999). Further, in order to make the model operational the applied statistic material often impose severe limitations on the models ability to verify a given theory.

Instead of pretending that model building takes place in this way it should be accepted that empirical models tries to describe a data generating process (DGP). It is not possible ever to find the true DGP, but one can hope to obtain a better approximation than provided by a theoretical model.

The obvious presence of seasonality in tourism with peaks during the holiday season clearly makes it difficult for econometricians to estimate and evaluate models. This problem is further complicated by the fact that peoples preferences for holidays have changed. Today people are more inclined to separate their holidays into several sub-periods. Using monthly observations, Sørensen (1999,2001) found that a varying and changing seasonal component are a common phenomenon in many time series for hotel nights in Denmark.

The purpose of the present paper is to focus on three univariate forecasting representations for monthly time series. Statistics on hotel nights in Denmark are divided into 11 nationalities covering more than 90 percent of total hotel nights. Models are estimated for the period from 1970.1 to 1997.12, whereas the period from 1998.1 to 2000.12 has been used for evaluating forecast performance. Two representations attempt to model DGP by use of statistics for hotel nights only, whereas the third representation models tourism demand by use of an error correction mechanism (ECM) model thereby providing insightful information on the short and long run elasticities of tourism demand.

The first presentation considered assumes that seasonality is governed by stochastic trends at the seasonal frequencies, and was introduced by Hylleberg, Engle, Granger and Yoo (1990) (hereafter HEGY). The second model assumes approximate deterministic seasonality, which is termed the »seasonal cycle«. This approach was introduced by Barsky and Miron (1989). The final model is a traditional demand set up, where hotel nights are assumed to be a function of economic as well as climatic variables. Although obvious to examine the impact of a weather index in relation to tourism, this approach has only been considered earlier by Barry and O'Hagan (1972).

2. Empirical Observations

A data bank has been set up by use of material from the regular publications on hotel nights supplied by Statistics Denmark. Table 1 give some summary statistics.

In 2000 the total number of hotel nights in Denmark amounted to about 13.3 million, equivalent to an increase of about 86% since 1970. With regard to nationality, half the hotel nights are attributable to Danes, and half to foreign nationals. The division of hotel nights between Danes and foreigners has remained relatively constant up to the mid-1990s. Hereafter the number of foreign tourists seems to have stagnated. The number of hotel nights has been affected by external events such as the slump in the mid-1970s. From the mid-1980s the pattern of hotel nights consumed by foreigners have become more volatile, perhaps due to variations in taste or fluctuations in the exchange rate for some important foreign visitors.

The most important foreign consumers of hotel nights in Denmark are Germans, Swedes and Norwegians; in other words people from the neighbouring countries. Tourists from these three nations account for more than half the foreign hotel nights. Overall, the number of hotel nights consumed by foreigners has increased by 75 percent over the period whereas it has doubled for Danes.

Overall, the coefficient of variation for foreigners exceeds that for Danes for the full period as well as for the sub-periods. This is not surprising. Foreigners' demand for hotel nights is high during the holiday season, and is also affected by their domestic economy, travels distance etc., whereas Danes can easily visit another part of the country throughout the year. For Danes and Foreigners' the coefficient of variation has remained fairly constant within the sub-periods.

Table 1. Summary statistics on hotel nights in by nationality. 1970.1 to 2000.12.

	Share	Growth 2000	Nights 2000	Coefficient of variation, CV			
Nationality:	Average Percent	Index	Relative to home	1970.1 - 1979.12	1980.1 - 1989.12	1990.1 - 2000.12	1970.1 - 2000.12
Total	100.0	186		0.49	0.46	0.44	0.51
Danes	49.6	197	1.132	0.25	0.24	0.27	0.36
Foreigners, total	50.4	175		0.72	0.71	0.67	0.73
Swedes	11.1	250	0.143	0.65	0.65	0.65	0.74
Norwegians	6.3	479	0.197	0.91	1.19	1.15	1.30
Finns	1.0	688	0.016	0.59	0.79	0.75	0.74
Germans	12.8	158	0.011	1.10	0.96	0.87	0.98
English	3.3	192	0.006	0.42	0.50	0.43	0.48
Dutch	1.5	203	0.012	0.83	0.68	0.88	0.92
French	0.9	93	0.001	0.71	0.57	0.46	0.59
Italians	1.1	201	0.002	0.62	0.94	0.98	0.99
Americans (USA)	4.7	41	0.001	0.86	0.75	0.70	0.87
Japanese	1.0	205	0.001	0.46	0.52	0.65	0.59

Note: The percentage shares are computed on the mean values for the full period. The computations in the second and third columns are on annual data additively aggregated from the monthly observations. In the second column the growth index is with 1970 set equal to 100. In the third column the figures measure the number of hotel nights relative to the total population in the home country. In the four last columns CV is the coefficient of variation computed using the non-transformed data and defined as: CV = standard deviation divided by the mean.

Divided by nationality for the full period, the highest coefficient of variation is found for Norwegians, Germans and Italians. Considering the sub-periods, the coefficient of variation has increased in four cases, i.e. Norwegians, Finns, Italians and Japanese, remained constant in three cases, i.e. Swedes, English, Dutch, and decreased in three cases, ie. Germans, French and Americans. An increase could be taken as an indicator of a more significant season - i.e. that people only want to visit Denmark during the summer, at Easter, and so on. This is the case for Norwegians and Finns, but also for visitors from a number of countries further away, including Japan and Italy. As for German, French and American visitors, the season seems to have become longer and more diversified.

3. Stochastic Seasonality

A time series model with seasonal unit roots is an approximation that allows for changes in the seasonal pattern. A test for seasonal unit roots is developed for the quarterly case by HEGY, and extended by Franses (1991) to the monthly case.

For a given time series variable x_t a univariate model integrated at all the seasonal frequencies as well as the long-run frequency is $(1\text{-}B^{12})x_t = \varepsilon_t$ - $iid(0,\sigma^2)$, $t = 1,2,..,T$, where B is the lag operator defined as $B^n x_t = x_{t-n}$. It can be proved that an AR(p) process of the form $\varphi(B)x_t = \varepsilon_t$, using a proposition given by HEGY defining the form of $\varphi(B)$, in the monthly case can be written as

$$
\begin{aligned}
(1)\qquad (1\text{-}B^{12})x_t = y_{8t} &= \pi_1 y_{1t-1} + \pi_2 y_{2t-1} + \pi_3 y_{3t-1} + \pi_4 y_{3t-2} \\
&+ \pi_5 y_{4t-1} + \pi_6 y_{4t-2} + \pi_7 y_{5t-1} + \pi_8 y_{5t-2} \\
&+ \pi_9 y_{6t-1} + \pi_{10} y_{6t-2} + \pi_{11} y_{7t-1} + \pi_{12} y_{7t-2} + \varphi^{\Delta}(B) y_{8t} + \mu_t + \varepsilon_t
\end{aligned}
$$

In this equation x_t is a linear combination of the variables y_i, $i=1,...,8$. These variables are all various transformations of x_t. In all there are 12 coefficients π_j, $j=1,...,12$ to be estimated by applying OLS to equation (1). The coefficients of the π's correspond to the 12 solutions of the equation $(1\text{-}B^{12})=0$ all lying on the unit cycle.

The test for monthly seasonal unit roots can then be performed as a test of whether or not the coefficients are lying on the unit cycle. There will be *no* seasonal unit root if the coefficient to a given π is significantly different from zero. If the coefficient is insignificant, a unit root is present showing a changing seasonal component. A survey of the data transformations and test hypothesis is given in Table 2.

Table 2. Testing for seasonal unit roots in monthly data.

No.	Frequency			Transformation	H_0: Unit Root	H_1: No Unit Root
0	0		long-run	$y_{1t} = (1+B+B^2+..+B^{11})x_t$	$\pi_1 = 0$	$\pi_1 < 0$
1	6/12		semi-annual	$y_{2t} = -(1-B)(1+B^2)(1+B^4+B^8)x_t$	$\pi_2 = 0$	$\pi_2 < 0$
2	3/12	(9/12)	quarterly	$y_{3t} = -(1-B^2)(1+B^4+B^8)x_t$	π_3 1 π_4 = 0	π_3 C π_4 ... 0
3	5/12	(7/12)	monthly	$y_{4t} = -(1-B^4)(1\,!\,3\%B+B^2)(1+B^2+B^4)x_t$	π_5 1 π_6 = 0	π_5 C π_6 ... 0
4	1/12	(11/12)	monthly	$y_{5t} = -(1-B^4)(1+3\%B+B^2)(1+B^2+B^4)x_t$	π_7 1 π_8 = 0	π_7 C π_8 ... 0
5	4/12	(8/12)	monthly	$y_{6t} = -(1-B^4)(1\,!\,B+B^2)(1\,!\,B^2+B^4)x_t$	π_9 1 π_{10} = 0	π_9 C π_{10}... 0
6	2/12	(10/12)	monthly	$y_{7t} = -(1-B^4)(1+B+B^2)(1\,!\,B^2+B^4)x_t$	π_{11} 1 π_{12} = 0	π_{11} C π_{12}... 0

Note: Critical values can be obtained from P.H. Franses and B. Hobijn, 'Critical values for unit root tests in seasonal time series', *Journal of Applied Statistics*, Vol 24, 1997, pp 25-47.

The test for seasonal unit roots at the relevant seasonal frequencies can then be performed either as t-tests for the estimates of $\pi_3,...,\pi_{12}$ or as joint F-tests for sets of parameters, i.e. π_3 and π_4 etc. All critical values of the test have a non-standard distribution and have to be simulated by Monte Carlo experiments.

Two more points concerning equation (1) should be made. First, μ_t is the included

Table 3. HEGY-tests for monthly seasonal unit roots in hotel nights in Denmark by nationality. 1970.1 to 1997.12.

	Total	Danes	For-eigners	Swedes	Norwegians	Finns	Germans	English	Dutch	French	Italians	Americans	Japanese
t-statistic:													
t_1: π_1	!2.26	!2.72	!1.84	!1.94	!1.21	!2.27	!1.54	!2.19	!1.21	!2.17	!2.31	!2.96	!4.42*
t_2: π_2	!4.08*	!4.42*	!4.23*	!5.52*	!2.76	!3.03*	!3.19*	!2.74	!1.73	!4.59*	!1.74	!4.54*	!2.91
t_3: π_3	1.33	1.99	1.06	1.01	1.67	!0.51	2.23	0.23	!0.01	!0.64	2.26	!0.14	0.05
t_4: π_4	0.01	2.06	!0.79	!0.38	1.82	0.38	!1.92	!0.24	!0.38	!1.55	0.43	!0.49	!1.78
t_5: π_5	!4.01*	!5.99*	!4.59*	!4.96*	!4.14*	!5.43*	!5.54*	!4.94*	!2.45	!4.39*	!4.48*	!6.77*	!3.91*
t_6: π_6	!4.89*	!5.39*	!5.14*	!4.49*	!5.07*	!5.87*	!5.20*	!5.31*	!2.83	!4.17*	!4.78*	!6.87*	!3.99*
t_7: π_7	2.15	3.33	2.38	2.66	3.07	3.40	3.42	2.26	3.64	2.61	1.74	1.54	2.76
t_8: π_8	!1.61	!2.39	!2.26	!2.55	!2.35	!2.62	!3.30*	!0.71	!2.77	!2.68	!1.76	!1.49	!1.71
t_9: π_9	!2.17	!2.42	!3.24	!2.47	!3.78*	!2.36	!2.69*	!3.30*	!0.11	!4.39*	!3.39*	!3.02*	!-5.66*
t_{10}: π_{10}	!3.84*	!3.16	!5.59*	!3.76*	!4.14*	!3.59*	!5.99*	!5.79*	!2.04	!5.84*	!4.88*	!4.97*	!4.44*
t_{11}: π_{11}	!2.08*	0.72	3.01	0.29	4.74	1.21	3.80	0.25	1.54	2.03	2.15	1.95	0.35
t_{12}: π_{12}	!3.27*	!2.64	!4.51*	!2.41	!4.25*	!3.82*	!5.47*	!2.60	!3.93*	!6.47*	!2.08	!5.87*	!3.66*
F-test:													
F: π_3 1 π_4	0.92	3.33	0.96	0.65	2.35	0.23	4.53	0.08	0.07	1.31	2.57	0.12	1.65
F: π_5 1 π_6	12.13*	18.04*	13.22*	12.43*	12.95*	17.58*	15.77*	14.35*	4.01	9.95*	11.68*	25.17*	8.39*
F: π_7 1 π_8	2.37	5.76	2.94	3.73	4.73	5.86	6.09	3.99	6.74*	3.81	1.66	1.27	4.04
F: π_9 1 π_{10}	7.44*	5.45	15.75*	7.31*	11.02*	6.71*	17.97*	16.82*	2.71	18.79*	12.42*	12.59*	17.90*
F: π_{11} 1 π_{12}	6.46*	3.73	10.54*	7.34*	14.13*	7.54*	16.07*	4.19	7.98*	22.67*	3.01	17.95*	8.39*
R^2	0.41	0.50	0.43	0.48	0.41	0.38	0.39	0.51	0.36	0.46	0.33	0.59	0.60
DW	2.02	2.00	2.03	1.96	2.12	2.06	2.05	2.00	2.03	1.97	2.00	1.93	1.99
σ	0.039	0.036	0.073	0.082	0.119	0.156	0.229	0.085	0.119	0.118	0.215	0.122	0.128
Augmentation	2,11, 12,14	1,2,7,10	2,12,23	2,4	2,3,12	2,5	11	1,3	1,4,12, 13,17	6	9,12,21	7	1,3,14, 24

Note: Auxiliary regression including a constant term, trend, and 11 seasonal dummies. A star indicates that the unit root hypothesis is rejected at the 5 percent level. Critical values are taken from Franses and Hobijn (1997). σ is the standard error of estimate. DW is the Durbin-Watson statistic.

deterministic component such as a constant, a trend, and seasonal dummy variables. Second, equation (1) has to be augmented by lagged y_{8t}. This is done in order to make the residuals white noise, and leaves the asymptotic distribution unaffected. The power is negatively affected if too many nuisance parameters are used in the augmentation. In the present case, a strategy to determine $\varphi^{A}(B)$ has been to start with 24 lags, and then test down the significant augmented variables.

Results of the applying the auxiliary regression (1) are displayed in Table 3. It is evident that a varying and changing seasonal pattern is a common phenomena. The seasonal pattern has been most varying for Danes, Norwegians, Englishmen, Dutch and Italians. The most stable pattern is observed for Swedes, Finns, Germans, Frenchmen, Americans, and Japanese using hotels in Denmark.

4. Deterministic Seasonality

Barsky and Miron (1989) analyses the economics of the seasonal cycle by considering a deterministic seasonal model of the form

$$(2) \qquad (1-B^{12})x_t = y_{8t} = \psi_1 D_{1,t} + \psi_2 D_{2,t} + ... + \psi_{11} D_{11,t} + \psi_{12} D_{12,t} + \sum_{n=1}^{m} a_n y_{8t} + \varepsilon_t$$

where ε_t *-nid*$(0,\sigma^2)$, $t = 1,...,T$, and $D_1,...,D_{12}$ are 12 seasonal dummy variables. In order to render the residuals white noise augmented values of y_{8t} is included. Since the ψ_s parameters are assumed fixed over the sample seasonality is of a constant non-changing nature. The interpretation of coefficients of the dummy variables is the elasticity of change from the month in the previous year.

Results from estimation of (2) by OLS by nationality are given in Table 4. As before 24 lags of y_{8t} were used. Overall the results are rather poor. For Finns, Englishmen, Frenchmen, Italians, Americans and Japanese not even one single significant coefficient is observed. Consequently, these formulations are fully determined by the lagged endogenous variables.

With regard to the diagnostics the deterministic seasonal model also performs more poorly than the HEGY model. Overall it is concluded that the DGP of seasonality is stochastic rather than deterministic.

5. A Seasonal ECM-model of Tourism Demand

A more traditional way of describing the DGP of hotel nights is to estimate a monthly time series demand model of the form $HN = f(TP,W,Z)$, where HN is the number of hotel nights for a given nationality, TP is tourist prices, W is a weather index, and Z is a vector capturing omitted variables like e.g. the income level which is not available at the monthly frequency.

Table 4. Estimation of seasonal monthly deterministic model of hotel nights in Denmark by nationality. 1970.1 to 1997.12.

	Total	Danes	For-eigners	Swedes	Norwegians	Finns	Germans	English	Dutch	French	Italians	Americans	Japanese
Coefficient:													
January ψ_1	.012	.005	.019	.001	.013	! .009	.032	.010	.024	.003	.025	! .015	! .004
Feb. ψ_2	.019**	.011	.020	.010	.027	! .009	.029	.020	.018	.024	.042	! .013	.018
March ψ_3	.020**	.012	.027*	.012	.029	! .009	.027	.004	.013	.010	! .001	! .018	.003
April ψ_4	.014*	.009	.021	.009	.045*	.001	.036	! .001	.018	! .003	! .019	! .032	! .011
May ψ_5	.017*	.012	.021	.023	.036	.002	.034	.006	.023	! .013	! .020	! .027	.014
June ψ_6	.012	.003	.019	.010	.072***	.002	.018	.008	! .019	! .001	! .004	! .023	.027
July ψ_7	.016*	.005	.027*	.016	.101***	.012	.007	.001	.006	! .020	.034	! .028	.026
August ψ_8	.013	.004	.017	! .006	.056**	.004	.013	.012	.046*	! .033	.034	! .029	! .009
Sept. ψ_9	.013	.014*	.005	.004	.020	.002	.014	.008	.001	.001	.019	! .039	! .002
October ψ_{10}	.027**	.018**	.035**	.029*	.039*	.021	.092**	.017	.036	.022	.018	! .035	! .001
Novem. ψ_{11}	.021**	.012	.027*	.010	.035	.005	.004	.025	.018	.022	.022	! .008	.009
Decem. ψ_{12}	.012	! .004	.031*	! .009	.018	.001	.075*	.018	.012	.027	.063	! .006	! .002
R^2	0.25	0.36	0.32	0.35	0.35	0.19	0.33	0.37	0.25	0.32	0.21	0.44	0.47
DW	1.97	2.08	1.86	2.13	1.75	2.04	1.92	2.09	2.03	2.02	2.00	2.12	2.06
σ	0.044	0.040	0.078	0.090	0.123	0.176	0.239	0.094	0.126	0.131	0.230	0.141	0.145
Augmentation	1,2, 12	1,3,10, 12,13	1,2,4, 11,12, 24	1,2, 12,13	3,4,9, 11,12, 23,24	1,2,5, 12,13	3,10,11, 12,15, 23,24	1, 12,13, 24	1,4,12, 13,17, 20,24	1,3, 12,14, 24	1,4,12, 13,16,24	1, 12,13, 24	1, 12,13

*Note: A *** indicates that the variable is significant at the 1 percent level, a ** at the 5 percent level, and a * at the 10 percent level. DW is the Durbin-Watson statistic. Information with regard to the augmentation of the estimates is explained in the text. T-statistics are omitted due to lack of space. σ is the standard error of estimate.*

As proxies for tourist prices *TP,* currency exchange statistics and consumer price (CPI) statistics for the home land and the host land are used. The price level in substitute destinations is not considered. The Danish CPI denoted *PDK* is taken to be a proxy for the cost of tourism. The foreign CPI denoted *PFO* is taken to be a proxy for cost of living at home. The exchange rate denoted *EXC* is included as suggested by Sørensen (1999,2001).

A high Danish price level is expected to have a negative impact on the number of hotel nights, whereas a high foreign price level is supposed to have the opposite effect. A low exchange rate is also expected to increase the number of hotel nights.

As proxy for the weather index *W* four variables enter the demand function unweighted. *TEMP* is the monthly mean temperature, *SUN* is the monthly number of bright sunshine hours, *CLOUD* is the monthly average cloud coverage, and *RAIN* is the monthly relative humidity. High temperature and many sunshine hours are expected to increase the number of hotel nights, whereas cloudy and rainy weather is expected to decrease the number of hotel nights. For these variables multicollinearity problems may appear between *TEMP* and *SUN*, and also between *CLOUD* and *RAIN*. The data sources are for the climatic variables Statistics Denmark and for the exchange rate and prices the Danish National Bank.

Over the long run, i.e. in a steady state the relation between hotel nights and for example the exchange rate is constant on any given growth path. This is the golden rule of growth known from growth theory, and should also be expected to be valid with regard to the other variables included in this model formulation. In the short run, however, fluctuations around the growth path may exist. The error correction mechanism (ECM) captures these problems. Based on the long run relation

$$(3) \quad logHN_t = \alpha_0 + \alpha_1 logPDK_t + \alpha_2 logPFO_t + \alpha_3 logEXC_t + \alpha_4 logTEMP_t + \alpha_5 logSUN_t + \alpha_6 logCLOUD_t + \alpha_7 logRAIN_t$$

an ECM-model, written with a one year lag for a given foreign nationality takes the form

$$(4) \quad \Delta logHN_t = \beta_0 + \beta_1 \Delta logPDK_t + \beta_2 \Delta logPFO_t + \beta_3 \Delta logEXC_t + \beta_4 \Delta logTEMP_t + \beta_5 \Delta logSUN_t + \beta_6 \Delta logCLOUD_t + \beta_7 \Delta logRAIN_t - \beta_8 (logHN_{t-12} - \beta_9 logPDK_{t-12} - \beta_{10} logPFO_{t-12} - \beta_{11} logEXC_{t-12} - \beta_{12} logTEMP_{t-12} - \beta_{13} logSUN_{t-12} - \beta_{14} logCLOUD_{t-12} - \beta_{15} logRAIN_{t-12}) + \sum_{i=1}^{n} a_i \Delta logX_{i,t-i} + \varepsilon_t$$

where $\varepsilon_t \sim nid(0,\sigma^2)$, $t = 1,...,T,$ and Δ is the 12th difference. In order to render the residuals white noise the variables are augmented with exogenous as well as endogenous variables.

In short, the coefficients can be given the following interpretation: β_0 is the constant term. β_1 through β_{15} are the impact effects, i.e. short run elasticities of hotel nights with regard e.g. the Danish price level β_1, etc. β_8 is the error correction term measuring the feedback effect, i.e. the speed of adjustment. Finally, β_9 through β_{15} are the long run responses, i.e. the long run elasticities. Notice that β_8 is not restricted to unity. For the long run analysis the β_9 through β_{15} ! coefficients should be used, whereas the remaining β ! coefficients should be used for the short run analysis.

Because β_8 is not restricted to any specific value, a two-step procedure needs to be undertaken in order to estimate all the coefficients in (4). First the long run coefficients are estimated. Second, by use of values of the error correction form the short run coefficients are estimated. Full model estimates by OLS on the elasticities in (4) are given in Table 5. To obtain these results a general to specific modelling strategy was conducted. Estimates for each nationality shows on the left side the short run coefficients (4), and on the rights side the long run elasticities (3). First, the long run relation was estimated, and then the long run solution was found. The final lag augmentation of the short run estimate is reported in the right side at the bottom. For example "H: 10,12" means that hotel nights lagged 10 and 12 periods (months) were included in the lag structure. *PK* is the Danish price level. *P* is the foreign price level etc. Because the long-run estimates are a result of the dynamic solution, it is not relevant to report the lag structure. With regard to the lag structure all variables were lagged with 24 lags, and then tested down.

In most of the cases the Danish price level has a negative influence on the demand for hotel nights. This reflects that Denmark is a costly, holiday nation. For Norway and the Netherlands the long run coefficient of the Danish price level is positive. This could reflect that the costs of a hotel night in these nations are even more expensive than in Denmark. Whereas the influence of the host country price level is positive as well as negative it seems that the exchange generally has a positive impact on foreigners demand for hotel nights. When it is cheap to buy the Danish currency the demand for hotel nights increases. This is especially the case for Americans.

The inclusion climatic variables clearly improve the results especially for the long run estimates. Generally, the signs of the significant variables are as expected. Some exceptions are although observed, e.g. for Americans an increase in the number of clouds will increase the demand for hotel nights! Here the influence of economic factors may have out weighted the influence of the climatic variables. Also factors not included like distance may influence.

The *ECM*-term is in general significant with Sweden, Germany, England and the Netherlands as an exception. Although several short run models have few significant coefficients, e.g. Sweden, Germany and the Netherlands, in all models there are several significant augmented variables. Especially lag 12 is frequently significant both with regard to the price level and with regard to the climatic variables. This could be a result of a "memory" behaviour. For example if the weather last summer was good then this positive memory will influence on the demand for hotel nights this summer, etc.

Table 5. Results from estimation of ECM-models of hotel nights in Denmark by nationality.

	USA		Japan		Sweden		Norway		Finland	
	Short run	Long run	Short run	Long run	Short run	Long run	Short run	Long run	Short run	Long run
	β_0 to β_8	β_9 to β_{15}	β_0 to β_8	β_9 to β_{15}	β_0 to β_8	β_9 to β_{15}	β_0 to β_8	β_9 to β_{15}	β_0 to β_8	β_9 to β_{15}
Constant	0.017		0.019		0.011		0.047***		0.026	
	(0.016)		(0.016)		(0.012)		(0.018)		(0.019)	
PDK, *Danish price*	!0.258	0.918***	!0.101	0.337	!0.316*	!0.653***	!0.240	7.927**	!1.581***	!0.630***
	(0.373)	(0.455)	(0.329)	(2.250)	(0.184)	(1.428)	(0.348)	(5.127)	(0.512)	(0.680)
PFO, *foreign price*	!1.456***	!1.814***	!0.406	!3.234***	0.417**	0.062*	0.687*	!7.541*	!0.084	0.706
	(0.453)	(0.555)	(0.621)	(3.102)	(0.224)	(1.509)	(0.354)	(5.222)	(0.468)	(0.624)
EXC, *exchange rate*	0.096	0.914***	0.019	1.730	0.218***	1.373***	0.320**	0.847	0.273*	1.183***
	(0.065)	(0.131)	(0.072)	(1.634)	(0.069)	(1.014)	(0.139)	(2.081)	(0.141)	(0.195)
TEMP, *temperature*	0.074*	!0.010***	0.149**	1.789***	0.038	0.785***	0.071*	0.986***	0.009	!0.193**
	(0.042)	(0.193)	(0.051)	(1.106)	(0.030)	(0.238)	(0.041)	(0.616)	(0.052)	(0.164)
SUN, *sunny hours*	0.095***	0.651***	!0.012	!2.611***	!0.016	0.295*	!0.011	0.613	0.008	0.136*
	(0.032)	(0.104)	(0.035)	(1.914)	(0.021)	(0.145)	(0.029)	(0.401)	(0.038)	(0.086)
CLOUD	0.231***	0.531***	0.096	1.888***	!0.017	!0.920***	0.078	1.492	0.167**	0.611**
	(0.064)	(0.167)	(0.069)	(1.661)	(0.045)	(0.635)	(0.057)	(1.166)	(0.079)	(0.236)
RAIN	0.001	0.073*	!0.014	!0.698***	0.009	0.050*	!0.019*	!0.261	!0.025	!0.016
	(0.011)	(0.046)	(0.012)	(0.570)	(0.070)	(0.133)	(0.011)	(0.252)	(0.014)	(0.680)
ECM	0.027*		0.026*		0.008		0.025*		0.036**	
	(0.016)		(0.014)		(0.015)		(0.014)		(0.014)	
Short run	Diagnostics	Augmentation	Diagnostics	Augmentation	Diagnostics	Augmentation	Diagnostics	Augmentation	Diagnostics	Augmentation
R^2	0.52	H: 1,12	0.48	H: 1,7,8,12	0.43	H:1,2,11,12	0.35	H: 9,10,12	0.32	H: 2,5,12
σ	0.128	PK: 12	0.143	P. 3,5	0.084	T: 12	0.122	T: 1,2,6	0.161	PK: 9
DW	2.03	E: 9	1.82	T: 3,12	2.00	S: 10	1.78	R: 1,8	1.80	E: 9
AR	0.72	R: 3	3.31***	R: 4	0.94	C: 2,10	2.40**		0.78	T: 6,10
ARCH	1.91*		0.92		1.10		3.60***		2.05**	C: 12
Normality	15.70***		1.52		0.06		8.47***		119.74***	

*Note: The number in the parenthesis is the standard error. A *** indicates that the variable is significant at the 1 percent level, a ** at the 5 percent level, and a * at the 10 percent level. The significance of the short run estimates are tested by use of a t-test, whereas the significance of the long run estimates is tested by use of a F-test. σ is the standard error of estimate. DW is the Durbin-Watson statistic. AR is a LM test of 1st to 7th order autocorrelation with an F-distribution. ARCH is a test for autoregrerssive conditional heteroscedasticity also of 7th order. Finally, a χ^2-test for normality is presented. See also Hendry and Doornik (2001).*

Table 5. (Continued) Results from estimation of ECM-models of hotel nights in Denmark by nationality.

	Germany		England		Netherlands		France		Italy	
	Short run	Long run	Short run	Long run	Short run	Long run	Short run	Long run	Short run	Long run
	β_0 to β_8	β_9 to β_{15}	β_0 to β_8	β_9 to β_{15}	β_0 to β_8	β_9 to β_{15}	β_0 to β_8	β_9 to β_{15}	β_0 to β_8	β_9 to β_{15}
Constant	0.021		0.013		0.047***		0.031**		0.141***	
	(0.031)		(0.010)		(0.016)		(0.013)		(0.028)	
PDK, *Danish price*	!5.282**	6.011**	!0.500**	!3.679***	!0.768**	9.214*	1.540***	1.564***	1.028	!6.145
	(2.217)	(3.634)	(0.214)	(1.333)	(0.353)	(8.285)	(0.513)	(1.525)	(0.828)	(5.349)
PFO, *foreign price*	6.364***	!6.851***	0.534	3.214***	0.669*	!8.070	!1.653***	!1.465	!1.325	4.316***
	(1.969)	(4.389)	(0.417)	(1.079)	(0.398)	(9.009)	(0.494)	(1.480)	(0.948)	(3.626)
EXC, *exchange rate*	!0.067	!5.245***	!0.106*	0.393	0.048	!11.030	0.531***	1.669**	0.654***	1.024
	(0.494)	(3.994)	(0.060)	(1.079)	(0.297)	(8.742)	(0.203)	(0.897)	(0.181)	(1.142)
TEMP, *temperature*	0.016	2.225***	0.013	!0.579***	0.033	0.901	0.047	!0.096	!0.058	!0.291
	(0.081)	(0.839)	(0.030)	(0.339)	(0.043)	(0.954)	(0.044)	(0.242)	(0.072)	(0.561)
SUN, *sunny hours*	!0.007	!1.335***	0.028	0.201	0.028	!0.969***	!0.011	0.419***	0.005	0.228***
	(0.007)	(0.708)	(0.022)	(0.152)	(0.032)	(1.608)	(0.029)	(0.133)	(0.052)	(0.555)
CLOUD	0.171	!0.189	0.074	1.024***	0.101	!3.161***	!0.018	!0.291***	0.133	1.334
	(0.133)	(0.881)	(0.046)	(0.554)	(0.063)	(3.863)	(0.059)	(0.286)	(0.109)	(1.075)
RAIN	!0.013	!0.339*	!0.009	!0.196**	!0.013	!0.030	0.017	0.046	0.009	0.068***
	(0.021)	(0.202)	(0.008)	(0.115)	(0.011)	(0.299)	(0.011)	(0.055)	(0.018)	(0.308)
ECM	0.007		0.002		0.005		0.041**		0.049***	
	(0.012)		(0.008)		(0.009)		(0.021)		(0.017)	
Short run	Diagnostics	Augmentation	Diagnostics	Augmentation	Diagnostics	Augmentation	Diagnostics	Augmentation	Diagnostics	Augmentation
R^2	0.29	H: 10,12	0.40	H: 1,12	0.23	H: 1,4,12	0.38	H: 1,10,12	0.29	H: 11,12
σ	0.241	PK: 1	0.090	P: 3,4	0.130	T: 1	0.125	E: 12	0.217	PK: 11
DW	1.93	P: 6,9	1.92	E: 7	1.93	S: 8	1.92	T: 7,11,12	1.77	P: 6,7
AR	2.47**	S: 1	1.87*	T: 6,11	1.13	C: 8	1.62	C: 8	1.90	S: 3,6
ARCH	0.17	C: 1	0.89		0.68		0.87	R: 7	2.68**	C: 6,8,12
Normality	395.01***	R: 2	10.00***		38.04***		10.31***		3.47	

*Note: The number in the parenthesis is the standard error. A *** indicates that the variable is significant at the 1 percent level, a ** at the 5 percent level, and a * at the 10 percent level. The significance of the short run estimates are tested by use of a t-test, whereas the significance of the long run estimates is tested by use of a F-test. σ is the standard error of estimate. DW is the Durbin-Watson statistic. AR is a LM test of 1st to 7th order autocorrelation with an F-distribution. ARCH is a test for autoregrerssive conditional heteroscedasticity also of 7th order. Finally, a χ^2-test for normality is presented. See also Hendry and Doornik (2001).*

Overall with regard to diagnostics the ECM approach is sufficient although several estimations have problems especially with the assumption of normality, and the test for autocorrelation.

6. Evaluation of Forecasting Performance

In order to evaluate the ex post forecasts over the period from 1998.1 to 2000.12, i.e. 36 months of the three presented models consider Table 6 presenting for each of the three representations considered several indicators of the forecasting performance.

First, *RCV* is the ratio of the coefficient of variation between the observed series and the forecasted series used as an indicator of the true and the forecasted variation. If $RCV>1$ the observed variation exceeds the forecasted variation. *RCV* should be as close to unity as possible. The remaining measures of forecasting performance are primarily based on the forecasting error defined as $e_t = Y_t - \hat{Y}_t$ where Y_t is the observed value and $\hat{Y}_t$ is the forecasted value. *ME* is the mean error, *SDE* is the standard deviation of the error, *MAE* is the mean absolute deviation or mean absolute error and defined as

$$(5) \qquad MAE = \frac{1}{m}\sum_{t=1}^{m} |e_t|$$

where *m* is the length of the forecasting horizon. *MaxAE* and *minAE* are the minimum and maximum value of the absolute error respectively. The mean absolute percentage error, *MAPE*, is defined as

$$(6) \qquad MAPE = \frac{\sum_{t=1}^{m} \frac{|e_t|}{Y_t}}{m}$$

The root mean squared percentage error, RMSPE, is defined as

$$(7) \qquad RMSPE = \sqrt{\frac{1}{m}\sum_{t=1}^{m}\left(\frac{e_t}{Y_t}\right)^2}$$

Finally, *M* denotes the percentage of times the observed value exceeds the forecasted value. If *M* takes a value around 50 the variation of the forecasted value should be efficient. If $M>50$ the model underestimates the true DGP and vice versa.

What is a good forecast? If a perfect representation is found then the observed and forecasted values should be equal and $e_t=0$. Therefore, all indicators should be as small as possible. Notice finally that if $Y_t=0$ MAPE and RMSPE will not be defined. These cases have consequently been omitted from the present analysis.

It is evident from the statistics that it is very difficult to judge on the forecasting performance from a given representation to another. In order to obtain a subjective impression of the

Table 6. Forecasting performance evaluation by presentation. 1998.1 to 2000.12.

	USA			Japan			Sweden			Norway			Finland		
	HEGY	Dummy	ECM	HEGY	Dummy	ECM	HEGY	Dummy	ECM	HEGY	Dummy	ECM	HEGY	Dummy	ECM
RCV	0.831	0.286	1.809	1.898	1.867	0.768	2.124	1.815	3.597	2.410	1.595	2.461	2.267	1.013	3.389
ME	0.042	0.053	0.030	!0.001	0.002	!0.030	!0.002	0.001	!0.006	!0.003	!0.002	0.008	!0.010	0.014	!0.029
SDE	0.124	0.121	0.119	0.149	0.127	0.178	0.117	0.122	0.111	0.125	0.135	0.150	0.159	0.164	0.166
MAE	0.091	0.100	0.097	0.108	0.091	0.142	0.094	0.088	0.084	0.095	0.099	0.119	0.110	0.125	0.123
*max*AE	0.285	0.325	0.231	0.406	0.352	0.420	0.282	0.313	0.294	0.342	0.462	0.302	0.470	0.425	0.430
*min*AE	0.001	0.009	0.009	0.001	0.002	0.023	0.004	0.001	0.009	0.005	0.001	0.004	0.001	0.002	0.001
MAPE	0.348	0.187	0.277	0.079	!0.007	!0.431	0.071	!0.085	!0.226	!0.384	!0.095	!0.317	0.156	0.259	0.019
RMSPE	1.164	1.758	1.434	1.380	1.234	1.939	1.853	1.364	1.646	1.527	1.778	1.481	1.354	1.398	1.403
M	63 %	74 %	61 %	53 %	53 %	48 %	52 %	47 %	43 %	33 %	39 %	58 %	52 %	53 %	44 %
Rank	2	3	1	2	1	3	2	1	3	2	1	3	1	2	3
	Germany			England			Netherlands			France			Italy		
	HEGY	Dummy	ECM	HEGY	Dummy	ECM	HEGY	Dummy	ECM	HEGY	Dummy	ECM	HEGY	Dummy	ECM
RCV	0.172	0.584	0.753	0.692	1.025	1.672	1.832	1.836	1.607	1.147	0.709	0.736	5.723	2.163	2.986
ME	!0.102	!0.131	!0.179	0.040	0.038	0.012	!0.003	!0.012	!0.060	!0.009	!0.022	!0.091	!0.041	!0.006	!0.012
SDE	0.201	0.181	0.197	0.094	0.075	0.079	0.205	0.230	0.239	0.110	0.099	0.125	0.134	0.132	0.137
MAE	0.182	0.173	0.204	0.084	0.067	0.064	0.149	0.162	0.180	0.087	0.087	0.127	0.114	0.098	0.156
*max*AE	0.570	0.506	0.717	0.265	0.221	0.175	0.540	0.645	0.587	0.270	0.203	0.304	0.286	0.349	0.345
*min*AE	0.026	0.027	0.017	0.001	0.003	0.009	0.003	0.013	0.003	0.005	0.008	0.020	0.002	0.008	0.004
MAPE	!0.999	!1.099	!1.268	0.445	0.255	!0.094	0.050	0.116	!0.174	!0.521	!0.723	!0.587	!0.665	!0.299	!0.930
RMSPE	2.286	1.919	1.597	1.127	0.924	1.286	0.891	0.990	0.982	1.388	1.615	1.204	1.670	1.095	2.024
M	28 %	25 %	25 %	67 %	71 %	67 %	46 %	46 %	44 %	49 %	37 %	31 %	44 %	44 %	28 %
Rank	2	1	3	3	2	1	1	2	3	1	2	3	2	1	3

Note: RCV is the numeric value of the ratio of the coefficient of variation between the observed and the forecasted observations. The coefficient of variation is defined as the standard deviation divided by the mean. The forecast error is defined as the observed value minus the forecasted value. ME is the mean error, SDE is the standard deviation of the error, MAE is the mean absolute error, maxAE and minAE are the minimum and maximum value of absolute error respectively, MAPE is the mean absolute percentage error, RMSPE is root mean squared percentage error, and finally M denotes the number of times the observed value exceeds the forecasted value. Detailed definitions and discussion of the "rank" are found in the text.

forecasting performance consider the row “rank”. Here the rules above have been imposed and for a given nationality the three representations have “competed” with each other. The representation with the minimum points has been accessed with “1” etc. For several reasons this is a very doubtful method. The choice of measures may not be optimal, each measure is being accessed with equal weight etc. Here this procedure only serves to make the model performance more visible.

With regard to forecast performance the “rank” results are surprising. Out of the 11 cases considered the HEGY representation performs best in three cases (Finns, Dutchmen and Frenchmen), the dummy representation performs best in five cases (Japanese, Swedes, Norwegians, Germans and Italians), whereas the ECM representation performs best in two cases (Englishmen and Americans). Overall the ECM representation is the less efficient model for forecasting purposes. This is the case in eight of the cases. Notice also that the most inefficient forecasting performance regardless of the representation considered is observed for the nationalities less significance for the overall demand for hotel nights such as Italy and France. Surprisingly the performance for a significant demander such as Germany is poor.

What is the interpretation of these results? In the previous sections it was observed that the nature of seasonality generally is stochastic rather than deterministic. Further, it was found that the ECM representation in many aspects proved insightful information with regard to the way that economic and climatic factors influenced on the demand for hotel nights.

The answer to this question depends of the purpose of the analysis in consideration. If the aim is to forecast the demand for hotel nights in the short run, it will be most efficient to rely on the dummy representation. This representation is static by nature and the message embodied in the model is “that things remains the same”. However, recall from Section 4 that several of the estimations relied on insignificant coefficient estimates. All then depended on the lagged variables. In this case the dummy variable representation should be used with great care. On the other hand, if the aim is to obtain long run forecasts, the HEGY approach is superior. In the long run an increasing number of factors may influence on the DGP increases. This is also very visible from the ECM modelling. This model is the most appealing if the purpose of the analysis it to identify the factors of importance for the demand for hotel nights.

7. Conclusion

The present analysis has drawn attention to the forecasting performance for three different approaches of modelling the DGP of monthly hotel nights in Denmark by nationality.

The demand for hotel nights is especially of importance for neighbouring countries such as Germany, Sweden and Norway. Using observations ranging for thirty years it is evident that the nature of seasonality is more stochastic than deterministic. In addition, it has been shown that insightful information can be obtained by using an ECM approach to model hotel demand not only with regard to the influence of economic factors such as the price level or the exchange rate, but also from climatic variables such as the temperature or the number of sunny hours. Further, a "memory" behaviour is observed especially with regard to the climatic variables.

Moving to forecasting performance it is likely that the deterministic seasonal model outperforms the stochastic seasonal approach. This result should be interpreted with great care. Several of the estimated deterministic models have only a very few significant coefficients. This means that the augmented lagged variables determine the process, and it is likely to be reduced to some kind of an AR-function. The poorest performance on forecasting is obtained by the ECM-model. On the other hand this model reveals the most interesting information on elasticities etc.

When comparing the three different models abilities to forecast all indicators are generally very good and the differences between the methods are consequently very small.

References

Barry, K., and J. O'Hagan, 1972, An Econometric Study of British Tourist Expenditure in Ireland. *Economic and Social Review,* 3, p 143-161.

Barsky, R.B., and J.A. Miron, 1989, The Seasonal Cycle and the Business Cycle. *Journal of Political Economy*, 97, p 503-534.

Clements, M.P., and D.F. Hendry, 1999, *Forecasting Non-stationarity Economic Time Series.* MIT Press.

Franses, P.H., 1991, Seasonality, Non-Stationarity and the forecasting of Monthly Time Series. *International Journal of Forecasting*, 11, p 447-475.

Hendry, D.F., and J.A. Doornik, 2001, *PcGive 10.*. Timberlake Consultans Ltd.

Hylleberg, S., R.F. Engle, C.W.J. Granger, and B.S. Yoo, 1990, Seasonal Integration and Cointegration. *Journal of Econometrics*, 44, p 215-238.

Song, H. And S.F. Witt, 2000, *Tourism Demand Modelling and Forecasting.* Pergamon Press.

Sørensen, N.K., 1999, 2001, Modelling the Seasonality of Hotel Nights in Denmark, *Tourism Economics*, 5, p 9-25 or Chapter 5 in Lundtorp, S. And T. Baum (ed), *Seasonality in Tourism*. Pergamon Press, Elsevier Science.

Witt, S.F. and C.A.Witt, 1992, *Modelling and Forecasting Demand in Tourism.* Academic Press.

Witt, S.F. and C.A. Witt, 1995, Forecasting Tourism Demand: A Review of Empirical research. *International Journal of Forecasting*, 11, p 447-475.